

INTERNATIONAL HANDBOOK
OF DEVELOPMENT ECONOMICS
VOLUME ONE

International Handbook of Development Economics

Volume One

Edited by

Amitava Krishna Dutt

University of Notre Dame, USA

and

Jaime Ros

University of Notre Dame, USA

Edward Elgar

Cheltenham, UK • Northampton, MA, USA

© Amitava Krishna Dutt and Jaime Ros 2008

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical or photocopying, recording, or otherwise without the prior permission of the publisher.

Published by
Edward Elgar Publishing Limited
The Lypiatts
15 Lansdown Road
Cheltenham
Glos GL50 2JA
UK

Edward Elgar Publishing, Inc.
William Pratt House
9 Dewey Court
Northampton
Massachusetts 01060
USA

A catalogue record for this book
is available from the British Library

Library of Congress Control Number: 2008927966

ISBN 978 1 84542 327 8 (2 volume set)

Printed and bound in Great Britain by MPG Books Ltd, Bodmin, Cornwall

Contents

<i>List of contributors</i>	viii
<i>Preface</i>	xiii

PART I INTRODUCTION

1	The meaning and measurement of development	3
	<i>Paul Streeten</i>	
2	Historical perspectives on development	16
	<i>Amiya Kumar Bagchi</i>	
3	Empirics of growth and development	32
	<i>Steven N. Durlauf, Andros Kourtellos and Chih Ming Tan</i>	
4	Structural change and development	48
	<i>Moshe Syrquin</i>	
5	Interdisciplinary approaches to development: the 'institutional' turn	68
	<i>Peter Evans</i>	
6	Data problems and empirical modeling in developing economies	83
	<i>Bill Gibson</i>	

PART II ALTERNATIVE APPROACHES TO DEVELOPMENT

7	Historical antecedents of development economics	97
	<i>Gianni Vaggi</i>	
8	Classical development theory	111
	<i>Jaime Ros</i>	
9	Theories of dependency	125
	<i>José Gabriel Palma</i>	
10	Structuralism	136
	<i>José Gabriel Palma</i>	
11	Marxism and development	144
	<i>Bob Sutcliffe</i>	
12	Institutionalist development economics	162
	<i>Kenneth P. Jameson</i>	
13	Neoclassical development economics	176
	<i>Salim Rashid</i>	

PART III MACROECONOMICS OF GROWTH
AND DEVELOPMENT

14	General long-run approaches to growth and development <i>Amitava Krishna Dutt</i>	189
15	New growth theory and development economics <i>Heinz D. Kurz and Neri Salvadori</i>	207
16	Short-run macroeconomic issues in development <i>Peter J. Montiel</i>	223
17	Sectoral interactions in development <i>Jorn Rattsø</i>	238
18	Open-economy issues in development <i>José Antonio Ocampo</i>	252

PART IV FACTORS IN DEVELOPMENT

19	Savings, investment and capital accumulation <i>Andrés Solimano and Mario Gutierrez</i>	269
20	Role of finance and credit in economic development <i>Philip Arestis and Santonu Basu</i>	290
21	Physical infrastructure <i>Gregory K. Ingram and Marianne Fay</i>	301
22	Population and development <i>Dennis Ahlburg and Robert Cassen</i>	316
23	Labor markets in developing countries <i>Albert Berry</i>	328
24	Education and human capital <i>George Psacharopoulos and Harry Anthony Patrinos</i>	341
25	Health and nutrition and economic development <i>Harold Alderman, Jere R. Behrman and John Hoddinott</i>	356
26	Entrepreneurship and development <i>E. Wayne Nafziger</i>	371
27	Natural resources and development <i>Richard Auty</i>	388
28	The environment and development <i>John McPeak</i>	404
29	Technical choice and technological change in development <i>Howard Pack</i>	417

PART V SECTORS IN DEVELOPMENT

30	Factor market imperfections in poor agrarian economies <i>Parikshit Ghosh and Ashok Kotwal</i>	435
31	The Green Revolution <i>Robert E. Evenson</i>	453
32	Industry and industrial policy <i>Helen Shapiro</i>	470
33	The informal sector <i>Victor E. Tokman</i>	483
34	Services and development <i>Dilip Dutta</i>	497
35	Urbanization and rural–urban migration <i>Charles M. Becker</i>	516
	<i>Index</i>	533

Contributors

Ahlburg, Dennis: Leeds School of Business, University of Colorado, Boulder, Colorado, USA

Alderman, Harold: World Bank, Washington, DC, USA

Arestis, Philip: Department of Land Economy, University of Cambridge, Cambridge, UK

Auty, Richard: Department of Geography, Lancaster University, Lancaster, UK

Bagchi, Amiya Kumar: Institute of Development Studies, Kolkata, India

Baland, Jean-Marie: Centre for Research in the Economics of Development (CRED), University of Namur, Namur, Belgium

Balasubramanyam, V.N.: Department of Economics, Lancaster University, Lancaster, UK

Bardhan, Pranab: Department of Economics, University of California, Berkeley, California, USA

Basu, Santonu: Queen Mary, University of London, London, UK

Becker, Charles M.: Department of Economics, Duke University, Durham, North Carolina, USA

Behrman, Jere R.: Department of Economics, University of Pennsylvania, Philadelphia, Pennsylvania, USA

Berry, Albert: Department of Economics, University of Toronto, Toronto, Canada

Birdsall, Nancy: Center for Global Development, Washington, DC, USA

Brown, Graham: Department of International Development, University of Oxford, Oxford, UK

Bruton, Henry J.: Department of Economics, Williams College, Williamstown, Massachusetts, USA

Cassen, Robert: London School of Economics, London, UK

Dahi, Omar S.: School of Social Science, Hampshire College, Amherst, Massachusetts, USA

Damon, Amy: University of Minnesota, Minneapolis-St. Paul, Minnesota, USA

Darity, William A., Jr: Duke University, Durham, North Carolina, USA

Demir, Firat: Department of Economics, University of Oklahoma, Norman, Oklahoma, USA

Durlauf, Steven N.: Department of Economics, University of Wisconsin-Madison, Madison, Wisconsin, USA

Dutt, Amitava Krishna: Department of Economics and Policy Studies, University of Notre Dame, Notre Dame, Indiana, USA

Dutta, Dilip: Faculty of Economics and Business, University of Sydney, Sydney, Australia

Elson, Diane: Levy Economics Institute of Bard College, and Department of Sociology, University of Essex, Colchester UK

Evans, Peter: Department of Political Science, University of California, Berkeley, California, USA

Evenson, Robert E.: Department of Economics, Yale University, New Haven, Connecticut, USA

Fay, Marianne: World Bank, Washington, DC, USA

Fishlow, Albert: School of International and Public Affairs, Columbia University, New York, USA

FitzGerald, Valpy: University of Oxford, Oxford, UK

Ghosh, Jayati: Centre for Economic Studies and Planning, School of Social Sciences, Jawaharlal Nehru University, New Delhi, India

Ghosh, Parikshit: Indian Statistical Institute, Delhi, India

Gibson, Bill: Department of Economics, University of Vermont, Burlington, Vermont, USA

Glass, Amy Jocelyn: Department of Economics, Texas A&M University, College Station, Texas, USA

Glewwe, Paul: Department of Applied Economics, University of Minnesota, St Paul, Minnesota, USA

Gray, Cheryl W.: World Bank, Washington, DC, USA

Griffith-Jones, Stephany: Institute of Development Studies, University of Sussex, Brighton, UK

Gutierrez, Mario: International consultant, Paris, France

Hoddinott, John: International Food Policy Research Institute, Washington, DC, USA

Hoksbergen, Roland: Department of Economics, Calvin College, Grand Rapids, Michigan, USA

Huang, Yasheng: Sloan School of Management, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

Ingram, Gregory K.: Lincoln Institute of Land Policy, Cambridge, Massachusetts, USA

Jager, Henk: Amsterdam School of Economics, University of Amsterdam, Amsterdam, The Netherlands

Jameson, Kenneth P.: Department of Economics, University of Utah, Salt Lake City, Utah, USA

Jha, Raghendra: Division of Economics, Australian National University, Canberra, Australia

Jomo K.S.: Department of Economic and Social Affairs, United Nations, New York, USA

Kaminsky, Graciela L.: George Washington University and NBER, Washington, DC, USA

Knaack, Ruud: Amsterdam School of Economics, University of Amsterdam, Amsterdam, The Netherlands

Kotwal, Ashok: Department of Economics, University of British Columbia, Vancouver, Canada

Kourtellos, Andros: Department of Economics, University of Cyprus, Nicosia, Cyprus

Kozul-Wright, Richard: Department of Economic and Social Affairs, United Nations, New York, USA

Kurz, Heinz D.: Department of Economics, University of Graz, Graz, Austria

Masson, Paul R.: Rotman School of Management, University of Toronto, Toronto, Canada

McPeak, John: Department of Public Administration, Maxwell School, Syracuse University, New York, USA

Montiel, Peter J.: Williams College, Williamstown, Massachusetts, USA

Morley, Samuel A.: International Food Policy Research Institute, Washington, DC, USA

Nafziger, E. Wayne: Department of Economics, Kansas State University, Manhattan, Kansas, USA

Ndulu, Benno J.: Governor, Bank of Tanzania, Dar-es-Salaam, Tanzania.

Ocampo, José Antonio: School of International and Public Affairs, Columbia University, New York, USA

O'Connell, Stephen A.: Department of Economics, Swarthmore College, Swarthmore, Pennsylvania, USA

Pack, Howard: Department of Business and Public Policy, Wharton School, University of Pennsylvania, USA

Palma, José Gabriel: University of Cambridge, Cambridge, UK

Patnaik, Prabhat: Centre for Economic Studies and Planning, School of Social Sciences, Jawaharlal Nehru University, New Delhi, India

Patrinos, Harry Anthony: World Bank, Washington, DC, USA

Platteau, Jean-Philippe: Centre for Research in the Economics of Development (CRED), University of Namur, Namur, Belgium

Psacharopoulos, George: University of Illinois, Champaign, Illinois, USA

Putterman, Louis: Department of Economics, Brown University, Providence, Rhode Island, USA

Radelet, Steven: Center for Global Development, Washington, DC, USA

Raffer, Kunibert: Department of Economics, University of Vienna, Vienna, Austria

Rashid, Salim: Department of Economics, University of Illinois, Urbana-Champaign, Illinois, USA

Rattsø, Jørn: Department of Economics, Norwegian University of Science and Technology (NTNU), Trondheim, Norway

Rayment, Paul: formerly, Economic Commission for Europe, United Nations, Geneva, Switzerland

Rivera-Batiz, Francisco L.: Columbia University, New York, USA

Rodríguez, Francisco: Department of Economics, Wesleyan University, Middletown, Connecticut, USA

Ros, Jaime: Department of Economics and Policy Studies and Kellogg Institute of International Studies, University of Notre Dame, Notre Dame, Indiana, USA

Saggi, Kamal: Department of Economics, Southern Methodist University, Dallas, Texas, USA

Salvadori, Neri: Department of Economic Science, University of Pisa, Pisa, Italy

Sapsford, David: Department of Applied Economics, University of Liverpool, Liverpool, UK

Sen, Anindya: Indian Institute of Management, Kolkata, India

Shapiro, Helen: Department of Sociology, University of California, Santa Cruz, CA, USA

Solimano, Andrés: UN-Economic Commission for Latin America and the Caribbean

Squire, Lyn: Global Development Network, New Delhi, India

Stewart, Frances: Department of International Development, University of Oxford, Oxford, UK

Streeten, Paul: Emeritus Professor, Department of Economics, Boston University, Boston, Massachusetts, USA

Sutcliffe, Bob: formerly at University of the Basque Country, Bilbao, Spain

Syrquin, Moshe: School of International Studies, University of Miami, Coral Gables, Florida, USA

Tan, Chih Ming: Department of Economics, Tufts University, Medford/Somerville, Massachusetts, USA

Tokman, Victor E.: Economic Advisor to the President of Chile, Santiago, Chile, formerly at the International Labour Organization, Geneva, Switzerland

Triplett, Russell E.: University of North Carolina at Chapel Hill, North Carolina, USA

Vaggi, Gianni: Department of Economics, University of Pavia, Pavia, Italy

Wilber, Charles K.: Professor Emeritus, University of Notre Dame, Notre Dame, Indiana, USA

Preface

Development economics is a very large and growing subdiscipline or field of economics. Since it is concerned with the economic problems of economies which have been variously defined as underdeveloped, less-developed or developing, and where most of the poor and the underprivileged people of the world live, it is also arguably one of the most important – if not the most important – field of economics. It is also a complex field which has been approached by scholars using different approaches, involving different definitions of development, different methods of analysis, different views of how economies function, and different recipes of what should be done to bring about development. In the course of the evolution of the field, even if one focuses only on its evolution since the end of World War II, some approaches have enjoyed greater popularity at certain times, when others have lost ground, with role reversals later on.

The field's size, complexity and transformations make it a difficult business to represent it with a handbook, especially when some excellent ones already exist. But these same features of the field provided us with some reasons to take on the task. It has been some years since some of the earlier handbooks were published. To the extent that handbooks survey recent literatures, they become outdated. Moreover, it can be argued that some of the earlier efforts did not sufficiently take into account the great diversity of approaches in the field and arguably stressed some approaches more popular at the time.

In line with our goals, we approached scholars who were experts in their area to participate in this project. We asked them to discuss what they believed to be some of the key issues concerning their topic, and of the major contributions to it, rather than writing exhaustive surveys. Moreover, we asked the contributors to examine analytical contributions, as well as the relation between these contributions and real-world and policy issues, although inviting them to choose the precise balance between these. We also asked contributors to attempt to cover contributions from alternative theoretical perspectives. To further have different views represented, especially views under-represented in mainstream development economics, we approached scholars using a variety of different approaches and devoted an entire section of the *Handbook* (Part II) to the discussion of alternative approaches.

The *Handbook* is divided into nine parts. Part I, which is introductory, discusses the meaning and measurement of economic development, historical

and interdisciplinary perspectives on development, empirical regularities in development, and data problems and empirical modeling in developing economies.

Part II, as mentioned earlier, deals with alternative approaches. It starts with earlier contributions to development economics, then discusses classical development theory of the early days after World War II, and then turns to different approaches to development economics, that is, the dependency and structuralist approaches, the Marxist approach, the institutionalist approach and the neoclassical approach. The different approaches can be characterized in different ways, and our contributors have chosen their own characterizations.

Part III examines the macroeconomics of growth and development. It starts with a discussion of general long-run approaches to growth from a theoretical perspective, and then discusses new growth theory in more detail. Then it turns to short-run macroeconomic issues. Next it turns to sectoral interactions, focusing on the interaction between agriculture and industry, and to general open economic issues in development.

Part IV discusses factors in development, not only in the narrow sense of inputs such as capital, labor and natural resources, but also in the broader sense which includes entrepreneurship, the environment and technological change. On capital, one entry examines savings, investment and capital accumulation in general, another entry focuses on finance and credit, and yet another discusses physical infrastructure. On labor, there are entries on population, labor markets, education and human capital formation, and health and nutrition. Next it turns to the role of entrepreneurship, and natural resources are then discussed. The environment is addressed next, not just as a factor of production, but also in terms of its sustainability. A final entry addresses technological issues, focusing on technical choice and technological change.

Part V examines specific sectors in development. On agriculture, there are entries on agricultural factor markets and institutions and on the so-called Green Revolution, which addresses the issue of technological change in the sector. Next, the discussion turns to the industrial sector, the informal sector, and services. Finally, there is an entry on urbanization and intersectoral migration.

Part VI turns to international issues. On international trade, it starts with a discussion of how free trade affects developing countries, then examines the role of the terms of trade, which has received much attention in the development literature, and then turns to trade policy, especially to the debate between import-substitution and export-promotion policies. On international capital flows, it examines direct foreign investment, debt and portfolio flows and addresses the issue of the volatility of capital flows, and then turns

to foreign aid. Next, international migration and the brain drain are examined. International technology transfers are addressed by focusing on one of the major modes of such transfers, foreign direct investment. The role played by international institutions in development is addressed by examining the Bretton Woods institutions and the World Trade Organization. This part ends with an examination of the relation between rich and poor countries, or what is usually referred to as North–South issues.

Part VII examines distributional issues. It first discusses the measurement and determination of income inequality, addressing how development affects inequality. It then turns to how income distribution affects growth and development. The measurement and determinants of poverty are addressed in the next entry, examining poverty in a narrow way as well as a broader, multidimensional way. Issues concerning gender are examined in the next entry which considers how the fruits of development are shared between the sexes and whether improving gender distribution and development are positively related. The economic conditions of children and the impact of growth of children's welfare are examined next. Finally, this part examines the measurement and conceptual issues relating to ethnic inequality, and discusses the relation between ethnic inequality and economic growth.

Part VIII examines the role of the state and other institutions in development. It commences with a general discussion of the two main institutions emphasized in economics, the state and the market. It next examines different aspects of government policy, that is, monetary policy, fiscal policy, stabilization policy and structural adjustment, planning and project appraisal, and state-owned enterprises and privatization. This is followed by a discussion of corruption. The role of law and legal institutions is examined next, followed by a discussion of the institution of property rights. Broader issues regarding culture and development are considered next. The part concludes with a discussion of the causes and consequences of wars, especially civil wars, in less-developed countries.

Finally, Part IX provides a review of the main issues concerning recent actual development experience. It opens with an overall discussion of development in less-developed regions from an international perspective. It then examines in turn Latin America and the Caribbean, sub-Saharan Africa, North Africa and the Middle East, China, South Asia, the East Asian newly industrialized countries, and the post-socialist transitional economies. The inclusion of the final region is explained both because these are often considered to be less-developed countries, and because their experience has important lessons for development.

The chapters just described, numbering 71, have been written by 90 contributors, who live in (or are from) all five continents. This – in addition to

the fact that it covers development issues relating to, and the experience of, countries all around the less-developed world – makes the *Handbook* truly international in scope.

A work such as this could not have been completed without the hard work of a large number of people. As editors, we would like to thank all the contributors, who have graciously devoted time and contributed their expertise to this project. We would also like to thank some other individuals for their comments and/or suggestions, including Chris Barrett, Kaushik Basu, Jagdish Bhagwati and Michael Ellerman. A few people whose work was to be included here – Sanjaya Lall, David Pierce and Richard Sabot – have passed away. We are grateful for their willingness to contribute and deeply saddened by their deaths.

The Editors

PART I

INTRODUCTION

1 The meaning and measurement of development

Paul Streeten

What do people want?

The great West Indian economist and Nobel Prize winner Arthur Lewis defined development as the enlargement of the range of people's choices. Following him the United Nations Development Programme's *Human Development Reports* chose the same definition. Some earlier definitions have run in terms of commodity bundles or specific needs satisfactions. In the book *First Things First* (Streeten et al., 1981) my co-authors and I say:

First, and most important, the basic needs concept is a reminder that the objective of the development effort is to provide all human beings with the opportunity for a full life. In the past two decades, those concerned with development have sometimes got lost in the intricacies of means . . . and lost sight of the end. They came near to being guilty, to borrow a term from Marx, of 'commodity fetishism'.

'Opportunity' is near in meaning to Amartya Sen's 'functioning' and 'capability'. In our basic needs work we tried hard to get away from the definition of development in terms of an aggregate of goods and services produced and consumed and its growth, of the detached objects people happen to possess, and to emphasize the end: people's full lives.¹

Amartya Sen's analysis has been in term of 'capabilities' and 'functionings', and not satisfactions, or happiness, or commodities.² Sen goes beyond the analysis of the commodities in terms of their characteristics (a shirt serves warmth and decoration, and if drip-dry saves ironing) which consumers value, and analyses the characteristics of the consumers; whether they have the capability to make use of the commodities. The same amount of food has a different significance: according to whether the consumer is healthy or has parasites in their stomach, in which case the basic needs of the worms rather than of the consumer are met; according to the rate of metabolism, the age, sex, size and work load of the consumer; according to the climate, according to whether she is pregnant or lactating; according to whether the consumer has acquired through education the knowledge of how to prepare the food; and according to whether they need the food for other uses than their own consumption, such as entertainment or ceremonies.

Sen also argues that human development cannot be judged only by end-states, and that the freedom to choose between different options is an important component of well-being. A given commodity bundle has a different significance to the consumer according to whether he or she has other options, though he or she does not exercise them, or whether that same bundle is the only one available. There is a difference between a starving pauper, a fasting monk and Gandhi on hunger strike, which is not reflected in the low calorie intake of all three. Only the starving pauper lacks capability. But Sen's capabilities cannot be observed, while achievements can. If failure of achievement is voluntary, it is acceptable. But some authors (like Frances Stewart)³ have argued that it is better to separate freedom of choice and look at poverty in terms of observable achievements. In this sense, all three are deprived. Sen lumps together achievement and freedom of choice in happiness 'capability'.

Happiness, as experienced by the individual, is not what human development can aim at or is mainly about. Not only can the government not deliver happiness;⁴ people may be miserably poor and yet be contented. Anita Brookner in one of her novels tells of a woman who was so modest that she did not even presume to be unhappy.⁵ And Susan Minot (1992) in her novel *Folly* writes: 'not only did she not think of making certain choices herself, she was completely unaware of having the desire to do so'. Indian women report being ill much less frequently than Indian men.⁶

The use of Sen's capabilities can be frustrated if the opportunities for their exercise do not exist or if individuals are deprived of these opportunities as a result of discrimination, obstacles or inhibitions: if there is no demand for their productive contributions so that people are unemployed, or if there are legal or social or conventional restrictions on their employment, or if they do not have enough leisure, or if political oppression or deprivation of human rights prevents them from full participation in the life of their communities. There can be 'jobless' growth, there can be 'voiceless' growth, there can be 'rootless' growth, and there can be jobless, voiceless and rootless non-growth. Different countries illustrate each of these cases.

Getting income is one of the options people would like to exercise. It is an important, but not an all-important option. Human development includes the expansion of income and wealth, but it includes many other valued and valuable things as well.

For example, in investigating the priorities of poor people, one discovers that what matters most to them often differs from what outsiders assume. More income is only one of the things desired by poor people. Adequate nutrition, safe water at hand, better medical services, more and better schooling for their children, cheap transport, adequate shelter, continuing

employment and secure livelihoods, and productive, remunerative, satisfying jobs do not show up in higher income per head, at least not for some time.

There are other non-material benefits that are often more highly valued by poor people than material improvements. Some of these partake in the characteristics of rights, both positive and negative; others in those of states of mind. Among these are good and safe working conditions, freedom to choose jobs and livelihoods, freedom of movement and speech, self-determination and self-respect, independence, mobility, liberation from oppression, violence and exploitation, less dependence on patrons, security from persecution and arbitrary arrest, not having to move in search of work, a satisfying family life, the assertion of cultural and religious values, a sense of identity, access to power or direct empowerment, recognition, status, adequate leisure time and satisfying forms of its use, a sense of purpose in life and work, the opportunity to join and participate actively in the activities of civil society, and a sense of belonging to a community. These are often more highly valued than income, both in their own right and as means to satisfying and productive work. They do not show up in higher income figures. No policy-maker can guarantee the achievement of all, or even the majority, of these aspirations, but policies can create the opportunities for their fulfillment.

Economic growth can be quite rapid without an improvement in the quality of life of the majority of the people, and many countries have achieved a high quality of life with only moderate growth rates of income. It has been observed that there is a positive correlation between income per head and the indicators of human development. Some have drawn the erroneous conclusion that it is only income that matters. But, first, this relationship is far from perfect, and the interesting questions are raised by the outliers and particularly by countries that have achieved high human development at low levels of income. Second, this relation depends entirely on the extra income that arises from growth being used for public education and health and for specific attacks on poverty. If these two conditions are absent, the correlation disappears.⁷ Much also depends on the initial distribution of assets. If land ownership is fairly equally distributed and mass education is widespread, the benefits of economic growth will be reflected in good human development.

Economic growth is often considered to be an essential component of human development. But growth (in the narrow sense of a continuing increase of the quantity of goods and services produced and consumed over time) is simply the inter-temporal dimension of any policy objective, although it has been wrongly monopolized by production and consumption: it should apply to poverty reduction, employment, investment, a

more equitable income distribution, environmental protection, leisure and, of course, also to income. But once you specify for income, consumption and production, the 'What?' 'To whom?' 'By whom?' 'For what?' and 'When?' growth becomes the incidental result, not the objective, of a sensible economic policy. Growth is too unspecified, abstract, aggregate and unbounded to be a sensible objective of policy. It also implies an infinite horizon, without limits to increases in income. What matters is the composition of the national income, to what uses it is put, its distribution among beneficiaries, now and for future generations; and with how much effort and in what conditions it is produced. If and only if the extra resources resulting from growth go largely to the poor, and if they are spent on public health and education, will a contribution to human development result.

The national income is a quite inadequate measure of human development for several reasons. It counts only goods and services that are exchanged for money, leaving out of account the large amount of work done inside the family, mainly by women, and work done voluntarily for children or older people or in communities. Public services are counted at their cost, so that doubling the wages of all public servants appears to double their contribution to welfare or development. National income accounting does not distinguish between goods and regrettable necessities, like military or anti-crime expenditure, products needed to combat 'bads'. Addictive eating and drinking is counted twice: when the food and the alcohol are consumed, and when large sums are spent on the diet industry and on cures for alcoholism. Much of what is now counted as economic growth is really either combating evils, and fixing blunders and social decay from the past, or borrowing resources from the future, or shifting functions from the community and household to the market.⁸

National income accounting does not add leisure gained by fewer working hours or an earlier retirement age, and does not subtract from the extra income-generated leisure lost if women are forced (or desire) to take on jobs outside the family, or men to take on a second job. Environmental degradation, pollution and resource depletion are not deducted, so that the earth is treated, it has been said, like a business in liquidation. Freedom, human rights and participation are ignored. Most important, the conventional measure does not allow for the distribution of the income, counting all goods and services at their market prices. Increasing the production of whiskey, bought by rich men, counts for much more than increasing the production of milk that would have gone to a starving child. Attempts have been made here and there to correct for these faults and omissions, but national income remains a quite inadequate measure of economic welfare or of development.

Some of these shortcomings can be removed by adjustments in the accounting methods. These concern those components of well-being that can be, in principle, brought into relation with the measuring rod of money. A monetary value can be attached to leisure time. Income distribution can be allowed for by attaching greater weights to the incomes and their growth of the bottom 20 percent, 30 percent or 40 percent of the population. Depletion of non-renewable raw materials can be evaluated and a measure for sustainable income can be designed.

For other components of choice and welfare, monetary measurement is much more difficult or may be impossible. The enjoyment we derive from an unspoiled wilderness, the satisfaction from work, political engagement that results from participation, the sense of community, brotherhood and sisterhood that grows out of social activities, the freedom, peace and sense of security that are common in a well-run society, these cannot easily be reduced to dollars and cents. Yet they form the essence of human development.

Human development: the latest stage

The contributing tributaries to human development can be grouped under five headings: (1) economic growth; (2) human resource development; (3) human rights and participation; (4) peace and security; and (5) sustainability. The role of culture falls under the heading of human rights and participation. Issues of equity, and in particular of gender equity, run through all five tributaries.

We now live in a 'risk society'. People are bombarded with assessments of the risks of decisions (from what they eat to whether they should build nuclear power stations). They have lost the old certainties about how their lives will turn out: no more jobs, or marriages, for life.

Human development is the end, the tributaries are the means; but they can also acquire end characteristics themselves. Environmental sustainability, peace, participation, human resources and, by some, even economic growth are valued in their own right. To the extent that they are ends, they all have to be included in human development. The five tributaries can augment each other, for example when human resources contribute to higher growth, or when respect for human rights advances peace. There are also feedbacks from achievements in human development to further improvements in human development. These may be indirect by improving the five components (economic growth, human resource development, human rights and participation, peace and security, and sustainability), or they may be direct. The latter occur within and between families when knowledge is passed on and when better education of mothers has an impact on their children. Several studies have shown that women's education, control over cash

income and access to power, in addition to being desirable in themselves, improve the health, nutrition and education of children, reduce fertility, reduce infant mortality, reduce health hazards of adults arising from low birth weight, raise productivity, reduce inequality, are beneficial for the environment, and increase the range and effectiveness of public debates.

Gender issues are particularly important for reproductive freedom – for people, especially women, to be able to choose the size of their families. There is now a wealth of evidence to show that given the opportunity to choose smaller families without adverse economic and social consequences, smaller families are indeed chosen. With human development – that is with the expansion of education, especially of girls and women, the reduction of infant mortality rates, and medical facilities (including the opportunity of birth control) – fertility rates have come down sharply. It may seem paradoxical that reduced infant mortality rates, more children surviving, should contribute to reduced population growth. But there is overwhelming evidence that parents try to overinsure themselves against the deaths of their children (particularly sons) and that more surviving children reduce the desired family size. Human development is the best way to reducing population growth, and reduced population growth advances human development. Human development, in addition to longer life expectancy, better education and securer lives, makes it possible for people to opt for smaller families.

It is thought that some of these links lend themselves more easily to measurement than others. The human resources of education can be captured under literacy rates and school enrolment rates, and the human resources of health under life expectancy and infant mortality. It is for this reason that more attention has been paid to these links than to others, such as that between participation and human development, not so readily brought into relation with a measuring rod. Some may have become victims of the fallacy that what cannot be counted does not count or even exist. But it may be questioned whether the quality of education or the attitudes that a good education instills, such as punctuality, discipline, teamwork, and so on, are caught under the conventional statistical social indicators. The same goes for health measures. Economic growth, based on increases in gross national product (GNP), has of course been the archetypal case of counting and has attracted the limelight of attention.

Human development goes beyond basic needs in that it is concerned with all human beings, not only the poor, not only poor countries, not only basic needs. Human development applies to the advanced, industrial countries, as much as to middle-income and low-income countries. The indicators are of course different though, alas, to John Kenneth Galbraith's complaint about private affluence amid public squalor has been added in many

advanced countries that of private affluence amid private squalor. A walk through the streets of New York or London provides plenty of evidence. But once nearly 100 percent literacy and average life expectancy of 78 years are reached, there is not much to distinguish one industrial country from another. Years of schooling have been included in the indicator for education as a differentiating characteristic between, say, Britain and the USA.

Sudhir Anand and Amartya Sen have suggested the division of all countries into three groups: low, medium and high levels of human development (see Anand and Sen, 1993, and 'Technical Notes 2 Human development index: a survey of recent reviews' in UNDP, 1993, pp. 104–14). For countries with a low value of human development the basic Human Development Index can be used to rank their performance. For countries with a medium value of human development Anand and Sen add one supplementary indicator to each of the three basic variables, life expectancy, literacy and log of gross domestic product (GDP) per head. In the longevity category they add infant and child mortality (under age five); in the education category they add secondary school enrollment; and in the income category they add the incidence of income poverty in the country.

For countries with a high level of human development they add a further supplementary indicator to the two already existing in each category in the medium group. To the survival (longevity) category they add the maternal mortality rate; to the education category they add tertiary enrollment; and to the income category they add Gini coefficient-corrected mean national income (that is, gross domestic product per head multiplied by $(1-G)$). The Table 1.1 illustrates the new additions.

Additional indicators of shortfalls from human development should be looked for elsewhere in the high human development countries: in homelessness, drug addiction and crime rates. Divorce rates and suicide rates are more controversial. They can be regarded as indicating more options and therefore positive achievements, particularly suicides of terminally ill elderly patients. On the other hand, they may be regarded as signs of the breakdown of the social fabric of a society, a failure of upholding what some regard as the moral values of the family or the sanctity of life.

A shorthand way of describing development is a variation of Abraham Lincoln's definition of government. It is development of the people, for the people, by the people. 'Of the people' implies adequate income generation through jobs, 'for the people' implies social services for those who need help, and 'by the people' means participation and democracy. It could also be interpreted as the economic, social and political dimensions of development.

The intellectual move from income to welfare or utility, to chosen bundles of goods and services, to characteristics of these goods and services, to needs that they meet, and finally to the enlargement of choices, has

Table 1.1 Indicators of human development

Human development level	Low	Medium	High			
Human Development Indicator	1.1	Life expectancy	1.1	Life expectancy	1.1	Life expectancy
			1.2	Under 5 mortality	1.2	Under 5 mortality
					1.3	Maternal mortality
	2.1	Adult literacy	2.1	Adult literacy	2.1	Adult literacy
			2.2	Secondary school enrollment	2.2	Secondary school enrollment
					2.3	Tertiary enrollment
	3.1	Log of GDP per head up to int poverty line	3.1	Log of GDP per head up to int poverty line	3.1	Log of GDP per head up to int poverty line
			3.2	Incidence of poverty	3.2	Incidence of poverty
					3.3	Gini coefficient corrected mean National Income

enriched our understanding. The enlargement of choices of one section should not be at the expense of the legitimate choices of another. This has two important implications: (1) in equity, so that one person's enlargement does not encroach on that of others; and (2) over time, so that our present choices do not encroach on the choices of future generations, or what has come to be known as 'sustainability'. This concern for the future should cover not only the physical environment – raw material exhaustion without technical substitution and pollution – but also resilience to outside shocks, debt and political sustainability.

Both equity and sustainability raise complex and difficult questions, not discussed here. 'To each according to his or her. . . .' Filling in the dots is highly controversial. Sustainability must refer to the constituents, not to the determinants of well-being. But maintaining the constituents of the well-being of future generations depends on population growth, and on changes in technology and in preferences, all uncertain.

A human development strategy stresses the importance of institutions for improving the human condition. Among these are not only the state, both as an agent to make markets work efficiently and to step in where they fail, and the market, but also the civil society: democratic political processes, the news media, non-governmental organizations, grassroots organizations, action groups and the public at large. It is in their interaction that the conditions for the good life should be found.

The Human Development Index: a political rallying point

The item in the UNDP's *Human Development Reports* (1990–2005) that has caught the public's eye and caused most controversy is perhaps analytically the weakest: it is the Human Development Index (more fully discussed below). It is clear that the concept of human development is much wider and richer than what can be caught in any index or set of indicators. This is true of other indicators, such as those of temperature. But, it might be asked, why try to catch a vector in a single number?

Yet, such indexes are useful in focusing attention and simplifying the problem. They have considerable political appeal. They have a stronger impact on the mind, draw public attention more powerfully, than a long list of many indicators, combined with a qualitative discussion. They are eye-catching. The strongest intellectual argument in their favor is that they show up the inadequacies of other indexes, such as GNP, and thereby contribute to an intellectual muscle therapy that helps us to avoid analytical cramps. They can serve as mental finger exercises. But it should be remembered that human development is a much richer concept than what can be caught in any index.

The Human Development Index comprises: (1) the logarithm of GDP per head, calculated at the real purchasing power, not at exchange rates, up to the international poverty line; (in subsequent *Reports* after that of 1990 this was modified in various ways); (2) literacy rates (and, since the 1991 *Report*, mean years of schooling); and (3) life expectancy at birth. These disparate items are brought to a common denominator by counting the distance between the best and worst performers and thereby achieving a ranking of countries. Critics have said that not only are the weights of the three components arbitrary, but also what is excluded, and what is included. Partha Dasgupta (2001) has pointed out that the index misrepresents concerns about the future, since it does not deduct capital depreciation; that it reflects only current well-being, and that it is an index only of human capital, leaving out natural capital. If these omissions are allowed for, what appears as a good human development performance turns out to be much worse.

As we have seen, one of the great drawbacks of average income per head is that it is an average that can conceal great inequalities. But, it may be

objected, the components of the Human Development Index (HDI), namely life expectancy and literacy, are also averages. They can conceal vast discrepancies between men and women, boys and girls, rich and poor, urban and rural residents, different ethnic or religious groups. The HDI has in fact been disaggregated by sex, region and ethnic groups for a few countries.

Another problem with the HDI is the implicit trade-off between life expectancy and income. For a country with an income per head less than the world average (\$5711 per year at 1993 purchasing power parity, which is about the income per head of Costa Rica) an increase of annual GDP per head of \$99 will exactly compensate for one year less of life expectancy, so as to keep the HDI constant.⁹ If the people in one poor country have one year less of life expectancy but \$100 higher GDP per head than in another country, this country will have a higher HDI. The value attached to longevity rises sharply with income. For a country with twice the average income (about the income per head of Malta), an extra year of life is valued at \$7482 in income per head. At three times the average (about the income in the United Kingdom) it is worth \$31631, about twice the country's income per head. At four times the average (about Switzerland's income) its value reaches \$65038, about three times actual income. The implication is that life is far less valuable in poor countries than in rich ones. The value judgments underlying these trade-offs have rightly been rejected. So 'human development' and the Human Development Index are not ultimate insights and other ideas will take their place. We are all free to guess what these will be.

There are, however, several reasons why human indicators are less misleading than income per head. First, the distribution of literacy and life expectancy is much less skewed than that of income. There is a maximum of 100 percent literacy. In spite of all the achievements of modern medicine, the maximum lifespan has not been extended so far, although there are some who predict that scientific progress will extend maximum life expectancy. Aubrey de Grey of the University of Cambridge predicts life expectancy in 2100 will be 5000 years. None of us will be around to check whether he is right (Nicholas D. Kristof, 2003). For income, on the other hand, the sky is the limit. A very few very high-income earners can raise the average. (The median or the mode would eliminate some of the distortions.)

Second, therefore, the average of the human indicators tells us something about the distribution. There cannot be high averages with too many people not participating. Since the non-poor have access to public services before the poor, reductions in infant mortality, and so on are indications of improvements for the poor. For life expectancy the average is actually better than a figure corrected for distribution between men and women. This is so

because the potential life expectancy of females is longer than that of males, if we start from the same life expectancy.¹⁰

Third, any upward move in a human indicator can be regarded as an improvement. Some might object if only the literacy of boys or the life expectancy of men is increased, but unless it can be shown that such increases worsen the fate of girls and women, by, for example, increasing the ability and desire to oppress, to object would smack of envy and dog-in-the-manger attitudes.

Fourth, whereas the high incomes of some can cause relative deprivation in others, this is not true for human indicators. If anything, the benefits in the health and education of anybody benefits the whole community.

Fifth, international income gaps, whether relative or absolute, may inevitably be widening, but to aim at reducing international gaps in human indicators is both sensible and feasible. In fact, looking at development in human terms presents a more cheerful picture than looking in income terms. Since 1960 average life expectancy has increased by 16 years, adult literacy by 40 percent, nutritional levels by over 40 percent, and child mortality rates have been halved. The international gap has closed. While average income per head in the South is 6 percent of that in the North, life expectancy is 80 percent, literacy 66 percent and nutrition 85 percent.

Sixth, human indicators show the troubles of overdevelopment or, better, mal-development, as well as of underdevelopment. Diseases of affluence can kill, just as the diseases of poverty can. Income, on the other hand, does not show up the destructive aspects of wealth.

Seventh, indicators that measure impact rather than inputs distinguish between goods and anti-bads (regrettable necessities) which bring us back to zero: unnecessary food requirements arising from unwanted pregnancies and feeding children that die; or from long walks to collect water and fuel; or from excess work or from long walks between unconsolidated plots or looking for work; for urban dwellers, high housing and transport costs.

Eighth, there is considerable political appeal in a simple indicator that identifies important objectives and contrasts them with other indicators.

A separate index covers aspects of human freedom and human rights, clearly an important aspect of human development. Life expectancy and literacy could be quite high in a well-managed prison. China shows remarkable progress on human development, but without political freedom.

Should the freedom index be integrated into the Human Development Index? There are some arguments in favor, but the balance of arguments is probably against. First, it might be said that freedom is so important (and, opportunity costs apart, costless) that no trade-off should be possible between its loss and gains in some of the other indicators.¹¹ Secondly, political conditions are much more volatile than changes in education and health.

Once a mother knows the importance of education for her children, or of hygienic behavior, this knowledge is not lost even when incomes drop. So human indicators tend to be fairly stable. Political indicators, on the other hand, can change overnight with a coup. A third argument against aggregating freedom with the positive aspects of human development is that grading is more subjective and less reliable than measuring life expectancy or literacy.

Finally, one of the most interesting questions is how freedom is related to human development more narrowly interpreted, or how negative and positive rights or freedom are associated. This can be done only if they are recorded by separate indexes, not components of the same.¹² Thus we might formulate a hypothesis that freedom, though not a necessary condition of human development, is entirely consistent with it even at quite low levels; and that human development, once it has reached a certain stage, leads inevitably to the call for freedom by the people. Here is a message of hope.

Notes

1. Sudhir Anand and Martin Ravallion (1993) criticize the basic need approach for being still 'firmly centered on commodity possession' (pp. 135–6).
2. Sen (1984, 1985, 1987).
3. Stewart (1993).
4. Keynes proposed the toast to the Royal Economic Society: 'to economics and economists, who are the trustees, not of civilization, but of the possibility of civilization'.
5. Brookner (1989).
6. In the film of Edith Wharton's *The Age of Innocence*, Newland Archer, seeing the futility of creating a soul mate of his wife, says, 'There's no point in liberating some one who does not realize she is not free.' Not everyone would agree. Against people's self-perception being above their real situation, Jodha (1988) found that people whose income had declined felt themselves better off by the criteria of independence (especially from patrons), mobility, security and self-respect.
7. Anand and Ravallion (1993).
8. Cobb et al. (1995).
9. Ravallion (1997).
10. See Anand and Sen (1993).
11. This objection could be mitigated by using a geometrical rather than an arithmetic average. With a zero weight for freedom, the total index becomes zero, however high the other components.
12. It could be said that the same argument applies to the relation between, for example, literacy and life expectancy, and that they should therefore not be lumped together in a single HDI.

References

- Anand, Sudhir and Martin Ravallion (1993), 'Human Development in Poor Countries: On the Role of Private Incomes and Public Services', *Journal of Economic Perspectives*, 7 (1), pp. 133–50.
- Anand, Sudhir and Amartya Sen (1993), 'Human Development Index: Methodology and Measurement', UNDP Occasional Papers No. 8.
- Brookner, Anita (1989), *Latecomers*, New York: Pantheon Books.
- Cobb, Clifford, Ted Halstead and Jonathan Rowe (1995), *The Genuine Progress Indicator*, San Francisco, CA: Redefining progress.

- Dasgupta, Partha (2001), *Human Well-being and the Natural Environment*, New York and Oxford: Oxford University Press.
- Jodha, N.S (1988), 'The poverty debate in India: A minority view', *Economic and Political Weekly*, Special Number, 23 November, pp. 2421–8.
- Keynes, John Maynard (1931), *Essays in Persuasion*, London: Macmillan, p. 176.
- Kristof, Nicholas D. (2003), 'Where is thy sting?', *New York Times*, 12 August.
- Minot, Susan (1992), *Folly*, New York: Washington Square Press.
- Ravallion, Martin (1997), 'Good and bad growth: The Human Development Reports', *World Development*, 25 (5), pp. 631–8.
- Sen, Amartya (1984), *Resources, Values and Development*, Oxford: Basil Blackwell, and Cambridge, MA: Harvard University Press.
- Sen, Amartya (1985), *Commodities and Capabilities*, Amsterdam: North-Holland.
- Sen, Amartya (1987), *The Standard of Living*, edited by G. Hawthorn, Cambridge: Cambridge University Press.
- Stewart, Frances (1993), 'Strategies to Fight Poverty', mimeo, revised draft, Oxford.
- Streeten, Paul with Javed Shahid Burki, Mahbub ul Haq, Norman Hicks and Frances Stewart (1981), *First Things First: Meeting Basic Human Needs in Developing Countries*, New York and Oxford: Oxford University Press.
- UNDP (1990–2005), *Human Development Reports*, New York and Oxford: Oxford University Press.
- UNDP (1993), *Human Development Report*, Oxford and New York: Oxford University Press.

2 Historical perspectives on development

*Amiya Kumar Bagchi*¹

Straightening the historical record

When and where did acceleration in rates of economic growth and human survival occur during the last 500 years? It was a maintained hypothesis of most mainstream economists (see, for example, Kuznets, 1960 [1965]) that the people of today's developing countries were already poorer than the Europeans when the Industrial Revolution began to transform the English economy. As a result of the work of Bairoch (1982) and eminent students of the economic history of China and India, such as Habib (1982), Raychaudhuri and Habib (1982), Li (1986, 1998), Frank (1998), Moosvi (2000), Pomeranz (2000) and Guha (2001), it is now clear that both China and India did better than the core countries of Western Europe in terms of human survival and growth of per capita incomes down to the mid-eighteenth century in the case of India, and down to the eve of the first Opium War, that is, 1840 or so in the case of China (Bagchi, 2004, 2005a). Population growth rates in India and China were higher than in the core countries of continental Western Europe (namely, England, France, Germany, Holland, Italy and Spain) between 1600 and 1750–1800 (Bagchi, 2005a, Chapter 5). We do not have reliable data on longevity for countries other than England and China before the nineteenth century. The available data on those two countries, with very unequal population numbers, indicate that before the middle of the eighteenth century the average Chinese citizen lived as long as his or her English counterpart (Bagchi, 2005a, Chapter 9).

Apart from the evidence on human survival, historians have also come to question the idea that the standard of living of ordinary people was higher in Europe than in India or China before the nineteenth century (Parthasarathi, 1998; Allen, 2004).

The evidence relating to economic growth and human development during the past 500 years throws up two axial ages, that is, two ages in which the countries were clearly separated into two groups by turning on an axis. The first was the period when the Industrial Revolution, originating in England, spread to continental Western Europe and the North Atlantic seaboard (and Australia and New Zealand), and thereby raised the rates of economic growth and changed the structure of national incomes and occupations in those lands from dominance by the primary to that by the secondary sector. This axial age started around 1760.

The second axial age can be dated from roughly the 1870s, when infant mortality rates began declining on a sustained basis in the industrializing countries and longevity went up beyond earlier historical records to 60 years and above, while the other countries continued to have high mortality rates and short lives.

The roots of the first axial age and, in particular, the origins of the industrial revolution in England and its diffusion in the nineteenth century to continental Western Europe and the USA (followed by Canada, Australia and Japan) have caused and continue to elicit scholarly controversy. The old imperialist idea, revived by North and Thomas (1973), Jones (1981 [1987]) and Landes (1998) that Europeans were always (or at least going back to, say, 1000 CE) different from and better off than Asians (who constituted the majority of the global population), has not stood up to serious scrutiny. The related idea that it was the protection of private property rights and freedom of trade and exchange that made Europe distinctive has also failed to meet the test of evidence. China and India had been both engaged in intensive intra-regional and intercontinental trade for centuries when, in 1498, Vasco da Gama navigated the route round the southern Cape of Africa to India. Moreover, property rights were as protected in China and India as in most of Europe down to the time of the establishment of European hegemony over these two most populous countries of the world (Bagchi, 2005a).

What then led to the ‘great divergence’ (Pomeranz, 2000) between the fortunes of today’s developed countries and the rest of the world? Undoubtedly, the Industrial Revolution – with its three major characteristics, namely, the exploitation of economies of scale, the introduction of progressively labour-saving production methods and the upsurge in the use of non-renewable resources of materials and energy – was the decisive process separating the rich and the poor nations of today. Following Wong (1997), we can distinguish between ‘Smithian growth’ in a market economy and growth in a fully capitalist economy. In the former, goods were produced for the market, occupations were specialized with a considerable degree of division of labour, there were dense networks of exchange supported by flows of money and credit; but property rights were regulated by law, regulations and custom, so nobody could accumulate unlimited amounts of capital or dispossess producers through the control of property. The commercialized economies of Western Europe, the lands bordering the Mediterranean, India, China, South-East Asia and Japan before the Meiji Restoration (in 1868) were all economies characterized by Smithian growth. However, China and India, with their diversifying agriculture and specializing manufactures in response to the growth of long-distance of trade and inflows of precious metals from Japan and the Americas were the

foremost exemplars of Smithian growth. The Indian system of manufacture for the market spreading out to the country has been likened to what was called 'proto-industrialization' in the context of Flanders, northern France, western Germany and England at around the same period (Mendels, 1972; Kriedte et al., 1981). In 1750, China produced 32.8 per cent and India produced 24.5 per cent of the global manufacturing output; by 1913, as a result of the working of imperialism and the industrialization of the leading capitalist economies, those shares had gone down to 3.2 per cent in the case of China and 1.4 per cent in the case of India (Bairoch, 1982; Simmons, 1985). Even with China being the fastest-growing economy through the 1990s and beyond, the proportion of China's manufacturing output to the world manufacturing output in 2003 was only 11.5 per cent and that of the Indian subcontinent (consisting of Bangladesh, India and Pakistan) was about 2.5 per cent (derived from World Bank, 2005). So the divergence still persists.

Neither proto-industrialization nor Smithian growth as such produced the Industrial Revolution and led to the great divergence between the earlier leaders of the world economy and the North Atlantic seaboard nations. The bourgeoisie needed to capture political power before they could dispossess the producers and freely use any natural resources they could claim as their property. These steps led to resource-intensive industrialization in which workers were subjected to the discipline of the machine. The bourgeoisie came to control the levers of state power first in the Netherlands and then in England. Capitalist relations as such did not allow the eruption of the Industrial Revolution in the Netherlands: it became too specialized in entrepôt trade and could not nurture 'infant industries', it became too greatly urbanized before the knowledge of prophylaxis and urban sanitation had spread. Moreover, it suffered a demographic crisis caused by an unsanitary urban environment, and losses of manpower in wars and voyages. In any case, the domestic market was too small to support the kind of economies of scale that became associated with the Industrial Revolution (Van Zanden, 1993; De Vries and Van der Woude, 1997; Bagchi, 2005a, Chapter 6).

In England, the enclosure movement created a proletariat even if it did little to raise agricultural productivity (Allen, 1992, 1994), and property-holders were allowed to control natural resources such as water sources, coal and iron ore with few restrictions. Moreover, infant industries such as cotton textiles were protected against foreign manufactures, for example, Indian calicos, even as old industries such as woollens and agriculture continued to enjoy protection. Meanwhile, the demand from the navy and the army created a market for more and more specialized and lethal guns and cannon, and set up a competition for machines that would work well with

only some training and would not be dependent on manual dexterity cultivated for generations. As capital and knowledge accumulated, they provided the base for further accumulation of capital, technology and skill, as Tucker (1774) had emphasized even before Smith's canonical treatise (Smith, 1776 [1976]) swept the field of political economy.

The Industrial Revolution can be seen as a process of cumulative causation. It is driven by static economies, learning by doing and learning through competitive emulation, capital accumulation, adaptation of complementary and competitive production and organizational structures, and innovations to overcome imbalances created by new vintages of equipment and organizational changes. These factors can be reinforced by network effects. The more a particular type of equipment or a particular way of doing business is adopted by a number of firms or users, the cheaper it becomes for the producer of that kind of equipment or the practitioner of that style of business to produce that equipment or intensify that particular style of doing business. The combination of dynamic economies of scale and network effects can then generate a path-dependence in production structures, consumption patterns and business organization and business behaviour (David, 1985; Arthur, 1989).

The Industrial Revolution had as its background a revolution in science and an atmosphere of readiness to apply useful knowledge for practical applications (Mokyr, 2005). But that revolution was no older than the seventeenth century, nor did it at once lead to the technological innovations that gave England and the follower countries of the North Atlantic seaboard a decisive advantage over India and China, the older manufacturing centres of the world. The predilection for measuring and quantifying has been seen as a peculiarly European virtue going back to the thirteenth century CE (Crosby, 1997). But again, even if it was, it did not confer its benefits for another 500 years in pushing the European economies to the frontier of technological advance. Finally, Francis Bacon, writing before the 1620s, has been seen to stand at the fountainhead of the conception of science as knowledge that has to be tested against empirical evidence and its practical utility. He also projected that innovations would radically change the way things worked and people lived (1986 [1627]). But again, it was another 100 years before innovations in civilian technology began to affect major areas of production in England. Bacon's imagined innovations also included weapons of war. Technological and organizational innovations associated with armed combat in Europe had a major role not only in showing the way towards setting up large-scale factories but also in enabling the early capitalist nations to aggrandize themselves by grabbing the resources, including labour power, of other countries.

In both the axial ages, the forging ahead of the Western European lands was facilitated by armed conquest and the establishment of European imperial hegemony over the rest of the world, and in particular on the two most populous countries, namely, China and India. The adverse impact of British rule on India was well known to many early political economists such as Lord Lauderdale, but it went into oblivion through the writings of propagandists such as James Mill (Bagchi, 1996). India and China had been the two greatest manufacturing nations of the world down to 1750 (Bairoch, 1982). India went into a phase of massive deindustrialization, when in some of the core regions producing manufactures with artisanal methods and exporting them worldwide, the proportion of the working force engaged in secondary industry declined from somewhere around 20 per cent to less than 10 per cent (Dutt, 1904 [1963]; Bagchi, 1976; Tilly, 1994; Clingingsmith and Williamson, 2004). Similar, though perhaps less catastrophic changes took place in most of the colonial economies and the newly independent countries of Latin America in the nineteenth century (Bagchi, 1982, Chapters 3 and 4).

Several misconceptions still cloud the analysis of the impact of imperialism on the colonial or semi-colonial countries. The first is to say that the disrupting effect of the invasion of the domestic market by machine-made manufactures was unstoppable. In fact, all the countries of Europe and the overseas settlements of the USA, Canada and Australia adopted measures to protect the domestic market against foreign manufactures and encourage the growth of domestic machine-driven industries (Sabel and Zeitlin, 1985). Moreover, contrary to conventional wisdom, industrial growth was faster in the industrializing countries in the age of protection than in earlier periods (Bairoch, 1993; O'Rourke, 2000). The non-white dependencies of Europe were prevented from adopting any similar measures. The newly liberated countries of Latin America could have adopted such measures. But the British support for their struggle against Spain and Portugal had been explicitly or implicitly conditional on their adopting a free trade policy – a policy that Britain did not adopt until the 1840s, two decades after the liberation of Latin America. Moreover, the liberation entrenched landlords generally exploiting unfree labour and abundant land as the ruling class of Latin America. That class saw its future as collaborators of the industrializing countries which would buy the products of their *latifundia*, rather than in industrializing their own countries. The depression of the 1930s forced these rulers to adopt the policy of import-substituting industrialization (Bagchi, 1982, Chapter 3), but the landlord-dominated social structure hobbled that effort.

Deindustrialization in countries like India was not compensated by vibrant agricultural growth either (*pace* Clingingsmith and Williamson,

2004). In fact, because of the disruption of earlier methods of irrigation and crop rotation, the extraction of a large surplus from India without any return and the consequent depression of domestic demand and, finally, the entrenchment of a landlord class as intermediaries in the tribute-extracting enterprise of British rulers, there was little growth in colonial India's agricultural productivity (Bagchi, 2005a, Chapter 10). There were devastating famines in China and India throughout the greater part of the nineteenth century. In the half-century before independence, the per capita agricultural output in India declined (*ibid.*).

Most of the economies of ex-colonial countries in Asia, Africa and Latin America began to grow again at a positive rate from the 1950s after their liberation from colonial rule or imperial domination. But their growth rates began to falter from the late 1970s largely under the onslaught of structural adjustment policies and financial liberalization forced on them by the International Monetary Fund, the World Bank and the rich capitalist countries led by the USA, often acting in collusion with rulers of the countries that had become indebted to the transnational banks through the corruption and profligacy of the same rulers.

Contrary to the impression conveyed by mainstream economists, the performance of the so-called developing countries was better in the period 1950–80 when these countries practised a state-promoted path of development than when they were forced to adopt a policy of freeing capital from all restraint, while depriving workers of the few social security benefits they had enjoyed earlier (Weisbrot et al., 2001; Bagchi, 2005a, Chapters 23–24). The countries or city-states of East Asia, such as Taiwan, South Korea, Singapore, Hong Kong and the People's Republic of China, bucked this trend: they were free of landlordism, their governments tried to universalize education, starting with elementary education and moving up towards higher levels, they practiced extensive state patronage and accessed foreign markets as a means of enabling their own firms to reap the benefits of specialization, acquiring new technologies and utilizing the economies of scale and scope. China, building on its socialist foundations but progressively increasing incentives for the producers and promoting competition among domestic enterprises by pushing them to adopt innovations on a country-wide scale and using her massive investments and continually improving industrial technologies, has become the fastest-growing economy of the world.

Economic development has any meaning only if it also leads to the development of human capabilities. As we have noted above, the surge in economic growth and the sustained development of longevity in industrializing Europe were separated by a century. In terms of political freedom also, outside the core countries of Britain and Scandinavia, most other

European members of the Organisation for Economic Co-operation and Development had to wait until the end of World War II before establishing formal democracy. The benefits of modern medicine began to diffuse to the developing countries after their independence. But from 1980 there has been a slowdown, if not a reversal in the decline in infant and adult mortality rates in some of the poorest countries of the world (Weisbrot et al., 2001; Bagchi, 2004; 2005a, Chapters 23–24; Deaton, 2005).

The world population of roughly 6.32 billion is now polarized between a small minority of high-income countries (with 948.3 plus million people), a group of middle-income countries with a population of 2748.6 million and a group of low-income countries with a population of 2614.5 million.² However, many of the middle-income countries depended on the export of oil for their income and saw their incomes falling over 1990–2003, as did many of the ex-Soviet countries of Eastern Europe and Central Asia (UNDP, 2005, ‘Human Development Indicators’, Table 14). Moreover, levels of mortality and longevity moved in an adverse direction in many ex-Soviet and sub-Saharan countries (UNDP, 2005, Chapter 1).

On the other hand, the experience of China and the industrialized regions of East Asia indicates that with appropriate changes in social structure, it is possible to close the gap rapidly with the more affluent countries in respect of most indicators of human development, and to use that higher level of development to try and narrow the gap in levels of income.

Thinking about economic development

In Europe, since the sixteenth century, publicists, merchants and advisors to governments had been putting forward their views about how to increase the wealth of the princes and countries they were concerned with. But the *locus classicus* of thinking about economic development was Adam Smith’s *Wealth of Nations* (Smith, 1776 [1976]). In this book, Smith put forward his view that economic development is driven by the division of labour and expanding markets in an economy which is free of unnecessary restraints on trade, and the freedom of economic agents to choose their professions and the fields in which to invest their capital. Smith’s analysis was historically grounded and he stressed institutional changes as much as market forces as the factors driving economic growth. Smith’s view that economic development is driven by capitalist accumulation was sharpened further by David Ricardo (1817 [1951]). He located profit as the basic income share out of which accumulation would be financed. In his scheme, there is a clear conflict between the interests of landowners and capitalists because, through the operation of diminishing returns, accumulation drives up the price of food grains, raises rents and lowers the profit share. Ricardo’s theory of comparative cost driving international trade and his theory of

profit-driven accumulation provided the arguments for abolishing the protection of British agriculture, since that would cheapen corn and increase the incentive and wherewithal for further accumulation. His free trade arguments also supplied the rationale for England's export-led growth of manufactures and capitalist development. But Ricardo is also the originator of economic analysis that is detached from particular historical contexts, and can therefore be regarded as the pioneer of ahistorical modes of analysis of economic growth.

Karl Marx (1867 [1886]) further developed the theory of capitalist accumulation by demonstrating that the conflict between an increasing wage share and capitalist accumulation would lead to class struggles over employment and wages, periodic cycles, and a long-term tendency of capitalists searching for and introducing labour-saving innovations. Marx's analysis was firmly grounded in the experience of the British Industrial Revolution and he saw capitalism as being driven by a continual need for expansion. Lenin (1899 [1964]) took up the theme of the continually expansive tendency of development under capitalism and introduced the idea of uneven development that has been used strategically by later analysts to explore the unevenness of development between nations, between regions, and between lagging and leading sectors of an economy.

The problem of economic development almost vanished from economists' discourse after the so-called marginalist revolution of the 1870s. Prodded by worries about what could be done in a war-devastated Eastern and South-Eastern Europe, and the soon-to-be-independent colonies of European powers, economic development entered into official and academic discourse again only from the 1940s. Rosenstein-Rodan (1943 [1958]) was a trailblazer in this direction, to be followed by a number of other writers trying to combine the lessons of Soviet industrialization with national income accounting that came into wide use after the Keynesian revolution (see, for example, Datta, 1952). Most of them agreed that the imperially imposed division of labour under which the underdeveloped countries were to specialize in agricultural commodities with low income elasticities of demand had to be overturned and a vigorous programme of industrialization had to be taken in hand if the poverty of these newly independent nations was to be seriously dented. The Prebisch–Singer thesis that the terms of trade of primary producers *vis-à-vis* the industrialized nations had been on a downward trend for most of the twentieth century added vigour to the industrializers' argument (Prebisch, 1950; Singer, 1950).

In mainstream economics, the analysis of economic development since the 1950s followed two parallel and mainly non-intersecting paths. The so-called neoclassical theories of growth associated with the names of Robert

Solow and Trevor Swan ended up with huge residuals that had to be explained away in an ad hoc fashion (for a summary, see Romer, 2006, Chapters 1–2). Moreover, they flew in the face of the Keynes–Kalecki theorem that in a market economy it is investment that drives growth, and saving is equated to investment through the working of the multiplier and changes in the relative shares of wages and profit. They also had nothing to say about the problem of underdevelopment of the economies that provided livelihoods to the majority of the global population. The new or endogenous growth theory in the neoclassical tradition makes a mystified use of the notion of human capital and totally fails to take account of contradictions in the growth process such as demand failures and unequal development between different countries and regions (for a summary of the theory, see Romer, 2006, Chapter 3; for a critique, see Bhaduri, 2006).

An alternative modelling trajectory was pursued by analysts who took the drives and institutions of capitalism seriously. This was to follow the path blazed by Kalecki in integrating the behaviour of a monopolistically competitive capitalist class with that of a working class struggling to maintain their real wages. The bargaining power of the working class waxed and waned as the labour market became tighter or slacker. Kalecki envisaged the possibility that business cycles might be caused by the ruling class in a capitalist economy in order to beat down the workers (Kalecki, 1971). Using a slightly different modelling strategy, Goodwin (1967 [1982]) treated the capitalist–worker relationship as a predator–prey interaction that generated a growth cycle.

Kaldor (1957) departed drastically from neoclassical models by treating accumulation of capital and the rate of technical progress as being organically linked. Here he was building on the work of Adam Smith, Allyn Young, Gunnar Myrdal and Josiah Tucker, who recognized the role of both static and dynamic economies of scale in raising productivity through increased division of labour and economies of agglomeration (Bagchi, 1998).

Following a Marx–Kalecki tradition and incorporating the Prebisch–Singer hypothesis that there is a basic asymmetry between the demand pattern of an agrarian and an industrialized economy – namely, the income elasticity of demand for primary products is generally less than one whereas the corresponding value is larger than one, especially for new industrial products – other economists have built up models of growth that generate systematic differentials between the industrialized North and the agrarian South (for a full-length treatment, see Dutt, 1990). The work of Steindl (1952) and Schumpeter, 1911/1934, 1942) has been used to illuminate processes of intra-capitalist competition and innovations arising out of, and in turn driving that competition (Bloch, 2000). Bhaduri (2006) has

built a model of endogenous growth in the Marx–Keynes–Kalecki tradition. In his model, capitalist growth is driven by intra-capitalist competition that tends to drive down the prices of commodities, and class struggles between capitalists and workers determine the share of profit in the aggregate output and thereby lead to fluctuations in the rate of investment. Intra-capitalist competition and the search for labour-saving innovations lead to growth in labour productivity, but the shares of wages and profits are all the time affected by the state of the class struggle. Along with these models, we need to incorporate the tendencies towards greater inequality and concentration of economic power that have resulted from financial deregulation and the exercise of naked military and political power of the USA and its allies since the fall of the Soviet bloc in 1989.

Kalecki was responsible for introducing the notion of the political business cycle that is generated by the deliberate political strategies of the capitalist class. When wages shoot up beyond levels considered tolerable by capitalists, individual capitalists and the capitalist class slow down investment and thereby generate higher levels of unemployment. That leads to a decline in the share of wages, a rise in rates of profit and a resumption of higher rates of investment. In understanding the actual working of the world economy, this political business cycle in the metropolitan core of the world economy has to be combined with the built-in tendency towards underdevelopment of the poor agrarian economies which have not undergone a social transformation of the kind that had happened in Western Europe after the French Revolution and in the East Asian economies from the late 1940s. The external pressure of imperialist forces and the internal working of a class structure in the unreformed agrarian societies – dominated as they are by landlords, speculative capital and other close collaborators of imperialism – ensure the continuation of underdevelopment (Baran, 1952 [1958]; Bagchi, 1982).

A number of economic historians (including some whose names have been already mentioned) tried to conceptualize, in their own ways, the major factors that catapulted the Western European economies and their overseas offshoots to the top ranks of global economic prosperity. Dobb (1946, 1951) stressed the changes in class structure since the Middle Ages of Europe that led to the growth of capitalism, and combined the insights thus gleaned with his studies of Soviet economic development to analyse the prospects of, and strategies for, lifting the ex-colonial countries from the slough of underdevelopment. As a challenge to Marxist modes of analysis, stressing changes in class structure and international differences in economic and military power, Rostow (1960) proposed a five-stage theory of economic growth consisting of: (1) traditional society; (2) the preconditions for take-off; (3) the take-off; (4) the drive to maturity; and

(5) the age of high mass-consumption. Rostow's theory was both Eurocentric and diffusionist and portrayed economic growth as a trickle-down process, involving no international contradictions. It failed to meet the test of historical plausibility, even as far as European lands were concerned (see, for example, the studies brought together in Rostow, 1963). On the basis of the experience of economic growth in Germany and Tsarist Russia, Gerschenkron (1952) put forward the hypothesis that countries that have fallen behind others in levels of economic prosperity forge new instruments and adopt a different strategy of industrialization to pull themselves forward. For example, they try and develop new instruments of long-term investment and foster capital goods industries ahead of demand, instead of passively following the market. However, this generalization would be valid for all major areas of Western Europe from the sixteenth century, as military and economic competition became fiercer. States often patronized schemes of partial industrialization, and new credit delivery systems were forged (Gille, 1973; Supple, 1973). But such projects of partial transformation did not lift either the numerous German principalities or eighteenth-century Russia, Austria or, for that matter, pre-revolutionary France out of their industrial backwardness compared with England. The state's capacity was limited by the social milieu in which it operated. French innovations in banking, technical education and state patronage for communication succeeded in raising French per capita income level to that of Britain after the *ancien régime* social structure had been transformed by the 1789 revolution (Cameron, 1961). Alexander Hamilton's pioneering prescriptions for protecting infant industries against the competition of British manufactures bore fruit in the USA because it was virtually free of all feudal institutional encumbrances (Hamilton, 1791; for a formalization of the idea that the domestic government might aid local producers against foreign firms enjoying monopoly power, see Bhattacharjea, 2002). Although the discourse of institutional innovations and their relative efficacy in different historical contexts was generated by the analysis of capitalist competition in Europe, it also throws light on successful industrialization in other parts of the world. For instance, Japan started its industrializing career by providing strong patronage to those industries that would increase its economic and political power. But the real acceleration of Japanese growth occurred only after World War II when peasants were finally freed of the incubus of landlord power. It also throws light on why apparently similar institutional innovations succeeded far more in the leading countries of East Asia than in most other parts of the developing world (Bagchi, 2005b).

The measurement of economic growth has attracted renewed attention from mainstream economists and economic historians (Maddison, 1991;

Barro and Sala-i-Martin, 1995). Their findings indicate that levels of economic as well as human development, as measured by the United Nations Development Programme, have converged among the countries of Western, Northern and Southern Europe, the USA, Australia, New Zealand and Japan. Among the factors mentioned have been a much greater two-way mobility of capital and diffusion of new technologies among these countries as against countries outside the club. However, it is rarely pointed out that the major countries of this group had been imperialist until and, in some cases, beyond World War II, and even those without colonies of their own benefited from a racialism that treated only white-settled countries as deserving of equal treatment in economic exchanges and political negotiations. Their processes of capital accumulation, knowledge acquisition and expansion of markets and sources of raw materials had benefited from the extraction of tributes and monopoly profits from the formal and informal colonies. The latter had been correspondingly impoverished and their social structures had been distorted towards favouring the collaborators of the imperial rulers (Baran, 1952 [1958]; Bagchi, 1982).

Many economists have only recently been sensitized to the fact that development involves many factors other than income as conventionally measured, and that among the causes of economic and human development are many influences that are deeply intertwined with patterns of living, patterns of social interaction and modes of governance in the economic, social and political spheres, and that the latter in turn are often largely shaped by geography, history and ecology. The Annales School of French historians have, however, taken on board all these influences in their practice for the last three-quarters of a century and helped both deepen the understanding of European history and widen the perspective of many historians to the world beyond Europe or Eurasia (outstanding examples of this kind of work can be found in Bloch, 1961 and Braudel, 1981–84). Economists' attention has also been directed in recent years towards issues of nutrition and human growth both as the result of economic growth and as influencing productivity in its turn. But anthropologists and specialists have been studying such issues for more than half a century, and it is important for social scientists to take account of this parallel work (for a good account of the discipline of auxology by a pioneer, see Tanner, 1981; see also Bagchi, 2005a, Chapters 1 and 22).

A final word of caution is needed about the sustainability of the current pattern of growth. Industrialization has been highly resource-intensive. An abnormally large percentage of the non-renewable resources of the world is concentrated in the hands of the rich in G7 countries. The wars in Afghanistan and Iraq, waged by the USA and its allies, show that to keep

control of those resources, they are prepared to adopt measures in violation of all international law and conventions. For the future of world peace and sustainable development, it is necessary that this resource-addiction of economic growth be mitigated by the widespread use of renewable energy resources and resisting the attempted monopolization of new technology and non-renewable resources by a few transnational corporations domiciled in G7 countries (Bagchi, 2005a; Klare, 2005).

Notes

1. The author is indebted to the editors for comments that helped improve the chapter. They are, of course, exculpated from all responsibility for errors.
2. The cut-off per capita income dividing high- from middle-income countries is US\$9386 or more and the cut-off point for low-income countries is \$765 or less, as given by the World Bank.

References

- Allen, R.C. (1992), *Enclosure and the Yeoman: The Agricultural Development of the South Midlands 1450–1850*, Oxford: Clarendon.
- Allen, R.C. (1994), 'Agriculture During the Industrial Revolution', in R. Floud and D. McCloskey (eds), *The Economic History of England since 1700, vol. 1: 1700–1860*, Cambridge: Cambridge University Press, pp. 96–122.
- Allen, R.C. (2004), *Mr Lockyer Meets the Index Number Problem: The Standard of Living in Canton and London in 1704*, Nuffield College, Oxford.
- Arthur, B. (1989), 'Competing technologies, increasing returns, and lock-in by historical events', *Economic Journal*, 99: 116–31.
- Bacon, Francis (1986), *The Great Instauration and New Atlantis*, ed. J. Weinberger, rev. edn with translation of Francis Bacon (1627), *Instauration Magna*, Arlington Heights, IL: Harlan Davidson.
- Bagchi, A.K. (1976), 'De-industrialization in India in the Nineteenth Century: Their Theoretical Implications', *Journal of Development Studies*, 12 (2): 135–64.
- Bagchi, A.K. (1982), *The Political Economy of Underdevelopment*, Cambridge: Cambridge University Press.
- Bagchi, A.K. (1996), 'Colonialism in Classical Political Economy: Analysis, Epistemological Broadening and Mystification', *Studies in History*, 12 (1): 105–36.
- Bagchi, A.K. (1998), 'Development', in H.D. Kurz and N. Salvadori (eds), *The Elgar Companion to Classical Economics*, Cheltenham, UK and Lyme, USA: Edward Elgar, pp. 230–35.
- Bagchi, A.K. (2004), 'The Axial Ages of the Capitalist World-System', *Review*, 27 (2): 93–134.
- Bagchi, A.K. (2005a), *Perilous Passage: Mankind and the Global Ascendancy of Capital*, Lanham, MD, Rowman & Littlefield.
- Bagchi, A.K. (2005b), *The Developmental State in History and in the Twentieth Century*, New Delhi: Regency Publications.
- Bairoch, P. (1982), 'International Industrialization Levels from 1750 to 1980', *Journal of European Economic History*, 11 (2): 269–334.
- Bairoch, P. (1993), *Economics and World History: Myths and Paradoxes*, London: Harvester Wheatsheaf.
- Baran, P.A. (1952), 'On the Political Economy of Backwardness', *Manchester School*, January; reprinted in A.N. Agarwala and S.P. Singh (eds) (1958), *The Economics of Underdevelopment*, Delhi: Oxford University Press, pp. 75–132.
- Barro, R.J. and X. Sala-i-Martin (1995), *Economic Growth*, New York: McGraw-Hill.
- Bhaduri, Amit (2006), 'Endogenous Economic Growth: A New Approach', *Cambridge Journal of Economics*, 30 (1): 69–84.

- Bhattacharjya, Aditya (2002), 'Infant Industry Protection Revisited: Entry Deterrence and Entry Promotion when a Foreign Monopolist has Unknown Costs', *International Economic Journal*, 16 (3): 115–33.
- Bloch, H. (2000), 'Schumpeter and Steindl on the Dynamics of Competition', *Journal of Evolutionary Economics*, 10: 343–53.
- Bloch, Marc (1961), *Feudal Society*, Vols 1–2, trans. from the French, London: Routledge.
- Braudel, F. (1981–84), *Civilization and Capitalism 15th–18th Century*, Vols 1–3, trans. from the French, London: Collins.
- Cameron, R.E. (1961), *France and the Economic Development of Europe 1800–1914: Conquests of Peace and Seeds of War*, Princeton, NJ: Princeton University Press.
- Clingingsmith, David and Jeffrey G. Williamson (2004), 'India's De-industrialization under British Rule: New Ideas, New Evidence', NBER Working Paper 10586, Cambridge, MA: National Bureau of Economic Research.
- Crosby, A.W. (1997), *The Measure of Reality: Quantification and Western Society, 1250–1600*, Cambridge: Cambridge University Press.
- Datta, Bhabatosh (1952), *The Economics of Industrialization*, Calcutta: World Press.
- David, P.A. (1985), 'Clio and the economics of QWERTY', *American Review*, 75: 332–37.
- Deaton, A.S. (2005), 'The Great Escape: A Review Essay on Fogel's "The Escape from Hunger and Premature Death, 1700–2100"', NBER Working Paper No. W11308, Cambridge, MA: National Bureau of Economic Research.
- De Vries, J. and A. Van der Woude (1997), *The First Modern Economy: Success, Failure, and Perseverance of the Dutch Economy, 1500–1815*, Cambridge: Cambridge University Press.
- Dobb, M. (1946), *Studies in the Development of Capitalism*, London: Routledge and Kegan Paul.
- Dobb, M. (1951), *Some Aspects of Economic Development*, Delhi: Delhi School of Economics.
- Dutt, A.K. (1990), *Growth, Distribution and Uneven Development*, Cambridge: Cambridge University Press.
- Dutt, R.C. (1904), *The Economic History of India, Vol. 1, Under Early British Rule*, London; reprinted (1963), Delhi: Manager of Publications, Government of India.
- Frank, A.G. (1998), *ReOrient: Global Economy in the Asian Age*, Berkeley, CA: University of California Press.
- Gerschenkron, A. (1952), 'Economic Backwardness in Historical Perspective', in B. Hoselitz (ed.), *The Progress of Underdeveloped Countries*; reprinted in A. Gerschenkron (1962), *Economic Backwardness in Historical Perspective*, Cambridge, MA: Harvard University Press, pp. 5–30.
- Gille, B. (1973), 'Banking and Industrialization in Europe 1730–1914', in C.M. Cipolla (ed.), *The Fontana Economic History of Europe: The Industrial Revolution*, London: Collins/Fontana, pp. 255–300.
- Goodwin, R.M. (1967), 'A Growth Cycle', in C.H. Feinstein (ed.), *Socialism, Capitalism and Economic Growth: Essays in Honour of Maurice Dobb*, Cambridge: Cambridge University Press; reprinted in R.M. Goodwin (1982), *Essays in Economic Dynamics*, London: Macmillan, pp. 163–70.
- Guha, S. (2001), *Health and Population in South Asia: From Earliest Times to the Present*, New Delhi: Permanent Black.
- Habib, I. (1982), 'Population', in T. Raychaudhuri and I. Habib (eds), *The Cambridge Economic History of India, Vol. I, c. 1200–1750*, Cambridge: Cambridge University Press, pp. 163–71.
- Hamilton, Alexander (1791), 'A Report on the Manufactures of the United States', www.juniversity.com/i_documents/ah_rom.htm, accessed 5 January 2005.
- Hobsbawm, E.J. (1980), 'Scottish Reformers of the Eighteenth Century and Capitalist Agriculture', in E.J. Hobsbawm, W. Kula, A. Mitra, K.N. Raj and I. Sachs (eds), *Peasants in History: Essays in Honour of Daniel Thorner*, Calcutta: Oxford University Press, pp. 3–29.
- Jones, E.L. (1981), *The European Miracle: Environments, Economies and Geopolitics in the History of Europe and Asia*; 2nd edn (1987), Cambridge: Cambridge University Press.
- Kaldor, N. (1957), 'A Model of Economic Growth', *Economic Journal*; reprinted in N. Kaldor (1960), *Essays on Economic Stability and Growth*, London: Duckworth, pp. 259–300.

- Kalecki, Michal (1971), *Selected Essays on the Dynamics of the Capitalist Economy, 1933–1970*, Cambridge: Cambridge University Press.
- Klare, M. (2005), ‘The Intensifying Global Struggle for Energy’, *TomDispatch.com*, 9 May.
- Kriedte, P., H. Medick and J. Schlumbohm (1981), *Industrialization before Industrialization*, Cambridge: Cambridge University Press.
- Kuznets, S. (1960), ‘Present Underdeveloped Countries and Past Growth Patterns’, in E. Nelson (ed.), *Economic Growth: Rationale, Problems and Cases*, Austin, TX: University of Texas Press; reprinted in S. Kuznets (1965), *Economic Growth and Structure: Selected Essays*, New York: W.W. Norton, pp. 176–93.
- Kuznets, S. (1965), *Economic Growth and Structure: Selected Essays*, New York: W.W. Norton.
- Landes, D.S. (1998), *The Wealth and Poverty of Nations: Why Some are so Rich and Some so Poor*, London: Little, Brown & Co.
- Lenin, V.I. (1899), *The Development of Capitalism in Russia*; translated into English and reprinted in V.I. Lenin (1964), *Collected Works*, Vol. 3, Moscow: Progress Publishers.
- Li, Bozhong (1986), *The Development of Agriculture and Industry in Jiangnan, 1644–1850: Trends and Prospects*, Hangzhou, Zhejiang, China.
- Li, Bozhong (1998), *Agricultural Development in Jiangnan, 1620–1850*, London: Macmillan.
- Maddison, A. (1991), *Dynamic Forces in Capitalist Development*, Oxford: Oxford University Press.
- Marx, Karl (1867/1886), *Capital, Vol. 1, The Process of Capitalist Production*, trans. from the German by S. Moore and E. Aveling; reprinted (1965), Moscow: Progress Publishers.
- Mendels, Franklin F. (1972), ‘Proto-Industrialization: The First Phase of the Industrialization Process’, *Journal of Economic History*, 31 (1): 241–61.
- Mokyr, J. (2005), ‘The Intellectual Origins of Modern Economic Growth’, *Journal of Economic History*, 65 (2): 285–351.
- Moosvi, Shireen (2000), ‘The Indian Economic Experience: 1600–1900: A Quantitative Study’, in K.N. Panikkar, T.J. Byres and U. Patnaik (eds), *The Making of History: Essays Presented to Irfan Habib*, New Delhi: Tulika, pp. 328–58.
- North, D.C. and R.P. Thomas (1973), *The Rise of the Western World: A New Economic History*, Cambridge: Cambridge University Press.
- O’Rourke, K.H. (2000), ‘Tariffs and Growth in the Late Nineteenth Century’, *Economic Journal*, 110: 456–83.
- Parthasarathi, P. (1998), ‘Rethinking Wages and Competitiveness in the Eighteenth-Century: Britain and South India’, *Past and Present*, 158: 79–109.
- Pomeranz, K. (2000), *The Great Divergence: China, Europe and the Making of the Modern World Economy*, Princeton, NJ: Princeton University Press.
- Prebisch, R. (1950), *The Economic Development of Latin America and its Principal Problems*, New York: United Nations.
- Raychaudhuri, T. and I. Habib (eds) (1982), *The Cambridge Economic History of India, Vol. 1, c.1200–c.1750*, Cambridge: Cambridge University Press.
- Ricardo, David (1817), *On the Principles of Political Economy and Taxation*; reprinted in P. Sraffa and M.H. Dobb (eds) (1951), *The Works and Correspondence of David Ricardo, Vol. 1*, Cambridge: Cambridge University Press.
- Romer, D. (2006), *Advanced Macroeconomics*, 3rd edn, New York: McGraw-Hill Irwin.
- Rosenstein-Rodan, P.N. (1943), ‘Problems of Industrialization of Eastern and South-Eastern Europe’, *Economic Journal*, 53, reprinted in A.N. Agarwala and S.P. Singh (eds) (1958), *The Economics of Underdevelopment*, Delhi: Oxford University Press, pp. 245–55.
- Rostow, W.W. (1960), *The Stages of Economic Growth: A Non-Communist Manifesto*, Cambridge: Cambridge University Press.
- Rostow, W.W. (ed.) (1963), *The Economics of Take-off into Sustained Growth*, London: Macmillan.
- Sabel, C. and J. Zeitlin (1985), ‘Historical Alternatives to Mass Production: Politics, Markets and Technology in Nineteenth-Century Industrialization’, *Past and Present*, 108: 133–76.
- Schumpeter, J.A. (1911), *The Theory of Economic Development*, transl. from the German by R. Opie (1934), Cambridge, MA: Harvard University Press.
- Schumpeter, J.A. (1942), *Capitalism, Socialism and Democracy*, London: Allen & Unwin.

- Simmons, C. (1985), ‘“De-industrialization”, Industrialization and the Indian Economy, ca. 1850–1947’, *Modern Asian Studies*, 19 (3): 563–622.
- Singer, Hans (1950), ‘The Distribution of Gains between Investing and Borrowing Countries’, *American Economic Review*, 40: 473–85.
- Smith, Adam (1776 [1976]), *An Inquiry into the Nature and Causes of the Wealth of Nations*, the Glasgow edition of the text of 5th edn, 1791, 2 vols, ed. R.H. Campbell, A.S. Skinner and W.B. Todd (1976), Oxford: Oxford University Press.
- Steindl, J. (1952), *Maturity and Stagnation in American Capitalism*, New York: Monthly Review Press.
- Supple, B. (1973), ‘The State and the Industrial Revolution 1700–1914’, in C.M. Cipolla (ed.), *The Fontana Economic History of Europe: The Industrial Revolution*, London: Collins/Fontana, pp. 301–57.
- Sylla, R.E. (1991), ‘The Role of Banks’, in R.E. Sylla and G. Toniolo (eds), *Patterns of European Industrialization: The Nineteenth Century*, London: Routledge, pp. 45–63.
- Tanner, James (1981), *A History of the Study of Human Growth*, Cambridge: Cambridge University Press.
- Tilly, Louise (1994), ‘Connections’, *American Historical Review*, 99 (1): 1–17.
- Tilly, R. (1991), ‘Germany’, in R.E. Sylla and G. Toniolo (eds), *Patterns of European Industrialization: The Nineteenth Century*, London: Routledge, pp. 175–96.
- Tucker, J. (1774), *Four Tracts on Political and Commercial Subjects*, 2nd edn, Gloucester and London.
- UNDP (2005), *Human Development Report 2005*, New York: Oxford University Press for United Nations Development Programme.
- Van Zanden, J.L. (1993), *The Rise and Decline of Holland’s Economy: Merchant Capitalism and the Labour Market*, Manchester: Manchester University Press.
- Weisbrot, M., D. Baker, E. Kraev and J. Chen (2001), *The Scorecard on Globalization 1980–2000: Twenty Years of Diminished Progress*, Centre for Economic Policy Research www.cepr.org, accessed 14 September 2005.
- Wong, R. Bin (1997), *China Transformed: Historical Change and the Limits of European Experience*, Ithaca, NY and London: Cornell University Press.
- World Bank (2005), *World Development Indicators 2005*, Washington, DC: World Bank.

3 Empirics of growth and development

*Steven N. Durlauf, Andros Kourtellos and
Chih Ming Tan*

For much of the post-war period, empirical work on economic growth focused on accounting exercises whose goal was to understand the relative roles of factor accumulation and technical change in explaining growth trends. This approach was initiated in Solow's seminal (1957) article; work by Denison (1974) represents a particularly sophisticated version of this approach. In contrast, modern growth empirics largely attempt to address issues of cross-country economic differences using regression or other statistical methods which permit the consideration of a host of different growth determinants. This new work has been facilitated by the availability of data for a broad cross-section of countries for the period 1960–2000 due to Summers and Heston (1988, 1991) and Heston et al. (2002).

What sorts of broad facts has the new growth empirics uncovered? Durlauf et al.'s (2005) extended survey suggests three classes of empirical findings that are especially salient.

Firstly, over the 40-year period from 1960 to 2000, most countries have grown richer, but vast income disparities remain and substantial heterogeneity exists across countries. The major countries of Western Europe have either maintained (UK) or substantially improved (Germany, Italy, France) their position relative to the USA. East and South-East Asia have unprecedented sustained growth rates. The weakest performers are predominantly located in sub-Saharan Africa, where many countries have barely grown at all, and some have become poorer. The record in South and Central America is distinctly mixed, with high output volatility and periodic output collapses.

Secondly, the international distribution of gross domestic product (GDP) per worker exhibits an emerging bimodality. While the distribution has shifted rightwards to reflect overall growth, there has been a noticeable thinning in the middle of the distribution. Further, when one explores the location of individual countries as the cross-section distribution evolves, there is little evidence of churning across the emerging twin modes so that poor countries from the left mode are extremely unlikely to 'transition' into the right mode (and vice versa).

Thirdly, there exist a host of factors that appear to affect growth beyond the factor accumulation and exogenous technical change that drive the

Solow model. These determinants include a range of economic, political, geographic and social factors. There also appears to be significant evidence of non-linearity and parameter heterogeneity in the way these factors enter into growth regressions.

These new classes of stylized facts have led growth economists to pose three major sets of formal statistical questions. The first revolves around the question of convergence. That is, are contemporary differences in aggregate economies transient over sufficiently long time horizons, or are these differences in fact permanent? If they are permanent, does that permanence reflect structural heterogeneity or the role of initial conditions in determining long-run outcomes? The second set of questions considers the properties of the cross-section income distribution. What probability density describes current incomes and how is this density evolving? The third set of questions surrounds the identification of growth determinants. Which factors seem to explain observed differences in growth? Can these growth determinants be organized into theoretically and empirically useful categories: what are the fundamental (as opposed to proximate) determinants of growth?

The field of growth econometrics has emerged through efforts to interpret and understand the above stylized facts in terms of simple statistical models, and in the light of predictions made by alternative statistical structures. For questions of convergence and the determination of which growth factors are salient, these alternative statistical structures usually represent variations on a baseline linear cross-country growth regression pioneered by Barro (1991), Kormendi and Meguire (1985) and Mankiw et al. (1992). For questions on the cross-section income distribution, the methods involve various techniques related to density estimation. The plethora of statistical methods that has been employed to study growth is examined in Durlauf et al. (2005).

Our review of the empirical growth literature will focus on growth differences between countries, as opposed to differences across individuals. For this reason, many of our claims concerning the evolution of international inequality and changes in the world distribution will mask how differences across individuals have evolved; in our analysis China and India will be given the same weight as countries with small populations such as Cyprus. Our reason for this focus is that our goal is to understand growth facts in the contexts of growth theories, theories which are defined at the country level. While individual-level incomes presumably matter more for normative evaluations, they are not required for the questions we address. Examples of individual-specific studies of world inequality include Sala-i-Martin (2002a, 2002b).

The convergence hypothesis

Much of the empirical growth literature has focused on the question of whether contemporary income differences between countries are transitory or permanent. Unconditional convergence is said to occur if the differences are transitory. Conditional convergence is said to occur if the differences are permanent and solely due to cross-country structural heterogeneity; see Galor (1996). The neoclassical (Solow) model predicts that once structural heterogeneity – such as exogenous differences in technology, population growth rates and the population's willingness to save – are controlled for, long-run economic outcomes are independent of initial values of state variables, and so the model predicts conditional convergence. On the other hand, if the differences are permanent and initial conditions determine, in part at least, long-run outcomes, then convergence clubs are said to arise.

Attempts to translate these economic notions of convergence into testable restrictions on cross-country growth data have given rise to a number of popular statistical approaches.

β-convergence

The most common statistical approach towards convergence relies on the properties of the coefficient of the logarithm of initial income in linear growth regressions. A general panel data growth regression for growth over K -year intervals is:

$$g_{i,t} = k + \beta \log y_{i,t} + \gamma Z_{i,t} + \varepsilon_{i,t} \quad (3.1)$$

where $g_{i,t}$ is real per capita growth between time t and $t+K$, $y_{i,t}$ is initial income at t , $Z_{i,t}$ is a set of additional control variables and $\varepsilon_{i,t}$ is an error. β -convergence in per capita income means that $\beta < 0$. β -convergence is readily interpretable in the context of the Solow growth model, since the property is implied (at least locally) by the dynamics of the model. The economic intuition is simple: when the marginal product of capital is decreasing, per capita growth becomes slower as per capita output rises, assuming constant savings and population growth rates. In turn, β -convergence is commonly interpreted as evidence against endogenous growth models of the type studied by Romer (1986) and Lucas (1988), since a number of these models specifically predict that high initial income countries will grow faster than low initial income countries, once differences in saving rates and population growth rates have been accounted for.

Findings of conditional β -convergence (that is, β -convergence in the presence of control variables typically including saving and population growth rates) are common in the cross-country growth literature. While this statistical evidence seems robust to the choice of control variables

(Doppelhofer et al., 2004; Fernandez et al., 2001), there is a conceptual problem with the conditional β -convergence literature: namely the absence of a tight theoretical relationship between β -convergence and the notion of convergence as an economic concept.

This problem can be seen initially when one considers the use of β -convergence as a test between the Solow model and a model with multiple stable steady states. Multiple stable steady states in a model clearly violate the economic idea of convergence, since long-run behavior in the model depends on the initial capital stock. A standard example of the latter type of model is due to Azariadis and Drazen (1990). In this model, there is a discontinuity in the aggregate production function for aggregate economies. This discontinuity means that the steady-state behavior of a given economy depends on whether its initial capital stock is above or below this threshold. Bernard and Durlauf (1996) show that it is possible for data generated by economies that are described by the Azariadis–Drazen model to exhibit β -convergence even when multiple steady states are present. The reason for this is that even if economies are converging to distinct steady states, those economies that are converging to a low steady state may still be growing faster than those converging to a higher one. More generally, the finding of β -convergence provides no insight as to whether the cross-section of countries exhibits growth and development differences that are, in fact, transient.

σ -convergence

Other studies have focused on changes in the dispersion of income differences across time. σ -convergence is said to occur if the cross-sectional standard deviation of per capita income is falling over time. The key finding, however, is that there is no evidence of σ -convergence when one examines a full cross-section sample of countries (Sala-i-Martin, 1996). In contrast, when one restricts the analysis to developed economies, σ -convergence appears to be present.

Some studies relate σ -convergence with β -convergence. One common assertion in the literature is that the finding of β -convergence implies σ -convergence, so that the finding of the former actually results in strong predictions regarding the diminishing dispersion of cross-country growth experiences with time. There is, however, in fact no clear relationship between the two concepts. These two convergence notions do not have any necessary implications for one another, that is, one may hold when the other does not. For our purposes, what is important is that σ -convergence is not an implication of β -convergence and so does not speak directly to the question of the transience of contemporary income differences. The erroneous assertion that β -convergence implies σ -convergence is known

as Galton's fallacy and was brought into the convergence debate by Friedman (1992) and Quah (1993a). One should also note that there is no well-defined notion of conditional σ -convergence, which renders the use of this convergence concept to adjudicate growth theories essentially impossible.

Time-series approaches to convergence

A third approach to convergence based on time-series ideas has been developed by Bernard and Durlauf (1995, 1996) and extended by Evans (1998) and Hobijn and Franses (2000) among others. Bernard and Durlauf (1995) define time series convergence in output in two economies to be the equality of their long-run output forecasts taken at a given fixed date. In practice, forecast convergence is tested by looking for unit roots or deterministic trends in the difference in per capita output between two countries, either of which implies long-run forecastability of output differences. As argued by Bernard and Durlauf (1996), these tests carry strong economic implications. Specifically, they assume that first differences in output for each country may be described as a process with a time-invariant autocorrelation function. This is inconsistent with economies that are still in their initial stages of development and converging towards a steady state. As a result, time-series convergence tests are only appropriate for developed economies that are near their steady states.

In general, time-series approaches have rejected convergence. For example, Bernard and Durlauf (1995) find that time-series forecast convergence can be rejected across all Organisation for Economic Co-operation and Development (OECD) economies based on long time-series data due to Maddison (1982, 1989). However, they find that some individual country pairs such as Belgium and the Netherlands do display such convergence. Hobijn and Franses (2000) similarly find little evidence of convergence across 112 countries taken from the PennWorld Tables for the period 1960–89. Pesaran (2004) confirms the findings of little convergence for both Maddison and PennWorld data sets.

Taken as a whole, the convergence literature is difficult to interpret. There is good evidence of conditional cross-sectional convergence given the β definition and some evidence of convergence given the σ definition. However, these definitions do not provide strong discriminatory power when one is comparing models with unique versus multiple steady states, and so do not speak to many of the fundamental questions that motivate endogenous versus neoclassical growth theory. Further, time series tests typically conclude that convergence is not present and so represent a challenge to cross-section and panel findings of convergence. While Michelacci and Zaffaroni (2000) propose a clever reconciliation of cross-section and

time-series evidence which supports convergence, the analysis requires that output levels obey long memory processes, which has far from been established as empirically correct. We therefore conclude that the convergence question is far from resolved.

The world income distribution

A second strand of empirical research has focused on the world income distribution. This work is motivated by interest in the question of whether this distribution exhibits bimodality, which is suggestive of permanently high degrees of cross-section inequality, as well as whether the distribution reflects multiple mixture components, which is suggestive of the presence of either multiple steady states or multimodality in certain growth determinants.

In terms of bimodality, a range of studies have described how the distribution of relative per capita income has changed from a unimodal shape to a bimodal ('twin-peaked') distribution from 1960 to 1989. Quah (1993b, 1996, 1997) studies the evolution of the entire cross-country income distribution by modelling the cross-section density as a dynamic process. Quah finds that the estimated transition probabilities imply a bimodal steady state. While Kremer et al. (2001) have questioned the robustness of Quah's methodology, his general conclusions have been confirmed by a number of authors using different methods. Bianchi (1997) uses kernel density estimations to construct statistical tests for multimodality in the international distribution of income. Paap and van Dijk (1998) analyze the distribution of real GDP per capita using a parametric two-component mixture model. Using the estimated mixture distributions, they analyze intra-distribution mobility to find that the main source of mobility occurs from rich to poor while the 'middle' group between poor and rich disappears.

Recently, Anderson (2003) has shifted the discussion from the analysis of multimodality and the twin peaks debate to polarization, that is, the extent to which gaps between the rich and poor are increasing. Using stochastic dominance techniques to construct measures of polarization of the income distribution, Anderson finds that between 1970 and 1995 polarization between rich and poor countries increased throughout the time period. An important methodological advantage of Anderson's approach is that it is non-parametric.

One important implication of the work on the evolution of the cross-section income distribution is that it implies that even if aggregate production functions exhibit decreasing marginal productivity of capital, other growth factors are sufficiently strong to produce increasing international inequality. What such findings cannot say is whether these other factors are themselves permanent or transitory.

Growth model determination

The major empirical effort in modern growth research consists of efforts to identify the relative importance of various growth determinants. As such, this work represents an extension of the original objective of empirical growth research to understand the respective roles of technological change and capital accumulation.

Growth accounting revisited

One strand in the literature has attempted to quantify the extent of various sources of convergence or divergence. The aim of growth accounting is to estimate the relative portions of variation in cross-country output per worker, or growth, which can be assigned to variation in factor accumulation rates and that which accrues to total factor productivity (TFP). As such, the literature extends the approach pioneered in Solow (1957).

The recent TFP literature produces two important claims. First, the bulk of cross-country variation in per capita income levels or in growth rates appears to derive from differences in TFP. Klenow and Rodríguez-Clare (1997) find that only about half of the cross-country variation in the 1985 level of output per worker is due to variation in human and physical capital inputs while a mere 10 percent or so of the variation in growth rates from 1960 to 1985 reflects differences in the growth of these inputs. These findings are consistent with Easterly and Levine (2001) who also find that differences in inputs are unable to explain observed differences in output.

Second, divergence in the form of the ‘twin peaks’ phenomena described above is more likely to be attributed to cross-country divergence in TFP than to factor accumulation rates. Feyrer (2003) finds that the long-run distributions of both output per capita and TFP are bimodal while those of both the capital–output ratio and human capital per worker are unimodal. Feyrer’s findings suggest that models of multiple equilibria that give rise to equilibrium differences in TFP are more promising than models that emphasize indeterminacy in capital intensity or educational attainment. Johnson (2005), however, shows that certain aspects of Feyrer’s analysis are not robust and that robust approaches to this decomposition suggest the presence of bimodality in the long-run distributions of both the capital–output ratio and TFP as well as in the long-run distribution of output per capita.

A key assumption of most TFP studies is that the aggregate production function is concave. Graham and Temple (2006), however, show that the existence of multiple steady states can increase the variance and accentuate bimodality in the observed cross-country distribution of TFP. It seems likely, therefore, that the imposition of a concave production function in this case will tend to exaggerate the measured differences in TFP and so

confound inferences about the importance of TFP variation in explaining cross-country variations in output per worker or growth. Extension of TFP analyses to richer specification of production functions appears to be an important next step.

Another problem of most TFP studies is that they ignore the possibility of spillovers between physical and human capital accumulation and productivity. These spillovers can take the form of technology spillovers from countries at the frontier to developing countries facilitated by human capital stocks, rule of law, openness, and so on. Aiyar and Feyrer (2002) analyze the causal links between human capital accumulation and growth in TFP. They find that TFP differences explain most of the cross-sectional (static) variation in GDP but at the same time they find that human capital plays a substantial role in determining the dynamic path of TFP. Their findings suggest the importance of further work on identifying the channel through which human capital affects productivity.

Growth determinants

The evidence of the importance of TFP in growth outcome may be linked to the general search for salient growth determinants in regression models. From the perspective of growth regressions, such as equation (3.1), many different candidates have been proposed for Z . The set of growth regressors that have been proposed as candidate growth determinants is large and growing. In a 1999 survey, Durlauf and Quah (1999) listed a total of 87 such potential growth determinants studied in the literature. By the time of Durlauf et al.'s 2005 survey, the number had risen to 145.

Recently, the growth literature has also begun to distinguish between determinants that are viewed to be 'fundamental' as opposed to being 'proximate' to growth. Many such fundamental determinants of growth have been proposed including economic institutions (North, 1990; Knack and Keefer, 1995; Hall and Jones, 1999; Acemoglu et al., 2001); legal and political systems (La Porta et al., 1999; La Porta et al., 2004); climate (Gallup et al., 1999; Masters and McMillan, 2001); geographic isolation (Radelet and Sachs, 1998; Frankel and Romer, 1999); ethnic fractionalization (Easterly and Levine, 1997; Alesina et al., 2003) and culture (Knack and Zak, 2001; Barro and McCleary, 2003; Tabellini, 2005).

This shift in the literature towards fundamental explanations of divergence is motivated in part by the desire to identify variables that are slow-moving and can be argued to be predetermined with respect to current growth rates in per capita income. The idea is that these fundamental determinants may not only provide interesting reduced form explanations for divergence, but may also constitute valid instrumental variables for (statistically) endogenous proximate causes. However, as Durlauf (2000) points

out, predetermined variables are not necessarily valid instruments. The difficulty is that with so many potential explanations for growth, it is hard to argue that simply because a variable is predetermined, it is also uncorrelated with omitted growth factors in growth regressions. Glaeser et al. (2004) have also questioned the direction of causality between certain fundamental determinants of growth and proximate factors of growth. They point out that some measures of economic institutions are themselves in reality choice variables of policy-makers who are in turn constrained by proximate factors such as the average level of initial human capital in the population.

The presence of so many potential growth regressors is unsurprising given the nature of new growth theories. As argued by Brock and Durlauf (2001), new growth theories are inherently open-ended. By 'theory open-endedness', Brock and Durlauf refer to the fact that typically the a priori statement that a particular theory of growth is relevant does not preclude other theories of growth from also being relevant. As a result, there is a great need for robust procedures that deal with model uncertainty by assessing the sensitivity of coefficient estimates and standard errors to choices of covariates.

An early attempt to develop ways to identify empirically salient growth determinants is that by Levine and Renelt (1992) who employed Leamer's (1983) extreme bounds analysis (EBA) to conclude that the only robust growth determinant among the set of growth determinants is the share of investment in GDP. However, from a decision-theoretic perspective, the extreme bounds approach is a problematic methodology. As discussed in detail in Brock and Durlauf (2001) and Brock et al. (2003), EBA corresponds to a very risk-averse way of responding to model uncertainty.

The limitations of EBA have led to a range of efforts to develop new tools for identifying robust growth determinants. Attempts to deal with the problem of model uncertainty include Sala-i-Martin's (1997) variants of extreme bounds analysis and the general-to-specific model selection approaches of Hendry and Krolzig (2004) and Hoover and Perez (2004). While these approaches avoid the implicit risk aversion found in extreme bounds analysis, they do not possess conventional statistical or decision theoretic justification.

An alternative approach has emerged that accounts for uncertainty in choice of growth regressors by systematically addressing the dependence of model-specific estimates on a given model. This method, known as model averaging was suggested by Leamer (1978) and has re-emerged in recent work in statistics; see Hoeting et al. (1999) for a survey. The idea of model averaging is to construct estimates of parameters of interest by aggregating information across all elements in a space of possible models. As such, the

method accounts for the fact that the true model is not known to the researcher, but rather presupposes that the true model is known to lie within some set. Model spaces can be constructed based on the choice of regressors as well as the way in which non-linearities or heterogeneity may appear in the growth process. Model averaging has been applied to cross-country growth data by Brock and Durlauf (2001), Fernandez et al. (2001), Brock et al. (2003), Doppelhofer et al. (2004), and Masanjala and Papageorgiou (2004), among others.

In terms of findings, these various approaches to identifying robust growth determinants conclude that at least two of the four canonical Solow variables – that is, initial income and the rate of physical capital accumulation – are robust determinants of growth. There is also some evidence that human capital accumulation as measured by secondary school education (Sala-i-Martin, 1997) and life expectancy (Fernandez et al., 2001) may be robust as well. Other variables that have been found to be relatively robust include measures of political stability (Hendry and Krolzig, 2004; Hoover and Perez, 2004), proxies for trade openness (Doppelhofer et al., 2004), as well as measures of culture as captured by the percentage of the population that is Confucian (Fernandez et al., 2001; Doppelhofer et al., 2004; Hendry and Krolzig, 2004; Hoover and Perez, 2004), and the percentage of the population that is Protestant (Hendry and Krolzig, 2004; Hoover and Perez, 2004). These last variables are difficult to interpret in terms of causality and indeed may reflect the absence of attention to parameter heterogeneity across countries.

Non-linearities and parameter heterogeneity

Another body of empirical growth analyses deals with the problems of parameter heterogeneity and non-linearities relative to the canonical cross-country growth regression equation (3.1). By non-linearity we mean that the determinants of economic growth enter the regression in a non-linear way, while by parameter heterogeneity we mean that the parameters of the model are explicitly allowed to vary across countries. The modeling assumptions of parameter heterogeneity can take various forms. The parameters can be assumed to vary in a systematic and/or non-systematic (random) fashion. When parameter heterogeneity is modeled in a systematic way, the parameters are thought to be parametric or non-parametric functions of dummy variables (for example a dummy for sub-Saharan countries) or more generally a subset of the determinants of economic growth (for example initial conditions). In this situation, one may view parameter heterogeneity as an interesting special case of non-linearity.

Concerns over non-linearity and parameter heterogeneity naturally arise when one considers theoretical growth models with multiple steady states.

A range of analyses have provided microfoundations for the emergence of multiple steady states and convergence clubs. Examples include human capital externalities (Azariadis and Drazen, 1990) or liquidity constraints (Galor and Zeira, 1993) in the accumulation of human capital and physical capital. More recently, several papers give a technological explanation for these growth anomalies. Howitt and Mayer-Foulkes (2005) show that a Schumpeterian approach that includes both innovation and technology implementation can give rise to convergence clubs. Acemoglu et al. (2006) show that institutional barriers can prevent a group of countries from using the same production function potential, thereby keeping each country in the group inside the group-level production possibility frontier. The equilibrium growth paths of these types of models are not well approximated by the linear growth model (3.1) in the way the neoclassical Solow growth model, or its Cass-Koopmans variation, is.

One approach to allowing for growth non-linearities is to use semi-parametric models. Liu and Stengos (1999) estimate a partially linear model to identify non-linear growth patterns. This approach allows one or more regressors in (3.1) to have additive but non-linear effects on growth. One of their findings is that the convergence hypothesis only holds for countries in the middle to upper range of initial income. Banerjee and Duflo (2003) use this same regression strategy to study non-linearity in the relationship between changes in inequality and growth. They find an inverted U shape between the growth rate and the change in the Gini coefficient.

Durlauf et al. (2001) extend this search for non-linearity to one for parameter heterogeneity and estimate a Solow growth model that allows the parameters for each country to vary as functions of initial income. In effect, this varying coefficient approach defines a distinct Solow regression at each initial income level. This approach reveals considerable parameter heterogeneity especially among the poorer countries. This work is extended in Kourtellos (2005) who finds parameter dependence on initial literacy, initial life expectancy, expropriation risk and ethnolinguistic fractionalization. The varying coefficient approach is also employed in Mamuneas et al. (2006) who analyze annual measures of TFP for 51 countries. One important finding is that, in general, the estimates of the elasticity of human capital with respect to output are positive and largest for high-income countries while the estimates for low-income countries are small and in some cases zero.

A conceptually different approach to modeling parameter heterogeneity and non-linearities has been taken by Durlauf and Johnson (1995), Bloom et al. (2003), Canova (2004), Masanjala and Papageorgiou (2004) and Tan (2005). These papers have employed statistical learning (specifically, sample

splitting and threshold regression) approaches that emphasize pattern recognition in order to uncover evidence of multiple steady states or ‘convergence clubs’ across countries. Durlauf and Johnson find evidence for convergence clubs that depend on initial values for state variables such as initial adult literacy rates and initial income. Masanjala and Papageorgiou (2004) find similar results using models that employ the constant elasticity of substitution (CES) production function. Further evidence of multiple regimes is also found by Bloom et al. using mixture distribution methods, and Canova using a Bayesian approach that differentiates multiple regimes and parameter heterogeneity. Most recently, Tan employs classification methods to adjudicate divergent claims on the importance of different fundamental growth determinants and finds strong evidence that measures of institutional quality and ethnic fractionalization define convergence clubs across a wide range of countries.

This discussion suggests that the assumptions of linearity and invariant parameters such as found in equation (3.1) are likely to be inappropriate in the analysis of cross-country growth data. That being said, no consensus yet exists on which types of non-linearity and heterogeneity are empirically most important, and so researchers will need to exercise judgment as to how to allow for these when analyzing a particular data set.

Summary and conclusion

Modern growth economics has led to a rich and wide-ranging empirical literature replete with many new methodologies and many new findings. Yet in comparing the modern empirical literature to the traditional growth accounting analyses of the 1960s and 1970s, one cannot help but be struck by the relative lack of progress on substantive conclusions. The critical role of TFP found in recent work is consistent with claims as far back as Solow. Evidence of statistical notions of convergence represents a new set of stylized facts but suffers from a lack of connection to economically interesting notions of convergence. The search for empirically successful growth models has provided a range of candidate growth determinants that lie far outside the domain of the neoclassical growth model, but efforts to search for robust determinants have had mixed results, outside of the finding that physical capital accumulation affects growth, which is no surprise given the earlier literature. Evidence of non-linearities and parameter heterogeneity is suggestive of multiple steady states and richer growth dynamics than neoclassical theories, but this evidence has yet to be integrated into a consistent whole. Together, this suggests that the next step in empirical growth research should be the unification of the vast array of statistical claims into a unified growth picture combined with efforts to link this picture more tightly with growth theories.

References

- Acemoglu, D., P. Aghion and F. Zilibotti (2006), 'Distance to Frontier, Selection, and Economic Growth', *Journal of the European Economic Association*, MIT Press, 4 (1): 37–74, 03.
- Acemoglu, D., S. Johnson, and J. Robinson (2001), 'The Colonial Origins of Comparative Advantage', *American Economic Review*, 91 (5): 1369–1401
- Aiyar, S. and J. Feyrer (2002), 'A Contribution to the Empirics of Total Factor Productivity', mimeo, Dartmouth College.
- Alesina, A., A. Devleeschauwer, W. Easterly, S. Kurlat and R. Wacziarg (2003), 'Fractionalization', *Journal of Economic Growth*, 8 (2): 155–94.
- Anderson, G. (2003), 'Making Inferences About the Polarization, Welfare, and Poverty of Nations: A Study of 101 Countries 1970–1995', *Journal of Applied Econometrics*, 19 (5): 537–50.
- Azariadis, C. and A. Drazen (1990), 'Threshold Externalities in Economic Development', *Quarterly Journal of Economics*, 105 (2): 501–26.
- Banerjee, A. and E. Dufo (2003), 'Inequality and Growth: What Can the Data Say?' *Journal of Economic Growth*, 8 (3): 267–300.
- Barro, R.J. (1991), 'Economic Growth in a Cross-Section of Countries', *Quarterly Journal of Economics*, 106 (2): 407–43.
- Barro, R. and R. McCleary (2003), 'Religion and Economic Growth', *American Sociological Review*, 68: 760–81.
- Bernard, A. and S. Durlauf (1995), 'Convergence in International Output', *Journal of Applied Econometrics*, 10 (2): 97–108.
- Bernard, A. and S. Durlauf (1996), 'Interpreting Tests of the Convergence Hypothesis', *Journal of Econometrics*, 71 (1–2): 161–73.
- Bianchi, M. (1997), 'Testing for Convergence: Evidence from Nonparametric Multimodality Tests', *Journal of Applied Econometrics*, 12 (4): 393–409.
- Bloom, D., D. Canning and J. Sevilla (2003), 'Geography and Poverty Traps', *Journal of Economic Growth*, 8: 355–78.
- Brock, W. and S. Durlauf (2001), 'Growth Empirics and Reality', *World Bank Economic Review*, 15 (2): 229–72.
- Brock, W., S. Durlauf and K. West (2003), 'Policy Evaluation in Uncertain Economic Environments (with Discussion)', *Brookings Papers on Economic Activity*, 1: 235–322.
- Canova, F. (2004), 'Testing for Convergence Clubs in Income Per Capita: A Predictive Density Approach', *International Economic Review*, 45 (1): 49–77.
- Denison, E.F. (1974), *Accounting for United States Growth, 1929–1969*, Washington, DC: Brookings Institution.
- Doppelhofer, G., R. Miller and X. Sala-i-Martin (2004), 'Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach', *American Economic Review*, 94 (4): 813–35.
- Durlauf, S. (2000), 'Econometric Analysis and the Study of Economic Growth: A Skeptical Perspective', in R. Backhouse and A. Salanti (eds), *Macroeconomics and the Real World*, Oxford: Oxford University Press.
- Durlauf, S. and P. Johnson (1995), 'Multiple Regimes and Cross Country Growth Behaviour', *Journal of Applied Econometrics*, 10 (4): 365–84.
- Durlauf, S., P. Johnson and J. Temple (2005), 'Growth Econometrics', in P. Aghion and S. Durlauf (eds), *Handbook of Economic Growth*, Amsterdam: North Holland, Vol 1, Part 1, pp. 555–677.
- Durlauf, S., A. Kourtellos and A. Minkin (2001), 'The Local Solow Growth Model', *European Economic Review*, 45 (4–6): 928–40.
- Durlauf, S. and D. Quah (1999), 'The New Empirics of Economic Growth', in J. Taylor and M. Woodford (eds), *Handbook of Macroeconomics*, Amsterdam: North Holland, pp. 235–308.
- Easterly, W. and R. Levine (1997), 'Africa's Growth Tragedy: Policies and Ethnic Divisions', *Quarterly Journal of Economics*, 112 (4): 1203–50.
- Easterly, W. and R. Levine (2001), 'It's Not Factor Accumulation: Stylized Facts and Growth Models', *World Bank Economic Review*, 15: 177–219.

- Evans, P. (1998), 'Using Panel Data to Evaluate Growth Theories', *International Economic Review*, 39 (2): 295–306.
- Fernandez, C., E. Ley and M. Steel (2001), 'Model Uncertainty in Cross-Country Growth Regressions', *Journal of Applied Econometrics*, 16 (5): 563–76.
- Feyrer J. (2003), 'Convergence By Parts', mimeo, Dartmouth College.
- Frankel, J. and D. Romer (1999), 'Does Trade Cause Growth?', *American Economic Review*, 89 (3): 379–99.
- Friedman, M. (1992), 'Do Old Fallacies Ever Die?' *Journal of Economic Literature*, 30: 2129–32.
- Gallup, J., J. Sachs and A. Mellinger (1999), 'Geography and Economic Development', *International Science Review*, 22: 179–232.
- Galor, O. (1996), 'Convergence? Inferences from Theoretical Models', *Economic Journal*, 106: 1056–69.
- Galor, O. and J. Zeira (1993), 'Income Distribution and Macroeconomics', *Review of Economic Studies*, 60 (1): 35–53.
- Glaeser, E., R. La Porta, F. López de Silanes and A. Shleifer (2004), 'Do Institutions cause Growth?', *Journal of Economic Growth*, 9: 271–303.
- Graham, B. and J. Temple (2006), 'Rich Nations, Poor Nations: How Much Can Multiple Equilibria Explain?', *Journal of Economic Growth*, 11 (1): 5–41.
- Hall, R. and C. Jones (1999), 'Why Do Some Countries Produce So Much More Output Per Worker Than Others?', *Quarterly Journal of Economics*, 114: 83–116.
- Hendry, D. and H.-M. Krolzig (2004), 'We Ran One Regression', *Oxford Bulletin of Economics and Statistics*, 66 (5): 798–810.
- Heston A., R. Summers and B. Aten (2002), *Penn World Table Version 6.1*, Center for International Comparisons at the University of Pennsylvania (CICUP), October.
- Hobijn, B. and P.H. Franses (2000), 'Asymptotically Perfect and Relative Convergence of Productivity', *Journal of Applied Econometrics*, 15: 59–81.
- Hoeting, J.A., D. Madigan, A.E. Raftery and C.T. Volinsky (1999), 'Bayesian Model Averaging: A Tutorial (with Discussion)', *Statistical Science*, 14: 382–401.
- Hoover, K. and S. Perez (2004), 'Truth and Robustness in Cross-Country Growth Regressions', *Oxford Bulletin of Economics and Statistics*, 66 (5): 765–98.
- Howitt, P. and D. Mayer-Foulkes (2005), 'R&D, Implementation and Stagnation: A Schumpeterian Theory of Convergence Clubs', *Journal of Money, Credit and Banking*, 37: 147–77.
- Johnson, P. (2005), 'A Continuous State Space Approach to "Convergence by Parts"', *Economic Letters*, 86, pp. 317–21.
- Klenow, P. and A. Rodríguez-Clare (1997), 'The Neoclassical Revival in Growth Economics: Has it Gone Too Far?' in B. Bernanke and J. Rotemberg (eds), *Macroeconomics Annual 1997*, Cambridge, MA: MIT Press, pp. 73–102.
- Knack, S. and P. Keefer (1995), 'Institutions and Economic Performance: Cross Country Tests Using Alternative Institutional Measures', *Economics and Politics*, 7 (3): 207–27.
- Knack, S. and P.J. Zak (2001), 'Trust and Growth', *Economic Journal*, 111 (470): 295–321.
- Kormendi, R. and P. Meguire (1985), 'Macroeconomic Determinants of Growth: Cross-Country Evidence', *Journal of Monetary Economics*, 16 (2): 141–63.
- Kourtellos, A. (2005), 'Modeling Parameter Heterogeneity in Cross-Country Growth Regression Models', mimeo, University of Cyprus.
- Kremer, M., A. Onatski and J. Stock (2001), 'Searching for Prosperity', *Carnegie-Rochester Conference Series on Public Policy*, 55: 275–303.
- La Porta, R., F. López-de-Silanes, C. Pop-Eleches and A. Shleifer (2004), 'Judicial Checks and Balances', *Journal of Political Economy*, 112: 445–70.
- La Porta, R., F. López-de-Silanes, A. Shleifer and R. Vishny (1999), 'The Quality of Government', *Journal of Law, Economics and Organization*, 15: 222–79.
- Leamer, E. (1978), *Specification Searches: Ad-Hoc Inference with Non-Experimental Data*, New York: John Wiley.

- Leamer, E. (1983), 'Let's Take the Con Out of Econometrics', *American Economic Review*, 73 (1): 31–43.
- Levine, R. and D. Renelt (1992), 'A Sensitivity Analysis of Cross-Country Growth Regressions', *American Economic Review*, 82 (4): 942–63.
- Liu, Z. and T. Stengos (1999), 'Non-Linearities in Cross Country Growth Regressions: A Semiparametric Approach', *Journal of Applied Econometrics*, 14 (5): 527–38.
- Lucas R., Jr. (1988), 'On The Mechanics of Economic Development', *Journal of Monetary Economics*, 22 (3): 3–42.
- Maddison, A. (1982), *Phases of Capitalist Development*, New York: Oxford University Press.
- Maddison, A. (1989), *The World Economy in the 20th Century*, Paris: OECD.
- Mamuneas, T., A. Savvides and T. Stengos (2006), 'Economic Development and Return to Human Capital: A Smooth Coefficient Semiparametric Approach', *Journal of Applied Econometrics*, 21: 111–32.
- Mankiw, N.G., D. Romer and D. Weil (1992), 'A Contribution to the Empirics of Economic Growth', *Quarterly Journal of Economics*, 107 (2): 407–37.
- Masanjala, W. and C. Papageorgiou (2004), 'The Solow Model with CES Technology: Nonlinearities and Parameter Heterogeneity', *Journal of Applied Econometrics*, 19 (2): 171–201.
- Masters, W. and M. McMillan (2001), 'Climate and Scale in Economic Growth', *Journal of Economic Growth*, 6: 167–86.
- Michelacci, C. and P. Zaffaroni (2000), 'Fractional (Beta) Convergence', *Journal of Monetary Economics*, 45: 129–53.
- North, D.C. (1990), *Institutions, Institutional Change and Economic Performance*, Cambridge: Cambridge University Press.
- Paap, R. and H. van Dijk (1998), 'Distribution and Mobility of Wealth of Nations', *European Economic Review*, 42 (7): 1269–93.
- Pesaran, M.H. (2004), 'General Diagnostic Tests for Cross-Section Dependence in Panels', mimeo, University of Cambridge.
- Quah, D. (1993a), 'Galton's Fallacy and Tests of the Convergence Hypothesis', *Scandinavian Journal of Economics*, 95: 427–43.
- Quah, D. (1993b), 'Empirical Cross-Section Dynamics in Economic Growth', *European Economic Review*, 37 (2–3): 426–34.
- Quah, D. (1996), 'Empirics for Economic Growth and Convergence', *European Economic Review*, 40 (6): 1353–75.
- Quah, D. (1997), 'Empirics for Growth and Distribution: Stratification, Polarization, and Convergence Clubs', *Journal of Economic Growth*, 2 (1): 27–59.
- Radelet S. and J. Sachs (1998), 'Shipping Costs, Manufactured Exports, and Economic Growth', mimeo, HIIID.
- Romer, P.M. (1986), 'Increasing Returns and Long-run Growth', *Journal of Political Economy*, 94 (5): 1002–37.
- Sala-i-Martin, X. (1996), 'The Classical Approach to Convergence Analysis', *Economic Journal*, 106: 1019–36.
- Sala-i-Martin, X. (1997), 'I Just Ran 2 Million Regressions', *American Economic Review*, 87 (2): 178–83.
- Sala-i-Martin, X. (2002a), 'The Disturbing "Rise" of Global Income Inequality', National Bureau of Economic Research Working Paper no. 8904.
- Sala-i-Martin, X. (2002b), 'The World Income Distribution', National Bureau of Economic Research Working Paper no. 8933.
- Solow, R.M. (1957), 'Technical Change and the Aggregate Production Function', *Review of Economics and Statistics*, 39: 312–20.
- Summers, R. and A. Heston (1988), 'A New Set of International Comparisons of Real Product and Price Levels Estimates for 130 Countries, 1950–1985', *Review of Income and Wealth*, 34: 1–25.
- Summers, R. and A. Heston (1991), 'The Penn World Table (Mark 5): An Expanded Set of International Comparisons, 1950–1988', *Quarterly Journal of Economics*, 106 (2): 327–68.

- Tabellini, G. (2005), 'Culture and Institutions: Economic Development in the Regions of Europe', mimeo, IGER, Bocconi.
- Tan, C.M. (2005), 'No One True Path: Uncovering the Interplay between Geography, Institutions, and Fractionalization in Economic Development', Tufts University, Department of Economics Working Paper No. 2005-12.

4 Structural change and development

Moshe Syrquin

Structural change is at the center of Modern Economic Growth (MEG), the term applied by Simon Kuznets (1966) to characterize the economic epoch of the last 250 years distinguished by the pervasive application of science-based technology to production. The principal characteristic of this economic epoch is ‘a sustained increase in per capita or per worker product, most often accompanied by an increase in population and usually sweeping structural changes’ (Kuznets, 1966, p. 1).

What is ‘structural change’?

There are many uses of the concepts of structure and structural change in economics. Some of them have a clear meaning while others are vague or worse.¹ The most common one refers to long-term persistent changes in the composition of an aggregate. In development and in economic history it usually refers to the relative importance of sectors in the economy and to changes in the location of economic activity (urbanization) and other concomitant aspects of industrialization (demographic transition, income distribution). The interrelated processes of structural change that accompany economic development are jointly referred to as the ‘structural transformation’.

A broader measure would also consider changes in institutions by which structural change is achieved. This wider framework is often acknowledged, though seldom represented in empirical work on development.²

The principal uniformities in the process of development identified in studies of the long-term experience of the industrialized countries and on the postwar experience of LDCs up to the mid-1980s were the subject of various studies in the late 1980s.³ There has not been much systematic comparative work since then except for growth regressions which have ignored structure and are of limited relevance for country experience over time. In this chapter I present some stylized facts derived from the Kuznets research program updated to the early 1980s. Simple calculations show that in general (important exceptions to be noted below) there have not been major changes in the main patterns. This says rather little as the patterns are long-term trends and are expected to have a high degree of persistence. The last 25 years have been chaotic and will leave their mark on long-term patterns but only at a later stage. Events that are expected to affect the structure of

the world economy and the economic structures within nations include: the intensification of globalization, the information technology (IT) revolution, the emergence of Asia, theoretical advances and the accumulation of disaggregated information on structure.

Why care?

Growth and structural change are strongly interrelated. Once we abandon the fictional world of homothetic preferences, neutral productivity growth with no systematic sectoral effects, perfect mobility and markets that adjust instantaneously, structural change emerges as a central feature of the process of development and an essential element in accounting for the rate and pattern of growth. It can retard growth if its pace is too slow or its direction inefficient, but it can contribute to growth if it improves the allocation of resources by, for example, reducing the disparity in factor returns across sectors or facilitating the exploitation of economies of scale. Policy can try to anticipate structural change, facilitating it by removing obstacles and correcting for market failures. Structural change can also hamper growth by blocking the required changes in structure or by attempting to dictate them. As the Soviet experience showed, forced industrialization can accelerate recorded growth, but only for a while and at very high cost.

Structural change is not just a theoretical construct. Structural change is a conflictive process that requires individual and societal adaptations and a large reallocation of population from rural traditional places to modern urban ones, especially in the early stages of development. These changes require mechanisms for conflict resolution. In the past, the state often emerged as the arbiter among group interests and as mitigator of the adverse effects of economic change.

Structural change in models of growth

Most growth models of the 'old' or 'new' vintages are aggregate models and by definition ignore structure.⁴ Sectoral models, of the von Neumann balanced growth type, also exclude structural change by definition. Those models gave us elegant duality relations and turnpike theorems with intriguing suggestions for planning. All of them ended up being irrelevant for development while fostering social engineering interventions, seldom positive. In many of these models relative prices remain unchanged and, since the identity of sectors in them is irrelevant, then by Hicks composite-good theorem we can, without loss, collapse the various sectors into a one-good model.

If in the growth literature sectors are not essential, in the development literature we have models, mostly not well formalized, where identity (of sectors) is destiny. A generic example is the staples approach where the

characteristics of the dominant staple determine the fortunes of the economy. The staples are resource-based commodities produced primarily for the external market, oil being today's predominant example. Linkages and flexibility emerge then as key for further transformation and continuing growth beyond the heavy dependence on the staple.

Recently, some formal models attempt to replicate the basic patterns of structural change by modifying some of the usual assumptions in a simulation framework. Echevarria (1997), for example, allows for non-homotheticity. This is a promising development, especially if it is integrated with empirical implementation.

Arguably the most important contribution to the early development literature was Lewis's (1954) model of dualistic development. Development was seen as a gradual replacement of traditional by modern sectors and techniques – structural change fueled by capital accumulation in the expanding modern sector. Other approaches going back to Marx stressed the composition of capital or of demand (consumption and investment) as crucial.

Fisher and Clark focused on sectors of economic activity, the former to draw attention to 'growing points' in the economy, and the latter to point out the association of level of development with structure. Kuznets embedded this in a more comprehensive approach. He regarded the structural shifts as a requirement for the high rates of growth and in turn saw the changes in economic structure as requiring 'shifts in population structure, in legal and political institutions, and in social ideology. [Not] all the . . . shifts in economic and social structure and ideology are *requirements*, [but] . . . *some* structural changes, not only in economic but also in social institutions and beliefs, are required without which modern economic growth would be impossible' (Kuznets, 1971, p. 348).

The 1930s saw two interesting and totally independent developments which presaged the break between growth and development. One was the Von Neumann model of growth and the second was the publication of Kuznets's (1930) *Secular Movements in Production and Prices*. Von Neumann's was an elegant parsimonious representation of equilibrium in a multi-sector expanding economy. It took more than a decade for it to be translated and interpreted in the economic literature and another decade to be appreciated as a complete exposition of duality, minimax, and so on. It became the canonical multi-sectoral balanced growth model.

Kuznets is often seen as elaborating on Clark's sequence of sectoral change. However, 'Kuznets' experience in developing alternative approaches to the measurement of national income – by type of product, industry, factor share, and size of income – and his study of demographers' work on population and labor force and their components of change led to a much more comprehensive undertaking' (Easterlin, 2001). It is true that Kuznets

started his comprehensive project on the economic growth of nations not much before 1950; however, already in his earlier studies in the late 1920s he showed interest in growth and structural shifts. His 1930 book on secular trends looks at long-term movements in production and prices in many products in six countries. He first notes that the global ‘modern economic system is characterized by ceaseless change . . . a process of uninterrupted and seemingly unslackened growth’ (pp. 1, 3); yet at the sectoral or national level the picture is less uniform: leadership among nations shifts over time and, within a nation, leading sectors are continuously replaced as retardation inevitably reaches former leaders. Kuznets contrasts the secular retardation at the sectoral level ‘with our belief in the fairly continuous march of economic progress’ (p. 5) and asks why not balanced growth? The answer combines demand effects and technological change: progress of technique makes new goods available (tea, cotton, radios . . .) but eventually demand reaches saturation, the pace of technical change slackens, new goods emerge, and possibly also competition from younger nations. With this general retardation come shifts in the relation between capital and labor, in the distributive process, in the character of the market, in the type of business organization and in the roles of industry and agriculture. Here we have in a nutshell the sources of structural transformation which were to reappear several decades later with technical change and sectoral shifts as key elements of the process.

Trade theory and structure

Until very recently trade theory had little to say on the volume of trade, but presumably it would have much to say about sectoral composition as determined by comparative advantage. Unfortunately it was not as helpful as expected. Older Heckscher–Ohlin models would predict much more specialization in production than in consumption. Actually, the correlation between output and domestic demand tends to be very high, especially in large countries which have lower trade shares.

Trade models focus on factor abundance and factor intensity. A major problem in applying these models is the ‘surprisingly little attention . . . [paid to] the appropriateness of industry classification for testing trade theory’ (Schott, 2003, p. 692). The sectors belonging to a common industry in trade theory are supposed to have similar factor intensities. This is mostly assumed without questioning. An early study that did look into this found that factor intensities within three-digit SITC (Standard International Trade Classification) categories varied as much as among the categories (Finger, 1975).

Until the mid-1970s assuming a closed economy for large countries was not a bad assumption for predicting their economic structure. Very low

trade shares were the rule in very large countries. In 1970 the share of exports in gross domestic product (GDP) was a low 4 to 7 percent in Brazil, China, India, Mexico, Turkey and the USA. By 2000 the shares had doubled in the USA and Brazil, had tripled in India, and gone up by a factor of about five in the other countries.

Globalization is not a new phenomenon but, since about 1980, it has accelerated its pace. There have been large increases in trade participation and more trade integration (as measured by price differentials which have declined), global capital markets have been established, the range of tradable commodities and services has expanded, and so on. And yet, as trade data with wide commodity and country coverage became available and empirical work on trade spread, the importance of local conditions and border effects and home biases were reaffirmed instead of vanishing.

With globalization trade has probably become a more important element in accounting for any transformation, together with technology and the evolution of domestic demand.

Structure and growth: stylized facts

In this section we refer to some very robust associations observed during the long-term process of development of today's industrialized countries and in the post-World War II experience of economies that traversed a large part of the transition from a low-income, primarily rural, economy with little use of modern technology to a richer, urban, technically advanced economy. Such stylized facts are useful in giving possible ranges of feasible paths and expected changes with development, but it is an open question how well they characterize economies at the lowest levels of development or even stagnant economies.

The elements of sectoral transformation are linked by a few accounting identities relating national product by sector to the final uses of income:

$$V = \sum V_j = Y = (C + I + G) + (E - M) = D + T \quad (4.1)$$

where V_j is value added of sector j , Y is gross domestic product, C is private consumption, G is government consumption, I is gross investment, E is exports, M is imports, D is domestic final demand and T is net trade.

At the sectoral level:

$$V_j = v_j X_j = v_j (W_j + D_j + T_j) \quad (4.2)$$

where X_j is gross output of sector j , W_j is intermediate demand for the output of sector j , (D_j and T_j are sectoral levels of domestic final demand and net trade) and v_j is the value-added ratio in sector j .

Structural transformation

Industrialization has to be analyzed in conjunction with changes in the structures of demand (final and intermediate) and trade. Results from an econometric study of the various elements of structural transformation are summarized in Table 4.1. The patterns of change in the table summarize the relationship that exists along observed growth paths where per capita income is the measure of development.

Table 4.1 Shares of economic structure associated with levels of per capita income (%)

Component of economic structure	Income per capita (1980 US dollars)				
	Actual ^a average <300	Predicted			Actual ^b average >4000
		300	1000	4000	
<i>Final demand</i>					
Private consumption	79	73	66	60	60
Investment	14	18	23	26	26
Exports	16	19	23	26	23
Food consumption	39	38	29	19	15
<i>Trade</i>					
Merchandise exports	14	15	19	21	18
Primary	13	14	15	12	07
Manufacturing	01	01	04	09	11
<i>Production (value-added)</i>					
Agriculture	48	40	23	10	07
Manufacturing	10	12	18	24	28
Utilities and Construction	10	11	14	15	17
Services	31	32	37	45	47
<i>Labor force</i>					
Agriculture	81	75	52	24	13
Industry	07	09	19	33	40
Services	12	16	29	43	47

Notes:

The numbers in the table are shares of GDP, except for the labor variables which are expressed as shares of total labor force.

^a Average for countries with per capita income less than \$300 in 1970; mean \$180.

^b Average for countries with per capita income greater than \$4000 in 1970; mean \$7300.

Source: Syrquin and Chenery (1989).

Demand

The best-established trends in the composition of final uses of output are the rise in the share of resources allocated to investment and the decline of the share of food in consumption. The latter (Engel's Law), is among the most robust empirical relationships in economics (see Table 4.2), but its implication of non-homothetic preference is rarely acknowledged in theories of growth and international trade.

During the process of development, the use of intermediates relative to total gross output tends to rise. A measure of this change is an increase in the density of the input–output matrix which reflects the evolution to a more complex system with a higher degree of fabrication, and the shift from handicrafts to factory production. A related robust trend is the significant increase with the level of income of the share of purchased intermediates in the total value of output in agriculture (Deutsch and Syrquin, 1989).

Trade

The rise in the ratio of capital (human and physical) to labor, and the observed higher rate of productivity growth in the more modern sectors of the economy, tend to shift the comparative advantage from primary activities to manufacturing. Accordingly we find the composition of exports shifting systematically from primary products to manufactures, but mostly in the upper levels of the transition (Balassa, 1979).

Table 4.2 Structure of British gross domestic expenditure, 1688 and 1996

	1688	1996
<i>Essentials</i>	74.5	23.5
(of which, Food, Beverages and Tobacco)	(25.7)	(6.5)
<i>Other</i>	9.8	49.9
Total private consumption (<i>Essentials</i> + <i>Other</i>)	84.2	73.4
Government consumption (except education & health)	9.0	10.9
Gross capital formation	6.8	15.8
Total gross domestic expenditure	100.0	100.0
Level of per capita GDP (1990 international dollars)	1 411	17 891

Notes:

1688 refers to England and Wales, 1996 to the UK.

Essentials include food, beverages and tobacco, clothing and footwear, light, fuel and power, furniture, furnishings and household equipment, and personal services; *Other* includes rent and imputed rent, education, health, recreation and entertainment, transport and communication, and other.

Source: Based on Table 3 in Maddison (2004).

Productivity growth

In most countries with available long-term sectoral information, total factor productivity (TFP) tends to be higher in manufacturing than in agriculture for extended periods (but see Martin and Mitra, 2001, discussed below under 'Reallocation'). Unbalanced productivity growth is one of the reasons on the supply side behind the shift in comparative advantage and the transformation of the structure of production during the transition discussed in the following section. The imbalance of TFP notwithstanding, successful industrialization has always been preceded or accompanied by a significant rise in productivity in agriculture. This point appears to be a most general and very significant result. While there are significant differences among the sectoral rates of TFP, these rates tend to be uniformly higher across sectors in countries with good average performance as well as within countries in periods of rapid growth of aggregate productivity. This finding suggests that the overall economic environment, which includes macroeconomic and trade policies, is an important factor in explaining differences in productivity growth.

Changes in the structures of production and employment

Changes in demand and trade reinforce each other. They combine with productivity growth to produce a more pronounced shift in the structures of production and labor use. The share of value added in agriculture declines sharply over the transition, whereas manufacturing, construction and utilities significantly increase their share. The decline in the share of agriculture in employment is more pronounced than in production, but since employment starts from a much higher level and its decline takes place at a relatively higher income level, it leads to a decline in the relative productivity of labor in agriculture. Only at higher levels of income does the trend reverse itself and the gap in average productivity begins to narrow.

Table 4.3 illustrates for the last century the large change in the structure of employment away from agriculture. By 1900 less than half of the labor force in the group of developed countries was still employed in agriculture, a point reached only recently by the less-developed countries (LDCs) as a group. Within this group there are large differences going from a high of 64 percent in sub-Saharan Africa to less than 18 percent in Latin America (ILO, 2006).

From the late 1960s in the more advanced countries and in some of the middle-income countries there has been an ongoing process of deindustrialization. In every single rich country, the share of industry in total output and employment has been going down for several decades now. The trend has agitated politicians and affected sectors even if the welfare implications for the nation have probably been nil. It is not the result of cheap Chinese

Table 4.3 *Sectoral distribution of employment 1900–2004*

	1900	1960	2004
<i>World</i>			
Agriculture	72	58	41
Industry	13	19	20
Services	15	23	39
<i>Developed countries</i>			
Agriculture	48	23	4
Industry	29	36	25
Services	23	41	71
<i>Less-developed countries</i>			
Agriculture	78	71	48
Industry	10	12	20
Services	12	17	32

Sources: 1900 and 1960 Bairoch and Limbor (1968), 2004 ILO (2006).

products but rather of the long-term workings of the normal transformation for an economy that, as it becomes richer, moves from manufacturing into services.

In LDCs structural change before 1960 was minimal (again, there were significant variations within the group). After 1960 there was a substantial decline in agriculture's share, taken up by industry and more so by services. The relative fast increase in services employment has often been interpreted as a failure of the industrial sector to expand employment; in fact the pattern is quite similar to the experience of the advanced countries at similar stages of development (Kuznets, 1957).

Variations in patterns of resource allocation

Divergences from the average patterns described above reflect primarily comparative advantage and its interaction with policy. A relative abundance of natural resources that are economical to exploit at given prices and technology is expected to lead to a high share of primary exports. Although it is difficult to measure the availability of resources, a simple proxy for the proportion of resources to population is the density of the population. A high density has been shown to be significantly associated with lower trade shares and a higher share of manufactured goods in total exports (Perkins and Syrquin, 1989).

Besides the availability of resources, there are various other initial conditions that can influence the patterns of development. The size of the

economy is significantly associated with the share of trade in output. Small countries are generally dependent on trade to a greater extent than are large countries, and they also tend to have higher degrees of concentration in production. The type of specialization in small countries is determined largely by the availability of natural resources and by the policies adopted. The pattern of specialization affects the timing of the transformation, but less so its overall nature. Thus, for example, resource-rich countries also industrialize but with a delay.

The growth experience of the last few decades has highlighted the paradox of resource-rich countries that seem to suffer from a resource curse, while resource-poor countries, such as the East Asian economies, manage to leapfrog and develop successfully without resources. The curse of resources is not located in the availability of resources (*ceteris paribus*, more is still better than less), but in the political economy or rent management. In contrast, the resource-poor East Asian economies succeeded in substituting for the lack of natural resources by their abundance of labor and good policies through the channels of international trade.

Proximate sources of structural transformation

What accounts for the observed changes in industrial structure? The principal proximate factors accounting for this central feature of structural transformation are changes in domestic final demand (Engel effects), the growing intermediate use of industrial products, unbalanced productivity growth and the evolution of comparative advantage as factor proportions change (see Syrquin, 1988, and references there). The fall in the primary share is mostly due to Engel effects at low income levels, and to trade effects afterwards.

The rise in the manufacturing share owes less to high income elasticities than to trade and technology. A more disaggregated analysis would show early import substitution in consumer goods, shifting to producer and capital goods at higher levels of development. The little-noted increase in the overall density of the input–output matrix that accompanies development is especially important in heavy industry (Deutsch and Syrquin, 1989).

In a relatively closed economy the structure of production has to conform closely to the structure of demand, as stressed in the balanced-growth approach of the 1950s. The extent of a country's participation in the international economy is only weakly related to the level of development across countries. The variable most correlated to the share of trade in income across countries is the size of the economy. This relation, among the more robust of the empirical regularities, has until recently been all but ignored by trade models. In small countries the share of trade in GDP is relatively

high, domestic markets relatively small, and the production structure, therefore, tends to be more specialized than in larger countries.

The evolution of comparative advantage and the bias in commercial policies have combined to create an export pattern that reinforces the shift from primary goods into industry, implicit in the pattern of domestic demand.

Resource shifts and productivity growth

The shift of resource among sectors is one of the most important elements of structural transformation. In any case where the values of marginal products of factors are not equal across sectors resource shifts can contribute to aggregate productivity growth and its acceleration at middle-income levels. Paradoxically, the slack in the economy when resources are not allocated efficiently becomes a potential source of growth, and the exhaustion of such slack may explain a slowdown in productivity growth.

Resource shifts are mostly ignored in old and new growth theories. These were central to the empirical analyses of growth of Kuznets (1966) and Denison (1967) during the 1960s. Their studies considered only partial measures of the contribution of resource shifts based on labor productivity. In the more general approach that considers total factor productivity⁵ it can be shown that the measured rate of aggregate productivity growth (λ^*) equals a weighted average of the sectoral rates (λ_i) with output weights ($\rho_i = V_j/V$), plus a factor measuring the effect of intersectoral resource shifts ($RE =$ reallocation effect):

$$\lambda^* = \sum \lambda_i \rho_i + RE \quad (4.3)$$

A positive reallocation effect shows the increase in efficiency when factors move from sectors with lower to sectors with higher marginal productivity, reducing the extent of disequilibrium. The reallocation effect is sometimes referred to by other labels; the 'structural bonus hypothesis' for example (Timmer and Szirmai, 2000).

In the early stages of development the growth of productivity in agriculture lags behind that of other sectors, further widening the productivity gap. The low mobility of resources lies behind the persistence of disequilibrium phenomena such as surplus labor in agriculture and other low-productivity activities, including handicrafts and services. When the industrial sector accelerates its growth in response to domestic demand and to changes in comparative advantage (usually with some help from commercial policies), the productivity gap tends to increase. Labor shifts out of agriculture, first in relative terms and eventually in absolute terms, but with a lag. Since productivity in agriculture rises even at this stage, a surplus of labor results.

The productivity gap between primary production on the one hand and industry and services on the other is greatest in the middle-income range, which is typically the period of greatest inequality of income.

Once migration and capital accumulation have significantly reduced the surplus labor, relative wages in agriculture increase and a catch-up process takes place whereby agriculture begins to reduce the productivity gap. Capital intensity then increases faster in agriculture than in other sectors, and the same seems to have been true of factor productivity in the recent past. Martin and Mitra (2001) report faster TFP in agriculture than in manufacturing for a panel of around 50 countries over the period 1967–92.

In advanced countries since the late 1960s the shift out of agriculture is no longer a potential source of growth. A different – and probably negative – allocation effect has been suggested as labor now shifts into services.

Most empirical studies find the contribution of resource reallocation to productivity and growth to be significant. But most also note with surprise the lower than expected magnitude of the effect. The principal reason, as Kuznets has already pointed out, is the inadequacy until recently of the available information – a seemingly technical issue with deep significance, as we shall see.

There are two aspects to this problem: insufficient disaggregation and ignoring quality changes, in particular, new goods and varieties. The broad definition of sectors, even in fairly disaggregated studies, hides all factor reallocations within those sectors. In a study about Taiwan, Kuznets (1979) made the convincing argument that the high rate of growth of product per worker, required ‘a much greater rate of shift [than the] one now suggested in the three-sector classification and that the shifts from old to new subbranches within these sectors are particularly neglected’ (p. 73). Aggregation is a problem, but not the only one. On the issue of quality, Kuznets argued that: ‘it is frustrating that the available sectoral classifications fail to separate new industries from old, and distinguish those affected by technological innovations . . . As a result, both the true rate of shift in production structure and its connection with the high rate of aggregate growth are grossly underestimated’ (Kuznets, 1971, p. 315). New products do not just substitute for old ones, but they tend to increase the variety of similar goods commonly grouped under the same classification.

Insufficient disaggregation: ‘Creative Destruction’

A fact not much noted is that most of the data on structural change, even at a disaggregated sectoral level, are net changes and hide the enormous turmoil at the firm or job level. For example, jobs are continuously being created and destroyed, but the figures we mostly work with give only the net change in employment which is often a small fraction of job turnover.

Recently, thanks to the growing availability of large micro data sets on firms and employment and to the greatly expanded computational capacity (Clark and Kuznets worked with desk calculators, at best), our information about the dynamics of the process of creative destruction has steadily expanded. Originally it focused on labor economics and macroeconomics of developed countries, spreading recently to developing countries spurred by the interest in the micro turmoil accompanying adjustment, transition and liberalization. These studies clearly illustrate the importance of resource reallocation for growth and the large degree of underestimation of those effects when we deal with relatively aggregate data.

The following is a summary of results, many for the USA and other developed countries, relevant for structural change.

Gross flows are large

Focusing on total employment and unemployment ignores the fact that changes in employment are the tip of the iceberg of what is a most active process of job creation and job destruction. For example, in the USA during the 1990s about 30 million jobs were destroyed each year but more than 32 million jobs were newly created every year. The sum of job creation and job destruction as a share of average employment is defined as the turnover rate. Very large figures for gross job flows and turnover rates have now been documented for various countries, industries and time periods (see Davis et al., 1996).

In a sample of 12 Latin American countries turnover rates ranged from 16 to 35 percent (IADB, 2004, Chapter 2). Somewhat lower rates, but still evidencing substantial churning, were found for manufacturing in post-socialist countries during the 1990s (De Loecker and Konings, 2006).

Reallocation within sectors may be more important than between sectors

Much of the reallocation of factors across producers takes place within sectors rather than between sectors. While this depends on the level of aggregation, reallocation rates are high within four-digit industries, regions and other subdivisions considered (Davis et al., 1996). There is even some evidence that among large firms in the USA the reallocation is intra-firm (Schuh and Triest, 1999).

Costs

Labor reallocation in the early stages of Modern Economic Growth involved massive displacement of population associated with the process of urbanization. Reallocation today does not necessarily imply even a change in housing, but this does not mean that adjustment costs are low. When focusing on the net flows and on unemployment we underestimate the costs

associated with gross flows which, as mentioned above, are an order of magnitude larger. Also, not every shifter is a winner, especially where workers are not well protected.

Growth and productivity

Modern economies need constantly to reallocate labor and other resources from old to new firms and products; the process of trial and error is important for growth. In a sample of ten countries, eight developed plus Chile and Colombia, labor reallocation from existing firms to more productive new entrants accounts for between 20 and 40 percent of total productivity growth (IADB, 2004, Chapter 2). Similarly, the creative destruction process in the Slovenian manufacturing sector was more the result of new, more productive firms replacing unproductive ones than of reallocation of employment between existing firms (De Loecker and Konings, 2006).

Encouraging firm entry and exit is important to enhance aggregate productivity but it conflicts with job security. Blanchard (2005), in analyzing labor markets in Europe, points out that governments instead of protecting workers have protected jobs and thus increased labor market rigidities and slowed the reallocation process. This resulted in more unemployment and lower productivity growth.

The new research on the dynamics of labor markets has contributed to clarify the question of whether recessions are conducive to higher productivity given the importance of reallocation from less to more efficient firms. Caballero and Hammour (2000), refute the idea of 'cleansing recessions', arguing that jobs destroyed during recessions flow mostly into unemployment or the informal sector, not into newly created jobs. Crises freeze the restructuring process because of the tight financial conditions which reduce the ability to finance creation of new units.

The last point is worth stressing for the general case of reallocation: reallocation may contribute to higher productivity when the shift is from lower- to higher-productivity units but the gains are not automatic. A sectoral gap in labor productivity, for example, indicates a potential gain from reallocation, but the gain will not materialize by the simple shift of a worker among sectors without complementary investments in human and other capital. That is, the resource shifts may not be feasible without high rates of investment and growth.

Quality changes

New goods

If we compare the typical basket of a consumer today with the typical basket in 1900, for example, we would find that well over 50 percent of the

commodities we consume today did not exist then at all. What exactly is the meaning of models of unchanged sectors when the type or identity of the output is changing drastically? This process underlies the explanation of Kuznets (1930) and Burns (1934) of their finding of retardation at the product or sectoral level with continuous replacement and the semblance of stability at the aggregate level.

Nordhaus (1997) divides today's consumption bundle into three categories according to how great and drastic the quality change has been since the beginning of the nineteenth century. The three groups and their weight in today's basket are:

1. Relatively small changes (clothing, haircuts, furniture): 27.7 percent
2. Major changes but still recognizable (housing and watches): 35.8 percent
3. Radical changes – new goods (appliances, medical care, transportation): 36.6 percent

The increase and variety has led to large increases in welfare. A recent rough estimate for the USA over the last two centuries puts the annual gain due to new consumer goods associated with technological progress at between 1.5 percent and 10 percent annually (Greenwood and Uysal, 2004).

The previous sections illustrate the difficulty of separating the growth in productivity from structural changes. In a modern economy, the former is virtually impossible in the absence of structural shifts. This is relevant to the argument, prevalent in the mid-1990s, that the success of the East Asian economies could be ascribed to factor accumulation and not to the growth in factor productivity. An alternative to this 'accumulation' approach is an 'assimilation' explanation whereby fast growth comes through the replacement of older processes by newer processes, which to become assimilated require an upgrading of skills and substantial structural changes (see Nelson and Pack, 1999). Once again, structural changes require high rates of physical and human capital.

Varieties

The Schumpeterian notion of creative destruction, when formally modeled, has often been incorporated into one sector growth models where all the drama is one of replacement within set categories.

There is ample information on quantities of goods produced and consumed, but little on product variety. Changes in technology in production, information, and distribution and the fall in transaction costs have increased the ability to customize products to consumers' tastes (Cox and Alm, 1998). The differentiation of products implied by the increase in

varieties is, together with economies of scale, the basis for explaining the increase in intra-industry trade in manufacturing among industrial countries. With the intensification of globalization, international trade has rapidly expanded into more products. But much of the structural change involved is still missed by stopping the degree of disaggregation at the firm or enterprise level.

In a remarkable example of the trend toward disaggregation in industrial economics and international trade, Bernard et al. (2005) show that ‘much of the expansion and contraction of firms is accounted for by the extensive margin of adding and dropping products rather than the intensive margin of changing output of existing products’ (p. 3). They then argue that:

existing measures of creative destruction may understate the true amount of reallocation between activities because they focus on firm rather than product market entry and exit. Because reallocation of resources within firms may involve lower transactions costs than a re-matching of factors between firms, product switching may play an important role in allowing industries and economies to adapt more efficiently to external shocks. (p. 30)

What sectors?

To conclude, I want to mention two conceptual issues not often discussed which will probably become more important in the future. How do we define the components of the aggregate and what should go into that aggregate? Or, what are sectors and is GDP still relevant as our aggregate measure?

Sectors

The division into components must have an analytical basis and the ‘sectors’ must differ significantly from each other. Dutt (1990) is still a very useful survey of various definitions of sectoral balance in development. Among the criteria for distinguishing sectors we find demand characteristics (elasticities), production (factor intensities, final or intermediate), tradability, location, and so on. The industrial classification in this paper can be traced back to Petty in the seventeenth century at least; however, in its modern version it probably originates with Colin Clark’s three-way division. The value of this sectoral classification was questioned for low-income countries primarily by Bauer and Yamey (1951), who argued that in the early stages of development there is insufficient differentiation as individuals are engaged in a variety of trades, and it is therefore difficult to assign them to any one occupation. With modern economic growth the size of the market increases and with it the degree of specialization and the differentiation of occupations. The difficulty – an empirical one – was expected to diminish as development proceeds. Fifty years later the

argument of insufficient differentiation has resurfaced. It is argued now that the line between manufacturing and services is blurred and is expected to become more so. With new technologies much of what used to be 'services' is becoming part of what used to be 'manufacturing'. But equally, much of employment growth in services reflects contracting out (outsourcing) of work previously done by manufacturing (see *The Economist*, 1998).

The line has sometimes also blurred between agriculture and the combined services–manufacturing sector. Thus much of the primary exports of Chile, for example, reflect sophisticated processing and marketing techniques. If exports of manufactures used to be the hallmark of a successful transformation, this needs now to be modified to take such developments into account.

Not only is the identity of a sector becoming cloudy, but determining its national location is also becoming more difficult if not impossible. As *The Economist* (1998) emphasizes, whole industries no longer migrate, as ship-building did from Europe to Asia in the 1970s; manufacturing is becoming a genuinely international affair.

The denominator: GDP?

For a long time the principal measure of the level of development of a country has been GDP per capita. It was always clear that this measure was, at best, a partial measure but arguably the best partial measure available. Given that GDP does not capture all benefits, such as increases in leisure and life expectancy, nor all costs such as environmental degradation, and not even all production such as subsistence production, how should we proceed?

There is a need to re-evaluate the Kuznetsian themes of delimiting what enters into the economic calculation and where to draw the dividing line between final and intermediate outputs. This has a close relation to the question of what are the benefits and the costs of economic growth. For example, it has been argued that the increased life expectancy and health status over the twentieth century in advanced countries has raised welfare as much as the increase in purchasing power.

In his 1971 *Economic Growth of Nations*, Kuznets refers to non-conventional costs and argues that a significant part of these are due to deficiencies in the 'conventional national economic accounting that treats some outputs that are really cost of production as *final* rather than as *intermediate* products' (1971, p. 75). We have reached this point, in part, by relying on national income measures well suited as indexes of short-term changes but not adequate as gauges of economic growth. In addition to reclassification of measured quantities, Kuznets considers various 'hidden' costs such as the time spent in commuting to work, and air and water

pollution. We could go back to Kuznets and other pioneers of national income accounting; some of these issues were thoroughly discussed by them at the time. Instead, there is a tendency to denigrate GDP and related measures as indicators of progress or achievement.

Structure is still an essential ingredient in studying development. But we urgently need to clarify 'structure of what'? Replacing GDP with a multi-dimensional index may make the concept vacuous or could politicize the choice of indicators. The three indicators in the Human Development Index (HDI), while highly correlated, yield an easily comprehended measure and are not really controversial. Not so more recent attempts to broaden substantially the composite measure of progress or happiness. Recent suggestions for broadening the HDI can clearly be identified as 'left' or 'right'; the former lists among the determinants of happiness or progress union participation and various measures of empowerment and inclusion, while the latter would consider family, community and faith.

Notes

1. Machlup (1963) is still the best source for the various ways in which the terms have been used and abused in economics.
2. But see Morris and Adelman (1988) and the recent work of Acemoglu et al. (2002). North (1981) interprets structural change in economic history as institutional change, but almost completely omits shifts in the structure of production and factor use.
3. See Chenery et al. (1986), Syrquin (1988) and Syrquin and Chenery (1989).
4. A notable exception is Pasinetti's classical model of growth and transformation (1981). It remains to be implemented empirically.
5. See Syrquin (1984) for a comparison of approaches to measure the contribution of inter-sectoral resource shifts.

References

- Acemoglu, Daron, Simon Johnson and James A. Robinson (2002), 'Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distribution', *Quarterly Journal of Economics*, 117: 1231–94.
- Bairoch, P. and J.-M. Limbor (1968), 'Changes in the Industrial Distribution of the World Labour Force, by Region, 1880–1960', *International Labour Review*, 98 (4): 311–36.
- Balassa, B. (1979), 'A "Stages Approach" to Comparative Advantage', in I. Adelman (ed.), *Economic Growth and Resources, Vol. 4: National and International Policies*, London: Macmillan.
- Bauer, P. and B.S. Yamey (1951), 'Economic Progress and Occupational Distribution', *Economic Journal*, 61: 741–55.
- Bernard, Andrew B., Stephen Redding and Peter K. Schott (2005), 'Multi-Product Firms and the Dynamics of Product Mix', Tuck School of Business at Dartmouth, mimeo, July.
- Blanchard, Olivier (2005), 'European Unemployment: The Evolution of Facts and Ideas', NBER Working Paper 11750, Cambridge, MA.
- Burns, Arthur (1934), *Production Trends in the United States since 1870*, New York: National Bureau of Economic Research.
- Caballero, Ricardo J. and Mohamad L. Hammour (2000), 'Creative Destruction and Development: Institutions, Crises, and Restructuring', NBER Working Paper 7849.
- Chenery, Hollis B., Sherman Robinson and Moshe Syrquin (1986), *Industrialization and Growth: A Comparative Study*, New York: Oxford University Press.

- Cox, W. Michael and Richard Alm (1998), 'The Right Stuff: America's Move to Mass Customization', in *1998 Annual Report*, Federal Reserve Bank of Dallas.
- Davis, Steven, John Haltiwanger and Scott Schuh (1996), *Job Creation and Destruction*, Cambridge, MA: MIT Press.
- De Loecker, Jan and Jozef Konings (2006), 'Job Reallocation and Productivity Growth in a Post-Socialist Economy: Evidence from Slovenian Manufacturing', *European Journal of Political Economy*, 22: 388–408.
- Denison, Edward F. (1967), *Why Growth Rates Differ: Postwar Experience in Nine Western Countries*, Washington, DC: Brookings.
- Deutsch, Joseph and Moshe Syrquin (1989), 'Economic Development and the Structure of Production', *Economic Systems Research*, 10: 447–64.
- Dutt, Amitava Krishna (1990), 'Sectoral Balance in Development: A Survey', *World Development*, 18: 915–30.
- Easterlin, Richard A. (2001), 'Review of Simon Kuznets *Modern Economic Growth: Rate, Structure and Spread*', *Economic History Services*, 29 October, URL : <http://eh.net/bookreviews/library/easterlin.shtml>.
- Echevarria, Cristina (1997), 'Changing Sectoral Composition Associated with Economic Growth', *International Economic Review*, 38: 431–52.
- The Economist* (1998), 'The World as a Single Machine', *A Manufacturing Survey*, 18 June, s3–s6.
- Finger, J. Michael (1975), 'Trade Overlap and Intra-Industry Trade', *Economic Inquiry*, 13: 581–9.
- Greenwood, Jeremy and Gokce Uysal (2004), 'New Goods and the Transition to a New Economy', NBER Working Paper 10793.
- IADB (Inter-American Development Bank) (2004), *Good Jobs Wanted: Labor Market in Latin America*, Washington, DC: Inter-American Development Bank.
- International Labor Office (ILO) (2006), *Global Employment Trends Brief*, January, Geneva.
- Kuznets, Simon (1930), *Secular Movements in Production and Prices: Their Nature and their Bearing upon Cyclical Fluctuations*, Boston, MA and New York: Houghton Mifflin.
- Kuznets, Simon (1957), 'Quantitative Aspects of the Economic Growth of Nations: II. Industrial Distribution of National Product and Labor Force', *Economic Development and Cultural Change*, Supplement to 5 (4): 3–111.
- Kuznets, Simon (1966), *Modern Economic Growth*, New Haven, CT: Yale University Press.
- Kuznets, Simon (1971), *Economic Growth of Nations: Total Output and Production Structure*, Cambridge, MA: Harvard University Press.
- Kuznets, Simon (1979), 'Growth and Structural Shifts', in W. Galenson (ed.), *Economic Growth and Structural Change in Taiwan*, Ithaca, NY: Cornell University Press, pp. 15–131.
- Lewis, W. Arthur (1954), 'Economic Development with Unlimited Supplies of Labor', *Manchester School*, 22: 139–91.
- Machlup, F. (1963), *Essays in Economic Semantics*, Englewood Cliffs, NJ: Prentice-Hall.
- Maddison, Angus (2004), 'Macromasurement Before and After Colin Clark', available at <http://www.ggdc.net/Maddison/maddison.htm>. This is a longer version of 'Quantifying and Interpreting World Development: Macromasurement Before and After Colin Clark', *Australian Economic History Review*, March 2004: 1–34.
- Martin, Will and Devashish Mitra (2001), 'Productivity Growth and Convergence in Agriculture and Manufacturing', *Economic Development and Cultural Change*, 49: 403–22.
- Morris, Cynthia T. and Irma Adelman (1988), *Comparative Patterns of Economic Development, 1850–1914*, Baltimore, MD: Johns Hopkins University Press.
- Nelson, Richard and Howard Pack (1999), 'The Asian Miracle and Modern Growth Theory', *Economic Journal*, 109: 416–36.
- Nordhaus, William (1997), 'Do Real Output and Real Wage Measures Capture Reality? The History of Lighting Suggests Not', in Timothy Breshnahan and Robert Gordon (eds), *The Economics of New Goods*, Chicago, IL: University of Chicago Press, pp. 29–70.
- North, Douglas C. (1981), *Structure and Change in Economic History*, New York: Norton.
- Pasinetti, Luigi (1981), *Structural Change and Economic Growth*, Cambridge: Cambridge University Press.

- Perkins, Dwight H. and Moshe Syrquin (1989), 'The Development of Large Countries', in H.B. Chenery and T.N. Srinivasan (eds), *Handbook of Development Economics*, Vol. 2, Amsterdam: North-Holland, pp. 1691–1753.
- Schott, Peter K. (2003), 'One Size Fits All? Heckscher-Ohlin Specialization in Global Production', *American Economic Review*, 93: 686–708.
- Schuh, Scott D. and Robert K. Triest (1999), 'Gross Job Flows and Firms', Federal Reserve Bank of Boston Working Paper No. 99-10.
- Syrquin, Moshe (1984), 'Resource Reallocation and Productivity Growth', in Moshe Syrquin, Lance Taylor and Larry E. Westphal (eds), *Economic Structure and Performance: Essays in Honor of Hollis B. Chenery*, New York: Academic Press, pp. 75–101.
- Syrquin, Moshe (1988), 'Patterns of Structural Change', in H.B. Chenery and T.N. Srinivasan (eds), *Handbook of Development Economics*, Vol. 1, Amsterdam: North-Holland, pp. 203–73.
- Syrquin, Moshe and Hollis B. Chenery (1989), 'Three Decades of Industrialization', *World Bank Economic Review*, May: 145–81.
- Timmer, Marcel P. and Adam Szirmai (2000), 'Productivity Growth in Asian Manufacturing: The Structural Bonus Hypothesis Examined', *Structural Change and Economic Dynamics*, 11: 371–92.

5 Interdisciplinary approaches to development: the ‘institutional’ turn

Peter Evans

Introduction

Institutional approaches to the study of development now dominate the mainstream of development economics. In other social science disciplines they have long predominated. Concepts that used to play the dominant explanatory models within economics – like capital accumulation and technological progress – have been displaced by ‘institutions’. Acknowledging the magnitude of the shift is the first step toward analyzing its implications.

Pranab Bardhan (2005, p. 1) sums up the change parsimoniously:

In the field of development economics, earlier preoccupations with the forces of capital accumulation or technological progress have been widely replaced by a belief that the institutional framework of an economy is crucial for the understanding of the process of development.

In their contribution to the *Handbook of Economic Growth*, Acemoglu et al. (2005, p. 1) pull no punches: ‘differences in economic institutions are the fundamental cause of differences in economic development’. Dani Rodrik, in a co-authored paper (2004)¹ called ‘Institutions Rule’, is equally straightforward:

the quality of institutions ‘trumps’ everything else. Once institutions are controlled for, measures of geography have at best weak direct effects on incomes . . . Similarly, once institutions are controlled for, trade is almost always insignificant.

Easterly and Levine (2003) offer further support for the primacy of institutions. All of these arguments come from the methodological mainstream of modern economics. They are supported by other quite different, but equally mainstream, points of view. Hoff and Stiglitz, for example, take a parallel position but use the language of ‘organizations’. In their view (2001, p. 389): ‘Development is no longer seen primarily as a process of capital accumulation, but rather as a process of organizational change’.

There are, of course, still dissenters who resist the dominance of the institutional perspective. Jeff Sachs and his collaborators continue to push geography and disease as fundamental causes of differences in national

wealth and incomes (Gallup et al., 1998; Sachs, 2001). Engerman and Sokoloff (1997, 2002) are more restrained but argue that current explanatory frameworks have gone overboard in neglecting the way in which institutions are themselves shaped by natural factor endowments.

There is merit in these dissenting points of view, but the institutional turn is highly unlikely to be reversed. Even if endowments, geography and disease gained purchase at the level of cross-national analysis, which they do not seem to be doing, these approaches would still be at a fatal disadvantage. The logic of institutional analysis can be replicated at different levels of analysis, ranging from the very powerful district-level comparisons recently executed by Banerjee and Iyer (2002) using Indian data, to the carefully designed micro-level research of the new generation of empirically oriented development economists, like Stefan Dercon, Erica Field and Ted Miguel. The possibility of reinforcing macro, national-level comparisons with complementary results at the regional, community and organizational level makes the institutional approach theoretically richer and more empirically compelling than its rivals.

Institutional approaches also offer more fruitful forms of engagement with policy debates than natural endowment-based theories. Institutions can be constructed and reconstructed; natural endowments and geography must be lived with. Even if initial disadvantages are created by endowments (including 'negative endowments' like disease burden), it is implausible that strategies for ameliorating such disadvantages can be successful in the absence of institutional transformation. Current global policy paradigms overestimate the malleability of institutions, sometimes disastrously (see Evans, 2004), but endowment-based analyses cannot negate the policy centrality of institutional analysis.

If the persistent future dominance of institutional approaches to development within mainstream economics is a reasonable premise, then it follows that future debates, both theoretical and empirical, over the dynamics of development will take place on the terrain of institutional analysis, with mainstream economic versions of institutional analysis continuing to play a prominent role, probably a dominant one.

The institutional turn represents recognition that traditional economic models were not working (cf. Easterly, 2001). At the same time, the institutional turn represents a conviction that the formal analytical tools that have been developed within economics can fruitfully be applied to other institutions, particularly those involving political power.

My argument here is, first of all, that the institutional turn in its early 'Northian' manifestations was undertheorized. 'Property rights' was forced to carry far too heavy an explanatory burden. Reducing historical trajectories to 'property rights' can too easily represent thinly disguised

intellectual imperialism in which the simplest kind of economic rationality is smuggled back into the analysis under the guise of being an 'institution'. I will also argue, however, that the application of traditional analytical tools to new terrain can reveal the necessity of very different kinds of explanations. This kind of revelation, even if initially unacknowledged, holds promise for extending the institutional turn in ways that could substantially enhance our understanding of development from an interdisciplinary perspective.

First, I will look briefly at the problems of an undertheorized, property rights version of the institutional turn. Then I will turn to the way in which the property rights perspective becomes transformed in practice, when applied to historical data. I will focus particularly on the paradigmatic work of Acemoglu, Johnson and Robinson in this second part of the discussion. To show how this transformation can be further extended, I will use the interaction of the work of Robinson with the work of a political scientist and sociologist, James Mahoney, in the more specific historical context of nineteenth-century Central America. Finally, I will use Acemoglu, Johnson and Robinson's analysis of the case of Botswana, especially in contrast to the sociological analysis of Ann Swidler, to make the case for two additional extensions. I will close by summarizing some of the lessons and challenges for the extension of the institutional turn that emerge from this brief review of a sample of existing work.

Institutions, property rights and development

One of the burdens of the institutional approach is that no one from any discipline can offer a definition of institutions that offers clear guidance for a research program. We could turn, for example, to the definition that Douglass North (1994, p. 360), offered in his Nobel Prize lecture:

The rules of the game: the humanly devised constraints that structure human interaction. They are made up of formal constraints (such as rules, laws, constitutions), informal constraints (such as norms of behavior, conventions, self-imposed codes of conduct), and their enforcement characteristics.

Or, we could take the definition that Ha-Joon Chang and I (2005) offered in a recent paper:

Institutions are systematic patterns of shared expectations, taken-for-granted assumptions, accepted norms and routines of interaction that have robust and durable effects on shaping the motivations and behavior of sets of interconnected social actors.

The two definitions are quite different, but they are similarly unspecific. Neither offers, in itself, the basis for a research program. The innumerable

other definitions of institutions that have been produced by social scientists across a range of disciplines have the same character. We lack a coherent general theory of how institutions work, within which specific analyses of development institutions might be embedded.

In practice, cross-national institutional analyses of development use a 'double finesse' to work around their conceptual problems. On the one hand, they tend to use very simple, concrete empirical proxies to stand for complex combinations of institutions. Perhaps most popular are the various measures put out by commercial 'political risk' services such as those embodied in the ratings of the *International Country Risk Guides* (Knack and Keefer, 1995). Exactly what 'institutions' are reflected in these measures is difficult, indeed often impossible to figure out, but they are available for a full range of countries at varying points in time. This element of the finesse is the key to empirical feasibility.

The second element of the finesse is on the theoretical side. The specific concrete measures used are assumed to reflect 'institutions' at an abstract level, reflecting the aggregate character of a whole complex of institutions – usually defined as 'property rights institutions'. It is a generically plausible finesse and an essential one given the low face validity of the empirical measures used.

The theoretical finesse rests on a simple and very plausible logic in which propensities to make productive investments depend on the predictability of future rights to claim the returns from those assets. If people cannot count on maintaining future control of assets that they consider theirs, then investing in productive assets whose benefits are only accrued in the future makes less sense. Income consumed is hard to take away and hoarded assets are easier to defend than productive ones (which must be exposed to public view to reap their benefits). A combination of consumption and hoarding makes more sense than investment when assets are insecure.

The idea that the predictability of future societal rules and circumstances is essential if people are to engage in productive investments, makes sense. The idea that property rights can be thought of as a simple ordinal scale along which concrete historical circumstances can be arrayed is anything but plausible. Any initial allocation of rights to different kinds of property – ranging from land to the broadcast spectrum to the human genome – is not just complex but also disputable and somewhat arbitrary. Enforcement of rights once they have been allocated is equally so. Sending the National Guard to evict peasants who are growing crops on a landlord's otherwise unused land is enforcing property rights. So is shutting down a factory whose pollution is making the surrounding neighborhood unlivable. Development almost certainly depends on how property rights are allocated and what kind of property rights are enforced for what segments

of the population. Exactly how these complex patterns of allocation and enforcement are related, positively or negatively, to development can hardly be taken for granted.

In short, neither the empirical proxies for institutions that are used in most cross-national institutional analyses of development, nor the reliance on a simple notion of 'effective property rights' constitutes an ordinal scale on which societies can be arranged, or appears to offer promising foundations for the institutional turn. Nonetheless, even broad quantitative cross-national institutional analyses have managed to generate intellectually exciting debates which belie the apparent foundational weaknesses of the approach. The collaboration of Daron Acemoglu, Simon Johnson and James Robinson offers a paradigmatic example.

A paradigmatic example of the institutional turn

Acemoglu, Johnson and Robinson (hereafter AJR) have produced a prolific set of institutional analyses of development (for example, 2001, 2002, 2003, and 2005). Here, I will use their already classic 2001 article in the *American Economic Review* as a starting point, in part because it conforms to the general 'double finesse' model that I have just laid out, but more importantly because it demonstrates the tendency for high-quality analysis that begins from the double finesse to transcend it.

AJR (2001, p. 27) admit that in their paper 'Institutions are treated largely as a "black-box"'. Ironically, it is their effort to solve the methodological problem of endogeneity that leads them to undertake a much more historically oriented analysis than would have been the case had they focused simply on the contemporary relationship reflected by their primary measure. AJR have been stimulated by their historical instrumental variable to open up the institutional 'black box' in interesting and potentially fruitful ways. Their basic argument is that where there were large amounts of resources (mineral deposits or land suitable to crops in high demand on world markets) and large indigenous populations to exploit, colonialists created 'extractive institutions'. Where settlers had to survive largely on the basis of their own efforts, 'institutions of private property' emerged.

These basic concepts reoccur in AJR's subsequent work. For example in their next paper (AJR, 2002, p. 17), 'institutions of private property' are defined as 'a cluster of (political, economic and social) institutions ensuring that *a broad cross-section of society* has effective property rights' (emphasis added). In a 2003 paper they explicitly divide the requirements for effective property rights into two components. The first is the traditional Northian general provision of secure property rights. The second is the requirement that such rights are extended to a 'broad cross-section of the society'. Thus, they argue, a society in which a 'small fraction of the

population' monopolizes control of property does not fully qualify as having 'institutions of private property', 'even if the property rights of this elite are secure' (AJR, 2003, p. 5). The implications of this perspective become clearer when the analysis focuses in at the regional level.

A regional laboratory for comparative institutional analysis

Central America offers a fascinating comparative microcosm for examining questions of institutions and growth. Five countries share a similar colonial heritage, history of commodity exports and geopolitical position.² Yet, once cut loose from the formal control of the Spanish empire at the beginning of the nineteenth century, they have strikingly different institutional histories and levels of economic success.

Robinson's analysis of Central America (done jointly with Jeffrey Nugent) focuses on a paired comparison of four coffee producers: Costa Rica and Colombia on the one hand and Guatemala and El Salvador on the other. (Colombia is, of course, not technically speaking part of Central America, but its shared colonial history and the importance of coffee in its economy make it a reasonable addition.) Coffee became the major export crop for all four countries during the latter part of the nineteenth century. Yet, the first two ended up with roughly double the incomes, and much higher levels of human development than the other two. Nugent and Robinson's argument is straightforward. They start from the fact that the primary difference between the first and second sets of countries is that in the former pair of countries, smallholders play a major or even dominant role. In the latter pair, coffee production is dominated by large landholders. In short, in Guatemala and El Salvador AJR's second requirement for institutions of private property, the 'broad cross-section' requirement, is violated.

Nugent and Robinson do not pursue the institutional determinants of political competition in any depth, but the kind of institutions that they might have focused on had they done so is suggested by one of Robinson's other collaborative efforts. In a very elegant article on the economic consequences of the introduction of the secret ballot in post-World War II Chile, Baland and Robinson (2003) provide a concrete confirmation and extension of the Nugent and Robinson perspective. Baland and Robinson's analysis demonstrates that in the Chilean case, it was not just an initial allocation of land rights that was key to the economic returns of landlords, but also the persistence of specific political institutions (that is, the absence of the secret ballot) that reinforced landlords' control over those who worked the land.

Comparing Nugent and Robinson's analysis with a political science perspective on Central America reveals an additional challenge to the

elaboration of a convincing political complement to AJR's version of the institutional turn. In outlining the political dynamics of elite strategies in nineteenth-century Central America, Nugent and Robinson rely heavily on James Mahoney's 2001 book *Legacies of Liberalism*. It is, therefore, interesting to examine the way in which Mahoney's political science training results in a different reading of the process, based on essentially the same historical evidence. Mahoney focuses on Central America per se and therefore does not include Colombia. Mahoney also includes Honduras and Nicaragua. I will leave them out here in order to maximize the parallels between his analysis and Nugent and Robinson's. Mahoney's interpretation of the contrast between Costa Rica on the one hand and El Salvador and Guatemala on the other hand parallels Nugent and Robinson's paired comparison but also differs in key respects.

Mahoney has a different view of how institutional change works, one which emphasizes both political agency, especially during what he calls 'critical junctures', and the subsequent effects of the institutional legacies generated by choices made during these 'critical junctures'. Like Nugent and Robinson, Mahoney emphasizes the role of political competition, but he evaluates differently both its relative intensity in different countries and its effects. In his view, nineteenth-century liberals in both Guatemala and El Salvador became 'radicals', wiping out non-market forms of land tenure much more thoroughly and promoting the legal right of large landowners to control both land and labor much more aggressively than the 'reformist' liberal elites of Costa Rica. Up to this point he is consistent with Nugent and Robinson, but, in contrast to Nugent and Robinson, he sees the motivation of elite choices as not simply, or even primarily, the promotion of the interests of large landholders. These elites generally did promote the interests of large landholders, but, according to Mahoney, the primary attraction of radical (as opposed to more reformist) strategies was that radical strategies appeared to be the most effective way of gaining and consolidating political control in the face of staunch opposition from conservative forces such as the church and its traditionalist allies.

Mahoney also sees the construction of new national state apparatuses with vastly expanded powers of coercion as the heart of the institutional agenda of radical liberalism in Central America. Hence the results of radical liberalism were not simply the polarization of rural class structures but also the emergence of powerful military-coercive state apparatuses. These two features went together: enforcing extreme polarization required a larger, more coercive military, and the conflicts generated by polarization increased the centrality of the military's position within national political institutions.

While the construction of military coercive apparatuses were the hallmark of radical liberalizing regimes in Central America, Mahoney points

out that state-building was an essential element for the export-led growth projects of both radical and reformist liberalizing regimes.

For Mahoney, the contrasting institutional forms which emerged in nineteenth-century Central America cannot be read as simply transmission belts for previously defined interests, but must be considered as 'new facts' with causal weight of their own. Mahoney gives extra causal weight to institutions in a second way as well. He suggests that once created, institutions may take on a life of their own.

Mahoney argues that, once in place, Central America's military coercive apparatuses began to have a set of preferences that went beyond those of the economic elites to whom they were connected, preferences that focused particularly on the preservation of the military's own power and privilege. In addition to new interests, there were new capacities created. These new capacities made some future outcomes possible that otherwise would have been unlikely, and made other future outcomes, which otherwise would have been real possibilities, very difficult to achieve.

In the end, the dynamics of Mahoney's 'critical juncture-legacy' model of institutional change is very different from the 'institutional persistence' model of AJR and Nugent and Robinson. Mahoney's basic methodological perspective might be summarized as follows. Firstly, instead of being determined by prior constellations of endowments and interests, institutions emerge out of uncertain, politically motivated choices, made primarily during 'critical junctures' when developmental possibilities are in flux. It is, therefore, impossible to exclude consideration of agency from the analysis of institutions. Secondly, institutions become embodied in new organizations and sets of social actors, creating new interests and capacities. These embodiments are central to the long-run effects of institutional change but are unlikely to be anticipated when the institution initially emerges. The organizational embodiments of institutional change must, therefore, be considered as causal factors in their own right.

In addition to suggesting a different theoretical perspective on institutional change, Mahoney's analysis offers substantive contributions to AJR's characterization of 'good institutions', confirming one of AJR's propositions and adding two more. Mahoney confirms the centrality of AJR's 'broad cross-section' requirement. He adds two propositions regarding state-building. First, Mahoney's analysis suggests that state involvement and, therefore, state-building, was essential to liberal strategies of economic growth, thus, supporting a 'development state' perspective. Second, Mahoney suggests that the negative effects on subsequent growth of violating the 'broad cross-section' requirement depend on and are crucially reinforced by the hypertrophy of the military-coercive side of the state-building.

In the Central American context, neither the political choice component nor the ‘state-building’ side of Mahoney’s institutional analysis appear to have a counterpart in AJR’s version of the institutional paradigm. If, however, we turn our attention to another case that has been the focus of AJR’s work – the surprising economic success of Botswana – it is clear that state-building and political choices also play a central role in AJR’s understanding of ‘good institutions’.

The success of ‘institutions of private property’ in Africa

Like many other analysts of Botswana, AJR call it ‘An African Success Story’. The data certainly support this view. From the 1970s through to the end of the twentieth century, Botswana’s gross domestic product (GDP) per capita grew at a rate that made it look as though it was part of East Asia. Its performance is particularly striking when compared to that of its neighbors in Southern Africa. Its purchasing power parity (PPP) GDP is four times the average for Southern Africa. AJR attribute this success, not surprisingly, to ‘institutions of private property’.

Once again, the interesting question is, ‘What do we mean by “good institutions?”’ Providing appropriate incentives to local private investors seems to have had little to do with Botswana’s success. Botswana’s political leaders were able, early on, to secure a contract with a transnational diamond mining company that gave the government 50 percent of all export revenues. This, in turn, allowed the government to maintain a reasonably well-paid, meritocratic bureaucracy in which ‘probity, relative autonomy and competency have been nurtured and sustained’ (Parson, 1984, quoted in AJR, 2003). About 40 percent of all formal sector jobs are in public service, and the government invests a larger share of public expenditure in education than either the USA or Canada.

AJR’s Botswana narrative is much closer to Mahoney’s emphasis on the choices of political leaders who controlled the state apparatus. While AJR focus on the continuity of political institutions, they also emphasize that Botswana’s good institutions were ‘reinforced by a number of critical decisions made by the post-independence leaders, particularly Presidents Khama and Masire’ (AJR, 2003, p. 1). AJR’s Botswana story also sounds more like Mahoney’s when it comes to the importance of state-building. Of course, the Botswana government plays a much more central economic role because of the centralized character of the key resource endowments. If we look at AJR’s description of Botswana, it might be described as a ‘resource-based mini-developmental state’. Both its institutional character and policies are, of course, different from those of classic industrially based developmental states like Korea or Taiwan. Once again, endowments matter, but how they matter depends on institutions.

One way then of reading the Botswana case is that it demonstrates that a combination of political constraint, effective state-building and abundant resources can obviate the necessity of conforming to the 'broad cross-section' requirement. There may, however, be limits to the benefits of quietly maintained traditional hierarchies. The reversal of fortune which Botswana suffered in the 1990s as a result of HIV/AIDS suggests that quietly maintained hierarchies may result in a state apparatus that lacks practice and aptitude when it comes to effectively engaging a sufficiently broad cross-section of the population in new projects.

Botswana and AIDS: the apparent failure of 'good' institutions

Throughout the 1990s everyone was puzzled by Botswana's inability to deal with AIDS. It is not simply that Botswana has done poorly, it is that it has done worse than other African countries that seem to be much less well endowed – either in terms of material resources, or in terms of effective institutions. AJR (2003, p. 2) note that: 'Not everything in Botswana is rosy. Though the statistics are not fully reliable, Botswana has one of the highest adult incidences of AIDS in the world'. They add: 'This probably represents, above all else, a serious public policy failure', but this failure does not figure in their evaluation of Botswana's institutions. Ann Swidler (2004) provides a provocative, though preliminary, effort to fill this lacuna.

Swidler takes AJR's analysis of Botswana's successful institutions as her starting point. Like others (for example Allen and Heald, 2004), Swidler starts by noting that the public policy response in Botswana has been precisely what one would expect on the basis of AJR's institutional analysis: modern, competent and thorough. It included mounting an informational campaign, putting advertisements on the radio, expanding clinics and sending public health officials out into the countryside to persuade people to change their behavior.

Yet, the impact on people's behavior appears to have been minimal, or even perverse. Pre-AIDS cultural preferences and social behavior persist despite the government's efforts. For too many Botswana, avoiding the stigma of the disease still appears to be a more compelling motivation than engaging in treatment that requires publicly acknowledging having the disease. And the devastation continues. In short, Botswana's careful stewardship of property rights and resources did not give elites or their state the capacity to mobilize non-elites and transform their behavior, a capacity that was the key to confronting the developmental challenge of AIDS.

We might summarize Swidler's propositions as follows. Institutional capacities to reshape social preferences and generate new sets of culturally validated motivations can be more important than the capacities which enable people to achieve previously defined goals. Critical junctures reveal

hidden strengths and weaknesses in previous institutional constructions, and may correspondingly be the occasion for either reinforcing the persistence of existing institutions or undercutting previously successful institutions.

Extending the institutional turn

When we follow the trail of Acemoglu, Johnson and Robinson and juxtapose their analysis with that of others working on the same cases, the heuristic power of the institutional turn becomes apparent. No less apparent is the extent to which 'institutions of individual property rights' is a conceptual Procrustean bed, even for AJR themselves. As soon as more specific historical and institutional content is added to the analysis, the insufficiency of 'property rights' as a framing is clear. It is not that the property rights argument is 'wrong' or 'irrelevant', it is just that as soon as cases are examined in even slightly more detail, questions of politics and distribution, most especially contestation over the distribution of political rights and power, come to the fore, along with the legacies of the institutional constructions that ensue from prior contestation.

Redirecting the institutional turn away from incentives that are 'economic' in the strict sense of the term and focusing on political contestation and its institutional legacies is consistent with AJR's own admonition (2001, p. 1395) that the institutional turn should shift its focus toward institutions that are 'more fundamental'. It is also consistent with the more historically specific elements in their argument that have been reviewed here. AJR's version of the institutional turn recognizes the centrality of political institutions in multiple ways, but has yet to fully integrate this recognition back into their analysis of growth and development.

Perhaps the most interesting point of consensus among all of the analyses that have been considered here is the way in which the state and the legal institutions associated with it keep slipping into a central position, despite the fact that none of these analyses are ostensibly 'state-centric'. Some of the state's centrality is relatively conventional. The role of nineteenth-century Central American states as key providers of infrastructure in the development of new export capacity is novel only in that these are not the sort of regimes usually associated with state entrepreneurship.

The most salient role of the state in these accounts has not, however, been its direct economic role, but its role as the crucible within which legal rights are generated and its role as the agent through which these rights are enforced. A state which is predictably and consistently constrained by its own laws and norms is the obvious *sine qua non* of effective property rights. Put another way, these cases suggest the proposition that having property rights requires having political rights, both in the positive sense of having some degree of institutionalized control, however imperfect, over access to

state power, and in the negative sense of not being arbitrarily subjected to the coercive power of the state.

The question of political rights brings us in turn back to AJR's 'broad cross-section' requirement. A number of hypotheses might be entertained as avenues for realizing the potential gains from this insight. The most obvious is that, when effective property rights are restricted to a small minority, political control is more likely to depend on coercion. Consequently, political contests are more likely to take the form of struggles over controlling the means of domestic coercion, which is not the form of politics designed to focus attention on productive investment. A more economically oriented hypothesis might focus on the asset preferences of small minorities as opposed to broad cross-sections, extrapolating from the case of land, as analyzed by Nugent and Robinson (2001) and Baland and Robinson (2003). Land (as opposed to human capital) is an attractive asset for a small minority, since property rights to land are more easily enforced via coercion. At the same time, since land is a fixed resource with diminishing returns it is the worst kind of focus for long-run developmental success. Broad cross-sections on the other hand are likely to be more prone to invest in human capital, a much more productive sort of asset in the long run.

Thinking about the broad cross-section requirement also raises the question of how and why systems of property rights might change. The robustness of even economically ineffectual institutional structures is impressive. AJR (2001, p. 27) are careful to disavow the idea that their findings might 'imply that institutions today are predetermined by colonial policies and cannot be changed'. Nonetheless, the continuities that connect the 'extractive institutions' of half a millennium ago to contemporary economic misery suggest that the dynamics of transcending the legacies of ineffective institutional heritage should be a central part of the agenda of extending the institutional turn.

Questions of discontinuity and change bring us to the question of agency. One of the curious commonalities in the accounts that have been considered here is that, while spreading property rights to a broad cross-section of the citizenry is considered a key feature of successful institutional development, the possibility that this 'broad cross-section' might be able to exercise some agency in the process of institutional change is considered only tangentially. One suspects that non-elites may play a more creative and positive role in the construction and reconstruction of effective institutions than these analyses suggest, but it is not evident in the work that we have reviewed here – with the possible exception of Swidler.

This brings us to the final point: institutional change driven by mobilization and the transformation of preferences. By raising the proposition

that effective institutional change requires new definitions of desirable, culturally valued behavior, Swidler points toward processes of institutional change in which non-elites must play an active role. Alerted to this possibility by Swidler's extreme case, institutional analysts would do well to consider whether it may have broader applicability.

If we assume that the provision of public goods plays a growing role in economic growth, then the idea that people will automatically 'know what they want' is dubious. It seems more likely that mobilization is a likely requirement for avoiding the undersupply of public goods: not simply supplying information, but actually generating a new set of preferences with regard to the allocation of resources. Even more obviously, if changing an inefficient but self-reinforcing set of property rights is what is required, non-elite mobilization would seem a prime candidate for driving institutional change.

This agenda for the extension of the institutional turn is a sample of possibilities, not a roadmap. Overall, extending the institutional turn is an arduous but unavoidable task. As it currently exists, the institutional turn is a painfully incomplete edifice. At the same time, its very incompleteness adds to its intellectual attractiveness. It is not a paradigm awaiting minor refinements and adjustments. It is a project ripe for major reconstruction.

Acknowledgement

The original version of this chapter is published as 'Extending the "Institutional" Turn: Property, Politics and Development Trajectories' by Peter Evans in *Institutional Change and Economic Development*, edited by Ha-Joon Chang and co-published by UNU Press and Anthem Press. It is reproduced with the kind permission of the authors and UNU-WIDER.

Notes

1. Citations to this and most other recent papers are based on the versions available on authors' websites. Page numbers do not conform to those in the published versions and quotations may vary from published versions.
2. Panama and Belize are usually excluded from comparative analyses despite being geographically in Central America because they do not share to the same degree the historical features that unite the other five.

References

- Acemoglu, Daron and James Robinson (2000), 'Why Did the West Extend the Franchise? Growth, Inequality and Democracy in Historical Perspective', *Quarterly Journal of Economics*, 115: 1167–99.
- Acemoglu, Daron, Simon Johnson and James Robinson (2001), 'The Colonial Origins of Comparative Development: An Empirical Investigation', *American Economic Review*, 91: 1369–1401.

- Acemoglu, Daron, Simon Johnson and James Robinson (2002), 'Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distribution', *Quarterly Journal of Economics*, 117: 1231–94.
- Acemoglu, Daron, Simon Johnson and James Robinson (2003), 'An African Success Story: Botswana', in Dani Rodrik (ed.), *Search of Prosperity: Analytic Narratives on Economic Growth*, Princeton, NJ: Princeton University Press, pp. 80–122.
- Acemoglu, Daron, Simon Johnson and James Robinson (2005), 'Institutions as the Fundamental Cause of Long-Run Economic Growth', in Philippe Aghion and Steve Durlauf (eds), *Handbook of Economic Growth*, Amsterdam: North-Holland.
- Allen, Tim and Suzette Heald (2004), 'HIV/AIDS Policy in Africa: What has Worked in Uganda and What has Failed in Botswana?', *Journal of International Development*, 16: 1141–54.
- Baland, Jean-Marie and James A. Robinson (2003), 'Land and Power: Theory and Evidence from Chile', *American Economic Review*, CEPR (Centre for Economic Policy Research) London, Discussion Paper No. 3800.
- Banerjee, Abhijit and Lakshmi Iyer (2002), 'History, Institutions and Economic Performance: The Legacy of Colonial Land Tenure Systems in India', Unpublished manuscript, Department of Economics, MIT.
- Bardhan, Pranab (2005), *Scarcity, Conflicts and Cooperation: Essays in the Political and Institutional Economics of Development*, Cambridge, MA: MIT Press.
- Chang, Ha-Joon and Peter Evans (2005), 'The Role of Institutions in Economic Change', in Gary Dymski and Silvana de Paula (eds), *Reimagining Growth: Institutions, Development, and Society*, London and New York: Zed Books, pp. 99–140.
- Easterly, William (2001), *The Elusive Quest for Growth: Economists' Adventures and Misadventures in the Tropics*, Cambridge, MA: MIT Press.
- Easterly, W. and R. Levine (2003), 'Tropics, Germs, and Crops: How Endowments Influence Economic Development', *Journal of Monetary Economics*, 50: 3–39.
- Engerman, Stanley L. and Kenneth L. Sokoloff (1997), 'Factor Endowments, Institutions, and Differential Growth Paths among New World Economies', in Stephen Haber (ed.), *How Latin America Fell Behind*, Stanford, CA: Stanford University Press, pp. 260–305.
- Engerman, Stanley L. and Kenneth L. Sokoloff (2002), 'Factor Endowments, Inequality, and Paths of Development among New World Economies', *Economia*, 3: 41–88.
- Evans, Peter (2004), 'Development as Institutional Change: The Pitfalls of Monocropping and Potentials of Deliberation', *Studies in Comparative International Development*, 38 (4): 30–53.
- Gallup, John Luke, Jeffrey D. Sachs and Andrew D. Mellinger (1998), 'Geography and Economic Development', NBER Working Paper No. w6849, December.
- Hoff, Karla and Joseph Stiglitz (2001), 'Modern Economic Theory and Development', in G. Meier and J. Stiglitz, *Frontiers of Development Economics*, New York: Oxford University Press [World Bank], pp. 389–460.
- Knack, Stephen and Philip Keefer (1995), 'Institutions and Economic Performance: Cross-country tests using Alternative Institutional Measures', *Economics and Politics*, 7(3): 207–27.
- Mahoney, James (2001), *Legacies of Liberalism: Path Dependence and Political Regimes in Central America*, Baltimore, MD: Johns Hopkins University Press.
- Meier, Gerald and Joseph Stiglitz (eds) (2001), *Frontiers of Development Economics*, New York: Oxford University Press [World Bank].
- North, Douglass C. (1994), 'Economic Performance Through Time', (Nobel Prize lecture, given on 19 December 1993), *American Economic Review*, 84 (3): 359–68.
- Nugent, Jeffrey B. and James A. Robinson (2001), 'Are Endowments Fate? On the Political Economy of Comparative Institutional Development', CEPR (Centre for Economic Policy Research) Discussion Paper No. 3206.
- Parson, Jack (1984), *Botswana: Liberal Democracy and the Labor Reserve in Southern Africa*, Boulder, CO: Westview Press.

- Rodrik, Dani, A. Subramanian and F. Trebbi (2004), 'Institutions Rule: The Primacy of Institutions over Geography and Integration in Economic Development', *Journal of Economic Growth*, 9 (2): 131–65.
- Sachs, Jeffrey D. (2001), 'Tropical Underdevelopment', NBER Working Paper 8119.
- Swidler, Ann (2004), 'The Politics of AIDS in Sub-Saharan Africa', manuscript, Department of Sociology, University of California, Berkeley, CA.

6 Data problems and empirical modeling in developing economies

*Bill Gibson*¹

Introduction

This chapter reviews econometric and simulation models as applied to developing countries. Both micro- and macroeconomic models are discussed. Data problems common to econometric modeling in developing countries are severe and may explain the increased popularity that computable general equilibrium and household simulation techniques have recently enjoyed.

Empirical modeling

There are two general approaches to empirical models of developing countries: econometric and simulation modeling.² Econometric models derive their power from classical and Bayesian statistical theory. In classical econometric models, characteristics of a population are inferred from a sample of observations on values of random variables. Once a governing probability distribution is assumed, typically a Student's *t*- or normal distribution, rigorous conclusions can be drawn concerning the reliability of the inference. Nothing, however, can be said about the quality of the data.³

Sadoulet and De Janvry review a wide range of microeconomic policy models, including demand, profit function, supply response and various household models under a range of assumptions about agent behavior (Sadoulet and de Janvry, 1995). These models essentially tally social and private costs and benefits in an effort to guide sectoral or regional policymaking. The authors also consider models of international trade and distortions from a partial equilibrium point of view, as well as computable general equilibrium (CGE) models and other economy-wide models.

Economy-wide models are usually based on either aggregate data from national income and product accounts or more disaggregated input–output tables. Regional models may link regional input–output models, analogous to international trade models. The informal sector can also be treated in the same way, operating alongside the formal economy and trading with it (Gibson, 2005).

Econometric models

Econometric models have been applied to developing economies at both the micro- and macroeconomic levels. Microeconomic models describe individual consumer and producer behavior. Data on consumer behavior are often supplied by household, income and expenditure surveys, while producer data might be gleaned from a manufacturing census, input-output studies, tax records, or direct questionnaires administered by governments, non-governmental organizations (NGOs) and independent researchers.

Econometric models, as applied to developing countries, suffer from more extreme violations of the underlying assumptions of the classical linear regression model than in the more stable environment of advanced countries. Strictly speaking, time series econometric models would only apply to a self-replicating stationary state in which nothing of fundamental importance changed over the estimation period. In particular, the assumption of repeated samples drawn from independent and identical conditional probability distributions (i.i.d.) for each value of the independent variable is severely compromised. This is well known, of course, and tests and corrections for heteroskedasticity are widely available and widely applied. In time series models, the i.i.d. assumption implies structural stability and is violated as a matter of course in developing economies, since structural change, rather than stability, is the explicit objective of most development policies. Beyond the violation of the most fundamental assumption of structural stability and heteroskedasticity, econometric models suffer from simultaneity bias, omitted variables and other model misspecifications, selectivity bias, as well as measurement and censored and cluster error. Econometric models applied to developing economies often ignore, for example, structural rigidities such as foreign exchange and skilled labor shortages, and the presence of a large informal sector (Behrman and Hanson, 1979). Policy and coordination problems are sometimes also overlooked, as are various endogeneities peculiar to developing economies, such as credit flows, human capital formation and even monetary and fiscal policy when authorities lack independence. Another reason large econometric macro models have fallen out of favor is that the correlation and high t-statistics observed in earlier macro models were due to the lack of stationarity of the time series. Many macroeconomic time series are highly correlated as a result of a common time trend. Removing the trend by taking first differences puts the framework on a much more solid basis but weakens the predictive power of the models.

Together with the 'Lucas critique', fundamental problems caused some researchers to abandon macroeconomics altogether and refocus econometric attention on microeconomic models. Others turned to models with

little or no theoretical content in an effort to improve short-term forecasts. Models with lagged dependent variables and the even more radical vector autoregression (VAR) models performed well, not because they possessed desirable statistical properties (which they do not) but because they were arguably more realistic. As such, the VAR framework can be seen as the first step in abandoning the inferential approach for one more grounded in reality.

Microeconomic models also faced many of the same estimation problems, but researchers have, by and large, found ingenious ways to adapt their models and to correct for deficiencies. The corrective procedures for models with heteroskedasticity, bias introduced by pooling time series and cross-sectional data, selection, clustering or other data deficiencies can cause collateral damage to the inferential process. Often, bias disappears only when the sample size grows large. Developing-country data sets are therefore prone to biased estimation.

To combat these and other problems of estimation, researchers attempt to construct robust models. Robustness means that the same qualitative conclusion emerges from a variety of different model specifications. Robust conclusions are more credible and convincing to consumers of econometric studies and disable much of the criticisms leveled at the models. Ultimately, however, robustness is subjective, thereby widening the gap between classical statistical theory and useful model conclusions.

Simulation models

Simulation models take the last step and abandon classical statistical theory altogether. They therefore cannot be rigorously evaluated. Since no inference from sample to population is involved, it is meaningless to ask how well any given simulation model reflects its parent population relative to, say, some other simulation model. Simulation models instead rely on a less precise criterion of validity. The principal means of validation is its perceived realism; that is, whether it resembles the object it is supposed to simulate (Gibson, 2003). A model that accurately covers ever-expanding dimensions of the economy is better than a model that covers only a subset of the same data. This is only true when the models are non-recursive so that the model must be calibrated as a whole. In non-recursive models, errors in one component will propagate into the rest of the model, such that mistakes multiply rather than cancel out. In recursive models, the calibration procedure can mask error.

Simulation models are based on the notion that good models do not contain results that are widely at variance with reality in any of their computable properties. The notion of computable properties, as used here, is broader than the properties of a given model that might be presented as a

result of in- or out-of-sample forecast properties. Computable properties include derivatives of presented properties and may reveal otherwise undetected inaccuracies in the model. Rational actor models, for example, especially those that deal with expectations of the future, may well have some computable properties that differ substantially from the perceptions of how an actual economy behaves.

Empirical modeling based on a general equilibrium approach avoids some of the problems of aggregation. General equilibrium imposes consistency on the decisions of rational actors, but consistency is achieved in a wide variety of ways. It is probably fair to say that all models combine, in varying proportions, elements of agency and structure. Economists tend to favor agency, despite the deep problem of the self-validation of rational actor models. Policy-makers, however, typically place more weight on structure than agency in evaluating the realism of a model. Thus, models that overemphasize agency are not subject to the Lucas critique, and can be seen to lack realism. Models that overemphasize structure, on the other hand, are guilty of the opposite excess. Policy becomes unrealistically effective, simply because agents are assumed not to adjust their behavior. Planning models of the 1950s through the 1970s were too optimistic about the effects of government policy and are now considered to have failed.

Policymakers and other consumers often reject empirical models in which the causal mechanisms at work are obscure. They cannot be blamed for shying away from 'black box' models that even their authors fail to comprehend fully. When there are several adjustment mechanisms at work in the same economy, such as with competitive markets in some sectors and oligopolistic markets in others, formal as well as informal agricultural or service sectors or segmented labor markets, numerically calibrated simulation models can place explicit weights on each of the various mechanisms. Sensitivity analysis can then be undertaken with respect to not just agent behavior, but also the overall structure of the economy.

Does robustness play a role in simulation modeling? Simulation models are usually subjected to sensitivity analysis, a procedure that aims at robustness. Model conclusions that are dependent on one or two critical parameters are not as convincing as those which are robust to reasonable changes in those parameters. Model structure, is of course, a different matter; models with different closures, as discussed in the next section, can have entirely different comparative static and dynamic properties.

Macro simulation models

Computable general equilibrium models are usually multisectoral, economy-wide models, which may be static or dynamic.⁴ They are usually

calibrated to a social accounting matrix (SAM) and exhibit a wide range of adjustment mechanisms, from closed, purely competitive, Walrasian models to macro structuralist models in which foreign exchange availability determines the level of output in some key sectors. CGE models have even been compiled at the village level (Taylor et al., 1999), combining modeling with more anthropological approaches.⁵

Sen describes a simple macroeconomic accounting framework in which the number of equations is one short of the number of unknowns. Formally speaking the model cannot be solved, or closed, until an additional equation is found and justified as part of the macroeconomic system (Sen, 1963). Closure then refers to selection of parameters and variables, specifically around the relationship between savings and investment. In a neoclassical closure, for example, the quantity of savings determines the level of investment. In a Keynesian closure, an independent investment function is present and savings adjusts to it through changes in output. A foreign exchange-constrained closure is similar to the neoclassical, except that instead of the supply of factors of production as the ultimate constraint on production, it is rather the level of foreign exchange for imported intermediates and capital goods. Closure is related to but not the same thing as a 'gap'. In gap models there are specific targets for output and employment and either a savings, foreign or fiscal constraint binds (Bacha, 1990; Taylor, 1994). The gap is determined by the amount by which the constraint would have to be shifted so that internal and external policy objectives could be met.

An algebraically indeterminate system may also be closed by some maximization procedure, with the marginal equality that results providing the needed additional equation. Planning models, for example, may try to maximize employment by choosing a sectoral pattern of output consistent with a foreign exchange constraint or some other supply-side limitation. One of the most well-known models in economics endogenizes the savings rate in order to maximize the discounted value of future consumption (Ramsey, 1928). Formally speaking, this closure is as acceptable as any other, provided of course that it passes the test of realism.

Most of the earlier applied CGE models were static and reconciled flows of supply and demand in any one period of time. But with the availability of highly efficient microcomputer programs, such as the General Algebraic Modeling System (GAMS) and the General Equilibrium Modeling Package (GEMPACK), dynamic models have become much more common. They may be solved recursively or simultaneously and can be closed by way of some optimization criterion. They may be solved in level terms or in growth rates as did Johansen in the original CGE model (Johansen, 1960). In this regard, CGEs have come to compete directly with large econometric models.

Solving a dynamic model means finding a solution path for each of the endogenous variables of the model. The computational characteristics of the model may show that some endogenous variables, or their ratio, reach a steady state in which there is no further change. Many dynamic models of developing countries are calibrated to time periods far away from the steady state and the absence or presence of smooth convergence seems hardly to affect their prestige in the eyes of policymakers and other users. The issue is similar to the i.i.d. problem discussed above; since development is itself about changing the fundamental parameters underlying the economy, it hardly seems desirable to project a distant future based on current values. The transient or transitional phase of the solution to the dynamic model is of considerably more interest.

Calibrating empirical models and policy

Sadoulet and De Janvry note that there are two steps in using quantitative models for policy analysis: (1) calibration and verification; and (2) forecasts and analysis (Sadoulet and de Janvry, 1995, p. 7). The calibration phase can be done formally in econometric models, the coefficient of variation, R^2 , determining the goodness of fit. Despite admonition against the practice, policy-oriented econometric models are calibrated much in the same way as simulation models, with variables added, deleted, combined, lagged or algebraically transformed until the goodness of fit reaches an acceptable level. In the process of calibration, econometric models can lose rather than gain transparency, since the model itself changes. The changes are theoretically rationalized, but lead to subtle and complex interactions that result in computational characteristics that are omitted or suppressed in the presentation of results.

In contrast, the transparency of simulation models is usually (although not always) unaffected by the calibration procedure. Calibration has many pitfalls of its own, however, and is sometimes called 'guesstimation'. Guesstimation refocuses attention on the realism of the final product, not the secondary issues of inference as noted above. It is sometimes argued that econometric models also have their own brand of guesstimation, but this rather concerns the model specification, an inherently informal process.

Data problems

According to Deaton:

The news . . . is dismal. National income and growth comparisons across countries are plagued by conceptual index number problems, and by immense practical difficulties. Many frequently used data from LDCs are of poor quality, or only pretend to exist, having their only reality in the mind of bureaucrats in New York and Washington. (Deaton, 1995, p. 1814)

Data in developing countries can be reliable, noisy and/or unreliable according to whether there are errors in the data collection process and whether these errors tend to cancel out.⁶ Errors in the data collection process result from their being made up by interpolation, extrapolation and falsification. There is no econometric test for unreliable data.

Errors also result from changing definitions as well as the standard index number or aggregation problem. Populations tend to be more heterogeneous in developing countries, because of race, religion and ethnic identity. Income is often badly distributed. Thus, aggregating rich and poor can distort data in developing countries (at the top of the Kuznets curve) more significantly than in more egalitarian societies. Most fundamentally, aggregation problems are more likely to occur in developing countries because the social structure is rapidly changing. Apart from the processes involved in development, macroeconomic imbalances, stagnation and crisis can cause emigration or social conflict which biases or causes large gaps in data collection. Consistency problems are multiplied when regions differ significantly or when political structure is regionally fragmented.

Governments and non-governmental organizations (NGOs) often lack budgets to do an adequate job of collecting, cross-checking and validating data. In household surveys, for example, respondents should be chosen randomly, but some live in inaccessible areas and researchers may literally risk their lives in war-torn or crime-ridden regions. Representativeness problems would be less severe if data were even collected in a consistent fashion over time, effectively creating proxies, but they are usually not.

The existence of a large informal or traditional sector also causes significant problems for developing-country data. The informal sector in agriculture can make up more than half the economy and is typically understudied. Developing economies are often only semi-monetized, with auto-consumption and barter playing an important role, especially in the rural sector. Investment in the informal sector is particularly difficult to track, often appearing in the national accounts as consumption or missed altogether. Smaller on-farm construction projects such as clearing, informal roads and irrigation canals or terracing are missed by government officials who concentrate on licenses, building permits and capital import authorizations to estimate investment in the national accounts (Taylor, 1979, p. 23).

There are no reliable unemployment data for most countries, and when unemployment surveys have been conducted they tend to cover urban areas only. Estimates are therefore both practically unreliable and conceptually clouded: are members of the informal sector considered unemployed if they would be willing to abandon their kiosks when offered a formal sector job? To a first-order approximation, one could argue, there is full employment in developing countries, given the absence of social safety nets

and inherent limitation of transfers available from extended families, communities and church (Gibson and Kelley, 1994). Underemployment is similarly difficult to conceptualize and measure.

When technocrats are in short supply, data gathering may be hampered by poorly trained or untrained field workers. Deaton notes that holding precision constant, any cost-minimizing sample design will lead to oversampling of urban households (Deaton, 1995, p. 1790). Democratic institutions, which would support objective collection and analysis of data, are not always in place. Local accounting procedures may themselves be part of the problem.

Specific sampling problems include stratification and cluster bias, groups of individuals with similar unobservable characteristics, such as ability or entrepreneurship, tastes or other characteristics assumed to be randomly distributed across the population. There is also selectivity bias, non-random reasons why some individuals enter a given sample. Respondents may also incorrectly report data when civil or criminal liability is an issue, such as 'unregistered' labor contracts, or they may lie for privacy or political purposes or in an attempt to conform to perceptions of researchers' expectations.

Further, uncertainty and inefficiency in tax laws may cause inaccurate reporting. This occurs in two ways: first, if tax liability is presumed to increase, information will be withheld from government officials. Conversely, if there are no tax or regulatory implications of investment projects, government data collection is more likely to overlook the activity, thereby underestimating the conceptual category. There may also be principal-agent problems, in which respondents misrepresent their objective conditions when it is in their interest to do so. Finally, a subjective or post-modernist effect may cause some respondents to report conditions that vary greatly from others when all are attempting to be objective.

Data problems specific to macroeconomic models

Data is collected and processed by different agencies or ministries with different missions, budgets, effectiveness and capabilities. In principle, each agency is estimating a different aspect of the same economy and thus should report broadly consistent magnitudes. In practice the magnitudes can vary substantially.

Most developing countries base their gross domestic product (GDP) estimates on the production rather than the demand side. These estimates could be cross-checked by demand-side surveys or census data. In practice this is not often done and satellite measurement may well turn out to provide the most reliable estimates. The Central Statistical Office (CSO) is typically responsible for the national accounts in units of local currency. If

the estimates are based on 'flow of product' concepts, the underlying information will determine the accuracy of the final figure. Unfortunately, this will vary from sector to sector, with the reliability of the information dropping off with the square the distance between the CSO and respondents. In other words, rural data will be collected with less frequency and lead to more between-year extrapolation. Sectoral data based on industrial census may be refreshed the most often, with other data scaled to these results. In general demand-side data, based on flow of product, is weaker than data based on value added, for which there might be fiscal interests at stake (Taylor, 1979, p. 22).

Balance-of-payments data, compiled by the Ministry of Trade or Central Bank in hard currency, may not agree with national accounts for exports and imports because of rapidly changing exchange rates or other difficulties. These problems are more severe with currency controls and import licensing. Underinvoicing of exports and overinvoicing of imports can be important sources of foreign exchange in some developing countries, and transfer pricing as a tax-minimizing strategy can significantly distort foreign trade data.

The Ministry of Finance typically compiles government expenditure data with help from the World Bank, International Monetary Fund (IMF) or regional development banks. Ministry of Finance data for government expenditure may not agree with national accounts data for government expenditure. The Ministry of Interior or Labor may handle household surveys with help from the World Bank, International Labour Organization (ILO) or NGOs. Household surveys are often inconsistent with data for consumption in national accounts. Finally, the Environmental Ministry may also be relevant, especially if environmental problems are seen as linked to growth, international trade in tourism and income distribution.

Addressing data problems in developing countries

The two generally accepted methods of dealing with data problems in developing countries are cross-checking and correlation. Cross-checking exploits the dual nature of transactions. Correlation is more elaborate and integrates econometric methods into the process of consistent data generation. Purchasing power parity methods, which correct for the effect of asset demand on exchange rates, can be used for cross-country comparisons. SAMs themselves are not models but can be used to create data that are free of the inconsistencies arising from their various sources. SAM methodology is a simple, but thorough, example of the cross-checking method. Cross-checking can also be done from reports on working conditions and environmental problems from different sources, workers and

political entities, but requires a significant degree of qualitative knowledge. 'Patterns of growth' data can also sometimes be used to comb out inconsistencies.

A sequence of SAMs can be used to cross-check investment, depreciation rates and capital accumulation. Financial data from balance sheets from firms and central banks can also be used, although procedures are in their infancy. Data from agencies regulating financial practices, labor standards and environmental compliance may also be employed. Correlation in cross-sectional or panel data shows that some measures are better than others. Household surveys, on which poverty estimates are made, are often inconsistent with national income and product accounts, making it impossible to know whether poverty in a particular developing country has increased or not (Gruben and McLeod, 2002). In principle, a household survey would be representative of the cross-section of the consuming population. Hence the mean consumption per capita should be highly correlated with the consumption per capita in the national accounts. Bhalla (2002) points out that this is not the case. Familiarity with the institutional environment and the use of anthropological and other case studies can be used to evaluate economy-wide data of dubious quality.

Some final observations

The assumptions required to make valid inferences are strained by the very process of development. If repeated samples of i.i.d. random variables were possible, simulation models would no doubt be less widespread. Their popularity derives from the fact that opportunities for random sampling are rare and especially unavailable in developing economies. Moreover, models based on unreliable data are themselves unreliable, despite any other attractive properties they may possess. Unreliable data are data measured with error, but if the error is not random and does not cancel out, bias will result. Since data can be and often are produced by individuals who lack knowledge of proper sampling procedures, or indeed with political or self-interested motives, no corrective procedures are available. Biased data are bad data and must be recognized as such, but the definition of 'good' or 'reliable' data remains subjective.

Notes

1. Thanks are owed to Diane Flaherty, Amitava Dutt, Jaime Ros and Elaine McCrate for very useful comments and criticisms.
2. There are many excellent general references for the application of empirical models to developing economies. See, for example, Sadoulet and de Janvry (1995), Taylor (1979) and Blitzer et al. (1975).
3. The application of Bayesian statistical methods to developing countries is still in its infancy. For an example, see Sala-i-Martin et al. (2004). For Bayesian methods, see Greene (2003), Chapter 16.

4. The literature on CGE models is large. Particularly useful are Robinson (1995) and the introductory chapter to Taylor (1990). Some other references are Dervis et al. (1982), De Maio et al. (1999) and Gunning and Keyzer (1995).
5. For an introduction to anthropological methods in economics, see Gregory and Altman (1989).
6. See the special issue of the *Journal of Development Economics* devoted to data problems in developing countries. An overview is provided in Srinivasan (1994).

References

- Bacha, E. (1990), 'A Three-Gap Model of Foreign Transfers and the GDP Growth Rate in Developing Countries', *Journal of Development Economics*, 32: 279–96.
- Behrman, J. and J.A. Hanson (1979), 'The Use of Econometric Models in Developing Countries', in J. Behrman and J.A. Hanson (eds), *Short-Term Macroeconomic Policy in Latin America*, Cambridge, MA: Ballinger for the National Bureau of Economic Research, pp. 1–38.
- Bhalla, S.S. (2002), *Imagine There's No Country: Poverty, Inequality and Growth in the Era of Globalization*, Washington, DC: Institute for International Economics.
- Blitzer, C., P. Clark and L. Taylor (1975), *Economy-Wide Models and Development Planning*, New York and London: Oxford University Press.
- De Maio, L., F. Stewart and R. van der Hoeven (1999), 'Computable General Equilibrium Models, Adjustment and the Poor in Africa', *World Development*, 27 (2): 453–70.
- Deaton, A. (1995), 'Data and Econometric Tools for Development Analysis', in H. Chenery and T.N. Srinivasan (eds), *Handbook of Development Economics*, Volume 3A, Amsterdam: Elsevier, pp. 1785–1882.
- Dervis, K., J. DeMelo and S. Robinson (1982), *General Equilibrium Models for Development Planning*, Cambridge: Cambridge University Press.
- Gibson, B. (2003), 'An Essay on Late Structuralism', in A. Dutt and J. Ros (eds), *Development Economics and Structuralist Macroeconomics: Essays in Honor of Lance Taylor*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 52–76.
- Gibson, B. (2005), 'The Transition to a Globalized Economy: Poverty, Human Capital and the Informal Sector in a Structuralist CGE Model', *Journal of Development Economics*, 78 (1): 60–94.
- Gibson, B. and B. Kelley (1994), 'A Classical Theory of the Informal Sector', *Manchester School of Economics and Social Studies*, 62: 81–96.
- Greene, W.H. (2003), *Econometric Analysis*, (5th edn), New York: Prentice Hall.
- Gregory, C. and J. Altman (1989), *Observing the Economy*, London and New York: Routledge.
- Gruben, W. and D. McLeod (2002), 'Choosing Among Rival Poverty Rates', <http://www.dallasfed.org/latin/papers/2003/lawp0301.pdf>.
- Gunning, J.W. and M.A. Keyzer (1995), 'Applied General Equilibrium Models for Policy Analysis', in H. Chenery and T.N. Srinivasan (eds), *Handbook of Development Economics*, Volume 3A, Amsterdam: Elsevier, pp. 2025–2107.
- Johansen, L. (1960), *A Multisectoral Study of Economic Growth*, Amsterdam: North-Holland.
- Ramsey, F. (1928), 'A Mathematical Theory of Savings', *Economic Journal*, 38: 543–59.
- Robinson, S. (1995), 'Multisectoral Models', in H. Chenery and T.N. Srinivasan (eds), *Handbook of Development Economics*, Volume 2, Amsterdam: Elsevier, pp. 884–947.
- Sadoulet, E. and A. de Janvry (1995), *Quantitative Development Policy*, Baltimore, MD and London: Johns Hopkins University Press.
- Sala-i-Martin, X., G. Doppelhofer and R.I. Miller (2004), 'Determinants of Long-Term Growth: A Bayesian Average of Classical Estimates Approach (BACE)', *American Economic Review*, 94 (4): 813–35.
- Sen, A. (1963), 'Neo-Classical and Neo-Keynesian Theories of Distribution', *Economic Record*, 39: 54–64.
- Srinivasan, T.N. (1994), 'Data Base for Development Analysis: An Overview', *Journal of Development Economics*, 44 (1): 3–27.

- Taylor, E.J., A. Yunez-Naude and S. Hampton (1999), 'Agricultural Policy Reforms and Village Economies: A Computable General-Equilibrium Analysis from Mexico', *Journal of Policy Modeling*, 21 (4): 453–80.
- Taylor, L. (1979), *Macromodels for Developing Countries*, New York: McGraw-Hill.
- Taylor, L. (ed.) (1990), *Socially Relevant Policy Analysis*, Cambridge, MA: Massachusetts Institute of Technology.
- Taylor, L. (1994), 'Gap Models', *Journal of Development Economics*, 45: 17–34.

PART II

ALTERNATIVE APPROACHES TO DEVELOPMENT

7 Historical antecedents of development economics

Gianni Vaggi

Introduction

The historical antecedents of development economics offer a wide range of notions, views and theories which are still of great importance for today's development theories and policies. These notions range from division of labour to static and dynamic comparative advantages and from the link between investments and accumulation of capital to embodied technical progress and to increasing or decreasing returns. Many among the founding fathers of economics investigated developmental issues; later on, other aspects occupied the stage in economic science. The importance of the legacy of classical political economy to modern development theories is largely due to the fact that both England and France considered economic growth to be a major political issue; as it is today in so many developing countries. As a matter of fact, from the mid-eighteenth century England was considered to be the most advanced country in Europe. The importance of the legacy has to do with the fact that for the classical economists the central subject matter of economics was the theory of the increase in national wealth. The study of the economic growth rate is the central issue for developing countries, but it is also the crucial issue for many high-income countries.

Other notions which constituted the very foundations of classical political economy have been rather ignored in mainstream economics: from surplus and reproduction to structural change and to the distinction between productive and unproductive sectors. However these concepts have always played a central role in development economics and above all in development policies, even when the reference to them was not explicit; think of all policies designed to favour some particular sectors of the economy.

This chapter is organized in short sections in historical order, in order to examine the various economic concepts which have been introduced by the major classical authors. We will examine only some of the protagonists and we will see that some modern (post-1939) visions and ideas are in fact quite old.

Mercantilism: wealth through foreign trade

The first economists who produced a view of development and growth of the wealth of nations were the mercantilists. Mercantilism cannot be defined as either a coherent theory or an organized school; sometimes its supporters have contrasting views. However, two facts make these authors a good starting point in the analysis of development. First, this theory lasted almost three centuries and dominated the economic policies of most European states. Second, it is impossible to evaluate properly the significance and scope of the contribution of the classical economists to development views without considering mercantilism. In particular, Adam Smith's theory of the causes of economic growth is an alternative to mercantilist policies.

The mercantilists consider national wealth as the stock of precious metals (hard currency reserves): a country grows richer and more powerful if the quantities of precious metals in its 'Treasure' increase. For many mercantilists wealth consists also in the size of population and in the availability of both basic and luxury goods, but in any case precious metals are the best measure of national wealth. Gold and silver are the only type of goods which are accepted for payments at an international level; they represent the 'generalized purchasing power' of a country. Precious metals are the ideal way for measuring the relative power of each nation and also for establishing whether the country's wealth is either improving or declining.

The sixteenth century sees what can be called the 'first phase' of mercantilism. During this period the dominant view is that the increase of national wealth has to be achieved mainly by direct control of the flows of gold and silver across borders. In modern terminology we can say that wealth can only increase if there is an excess of capital inflows over capital outflows, that is to say a surplus in the capital account. The policies designed to achieve this aim are not dissimilar from those which can be employed today. High real interest rates are needed in order to favour net capital inflows, and these must also take into account interest rate differentials. The solidity or stability of the national currency is another way to improve the country's reserves; the so-called 'Gresham's law' according to which the good currency, gold, is being stocked away while the bad one, copper, is used in exchanges: obviously the currency must not undergo clipping or devaluation phenomena.

However, by the early seventeenth century it becomes clear that the surplus in the capital account can hardly be regarded as the ideal path to prosperity. For the whole of the sixteenth century Spain has been the state with the highest reserves, but it has gradually lost them all as a result of a trade deficit. This demonstrates that nations can also lose wealth, that development is not a one-way road: nations can have a rise-and-fall cycle.

The passing from the 'first phase' to the 'second phase' of mercantilism does not question the idea that foreign trade is always the only way to increase national wealth, but now economic improvements depend on a surplus in the balance of trade. Thomas Mun (see Mun, 1623 [1986]), one of the directors of the British East India Company, provides a full set of policies designed to lead to wealth and development via the management of the trade balance. The capital account flows simply mirror and depend upon the trade balance of a country, thus the capital account depends on the trade account; if exports are higher than imports this surplus will have to be paid for, and gold and silver will flow into the country.

The means to favour a positive balance of trade are not much different from those which are used nowadays, and they end up in the protection of domestic industry: export subsidies, import duties and tariffs and easy credit to the exporters. To Mun, interest rates must be low, in order to favour the British merchants; the same is true for wages, in order to keep the costs of production low. Even the exportation of British gold is no sin, provided this foreign investment will bring home higher returns. Contrary to the traditional mercantilist wisdom of 'buying cheap and selling dear', Mun says that it can be a good thing to reduce the price of exports if this will increase foreign sales in such a way that overall revenues will increase; a clear idea of the role of the price elasticity of demand. In the end a successful balance of trade is the result of the ability to produce at low costs. Given the success of these policies one cannot easily dismiss mercantilist views.

The Navigation Acts issued by Cromwell from 1651 establish that all trade to be carried on with Great Britain and her colonies must take place with British vessels; this makes the control of imports and exports much easier, but also allows the country to gain on freights. Not only is the surplus in the trade balance the clue to development, but a new notion enters the picture: what we now call the current account, or the goods and services balance.

From Petty to physiocracy: surplus and agriculture

In mercantilist theories, international trade is a zero-sum game. Nations can grow rich only at each other's expense and, even worse, mercantilist policies lead to commercial expansion and to wars, such as the continuous wars between France and England. At the middle of the eighteenth century the followers of the Enlightenment think that mercantilist views cannot lead to the prosperity of all nations, and this leads to a different view of trade which derives from a different theory of the rise and fall of the wealth of nations. The roots of this new view are to be found in the work of Sir William Petty, who almost a century before introduces two notions that will

have a major role in the overcoming of mercantilist views: division of labour and surplus. Petty believes that organized human societies are characterized by various levels of 'social division of labour': men can specialize in some activities and obtain from other people part of the goods necessary to satisfy their needs. This division of labour is possible because of the productivity of agricultural workers who produce more than they need, and thanks to this surplus they maintain the rest of the society (on Petty, see Asproumorgos, 1996). The notion of the 'agricultural surplus' as the basis of society will play an important part in classical development theories and it is widely considered in modern development views.

Between 1748 and 1776 classical economic thought was established. In his 'The Spirit of Laws' of 1748 Montesquieu establishes the traditional division of powers and of functions of a modern state: the judicial, the legislative and the executive power. He says that trade does not have to be a battlefield, a place of conflicts – trade can be 'sweet'; exchanges can lead to a situation of mutual interest and of peace and to the development of nations. This is a crucial point to understand the approach to trade of the founding fathers and has been extensively commented upon by the major development scholar Albert Hirschman (see Hirschman, 1977). In 1752 David Hume publishes several essays under the title of *Political Discourses*, the essays are clearly against mercantilist views as some of the titles make clear: 'Of Money' 'Of the Balance of Trade' 'Of Interest', and so on. We find the description of the 'flow-specie' mechanism, according to which no permanent gain can derive to a country from a trade surplus. If there is a trade surplus, gold enters the country and therefore domestic prices increase; as a result exports will decrease and imports will rise. In the end, once the mechanism has performed all its effects and provided that there are no artificial hindrances to trade, international markets will 'balance' the external accounts. A trade surplus is only a transitory phenomenon and cannot be the ultimate or permanent cause of prosperity and development.

If it is not foreign trade, then what is the cause of development? A first clear answer comes from Francois Quesnay and the physiocrats: the forces of prosperity have to be searched in the process of production and above all in agriculture. They lay the ground not only to a theory of growth, but indeed to a new vision of development, because for Quesnay contemporary France is a backward country when compared to England, which is getting richer and richer. Following Petty, Quesnay's analysis starts from the agricultural sector where the means of subsistence for the whole population are to be produced.

In France sharecropping prevails; this is an old form of agricultural production derived from the feudal period, which either does not yield a surplus, or the surplus is too small to be reinvested in production. The

sharecroppers are poor and do not have the means to invest in modern capital equipment and in the modern techniques of cultivation, where 'modern' implies that the technology is more productive. This is the 'small scale cultivation' (see Quesnay, 1756 [1958]). In England agriculture is developed and has high productivity, both per unit of land and per worker, and the tenant-farmers can invest in the 'best technology'. This is the 'large scale cultivation' (ibid.). Notice that in modern agriculture there is an annual surplus of corn; at the macro level, that is to say in the overall economy, this surplus can be measured in physical terms, because the output of corn exceeds the corn being used as an input. But the farmers also enjoy a surplus in value terms when the value of their output exceeds the sum of all their expenses, and this provides the funds and the incentive to invest in cultivation. The starting point of Quesnay's development policies is the introduction of the most modern production techniques into agriculture through the use of more advanced production tools. The accumulation of capital in agriculture leads to technical progress, that is to say to higher productivity, which is the key to the development.

Another important legacy by Quesnay is the fact that he faces directly the problem of the complexity of social change. The problem is the following: how to trigger the virtuous circle of development in a backward country, such as France at the time. This should be the outcome of a series of economic reforms that favour the French cultivators, induce them to accumulate capital in cultivation and transform them into rich farmers. One way to achieve this aim is by means of an appropriate tax policy: the physiocrats want to abolish all the taxes on peasants and put a single tax on landlords' rent, which is the only disposable part of the agricultural surplus, because contrary to farmers' profits these rents will not to be used for investments. This idea will leave an important mark even in today's development debates; reinvested profit must be exempted from taxation, the more productive sectors of the economy must enjoy some tax advantages, and so on.

For Quesnay, another essential way to trigger the development process is through the famous notion of *laissez-faire*, by which he means specifically the liberalization of the export of French corn. The reason is that the high foreign demand will lead to an increase in the wholesale price of corn in France; in turn these high exportation levels lead to high profits which can be reinvested by the tenant farmers, and the mechanism of accumulation of capital is put into action.

As we shall see, Quesnay's development chain is similar to that of Smith: it is a self-sustaining growth model in which technical progress is dragged along by the accumulation of capital. The message that the physiocrats leave us concerns the central role played by the primary goods

sector and the need to modernize it, a strategy which reminds us of many development policies from 'Green Revolution' to 'agriculture first' to 'basic needs'.

Of course in physiocracy there is one major limitation: advanced agriculture is the only sector in which there is a surplus; industry is sterile, because the value of its outputs is equal to the value of its inputs. According to Quesnay, in manufacturing there is no surplus because it simply transforms the primary commodities and the wage goods it receives from the peasants without adding any value to them. This view represents a very serious limitation of physiocracy.

With his *Tableau économique* Quesnay lays the ground for the analysis of development process by means of structural change in the sectors' composition of output, because in the *Tableau* the economy is made up of different sectors and it is not a 'single commodity world' (see Quesnay, 1766 [1962]). This is an essential legacy to following development theories and it opens the way to the analysis of development in terms of structural change as well as to planning and to the views of unbalanced growth. Unfortunately this sector type of analysis is largely ignored in neoclassical growth theory and also in the so-called 'new growth theory'.

Adam Smith: the division of labour

In his search for the origin and growth of national wealth, following Petty and Quesnay, Adam Smith highlights the division of labour as the core of the analysis. The concept of wealth that Smith introduces is the modern idea of annual product per head, and the division of labour explains how this magnitude can increase. The social division of labour explains that economic activities are and should be differentiated; different men can specialize into different branches of the economy. As in Petty, there are tradesmen, administrators and scholars, but now there are three major social classes: workers, landowners and capitalist entrepreneurs with clearly separated economic functions in the process of production and exchange.

In the first chapters of the *Wealth of Nations* (Smith, 1776) there is the notion of the 'rude state of society', which is opposed to a more civilized society. This view derives from the so-called 'four stages theory', which describes the evolution of human societies through different stages: hunting, pasturage, agriculture and commercial society (see Meek, 1976). This exercise in comparative economic history is crucial to development thinking. Moreover, this evolution through time is characterized by different 'modes of subsistence' but also by different institutions, which play a crucial role in the process of social change, a view that is now widely shared in modern development policies.

However, the major reason for the increase in national wealth is to be found in the ‘technological division of labour’, that is to say the subdivision of complex production operations into simple ones, a process which enables each worker to be more productive. In Chapter 1 of Book 1 of the *Wealth of Nations* there is the famous example of the production of ‘pins’; by assigning each of the 18 operations necessary to produce a pin to a different labourer, the average productivity per worker increases. This represents the foundation of what will be later called ‘increasing returns to scale’, which derives directly from the accumulation of capital in the productive activities. Capital accumulation implies that profits are reinvested in production by means of an increase in the wage bill, thus more workers can be employed and this leads to the possibility of ‘dividing labour’ and makes it more productive. It is the accumulation of capital, and of circulating capital in particular, which leads to the increase of average labour productivity. The endogenous growth theory has rediscovered the importance of having non-decreasing returns to the input which can be accumulated, which is an old lesson (see Kurz, 2003).

Smith accepts the idea that there are both productive and unproductive activities, or labour, but to him industry and manufacturing are no longer to be regarded as sterile occupations, the distinction is now related to the type of output. Only the sectors producing commodities which can be invested and accumulated for further production can be regarded as being productive; the service sector is sterile because it produces for consumption and its output cannot be accumulated. Agriculture still plays a fundamental role because it generates a surplus of subsistence goods, but manufacturing becomes the driving sector of the economy, because it is in this sector that the technical division of labour can show all its potential. This is a clear antecedent of the modern views according to which less-developed countries (LDCs) should not get trapped in the production of primary commodities, and for which the diversification of output and of exports is a crucial element in the process of development.

The key to the increase of labour productivity is the accumulation of capital in the productive sectors. According to Smith, in order to achieve development and prosperity there is a sort of ideal cycle of investments: first a country must invest in agriculture, making it productive and self-sufficient; then in manufacturing, where the technological division of labour brings about the greatest increase in labour productivity, thus leading to a booming economic phase. The above consideration may hint at the problem of uneven development and at the difficult balance between industry and agriculture in this process. The next step implies investing in domestic trade, for example in transportation, that favours and facilitates exchanges, and lastly the country should invest in foreign commerce (see

Smith, 1776 [1976], Book 2, Chapter 5). The latter two types of investments are motivated by the need to extend the market, or the so-called ‘vent for surplus’ argument’ (see Myint, 1977) in such a way that the productive potentialities of the division of labour are not constrained by lack of effectual demand. One is reminded of the recent development phenomena particularly in East Asia. For sure, Smith is no advocate of planning, but his view of the ‘natural order’ of investments is a very convenient and efficient substitute for some well-known policies which over the years have been largely used to support and to direct the process of development. These are policies which emphasize the role of infrastructures and that of an efficient and modern agriculture; they are also policies designed to favour the manufacturing sector. Many modern development theories suggest that it is worth investing where the yield is higher, that is to say in the sectors that are more productive than others; export-led policies are based on this idea, as is the distinction between tradable and non-tradable commodities.

Smith is often associated with liberal economic views; he supports free competition but in a very specific sense. The absence of monopolistic power and of exclusive privileges is designed to let the ‘natural order’ of investments unfold freely, thus leading to the highest increases in labour productivity. The capitalist entrepreneur becomes the fundamental figure in the control and organization of the production process; he is moved by the profit motive and he is able to introduce technical progress and innovations. Smith opposes the alliance between the big merchants, like the East India Company, and the state, which was characteristic of mercantilism, because this is an additional cost on productive activities and because it distorts the natural order of investments. In the *Wealth of Nations* the famous passage about the ‘invisible hand’ appears in Book 4 where Smith attacks mercantilist policies which by favouring foreign industry hinder investments in domestic industry (see Smith, 1776 [1976], Book 4, Chapter 2, para. 9).

The accumulation of capital in the productive activities of the economy is the key to the increases in labour productivity. The classical mechanism of growth can be summarized as follows (see more in Stathakis and Vaggi, 2006; see also Eltis, 1984):

Surplus \Rightarrow profits \Rightarrow savings \Rightarrow investments in the productive sectors [\Leftarrow expected rate of profit] \Rightarrow capital stock increases \Rightarrow (structural change *and* division of labour) [\Leftarrow extent of the market] \Rightarrow increases in labour productivity \Rightarrow increases in surplus and profits.

A few ad hoc assumptions render this sequence similar to some modern growth theories: from endogenous growth models to Kaldorian views of

industrialization-led growth. However, to Smith as to Quesnay, structural change is an essential component of the development process.

Into the nineteenth century: from Malthus and Ricardo to decreasing returns

At the turn of the century, the academic mood about development and growth becomes less optimistic. In his *An Essay on the Principle of Population* of 1798, Thomas Robert Malthus maintains that while population grows in a geometric proportion, the production of subsistence goods grows in an arithmetical proportion. The population grows in geometric proportion because when there are high salaries families live better and conceive more children; however this population growth leads to a decrease in wages which in the long term will be fixed at subsistence level.

The fact that wages are at subsistence level contributes to create the conditions for a lack of effective demand and a general glut; the profitability of investments decreases and the process of accumulation comes to an end. Because of the insufficient purchasing power to sustain the process of accumulation, the economy enters a period of crises.

In the early decades of the nineteenth century David Ricardo contributes to the view that in the long run the profit rate will tend to fall. Ricardo's analysis concentrates on the limited availability of the more fertile lands and is based on the theory of differential rent. The agricultural entrepreneurs want to produce on the more fertile lands, and to this purpose they are prepared to pay a rent to the landlords, but the profit rate depends on the productivity of labour on the land of lower fertility. The need to use more land is linked either to the increase in population or to the duties which limit the importation of corn from abroad; the result is that even the less fertile domestic lands are put to use. The workers' wage rate, w , is at subsistence level and cannot be decreased, but the diminished labour productivity in agriculture, P/L , leads to a decrease in the profit rate, r , which is shown by:

$$r = (P/L - w)/w$$

A fall in profit rate would have negative consequences for the whole economy, as the accumulation of capital would come to a halt; only technical progress can increase labour productivity and delay the fall in the profit rate.

Another solution to counteract the tendency of the rate of profit to fall is Ricardo's theory of 'static comparative advantages'. This theory suggests that every nation should specialize in producing the goods in which it has a comparatively higher productivity. The country should import the goods

for which it has lower productivity, independently of whether or not it could produce these goods. The goods produced by each nation are then exchanged in a competitive market so as to enable all countries to concentrate on those products for which they have the highest labour productivity. By sustaining productivity in all countries, international specialization delays the falling of the rate of profit and therefore supports the world's economic growth. In Ricardo's theory, the tendency for the profit rate to decrease derives from the fact that there are 'non-reproducible production inputs', these scarce inputs are the most fertile lands. For food-importing countries, free trade is another way of opposing the decrease in the rate of profit. The impact on food-exporting countries can be ambiguous. Therefore the decreasing returns in agriculture are closely linked to the existence of different types of land; decreasing returns emerge when there are different production techniques with different labour productivity, and there is at least one input which is non-reproducible and non-accumulable. As in Quesnay and in Smith, the relationship between agriculture and industry plays a key role in Ricardo's theory of profit. All the more so if we consider Ricardo's labour theory of value and the role of inter-sectoral relative prices, a theme to be taken up by Marx. These themes can be found in many modern theories which analyse the agriculture–industry interaction and the problem of terms of trade.

The idea that with the progress of capital accumulation there can be a fall in the profit rate has become part of the traditional theory of growth following the success of Solow's 1956 model. Decreasing profitability is not limited to agricultural production, but it takes place in any type of production process and not only in the cases where there is a problem of non-reproducible inputs. The production function used in many economic models is growing with capital per head but shows falling marginal increments. From this function derives the supposed 'convergence' in income per capita between rich and poor countries, the 'catching up' according to which poor countries grow faster than the rich ones. This is completely different from Ricardo's description and from his, and others', theory of differential rent, and in 1848 John Stuart Mill provides some considerations which can explain the emergence of this view. In his *Principles of Political Economy* he expands Ricardo's argument of the possibility of decreasing returns outside agriculture and outside the case of a non-reproducible input and of different techniques of production, thus extending the case of diminishing productivity also to capital, now considered as an input independent from labour and land. Mill indicates that in the long run the rate of profit will fall because of capital accumulation, thus leading the economy to a stationary state (see Vaggi and Groenewegen, 2003).

List and the infant industry argument

More or less during the same years of the full triumph of liberal economic views in England, at the middle of the nineteenth century in continental Europe, Friedrich List defends the reasons for protecting the industrial sector of the less-developed countries. In his book *The National Systems of Political Economy* of 1841 List says that free trade is beneficial among countries at a similar level of development and not so for less-developed countries whose manufacturing sector cannot sustain the competition of the similar sector in the rich economies (see Chang, 2002). Therefore in order to achieve development the 'latecomers' need to protect their 'infant industries', in the same way as the rich countries have done before. The argument is of extreme interest because it highlights two methodological points. First, it draws attention to the fact that when players are different not only because of different products but because of different income per capita, the trade game is a largely different story from that depicted in the ideal free trade comparative advantage view. Hence we are brought back to a clear case of development and not simply one of growth. Second, the historical conditions play an important role in determining the economic outcomes, and this will lead to the emergence of the German Historical School whose influence will last well into the twentieth century.

Some 50 years before List, Alexander Hamilton, a Secretary of the Treasury with George Washington, anticipates some of List's views. Hamilton does not believe in the advantages of free trade. He favours a very active economic policy by the government and in particular the establishment of subsidies for the support of domestic manufactures of the United States.

Marx and the mode of production

The idea of the fall in profit rate is taken up again by Karl Marx as part of his view of the intrinsic contradictions and the final collapse of the capitalist mode of production. In Marx's contribution one must notice the reappraisal of the method of describing the economy in terms of different sectors, as did Quesnay, this method is particularly clear in the reproduction schemes in Volume 2 of *Capital*.

There is another contribution by Marx that is very interesting for development theories: this is his view of the different modes of production. Marx places capitalist development in a multi-secular context in which there is a succession of modes of production: the ancient one, the feudal one and the capitalist one; they represent the different stages in the history of mankind. The idea that the history of mankind proceeds through specific stages, which are characterized mainly by their economic structure, can be found in several authors, including Smith, of the age of Enlightenment. This legacy

from the eighteenth century and Marx's contribution leaves an important mark in development thinking because it opens the way to a broader analysis of the differences among rich and poor countries and in particular it gives the possibility to analyse the process of transition from one stage, or mode of production, to the next. Countries at different levels of development are characterized by different per capita incomes and by a different economic structure, in the sense that they have a different composition of output and of exports. However, countries can also have completely different ways of organizing economic life, as is the case when different modes of production coexist alongside each other, even if the capitalist mode dominates the others. This is an extremely important phenomenon in development theories and policies; consider the role of land reforms, the whole process of colonization and de-colonization, and the various cases of rent-seeking activities, particularly when a country is rich in natural resources.

Marx provides another important contribution to development economics. In Volume 2 of *Capital* he adopts the 'reproduction schemes', which are a way to describe the economy by means of a multi-sectoral model. There are two major economic sectors: one sector produces the wage goods, the other sector produces various types of inputs, from raw materials to machines. These capital goods are commodities which have themselves been produced in previous periods and must be considered in a totally different way from the only two original inputs: labour and natural resources. This approach opens the way to modern input-output analysis and also to the views which emphasize the role of structural change during the process of development.

Many modern approaches to development derive from the Marxian tradition. One can recall authors such as André Gunder Frank and the 'dependency school', according to which underdevelopment is the result of the economic relationships between the centre and the periphery of the capitalist systems. Another interesting stream of thought is linked to the views of Immanuel Wallerstein and Samir Amin. In all these approaches developing countries are part of a wider capitalist mechanism which hinders and prevents a generalized process of development in the so-called 'Third World'. Marx's analysis of the capitalist system provides several arguments to the views which regard underdevelopment not as a simple problem of some countries which are 'latecomers', but as the other face of capitalist development in Europe and North America from the period of the Industrial Revolution (see Blomstrom and Hettne, 1984).

Conclusion

The legacy of the classical economists to development economics relates to more than the role of investments and the accumulation process. There are

two major features which go beyond the simple anticipation either of a notion or of a theory. First, the problem of the increase in national wealth is the central issue for some of the major economists from the seventeenth to the nineteenth century, and for most of them this is a true problem of development. Second, the search of rigorous economic explanations goes in parallel with the belief that social dynamics and institutional change are part of the analysis and of course of the policy prescriptions; thus development is a complex story. This latter feature of the classical approach has strongly re-emerged in development economics. The role of institutions in the process of development and notions such as those of trust and of social capital play a key role in the explanations both of success stories and of development failures, and are now part of the so called post-Washington consensus.

Among the legacy of the classical political economy we find four views which are used to a limited extent in modern development theory, particularly in its analytical apparatus, even if they appear quite often in the practice and in the policies of development. They are: the notions of surplus and reproduction, the role of inter-sectoral analysis and the process of structural change. Development thought will greatly benefit from making more use of them.

References

- Aspromourgos, T. (1996), *On the Origins of Classical Economics: Distribution and Value from William Petty to Adam Smith*, London: Routledge.
- Blomstrom, M. and B. Hettne (1984), *Development Theory in Transition*, London: Zed Books.
- Chang, Ha-Joon (2002), *Kicking away the Ladder-Development Strategy in Historical Perspective*, London: Anthem Press.
- Eltis, W. (1984), *The Classical Theory of Economic Growth*, London and Basingstoke: Macmillan.
- Hirschman, A.O. (1977), *The Passions and the Interest*, Princeton, NJ: Princeton University Press.
- Kurz, H. (2003), 'What could the "new" growth theory teach Smith or Ricardo?', in H. Kurz and N. Salvadori, *Classical Economics and Modern Theory*, London: Routledge, pp. 137–59.
- List, F. (1841 [1885]), *The National System of Political Economy*, London: Longmans and Co.
- Malthus, T.R. (1798 [1965]), *An Essay on the Principle of Population*, New York: Augustus M. Kelley.
- Marx, K. (1954), *Capital – A Critique of Political Economy*, 3 Vols, London: Lawrence and Wishart.
- Meek, R.L. (1976), *Social Science and the Ignoble Savage*, Cambridge: Cambridge University Press.
- Mill, J.S. (1848 [1970]), *Principles of Political Economy*, Harmondsworth: Pelican Books.
- Montesquieu, Charles Secondat Baron de (1748 [1949]), *The Spirit of Laws*, translated by T. Nugent (1752), New York: Macmillan.
- Mun, T. (1623), *England's Treasure by Forraign Trade*, London: Thomas Clark, 1664; reprinted (1986), New York: Augustus M. Kelley Reprints.
- Myint, H. (1977), 'Adam Smith's Theory of International Trade in the Perspective of Economic Development', *Economica*, 44, pp. 231–48

- Quesnay, F. (1756), *Fermiers*; in Institut Nationale d'Etudes Démographiques (1958), *Francois Quesnay et la Physiocratie*, Vol. 2, Paris: Presses Universitaires de France.
- Quesnay, F. (1766), *Analysis of the Arithmetical Formula of the Tableau Economique*; reprinted (1962), London: Meek, Allen & Unwin.
- Smith, A. (1776), *An Inquiry into the Nature and Causes of the Wealth of Nations*; reprinted (1976), ed. R.H. Campbell, A.S. Skinner and W.B. Todd, Oxford: Oxford University Press.
- Solow, R. (1956), 'A Contribution to the Theory of Economic Growth', *Quarterly Journal of Economics*, 50: 65–94.
- Stathakis, G. and Vaggi G. (2006), 'Economic Development and Social Change: The Classical View and the Moderns', in G. Stathakis and G. Vaggi (eds), *Economic Development and Social Change: Historical Roots and Modern Perspectives*, London: Routledge, pp. 1–25.
- Vaggi, G. and P. Groenewegen (2003), *A Concise History of Economic Thought*, Basingstoke: Palgrave, Macmillan.

8 Classical development theory

*Jaime Ros*¹

Some 50 years ago, a new field of economic theory emerged aiming to answer big questions in development economics, to address issues about the persistence of underdevelopment and to search for remedies to overcome poverty. The nature of these questions made it that, as Taylor and Arida (1988) put it in their survey of development thinking, the new field was ‘born macro’ and had to rely on a paradigm built upon notions of imperfect competition, increasing returns and labor surpluses that were not properly integrated into, and in some cases altogether alien to, the then established body of economic theory. Indeed, to the pioneers of development economics, underdevelopment appeared as a paradoxical situation characterized by a lack of capital – which was consistent with labor receiving lower wages than in developed countries – but also by a low rate of return to capital. For Nurkse, for example, the scarcity of capital was ‘at the very centre of the problem of development in economically backward countries. The so-called “underdeveloped” areas, as compared with the advanced, are underequipped with capital in relation to their population and natural resources’ (Nurkse, 1953, p. 1). This lack of capital resulted from a low capacity to save, given the low level of real income, but also from the ‘weakness of investment incentives’ that had its source in a low rate of return to capital (Nurkse, 1953, Chapter 1). The paradox of both capital and labor receiving lower returns, and the surprising conclusion that the lack of capital may have to be attributed to a low profit rate, understandably led to the search for a novel analytical framework, as anyone familiar with the modern controversies on neoclassical growth theory would probably agree.

The purpose of this chapter is to look at the early contributions to development economics, the contributions of classical development theory, especially the writings of Paul Rosenstein-Rodan (1943, 1961, 1984), Ragnar Nurkse (1952, 1953) and Arthur Lewis (1954). The chapter develops four themes. First, it looks at the distinctive assumptions of classical development economics as a growth theory – increasing returns to scale and surplus labor – and how they generate poverty traps at low income levels and virtuous circles of growth in a peculiar transition to a mature economy steady state. Second, it argues that classical development theory rejected the traditional dichotomy between the static analysis of resource

allocation and the dynamic analysis of growth as the outcome of two separate forces – factor accumulation and technological change – by emphasizing the close links between resource reallocation, factor accumulation and technological change. Third, it discusses the role of effective demand (or rather the lack of it) in the approach to development of the pioneers. Fourth, it argues, contrary to the neoclassical resurgence of the 1960s, that the scope of classical development theory goes far beyond the low levels of development and the closed economy problems to which it was initially applied. After a discussion of these themes in the first four sections the chapter turns to a brief empirical assessment of the broad implications of early development theory. Finally, the last section concludes by referring to the policy relevance of classical development economics to the current problems of developing countries.

The approach to growth in classical development theory

The relevance of early development theory of the 1940s and 1950s for the current state of growth theory is that, compared to the neoclassical extensions of the Solow model or to new growth theory, it offers a more general and more promising road away from the neoclassical model of growth. What one may call the growth model of early development theory can indeed be seen as a departure from the neoclassical growth model² that involves two basic ingredients.

The first refers to increasing returns to scale associated to: (1) pecuniary external economies at the aggregate level generated by internal economies of scale in modern industrial production – dramatized in Rosenstein-Rodan's (1943) example of the shoe factory which to operate profitably with modern technologies would have to employ no less than 20 000 workers – and by economies of scale due to indivisibilities and technical discontinuities in the provision of infrastructure, or as Rosenstein-Rodan put it, 'social overhead capital', which 'requires a minimum high quantum of investment which would serve, say, fifty factories but would cost far too much for one' (Rosenstein-Rodan, 1984, p. 208); (2) technological externalities associated to industrial training and arising from the incomplete appropriability of the social returns of this activity. In his 1943 article on the problems of industrialization in Eastern and South-Eastern Europe, and in later contributions, Rosenstein-Rodan was probably the economist that most vehemently departed from traditional theory in this respect.

The second ingredient refers to an elastic labor supply arising from the presence of surplus labor. It is not my purpose here to try to even briefly summarize the vast literature on the subject. I shall simply point out that the early views on underdevelopment as a situation characterized by a small capital endowment in relation to available labor supplies led to the

conclusion that the elasticity of the labor supply in these conditions was likely to be higher than in developed economies with a much higher capital–labor ratio. The reason was that with a low aggregate capital–labor ratio, the marginal product of labor at full employment in the capital-using sector would be so low that a large fraction of the labor force would remain employed in a non-capitalist or subsistence sector, using technologies with negligible capital intensity. It is worth quoting Lewis at length here:

The capitalist sector is that part of the economy which uses reproducible capital, and pays capitalists for the use thereof. (This coincides with Smith's definition of the productive workers, who are those who work with capital and whose product can therefore be sold at a price above their wages.) We can think, if we like, of capitalists hiring out their capital to peasants; in which case, there being by definition an unlimited number of peasants, only some will get capital, and these will have to pay for its use a price which leaves them only with subsistence earnings. More usually, however, the use of capital is controlled by capitalists, who hire the services of labor . . . The subsistence sector is by difference all that part of the economy which is not using reproducible capital. Output per head is lower in this sector than in the capitalist sector, because it is not fructified by capital . . . As more capital becomes available more workers can be drawn into the capitalist from the subsistence sector, and their output per head rises as they move from one sector to the other (Lewis, 1954, p. 147)

The key necessary condition for the coexistence of these two sectors is that the average product of labor in the non-capitalist sector be higher than the marginal product of labor that would prevail if the whole of the labor force were employed in the capitalist sector.³ And this is what usually happens when the economy-wide average endowment of capital per worker is low. As long as the two sectors coexist, the labor supply to the capitalist sector is bound to be more elastic than in a developed economy where the higher capital endowment per worker turns the use of subsistence technologies unprofitable. How much more elastic this labor supply is depends on the size of the subsistence sector (and thus on the economy-wide capital–labor ratio), the elasticity of substitution in demand between the goods produced by the two sectors and the nature of returns to labor in the subsistence sector.⁴ Under some special conditions (infinite elasticity of substitution between the goods produced in the two sectors and constant returns to labor in the subsistence sector), the supply of labor to the capitalist sector will be perfectly elastic as in Lewis's well-known model. But whether the labor supply is perfectly elastic, or only imperfectly so, is of no great importance to the growth model of early development theory.

Lewis was the economist that developed and emphasized the surplus labor assumption. Nurkse, drawing on Smith, Young and Rosenstein-Rodan,

was the economist who stressed the role of income effects associated to increasing returns. The two ingredients – increasing returns and surplus labor – were present from the ‘beginning’ in Rosenstein-Rodan (1943), as Rosenstein-Rodan rightly claimed in his 1984 contribution (Rosenstein-Rodan, 1984). I believe it fair to say that only Rosenstein-Rodan fully perceived the general equilibrium implications of these two assumptions taken together.⁵

As a number of recent contributions have made clear, bringing these two ingredients together – increasing returns and an elastic labor supply – can generate a model with multiple equilibria in which depending on initial conditions the economy can be stuck in a poverty trap that can only be overcome through a ‘big push’.⁶ A sketch of such a model is presented in what follows. This may be labeled a Rosenstein-Rodan–Lewis model which has as a special case the Lewis model (the case of constant returns to scale in the modern sector of the economy).

Consider an economy with two sectors (S and M) which produce the same good (or basket of goods). Sector S uses traditional production techniques that are labor-intensive (or, more generally, with low productivity owing to the limited use of capital). The other sector (M) uses a mass production technology subject to increasing returns to scale. The corresponding production functions are:

$$S = L_S \quad (8.1)$$

$$M = (K^\mu)K^a L_M^{1-a} \quad \mu > 0, a + \mu < 1 \quad (8.2)$$

where S and M are the levels of production in the two sectors, L_S and L_M are the labor inputs in each sector, K is the capital stock and K^μ reflects the existence of technological externalities associated with the aggregate capital stock accumulated in the past. A positive value of parameter μ guarantees that the capitalist technology exhibits increasing returns to scale. The restriction $a + \mu < 1$ implies the assumption of diminishing returns to capital in the production function of the capital-intensive sector.

Let us also assume that both sectors operate in competitive conditions. The assumption that the capitalist sector is profit-maximizing generates the following labor demand function:

$$L_M = [(1 - a)K^\mu/w_M]^{1/a}K \quad (8.3)$$

In addition, assuming that workers who do not find employment in the capitalist sector are employed in the traditional sector and that wages in the two sectors are equal owing to labor market competition, we have:

$$L = L_S + L_M \quad (8.4)$$

$$w_M = 1 \quad (8.5)$$

where L is the total labor force and w_M is the wage in sector M . We have chosen units so that $w_S = 1$, and since $w_S = w_M$, we have $w_M = 1$.

Using the production functions of the two sectors (equations 8.1 and 8.2), total output ($Y = S + M$) can be written as $Y = L_S + K^{a+\mu} L_M^{1-a}$. Using (8.4) to eliminate L_S from this expression and (8.3) to eliminate L_M (and also using equation (8.5) gives the following equation:

$$Y = L + a(1-a)^{(1-a)/a} K^{1+\mu/a} \quad (8.6)$$

Equation (8.6) shows that even though the capitalist sector's technology is subject to diminishing returns to capital ($a + \mu < 1$), the aggregate production function shows increasing returns to capital ($1 + \mu/a > 1$).⁷ This is so, of course, provided that the two sectors coexist (since (8.6) is derived from the assumption $w_S = w_M = 1$). Otherwise, if the traditional sector disappears, the aggregate production function is the same as that of the capital-intensive sector.

Increasing returns to capital during the phase in which the two sectors coexist are the result of interactions between an elastic labor supply for the capital-intensive sector ($w_M = 1$) and increasing returns to scale ($\mu > 0$). Increasing returns to scale strengthen the effects of capital accumulation on productivity, while the elastic labor supply weakens the effects of capital accumulation on real wages. The rates of profit and capital accumulation may thus be increasing functions of the capital stock. This has two implications. The first is that at very low income levels, the profit rate may be so low that the rate of accumulation falls below the depreciation rate and the capital stock contracts instead of expanding. The economy is then in a profitability trap in which the elastic labor supply and increasing returns interact negatively to block the expansion of the modern sector: the elastic labor supply sets a floor on the real wages which the modern sector has to pay and this, combined with the initial conditions of low productivity, prevents the profitable use of capital-intensive technologies with increasing returns. The inducement to invest is adversely affected so that the initial conditions of low productivity, capital scarcity and small market size persist.

The second implication is that the dynamics of growth are very different from the transition to long-term equilibrium in neoclassical models and from accumulation processes in endogenous growth models. In contrast to what happens in neoclassical models (and in line with what happens in endogenous growth models with increasing returns) at low income levels,

but beyond the profitability trap, the interactions between increasing returns to scale and an elastic labor supply are positive and counteract the influence of diminishing returns to capital in the technology of the capital-intensive sector. As a result, the growth rate may increase over a long period, generating a trend towards divergence in income levels. In contrast to what happens in endogenous growth models (and in line with what happens in neoclassical models), a reduction in the elasticity of the labor supply at higher income levels, as the ratio of capital to labor increases and the traditional sector disappears, tends to reduce the rates of profit and growth and, therefore, to generate convergence. Thus, the model implies transitional dynamics characterized by a pattern of conditional divergence followed by convergence, in which the highest rates of accumulation are found in the intermediate rather than the initial stages of the transition, as occurs in the neoclassical model, or in more advanced stages, as in models with increasing returns to capital.

The virtuous circle of growth may thus converge to a high-level equilibrium in which labor surpluses have been largely absorbed into the capitalist sector and the economy, with a large capital endowment, is able to generate high real wages. Lewis (1954), and following him Fei and Ranis (1964) viewed this high equilibrium as the end of the development process, or the final stage of the transition phase towards a mature economy in which the rate of growth would depend exclusively, as in the Solow equilibrium path, on technical progress and labor-force growth. In between the vicious and virtuous circles, there is a low-level and unstable equilibrium which has to be associated to Rosenstein-Rodan. For it corresponds, indeed, to that critical mass of investments which generates the externalities and scale economies required for a big push towards sustained economic development.

Resource reallocation, factor accumulation and growth

A second aspect of classical development theory refers to the links between resource reallocation, factor accumulation and technological change. The traditional division between the 'static' analysis of resource allocation and the 'dynamic' analysis of growth, as well as the analysis of growth as the outcome of two separate forces, factor accumulation and technical progress, become too artificial in the presence of increasing returns. A reallocation of resources (towards or away from the activities affected by increasing returns) may then have long-lasting effects on growth and growth itself has to be seen as a process of resource reallocation rather than of mere factor accumulation cum technical change.

Moreover, as Kaldor and others used to emphasize, the distinction between movements along a production function and technical progress

(shifts of the production function) becomes blurred under increasing returns to scale. With the expansion of output, more capital-intensive (or roundabout) methods of production become profitable and are adopted. This is so whether these techniques were already known, and not used because they were unprofitable at a lower scale of output, or truly new and become part of the stock of knowledge as the incentives for their invention appear with the expansion of the market. In developing economies, unlike those of developed countries, these technical changes mostly result from the adoption of technologies that were known elsewhere. From this perspective, they constitute a movement along a production function. Yet, their adoption, unlike the typical movement along a production function, is not the consequence of a change in factor prices leading to the substitution of capital for labor, but rather the result of these more capital-intensive techniques becoming profitable as the scale of output increases.

The links among resource reallocation, factor accumulation and technological change are evident in the process of economic growth over the last two centuries. This process has been marked by industrialization, understood as the expansion of the range of goods produced under increasing returns, and by the simultaneous sharp increase in the capital-labor ratio. These two aspects are intimately connected. Paraphrasing Allyn Young (1928), the division of a group of complex processes into a succession of simpler processes, that is made economical by the presence of increasing returns, lends itself to the use of 'roundabout' methods of production which imply the use of more capital in relation to labor.

This approach to growth as resource reallocation was present in classical development economics. The approach faded away, at least in the more theoretically oriented literature, with the triumph of the counter-revolution in development theory that has dominated the field since the mid-1960s.⁸ The neoclassical resurgence brought back the assumptions of constant returns to scale and perfect competition, and restored the traditional distinction between resource allocation and factor accumulation. The move coincided with, and perhaps contributed to, a declining interest in the analysis of growth during the 1970s. Endogenous growth theory has revived the interest in growth and has even brought increasing returns to scale back into the analysis. But it has remained largely within the framework of one-sector or quasi one-sector models, thus missing the links between growth and resource reallocation.

Effective demand and underdevelopment

Development economics was 'born macro', as Taylor and Arida (1988) phrased it in their survey of development theories; but it was not born

Keynesian or structuralist. In Lewis's view: 'from the point of view of countries with surplus labor, Keynesianism is only a footnote to neo-classicism albeit a long, important and fascinating footnote' (Lewis, 1954, p. 140). Nurkse was blunter:

We are here in the classical world of Say's law. In underdeveloped areas there is generally no 'deflationary gap' through excessive savings. Production creates its own demand, and the size of the market depends on the volume of production. In the last analysis, the market can be enlarged only through an all-round increase in productivity. Capacity to buy means capacity to produce. (Nurkse, 1953, pp. 8–9)

A comparison between Lewis labor surplus and Keynesian unemployment illustrates this point. While Lewis labor surplus is the result of a low level of the economy-wide capital–labor ratio, unemployment for Keynes derives from a low level of effective demand in the goods market. For Keynes, a deficiency of demand for goods keeps the real wage above the marginal product of labor at full employment and thus gives rise to an excess supply of labor. An increase in the effective demand for goods that reduces the real wage – by increasing the price level, given the nominal wage – will then cause an expansion of employment along the labor demand curve, thus reducing unemployment.

The similarity with Lewis is that in both cases the real wage is above the marginal product of labor at full employment in the capitalist sector. The difference is that in Lewis nothing can be done about it by increasing effective demand in the goods market. For employment in the capitalist sector to increase as a result of an expansion in the demand for goods, the real wage would have to fall below the wage in the subsistence sector (adjusted for the wage premium). This is prevented by competition in the labor market. The only way to reduce the labor surplus is by expanding not the aggregate demand for goods but the capital stock, an upward shift in the labor demand curve rather than a movement along it. In modern terminology, the Lewis model refers to a situation of labor market equilibrium with a labor surplus resulting from a 'real rigidity'. Keynesian unemployment is a situation in which labor market disequilibrium is associated with (even if it is not due to) a 'nominal rigidity'.

We need not take these warnings against the 'Keynesian temptation' of development economics too literally to recognize that, no matter how valid Keynes's insights and later contributions to development macroeconomics based on them, the development problems on which Rosenstein-Rodan, Nurkse and Lewis focused would remain even if Keynesian problems were successfully overcome. Increasing returns to scale are essential to the development problem, and irrelevant to the Keynesian argument. Despite some

similarities such as the presence of an elastic labor supply, which however need not arise as in Keynes from a low level of resource utilization, we should not confuse these development problems with the effective demand problems on which Keynes focused. Not much is lost, for example, by assuming Say's law when looking at income differences across countries: differences in resource utilization account for a very small fraction of the large gaps in income per capita across the world (Ros, 2000, Chapter 1).

In the case of differences in growth performance, which approach to take depends on the particular questions one is seeking to answer. Keynesian economics seems to have little to say on why Europe and Japan grew faster than the United States in the post-World War II period or why the East Asian newly industrializing countries (NICs) grew faster than the Latin American countries during the 1960s and 1970s. Yet, economies depart from the factor accumulation path, sometimes for prolonged periods of time, and Keynesian problems and structural constraints are not always successfully overcome. Abandoning Say's law seems essential in understanding why Latin America grew so little in the 1980s as compared to its long-run performance, just as it is essential to understand the poor performance of the United States economy during the inter-World War period or that of the Japanese economy in the 1990s.

There are thus a number of situations (in developing and developed countries alike) in which medium- or even long-term growth performance cannot be properly explained if one remains strictly within the framework of early development theorists. This was well recognized by the later structuralist contributions to development economics. The neglect of effective demand failures and structural constraints, while in the spirit of early development theory, can therefore be an important limitation under some circumstances.

The scope of classical development economics

A final theme refers to the scope of early writing on development theory. While inspired by poverty traps at low income levels in closed economies, this analytical framework can help us think about a much wider variety of development problems than those to which it was originally applied. Development traps can arise under a broad set of circumstances involving increasing returns, demand elasticities and factor supply elasticities. These circumstances are not confined to low levels of economic development. Because the slow rate of accumulation in the trap is due to a low rate of return to capital, the approach has greater generality than other poverty trap models which rely, for example, on vicious circles between income and savings or population growth. The framework can be fruitfully applied to any situation in which a combination of demand and factor supply

elasticities together with a dose of increasing returns in new industries interact to hold back the 'inducement to invest'.

Moreover, those circumstances are not confined to a closed system. Although sometimes formulated or illustrated with a closed economy, the argument survives the extension to the case of an open economy. More precisely, the big push argument does not survive in the case of horizontal pecuniary externalities (involving demand spillovers across sectors producing tradable goods) in a small open economy facing given terms of trade, because the profitability of the shoe factory does not depend any longer on the presence of industrial investments in other tradable goods sectors. This case – which, it should be noted, requires perfect tradability and infinite demand elasticities for tradable goods – is the one envisaged by Bhagwati (1985) in his critique of Rosenstein-Rodan (1943). However, the argument remains unaffected in the case of technological externalities – as in the model presented above – or in the case of vertical pecuniary externalities (involving demand interactions between producers of tradable goods and providers of non-tradable goods inputs subject to economies of scale). This second case has been analyzed by Rodrik (1994), Rodriguez-Clare (1996) and Skott and Ros (1997).

Thus, opening the economy to trade and capital movements introduces important differences and modifies the policy implications but does not make the coordination problems disappear. Coordination failures are likely to emerge, in particular, in the transition from old to new patterns of production and trade specialization. Arguably, this situation is characteristic of a number of semi-industrial 'sandwich economies' in which old comparative advantages in labor-intensive industries are being eroded and the new ones in capital and technology intensive activities are only slowly emerging. Thus, in contrast to the counter-revolution in development theory which denied the usefulness of the approach for the small open economy of a 'typical' developing country, the approach can be fruitfully applied to the development problems of open economies.

In fact, it is when applied to the interpretation of post-World War II development experience that the approach taken by early development theory shows its strengths and most useful insights.⁹ From this perspective, we can view the staggering success stories of East Asia's industrialization (and, to a lesser extent, of a few Latin American countries for some time before the 1980s) as a succession of policy interventions that accelerated the transition between different patterns of production and trade specialization. It is difficult to see how a primarily market-driven development model, that inspires many of today's policy recommendations to developing countries, could have traversed those transitions so successfully. This is not because market-based successes have been entirely absent (this is very

debatable). It is hard to see simply because sound theory suggests exactly the contrary: that market forces are unlikely to address effectively (or, at least, efficiently) the coordination problems of the transition.

Empirical assessment

The broad implications of classical development theory examined in the first section are quite consistent with the trends in per capita incomes since the early 1950s: convergence among the Organisation for Economic Co-ordination and Development (OECD) economies and a number of middle-income developing economies, together with an increasing heterogeneity among the developing countries. The model accommodates in particular two striking features of post-World War II development trends. First, the highest growth rates are found among developing countries and a number of industrial countries that were initially relatively less developed. For example, in the period 1965–85, the economies in the highest quintile of growth in the World Bank data reported by Barro and Lee (1993) were all developing economies, plus Malta, Japan, Portugal, Norway, Greece, Italy and Finland. Second, the lowest growth rates are typically found among the low-income countries.¹⁰ According, again, to the World Bank data reported in Barro and Lee (1993), most of the economies in the lowest quintile of growth for 1965–85 (16 out of 22 economies) were those of low-income countries in sub-Saharan Africa and Asia. Similar conclusions are reached by Ros (2000) for the period 1965–92.

This is surely an incomplete picture, but if too-strong convergence is the reason for abandoning the Solow model, the model of early development theory seems to offer an attractive and alternative worth exploring. For it offers, in particular, an initial answer to the question of why the lowest growth rates are typically found among the initially poorest countries. In this it provides a more attractive framework than the extension of the Solow model suggested by Mankiw et al. (1992) which consists of bringing human capital into the Solow model, along with labor and physical capital, or to Barro's analytical framework which incorporates, in particular, human capital and political risk (Barro, 1991, 1997) (see, on this topic, Ros, 2000).

The implications also seem more consistent with the broad trends of the post-World War II period than those that one may derive from endogenous growth models. For, without further revisions and extensions, the new growth models appear to imply an excessive degree of divergence, much greater than is suggested by historical experience.

Indeed, recent growth models that rely on increasing returns to scale to generate persistent growth are led to assume that the externalities deriving from the process of capital accumulation are so large as to generate

non-diminishing returns to capital in the aggregate production function. Taking a very long-term perspective, Romer (1991) finds this assumption attractive because it is consistent with the fact that the productivity growth rates of the technological leaders have been increasing over the centuries. But if we apply this analytical framework to explain cross-country differences in growth rates, we are faced with some difficulties. Just as diminishing returns to capital in the neoclassical model tend to generate too much convergence, so the assumption of increasing returns to capital tends to generate too much divergence; not only should the gaps in income per capita widen over time, but the differences in growth rates themselves should also become larger.

This implication is avoided in endogenous growth models that restrict the coefficient on capital in the aggregate production to unity (the AK model). With constant returns to capital, these models generate persistent growth at a constant rather than increasing rate. But they do so at the cost of further restricting the assumptions on technology with no support so far from the empirical evidence.¹¹ And again this difficulty is not present in the growth model of early development theory. This model can generate constant or even increasing growth rates over a long transition period without having to rely on restrictive assumptions on technology. The reason is simply that the forces generating constant or increasing growth rates are not exclusively technological. They are rooted in the interaction between a (moderate) dose of increasing returns to scale and a sufficiently elastic supply of labor.

Conclusion

Is this vindication of classical development theory also a policy rehabilitation? The answer is not clear-cut. Classical development economics focused on the coordination problems that would remain in an otherwise well-functioning market economy. One may criticize the associated policy prescriptions for having neglected other sources of malfunctioning and for an overoptimistic attitude towards government policy interventions. Yet these criticisms do not make these problems disappear. The aim of economic reforms in developing countries since the early 1980s has been to alleviate the malfunctioning of the market economy arising from policy distortions. Rather than reducing it, these reform processes may have enhanced the relevance of classical development economics: precisely because these other (policy) sources of malfunctioning are being removed, the focus may now have to shift again to the kind of market failures with which early development theory was concerned.

Notes

1. I am grateful to Amitava Dutt for comments on a previous version of this chapter.

2. *Avant la lettre*, one might add, since most of these writings preceded the neoclassical model of growth at least as formalized by Solow in the mid-1950s (Solow, 1956).
3. Strictly this is so in a model in which the two sectors produce the same goods, like the one Lewis seems to have in mind in the quote above.
4. A full formal treatment of this proposition and others below is presented in Ros (2000, Chapter 3).
5. Expanding on this topic would take us far beyond the scope of this chapter. The basic point is that Nurkse's description of the vicious circle is frequently 'too Smithian', overemphasizing the circularity between the size of the market and the level of productivity (division of labor). This has led to misunderstanding of the 'big push' argument (and to its incorrect assimilation to Nurkse's balanced growth doctrine).
6. See, for example, Murphy et al. (1989) and Krugman (1992, 1995).
7. It is worth noting that with constant returns to scale in the production of the capitalist sector ($\mu = 0$), the aggregate production function would feature constant returns to capital, as long as the two sectors coexist, exactly as in the Lewis model and modern AK models.
8. I use the term 'counter-revolution' in development theory, or neoclassical resurgence, to indicate the partial abandonment of the labor-surplus-increasing returns paradigm in development economics. Both of these terms are somewhat misleading, however, as there was no neoclassical development economics before the 1940s.
9. For an interpretation of South Korea and Taiwan take-offs in terms of a coordination failure model (see Rodrik, 1994).
10. On the subject, see also Sen (1993). Out of the 16 countries with the lowest growth rates reported by Sen (1993), 12 were low-income countries in sub-Saharan Africa.
11. As Solow has observed: 'If [this branch of the new growth theory] found strong support in empirical material, one would have to reconsider and perhaps try to find some convincing reason why Nature has no choice but to present us with constant returns to capital. On the whole, however, the empirical evidence appears to be less than not strong; if anything, it goes the other way' (Solow, 1994, p. 51).

References

- Barro, R.J. (1991), 'Economic Growth in a Cross Section of Countries', *Quarterly Journal of Economics*, 106: 407–44.
- Barro, R.J. (1997), *Determinants of Economic Growth*, Cambridge, MA: MIT Press.
- Barro, R.J. and J. Lee (1993), 'Losers and Winners in Economic Growth', *Proceedings of the World Bank Annual Conference on Development Economics 1993*, Washington, DC: World Bank.
- Bhagwati, J. (1985), *Essays on Development Economics*, Vol. 1, Cambridge, MA: MIT Press.
- Fei, J. and G. Ranis (1964), *Development of the Labor Surplus Economy*, Homewood, IL: R.D. Irwin.
- Krugman, P. (1992), 'Toward a Counter-Counterrevolution in Development Theory', in *Proceedings of the World Bank Annual Conference on Development Economics*, Washington, DC: World Bank.
- Krugman, P. (1995), *Development, Geography, and Economic Theory*, Cambridge, MA: MIT Press.
- Lewis, A. (1954), 'Economic Development with Unlimited Supplies of Labor', *The Manchester School of Economic and Social Studies*, May: 139–91.
- Mankiw, G., D. Romer and D. Weil (1992), 'A Contribution to the Empirics of Economic Growth', *Quarterly Journal of Economics*, 107: 407–37.
- Murphy, K., A. Shleifer and R. Vishny (1989), 'Industrialization and the Big Push', *Journal of Political Economy*, October: 1003–26.
- Nurkse, R. (1952), 'Some International Aspects of the Problem of Economic Development', *American Economic Review*, May: 571–83.
- Nurkse, R. (1953), *Problems of Capital Formation in Underdeveloped Countries*, New York: Oxford University Press.

- Rodriguez-Clare, A. (1996), 'The Division of Labor and Economic Development', *Journal of Development Economics*, 49: 3–32.
- Rodrik, D. (1994), 'Getting Interventions Right: How South Korea and Taiwan Grew Rich', NBER Working paper no. 4964.
- Romer, P.M. (1991), 'Increasing Returns and New Developments in the Theory of Growth', in W. Barnett (ed.), *Equilibrium Theory and Applications: Proceedings of the 6th International Symposium in Economic Theory and Econometrics*, Cambridge: Cambridge University Press.
- Ros, J. (2000), *Development Theory and the Economics of Growth*, Ann Arbor, MI: University of Michigan Press.
- Rosenstein-Rodan, P. (1943), 'Problems of Industrialization of Eastern and South-Eastern Europe', *Economic Journal*, June–September: 202–11.
- Rosenstein-Rodan, P. (1961), 'Notes on the theory of the big push', in H. Ellis (ed.), *Economic Development for Latin America*, Proceedings of a Conference held by the IEA, New York: St Martin's.
- Rosenstein-Rodan, P. (1984), 'Natura Facit Saltum: Analysis of the Disequilibrium Growth Process', in G. Meier and D. Seers (eds), *Pioneers in Development*, New York: Oxford University Press.
- Sen, A. (1993), 'Economic Regress: Concepts and features', *Proceedings of the World Bank Annual Conference on Development Economics 1993*, World Bank.
- Skott, P. and J. Ros (1997), 'The "Big Push" in an Open Economy with Non Tradable Inputs', *Journal of Post Keynesian Economics*, 20: 149–62.
- Solow, R.M. (1956), 'A Contribution to the Theory of Economic Growth', *Quarterly Journal of Economics*, 70: 65–94.
- Solow, R.M. (1994), 'Perspectives on Growth Theory', *Journal of Economic Perspectives*, 8: 1.
- Taylor, L. and P. Arida (1988), 'Long-Run Income Distribution and Growth', in H. Chenery and T.N. Srinivisan, *Handbook of Development Economics*, Vol. 1, Amsterdam: North-Holland.
- Young, A. (1928), 'Increasing Returns and Economic Progress', *Economic Journal*, 38: 527–42.

9 Theories of dependency

*José Gabriel Palma*¹

Dependency theories emerged in Latin America in the early 1960s as attempts to transform Marxist and Structuralist thinking radically about both the obstacles facing capitalist development in the periphery, and whether there still was an actual need for capitalist development as a necessary transition step towards socialism.

There can be little doubt that the Cuban Revolution was a turning point in Marxist analysis of capitalist development in the periphery. The events in Cuba gave rise to a new approach, of which most of the ‘dependency analyses’ form part, which argued against both the feasibility and the need of a capitalist ‘bourgeois-democratic’ revolution in the backward regions of the capitalist world. Consequently, this approach also argued against the politics of the popular fronts,² and in favour of a policy of immediate transition towards socialism in the periphery.

The pre-dependency, pre-Cuban Revolution approach still saw capitalism as historically progressive in the periphery, but argued that its necessary ‘bourgeois-democratic’ revolution was being inhibited by a new alliance between imperialism and the traditional elites. The ‘bourgeois-democratic’ revolution was the revolt of the forces of production against the old pre-capitalist relations of production. This revolution would be based on an alliance between the emerging bourgeoisie and the working classes; the principal battle line in this revolution would be between the bourgeoisie and the traditional oligarchies, between industry and land, capitalism versus pre-capitalist forms of monopoly and privilege. Because it was the result of the pressure of a rising class whose path was being blocked in economic and social terms, this revolution would bring not only political emancipation but economic progress as well.

As this ‘bourgeois-democratic’ revolution in the periphery was being hindered by a new ‘feudal-imperialist’ alliance, this pre-dependency approach identified imperialism as the main enemy – in one way or another the omnipresent explanation of every social and ideological process that occurred. The principal target in the struggle was therefore unmistakable: North American imperialism. The allied camp for this fight, on the same reasoning, was also clear: everyone, except those internal groups allied with imperialism. Thus, for this traditional approach the anti-imperialist struggle was at the same time a struggle for capitalist development and

industrialization. The local state and the 'national' bourgeoisie appeared as the potential leading agents for the development of the capitalist economy, which in turn was viewed as a necessary stage towards socialism.

The Cuban Revolution questioned the very essence of this approach, insisting that the local bourgeoisies in the periphery no longer existed as an active social force but had become 'lumpen', incapable of rational accumulation and rational political activity, dilapidated by their consumerism, and blind to their 'real' interest. It is within this framework, and with the explicit motive of developing theoretically and documenting historically this new form of 'dependency' analysis of the Latin American revolution, that André Gunder Frank appeared on the scene. At the same time, both inside and out of the Economic Commission for Latin America (ECLA), there began to develop two other major approaches to 'dependency'.

The general focus of all 'dependency' analyses is the development of peripheral capitalism (or lack of it). More specifically, these studies attempted to analyse the obstacles to capitalist development in the periphery from the point of view of the interplay between 'internal' and 'external' structures. However, this interplay was analysed in several different ways.

With the necessary degree of simplification that every classification of intellectual tendencies entails, I distinguish between three major approaches – not mutually exclusive from the point of view of intellectual history – in 'dependency' analysis. First is the approach begun by Frank, its essential characteristic being that it attempted to construct a comprehensive theory of the impossibility of capitalist development in the periphery. In these theories the 'dependent' character of peripheral economies is the crux on which the whole analysis of underdevelopment turns; that is, dependency is seen as causally linked to permanent capitalist underdevelopment.

The second approach is associated with one branch of the ECLA Structuralist School, especially Celso Furtado, Anibal Pinto and Osvaldo Sunkel. These writers sought to reformulate the classical ECLA analysis of Latin American development from the perspective of a critique of the obstacles to 'national' development. This attempt at reformulation was not a simple process of adding new elements (both political and social) that were lacking in the original Prebisch–ECLA analysis, but a thoroughgoing attempt to proceed beyond that analysis, adopting an increasingly different perspective.

The final approach, deliberately avoiding the formulation of a mechanico-formal theory of dependency – and specifically, a mechanico-formal theory of the inevitability of underdevelopment in the capitalist periphery based on its dependent character – concentrated on what has been called the study of 'concrete situations of dependency'. In the words of Fernando Henrique Cardoso:

The question which we should ask ourselves is why, it being obvious that the capitalist economy tends towards a growing internationalisation, that societies are divided into antagonistic classes, and that the particular is to a certain extent conditioned by the general, with those premises we have not gone beyond the partial – and therefore abstract in the Marxist sense – characterisation of the Latin American situation and historical process. (Cardoso, 1972, pp. 326–7)

What was needed therefore was the study of the concrete forms in which dependent relationships develop; that is to say, the precise forms in which the different economies and politics of the periphery have been articulated with those of the advanced nations at different times, and how their specific dynamics have thus been generated.

According to this view, one of the main problems with most Marxist analyses at the time (and probably since Lenin's death) was that they had not taken the diversity of actual experiences of peripheral countries seriously enough. What was clear was that without a considerable improvement in our knowledge of this diversity of historical experiences, general theories of capitalist development in the periphery were doomed to fall into the trap of 'abstract dialectical thought', or the working out upon itself of an abstract dialectic, unable to move from previously constructed concepts.

Dependency as a formal theory of the inevitability of capitalist underdevelopment: on cutting a knot that could not be unravelled

There is no doubt that the 'father' of this approach was Paul Baran. His principal contribution (1957) took up the approach of the Sixth Congress of the Comintern regarding the supposedly irresolvable nature of the contradictions between the economic and political needs of imperialism and those of the process of industrialization and economic development of the periphery.

To defend its interests, international monopoly capital would not only form alliances with pre-capitalist domestic elites intended to block progressive capitalist transformations in the periphery, but its activities would also have the effect of distorting the process of capitalist development in these countries. As a result, international monopoly capital would have easy access to peripheral resources and the traditional elites in the periphery would be able to maintain traditional modes of surplus extraction and continued monopoly on power. Within this context the possibilities for economic growth in dependent countries were extremely limited, or non-existent; the surplus they generated was largely expropriated by foreign capital, or otherwise squandered by traditional elites. This process would necessarily lead to economic stagnation and underdevelopment in the periphery. The only way out was political. At a very premature stage, capitalism had become a fetter on the development of the productive forces in

the periphery and, therefore, its historical role had already come to an early end.

Baran developed his ideas influenced both by the Frankfurt School's general pessimism regarding the nature of capitalist development and by Sweezy's proposition that the rise of monopolies imparts to capitalism a tendency towards stagnation and decay. He also followed the main growth paradigm of his time, the Harrod–Domar theory, which held that the size of the investable surplus was the crucial determinant of growth (together with the efficiency with which it was used: the incremental capital–output ratio).

Starting out with Baran's analysis, Frank attempted to prove the thesis that the only political and economic solution to capitalist underdevelopment was a revolution of an immediately socialist character. For our purposes we may identify three levels of analysis in Frank's model of the 'development of underdevelopment'. In the first, (arguing against 'dualistic' analyses) he attempted to demonstrate that the periphery has been incorporated and totally integrated into the world capitalist economy since the very early stages of colonial rule. In the second, he attempts to show that such incorporation into the world capitalist economy has transformed the countries in question immediately and necessarily into fully capitalist economies. Finally, in the third level, Frank tries to prove that the integration of these supposedly capitalist economies into the world capitalist system was achieved through an interminable metropolis–satellite chain, through which the surplus generated at each stage was successfully siphoned off towards the centre.

However, Frank never defines what he means by capitalism; he simply affirms that, since the periphery was never 'feudal' and has always been fully incorporated into the world capitalist system, then it must follow that it has been 'capitalist' from the beginning of colonial times, that is, from the very beginning of their integration into the world capitalist system. In turn, for Frank it is capitalism (and nothing else but capitalism), with its metropolis–satellite relations of exploitation, which has produced underdevelopment. The choice was clear: socialist revolution or continuing endlessly to underdevelop within capitalism. Therefore, '[t]o support the bourgeoisie in its already played-out role on the stage of history is treacherous or treachery' (Frank, 1967, p. xvii).

In my opinion, the real value of Frank's analysis is his critique of the supposedly dual structure of peripheral societies. Frank shows clearly that the different sectors of the economies in question are and have been, since very early in their colonial history, linked to the world economy. Moreover, he has correctly emphasized that this connection has not automatically brought about capitalistic economic development, such as 'optimistic'

models (derived from Adam Smith) would have predicted, in which the development of trade and the division of labour would inevitably bring about economic development. Nevertheless, Frank's error (shared by the whole tradition of which he is part, including Sweezy, Amin and Wallerstein among the better known) lies in his attempt to explain this phenomenon by using the same economic deterministic framework of the model he purports to transcend. In fact, he merely turns it upside-down: the development of the 'core' necessarily requires the underdevelopment of the periphery.

It is not surprising that this method leads Frank to displace class relations from the centre of his analysis of economic development and underdevelopment. Although it is evident that capitalism is a system where production for profits via exchange predominates, the opposite is not necessarily true: the existence of production for profits in the market is not necessarily an indication of capitalist production. For Frank, this is a sufficient condition for the existence of capitalist relations of production; thus he develops a circular concept of capitalism.

Although Frank did not go very far in his analysis of the world capitalist system as a whole, of its origins and its development, Wallerstein tackled this tremendous challenge in two remarkable books (1974, 1980).

The central concerns of Frank's theory of the 'development of underdevelopment' are addressed from a critical point of view by dos Santos, Marini, Caputo, Pizarro, Hinkelammert, and continued later on by Amin and many non-Latin American social scientists.³ The most thoroughgoing critiques of these theories of underdevelopment have come from Laclau, Cardoso, Lall, Warren, Brenner and Palma.

I would argue that the theories of dependency examined here are mistaken not only because they do not 'fit the facts', but also – and equally important – because their mechanico-formal nature renders them both static and ahistorical. Their analytical focus has not been directed to the understanding of how new forms of capitalist development have been marked by a series of specific economic, political and social contradictions, instead only to assert the claim that capitalism had lost, or never had, a historically progressive role in the periphery.

Now, if the argument is that the progressiveness of capitalism has manifested itself in the periphery differently than in advanced capitalist countries, or in diverse ways in the different branches of the peripheral economies, or that it has generated inequality at regional levels and in the distribution of income, and has been accompanied by such phenomena as underemployment and unemployment, and has benefited the elite almost exclusively, or again that it has taken on a cyclical nature, then this argument does no more than affirm that the development of capitalism in the

periphery, as in any other area and at all times, has been characterized by its contradictory and exploitative nature. The specificity of capitalist development in the Third World stems precisely from the particular ways in which these contradictions have been manifested, the different ways in which many of these countries have faced and temporarily overcome them, the ways in which this process has created further contradictions, and so on. It is through this process that the specific dynamic of capitalist development in different peripheral countries has been generated.

In this connection, we should recall that the whole of Lenin's analysis of the development of capitalism in Russia was a detailed study of the specific way in which capitalism in that country developed and temporarily and partially overcame its contradictions. It is important to remember that he specifically criticized the Narodniks for interpreting these contradictions as 'proof' that capitalism was impossible in Russia, and for failing to understand that the same contradictions were the very ones which were basic to capitalist development, and which took specific forms in Russia.

To deny, as these 'contemporary Narodniks' do, that capitalist development is taking place in many countries in the Third World, is no less than absurd. To recognize these changes, in turn, as Lenin told the Narodniks, 'is quite compatible with the full recognition of the many negative sides of capitalism, and is in no way an apology for it'.

Reading their political analysis, one is left with the impression that the whole question of what course the revolution should take in the periphery revolves solely around the problem of whether or not capitalist development is viable. In other words, their conclusion seems to be that if one accepts that capitalist development is feasible on its own terms, one is automatically bound to adopt the political strategy of awaiting and/or facilitating such development until its full productive powers have been exhausted, and only then to seek to move towards socialism. As it is precisely this option that these writers wish to reject, they have been obliged to make in their work a forced march back towards a pure ideological position to deny dogmatically any possibility of capitalist development in the periphery.

Dependency as a reformulation of the ECLA analysis of Latin American development

Towards the middle of the 1960s the ECLA approach to Latin American development suffered a gradual decline due to several key factors. Some statistics relating to Latin American development in the period after the Korean War presented an apparently gloomy picture, which was interpreted (including within ECLA) as a failure of ECLA-type policies. This new pessimism regarding the viability of capitalist development in the

periphery, including of the 'reformed' type, led structuralist thinkers to change their basic paradigm, as the Cuban Revolution had for the majority of the traditional Marxist Left. Furthermore, the first attempts to introduce into classical ECLA analysis a number of 'social' and 'political' aspects did not strengthen the analysis but instead revealed its fragility.

The process of import-substituting industrialization, which ECLA recommended, seemed to have aggravated balance-of-payments problems, instead of alleviating them. Foreign investment was not only partly responsible for that (as after a certain period of time there was a tendency for a net flow of capital away from the subcontinent), but also did not seem to have brought other positive effects that ECLA had expected. Income distribution was worsening in several countries. The problem of unemployment was also growing more acute, in particular as a result of rural-urban migration. Industrial production was becoming increasingly concentrated in products typically consumed by the elites, and was not having much of a 'ripple effect' upon other productive sectors of the economy, particularly the agricultural sector.

The apparently gloomy panorama of capitalist development in Latin America in the 1960s led to substantial ideological changes in many influential ECLA thinkers, and it strengthened the convictions of the Marxist 'dependency' writers reviewed earlier. The former were faced with the problem of trying to explain some of the unexpected consequences of their policies – particularly concerning industrialization. The latter were led to deny with the greatest possible vehemence the least possibility of dependent capitalist development.

Finally, by making a basically ethical distinction between 'economic growth' and 'economic development', their research followed two separate lines, one concerned with the obstacles to economic growth (and in particular to industrial and agricultural growth), and the other concerned with the perverse character taken by local 'development'. The fragility of this formulation lies in its inability to distinguish between a socialist critique of capitalism and the analysis of the actual obstacles to capitalist development in the periphery.

Dependency as a methodology for the analysis of concrete situations of development

In my critique of the dependency studies reviewed so far, I have described the fundamental elements of what I understand to be the third of the three approaches within the dependency school. This approach is primarily associated with the work of the Brazilian sociologist Fernando Henrique Cardoso and the Chilean historian Enzo Faletto, dating from the completion of their 1967 book.

Briefly, this third approach to the analysis of dependency can be summarized as follows. In common with the two other approaches to 'dependency' discussed already, this third approach sees the Latin American economies as an integral part of the world capitalist system, in the context of increasing internationalization of the system as a whole. It also argues that the central dynamic of that system lies outside the peripheral economies and that, therefore, the options which are open to them are limited (but not determined) by the development of the system at the centre. In this way the 'particular' is in some way conditioned by the 'general'. Therefore, a basic element for the analysis of these societies is given by the understanding of the 'general determinants' of the world capitalist system, which is itself rapidly changing. The analysis therefore requires an understanding of the contemporary characteristics of the world capitalist system. However, the theory of imperialism, which was originally developed to provide an understanding of the dynamics of that system, has had enormous difficulty in keeping up with the significant and decisive changes in the capitalist system since the death of Lenin. During this period, capitalism underwent substantial changes, and the theory totally failed to keep up with them properly.

One widely recognized characteristic of the third approach to dependency has been its ability to incorporate these transformations more successfully. For example, this approach was quick to grasp that the rise of the multinational corporations progressively transformed centre-periphery relationships, as well as relationships between the countries of the centre. As foreign capital became increasingly directed towards manufacturing industry in the periphery, the struggle for industrialization, which was previously seen as an anti-imperialist struggle, in some cases increasingly become the goal of foreign capital. Thus dependency and industrialization ceased to be necessarily contradictory processes, and a path of 'dependent development' for important parts of the periphery became possible.

The third approach has not only accepted, but also enriched the analysis of how developing societies are structured through unequal and antagonistic patterns of social organization, showing the social asymmetries, the exploitative character of social organization and its relationship with the socio-economic base. This approach has also given considerable importance to aspects of each economy like the effect of the diversity of natural resources, geographic location and so on, thus also extending the analysis of the 'internal determinants' of the development of the Latin American economies.

However, while these improvements are important, the most significant feature of this approach is that it attempts to go beyond these elements, and insists that from the premises so far outlined one arrives only at a partial, abstract and indeterminate characterization of the historical process in the periphery, which can only be overcome by understanding how the general

and specific determinants interact in particular and concrete situations. It is only by understanding the specificity of 'movement' in the peripheral societies as a dialectical unity of both these 'internal' and 'external' factors – that one can explain the particularity of social, political and economic processes in these societies.

Only in this way can one explain how, for example, the same process of mercantile expansion could simultaneously produce systems of slave labour, systems based on other forms of exploitation of indigenous populations, and incipient forms of wage labour. What is important is not simply to show that mercantile expansion was the basis of the transformation of most of the periphery, and even less to deduce mechanically that that process made these countries immediately capitalist. Rather, this approach emphasizes the specificity of history and seeks to avoid vague, abstract concepts by demonstrating how, throughout the history of backward nations, different sectors of local classes allied or clashed with foreign interests, organized different forms of the state, sustained distinct ideologies or tried to implement various policies or defined alternative strategies to cope with imperialist challenges in diverse moments of history.

The study of the dynamic of dependent societies as a dialectical unity of internal and external factors implies that the conditioning effect of each on the development of these societies can be separated only by undertaking a static analysis. Equally, if the internal dynamic of the dependent society is a particular aspect of the general dynamic of the capitalist system, it does not imply that the latter produces concrete effects in the former, but only that it finds concrete expression in that internal dynamic.

The system of 'external domination' reappears as an 'internal phenomenon' through the social practices of local groups and classes, who share the interests and values of external forces. Other internal groups and forces oppose this domination, and in the concrete development of these contradictions the specific dynamic of the society is generated. It is not a case of seeing one part of the world capitalist system as 'developing' and another as 'underdeveloping', or of seeing imperialism and dependency as two sides of the same coin, with the underdeveloped or dependent world reduced to a passive role determined by the other. Instead, in the words of Cardoso and Faletto:

We conceive the relationship between external and internal forces as forming a complex whole whose structural links are not based on mere external forms of exploitation and coercion, but are rooted in coincidences of interest between local dominant classes and international ones, and, on the other hand, are challenged by local dominated groups and classes. In some circumstances, the networks of coincident or reconciliated interests might expand to include segments of the middle class, if not even of alienated parts of working classes. In other

circumstances, segments of dominant classes might seek internal alliance with middle classes, working classes, and even peasants, aiming to protect themselves from foreign penetration that contradicts their interests (Cardoso and Faletto, 1967 [1977], pp. 10–11).

There are, of course, elements within the capitalist system that affect all developing economies, but it is precisely the diversity within this unity that characterizes historical processes. Thus the analytical focus should be oriented towards the elaboration of concepts capable of explaining how the general trends in capitalist expansion are transformed into specific relationships between individuals, classes and states, how these specific relations in turn react upon the general trends of the capitalist system, how internal and external processes of political domination reflect one another, both in their compatibilities and their contradictions, how the economies and politics of peripheral countries are articulated with those of the centre, and how their specific dynamics are thus generated.

However, as is obvious, this third approach to the analysis of peripheral capitalism is not unique to ‘dependency’ studies and as such, in time, has superseded them.

Notes

1. I am extremely grateful to Fiona Tregenna for many constructive comments on a previous draft.
2. The idea of the ‘popular fronts’ emerged from the Seventh Congress of the Comintern; the basic proposition of this analysis as far as developing countries were concerned was that the main obstacle for capitalist development in the periphery was the ‘feudal–imperialist’ alliance (see below).
3. For an analysis of the Latin American branch of this school of thought, see Kay (1989).

Bibliography

- Amin, S. (1972), ‘Underdevelopment and Dependence in Black Africa: Origins and Contemporary Forms’, *Journal of Modern African Studies*, 10 (4).
- Baran, P. (1957), Political Economy of Growth, Monthly Review.
- Cardoso, F.H. (1972), ‘Dependent Capitalist Development in Latin America’, *New Left Review*, 74.
- Cardoso, F.H. and E. Faletto (1967), *Dependencia y Desarrollo en América Latina*, Mexico, Siglo XXI Editores; English edition (with a new Preface) (1977), Berkeley, CA: University of California Press.
- CESO (1968), *Imperialismo y Dependencia Externa*, Santiago: CESO.
- Dos Santos, T. (1970), ‘The Structure of Dependence’, *American Economic Review*, 60 (2).
- ECLA (1969), *El Pensamiento de la CEPAL*, Santiago: Editorial Universitaria.
- Frank, A.G. (1967), *Capitalism and Underdevelopment in Latin America: Historical Studies of Chile and Brazil*, New York: Monthly Review Press.
- Furtado, C. (1966), *Subdesarrollo y Estancamiento en América Latina*, Buenos Aires: CEAL.
- Hinkelammert, F. (1970), *El Subdesarrollo Latino-americano: un caso de desarrollo capitalista*, Santiago: Ediciones Nueva Universidad.
- Kay, C. (1989), *Latin American Theories of Development and Underdevelopment*, London and New York: Routledge.

- Laclau, E. (1971), 'Feudalism and Capitalism in Latin America', *New Left Review*, May–June.
- Owen, R. and B. Sutcliffe (eds.) (1972), *Studies in the Theory of Imperialism*, London: Longman.
- Palma, J.G. (1978), 'Dependency, a Formal Theory of Underdevelopment or a Methodology for the Analysis of Concrete Situations of Underdevelopment?' *World Development*, 6 (7/8).
- Palma, J.G. (2008), Entries on Raúl Prebisch, Dependency and (with Stephanie Blankenburg and Fiona Tregenna) Structuralism, *The New Palgrave: A Dictionary of Economics*, Macmillan.
- Palma, J.G. (forthcoming), *Radical Theories of Development: A Critical Reappraisal*, Academic Press.
- Wallerstein, I. (1974), *The Modern World System I: Capitalist Agriculture and the Origins of the European World-Economy in the Sixteenth Century*, New York: Academic Press.
- Wallerstein, I. (1980), *The Modern World System II: Mercantilism and the Consolidation of the European World-Economy, 1600–1750*, New York: Academic Press.
- Warren, B. (1980), *Imperialism: Pioneer of Capitalism*, J. Sender, London: New Left Books.

10 Structuralism

José Gabriel Palma

Structuralism is basically a method of enquiry which challenges the assumptions of empiricism and positivism. This method is found in literary criticism, linguistics, aesthetics and social sciences, both Marxist and non-Marxist.

The principal characteristic of structuralism is that it takes as its object of investigation a 'system', that is, the reciprocal relations among parts of a whole, rather than the study of the different parts in isolation. In a more specific sense this concept is used by those theories that hold that there are a set of social and economic structures that are unobservable but which generate observable social and economic phenomena.

In anthropology, structuralism is particularly associated with Lévi-Strauss and Godelier. The main structuralist current in Marxist thought has its origins in Althusser and stands in opposition to the version of Marxist theory developed by Lukacs, Gramsci and the Frankfurt School. While structuralism seeks to explain social phenomena by reference to the underlying structure of the mode of production (hence, trying not to be 'humanistic' or 'historicist' in a teleological sense), the second group of Marxist theories stress the role of human consciousness and action in social life, with a concept of history in which (arguably) some idea of 'progress' is either implicit or explicit.

In economics structuralism is primarily associated with the school of thought originated in the ECLAC (United Nations Economic Commission for Latin America and the Caribbean), and in particular with the work of its first director, Raul Prebisch.

The key to the internal unity of the ECLAC thought lies in its early postulation of the original ideas and hypotheses around which its subsequent contributions would be organized. The starting point was the idea that the world economy was composed of two poles, the 'centre' and the 'periphery', and that the structure of production in each differed substantially. That of the centre was seen as homogeneous and diversified, that of the periphery, in contrast, as heterogeneous and specialized; heterogeneous because economic activities with significant differences as to productivity existed side by side, with the two extremes provided by an export sector with relatively high productivity of labour, and a subsistence agriculture in which it was particularly low; specialized because the export sector would

tend to be concentrated upon a few primary products, with production characteristically confined to an 'enclave' within the peripheral economic structure or, in other words, having very limited backward and forward linkage effects with the rest of the economy. It was this structural difference between the two types of economy which lay behind the different function of each pole in the international division of labour, and this in turn had the effect of reinforcing the structural difference between the two.

Thus the two poles were closely bound together, and were mutually and reciprocally conditioning. Therefore, the structural difference between centre and periphery could not be defined or understood in static terms, as the transformation of either pole would be conditioned by the interaction between them. Centre and periphery formed a single system, dynamic by its very nature.

The nucleus of ECLAC analysis was the critique of the conventional theory of international trade (as expressed in the Heckscher–Ohlin–Samuelson version of Ricardo's theory of comparative advantage); it aimed to show that the international division of labour which conventional theory claimed was 'naturally' produced by world trade was of much greater benefit to the centre (where manufacturing production is concentrated) than to the periphery (which was destined mainly to produce primary products, be they agricultural or mineral). The analysis of the ECLAC has a unity and an internal coherence which is not always perceptible at first sight, as its component parts are scattered through numerous documents published over a period of years (mainly in the 1950s and 1960s). Several contributions had their origins in the examination of specific problems, around which a series of theoretical arguments were articulated, in an attempt to isolate their causes and to justify the economic policy measures recommended to resolve them.

The ECLAC analysis turns on three tendencies which are considered inherent to the development of the periphery: unemployment of the labour force, external disequilibrium and the tendency to deterioration of the terms of trade (see Rodriguez, 1980, 2006).

Structural heterogeneity and unemployment

The problem of employment in the periphery has two facets: the absorption of additions to the active population, and the reabsorption of the labour force of the most backward areas into economic activities in which productivity is higher. As the ECLAC analysis assumes that demand for labour is proportionate to the level of investment (its rate of growth is directly related to the rate of capital accumulation), and this takes place almost exclusively in the modern sector, full employment of the labour force at adequate levels of productivity can only be achieved if the rate of

capital accumulation in the export sector and in import-substituting manufacturing activities is sufficient not only to absorb the growth in the whole of the active population, but also to reabsorb labour from the traditional sector. Thus the level of employment depends on the balance between the growth of the active population and the rhythm of the expulsion of labour from the traditional sector, and on the level of capital accumulation in the modern sector. It is from the heavy burden on the modern sector to provide full employment in the economy at an adequate level of productivity that the structural tendency towards unemployment in the peripheral economies is deduced.

Specialization in production and external disequilibrium

The structure of production in the periphery is specialized in a double sense: mainly primary products are exported, and the economies are in general poorly integrated. From this it follows that a significant proportion of the demand for manufactured products is oriented towards imports, and given that their income elasticity is greater than unity, imports tend to grow faster than the level of real income. The opposite is the case in the centre, as imports from the periphery consist essentially of primary products, for which income elasticity is usually less than unity; hence they grow less rapidly than real income.

Thus for a given rate of growth of real income in the centre, the disparity between the income elasticities of imports at each pole will impose a limit upon the rate of growth of real income in the periphery (unless the latter is able to diversify its productive structure). This will tend not only to be less than that of the centre, but also to be less in proportion to the degree of the disparity between the respective income elasticities of demand for imports. If the periphery attempts to surpass this limit, it will expose itself to successive deficits in its balance of trade; the only long-term alternative will be an increased effort to satisfy the highly income-elastic demand for manufactured products with internal production, and to diversify its export trade towards income-elastic products. Only a process of industrialization, given these assumptions, can allow that and enable the periphery to enjoy a rate of growth of real income higher than that determined by the rate of growth in the centre and the disparity between income elasticities of demand for imports.

As this process of industrialization also generates a need for imports which can exceed the availability of foreign currency deriving from the slow expansion of primary exports, the ECLAC argues in its documents that there is a role for foreign capital in the first stages of the process, both to remedy the shortage of foreign currency, and to complement internal savings.

Specialization, heterogeneity and the tendency to deterioration of the terms of trade

The explanation for the phenomena of the tendency to deteriorating terms of trade and the disparity in incomes which it brings with it are, in the thought of the ECLAC, a logical analytical deduction from the phenomena of specialization and heterogeneity. (It is not, as is usually assumed, the starting point of ECLAC thought, but – given its assumptions and hypotheses – a natural analytical deduction.)

There are, basically, a demand and a supply element behind this tendency to deterioration of the terms of trade of the periphery. The basic problem is the effect of economic growth on the terms of trade. From a demand point of view – given the problem of specialization and the differences in income elasticities for imports between the centre and periphery – the ‘consumption path’ of the periphery is biased towards trade (that is, as incomes grow the proportion of importables from the industrialized North in total consumption increases). From the point of view of supply – given the effect of heterogeneity on technological change and the differences in price elasticity of supply of exports between the centre and the periphery – the ‘production path’ of the periphery is also biased towards trade (that is, as output grows the proportion of exportables to the North in domestic production increases). The combined effect would be a tendency towards an increased demand for imports of manufacturing goods and an increased supply of primary products from the periphery. As income elasticities of the industrialized countries for Southern commodities is low, there would be an asymmetry in the international trade between these two groups of countries that, if left to the ‘invisible hand’ of international markets, would tend to push up prices of the periphery’s imports and push down prices of the periphery’s exports. Thus the tendency towards deterioration of the terms of trade of the periphery.

According to the ECLAC, it is possible to escape from this vicious circle through a process of transformation of the economic structure of the periphery capable, ideally, of providing those economies with a rapid and sustained rate of growth, and avoiding unemployment, external disequilibrium and the deterioration of the terms of trade. The central element in this structural transformation is the process of industrialization, which could provide those highly income-elastic importables and eventually also produce more price-elastic exportables; thus Prebisch summarizes the ECLAC’s task as having been that of ‘showing that industrialization was an unavoidable prerequisite for development’ (1980, p. viii). Furthermore, the article in question appears at times to use the concepts ‘industrialization’ and ‘development’ as synonyms.

In other words, to achieve accelerated and sustained economic growth in the periphery a necessary condition (and, some ECLAC writings seemed to suggest, a sufficient one) was the development of a process of industrialization. But this process could not be expected to take place spontaneously, for it would be inhibited by the international division of labour which the centre would attempt to impose, and by a series of structural obstacles internal to the peripheral economies. Consequently, a series of measures was proposed, intended to promote a process of deliberate or 'forced' industrialization; these included state intervention in the economy both in the formulation of economic policies oriented towards these ends and as a direct productive agent. Among the economic policies suggested were those of 'healthy' protectionism, exchange controls, the attraction of foreign capital into manufacturing industry, and the stimulation and orientation of domestic investment. The intervention of the state in directly productive activities was recommended in those areas where large amounts of slow-maturing investment were needed, and particularly where this need coincided with the production of essential goods or services.

The dimensions of the thought of the ECLAC are based then not only upon its breadth and internal unity, but also upon its structuralist nature. The three most important characteristics of the development of the economy in the periphery – unemployment, external disequilibrium and the tendency to deterioration of the terms of trade – are derived directly from the characteristics of the structure of production in the periphery; thus the possibility of tackling them is seen in terms of an ideal pattern of transformation, which indicates the conditions of proportionality which must hold if those features are to be avoided. This leads to the formulation, tacitly or explicitly, of the law of proportionality in the transformation, which will avoid heterogeneity and will thus allow full employment at adequate levels of productivity, avoid specialization and thus permit the escape from external disequilibria, and thus counteract the tendency towards deterioration of the terms of trade.

Nevertheless, it is also in this very structuralist nature that the limitations of ECLAC thought lie; at this level of analysis no consideration is given to the social relations of production which are at the base of the process of import-substituting industrialization, and of the transformation in other structures of society that this brings in its wake.

The ECLAC proposes an ideal model of sectoral growth – and hence of global growth – designed in such a way that the three tendencies peculiar to economic development of the periphery are not produced; from this are derived the necessary conditions of accumulation which will allow the proportionality required in the transformation of the different sectors of material production. Nevertheless, even when pushed to the limits of its

potential internal coherence, the structural approach is inadequate for the analysis of the evolution in the long term of the economic system as a whole, as it clearly involves more than the transformation of the structure of production alone. The theories of the ECLAC describe and examine certain aspects of the development of the forces of production (to the extent that they deal with the productivity of labour and the degree of diversification and homogeneity of the structures of production), but do not touch on relations of production, nor, as a result, on the manner in which the two interact.

Furthermore, the analysis of the inequalities of development cannot be carried out solely in terms of the patterns of accumulation necessary to avoid the creation of certain disproportions between the different sectors of material production, as inequalities of development are clearly linked to the possibility of saving and accumulation in each pole. That is to say, the requirements as far as accumulation is concerned are derived from those disproportions, but their feasibility depends more upon the general conditions in which accumulation occurs at world level than upon those disproportions. In other words, if the intention is to analyse the bipolarity of the centre-periphery system, it is not enough to postulate the inequality of development of the forces of production; it is necessary also to bear in mind that those forces of production develop in the framework of a process of generation, appropriation and utilization of the economic surplus, and that process, and the relations of exploitation upon which it is based, are not produced purely within each pole, but also between the two poles of the world economy.

It is not particularly surprising that the ECLAC should have attracted its share of criticism, particularly as it went beyond theoretical pronouncements to offer packages of policy recommendations. It was criticized from sectors of the Left for failing to denounce sufficiently the mechanisms of exploitation within the capitalist system, and for criticizing the conventional theory of international trade only from 'within' (see for example Frank, 1967 and Caputo and Pizarro, 1974). On the other hand, from the Right the reaction was immediate and at times ferocious: the ECLAC's policy recommendations were totally heretical from the point of view of conventional theory, and threatened the political interests of significant sectors. A leading critic in academic circles was Haberler (1961), who accused the ECLAC of failing to take due account of economic cycles, and argued that single factorial terms of trade would be a better indicator than the simple relationship between the prices of exports and imports (see also Baldwin, 1955).

On the political front, the Right accused the ECLAC of being the 'Trojan Horse of Marxism', on the strength of the degree of coincidence between

both analyses. In both cases the principal obstacle was located overseas (international division of labour imposed by the centre), and both share the conviction that without a strenuous effort to remove the internal obstacles to development (the traditional sectors) the process of industrialization would be greatly impeded.

Furthermore, the coincidence between crucial elements in the analysis of the two respective lines of thought is made more evident by the fact that the processes of reformulation in each occurred simultaneously. Thus when it became evident that capitalist development in Latin America was taking a path different from that expected, a number of the ECLAC members began a process of reformulation of the traditional thought of that institution, just at the time that an important sector of the Latin American Left was breaking with the traditional Marxist view that capitalist development was both necessary and possible in Latin America, but hindered by the 'feudal-imperialist' alliance. Moreover, both reformulations had one extremely important element in common: pessimism regarding the possibility of capitalist development in the periphery (see Chapter 9 in this volume, on the-ories of dependency).

Some of the ECLAC analysis re-emerged in the 1980s in some North American academic circles (see especially Taylor, 1983). This was mainly an attempt to formalize classical ECLAC thought through the reworking of the assumptions and hypotheses of the traditional ECLAC school using modern economic analysis. Although this exercise has proved to be an important contribution (and a much-needed one) to mainstream economics, it has not really succeeded as an attempt to use structuralism as a new method of enquiry into modern economic analysis.

Bibliography

- Bacha, E.L. (1978), 'An Interpretation of Unequal Exchange from Prebisch to Emmanuel', *Journal of Development Economics*, 5: 319–30.
- Baer, W. (1962), 'The Economics of Prebisch and ECLA', *Economic Development and Cultural Change*, 10: 169–82.
- Baldwin, R.E. (1955), 'Secular Movements in the Terms of Trade', *American Economic Review, Papers and Proceedings*, 45: 259–69.
- Bhagwati, Y. (1960), 'A Skeptical Note on the Adverse Secular Trend in the Terms of Trade of Underdeveloped Countries', *Pakistan Economic Journal*, 8: 235–48.
- Cardoso, F.H. (1977), 'The Originality of the Copy: CEPAL and the Idea of Development', *CEPAL Review*, 4: 7–40.
- Caputo, O. and A. Pizarro (1974), *Dependencia y relaciones internacionales*, Costa Rica: EDUCA.
- Di Marco, L.E. (ed.) (1972), *International Economics and Development: Essays in Honour of Raul Prebisch*, New York and London: Academic Press.
- ECLAC (1963), *El desarrollo social de América Latina en la postguerra*, Buenos Aires: E. Solar-Hachette.
- ECLAC (1964), *El desarrollo económico de América Latina en la post-guerra*, Santiago: UN (CEPAL).

- ECLAC (1965), *El financiamiento externo de América Latina*, Santiago: UN (CEPAL).
- ECLAC (1966), *El proceso de industrialización de América Latina*, Santiago: UN (CEPAL).
- ECLAC (1969), *El pensamiento de la CEPAL*, Santiago: Editorial Universitaria.
- ECLAC (1973), *Bibliografía de la CEPAL. 1948–1972*, Santiago: UN (CEPAL).
- Ellsworth, P.T. (1956), 'The Terms of Trade between Primary Producing and Industrial Countries', *Interamerican Economic Affairs*, Summer.
- Flanders, F.J. (1964), 'Prebisch on Protectionism: An Evaluation', *Economic Journal*, 74: 305–26.
- Frank, A.G. (1967), *Capitalism and Underdevelopment in Latin America: Historical Studies of Chile and Brazil*, New York: Monthly Review Press.
- Johnson, H.G. (1972), *Economic Policies Towards Less Developed Countries*, New York: Brookings.
- Haberler, G. (1961), 'Terms of Trade and Economic Development', in H.S. Ellis (ed.), *Economic Development of Latin America*, New York: St Martin's Press, pp. 275–97.
- Hirschman, A. (1961), 'Ideologies of Economic Development', reprinted in A. Hirschman (1971), *A Bias for Hope*, New Haven, CT: Yale University Press.
- Kindleberger, C.P. (1956), *The Terms of Trade: A European Case Study*, New York: Technology Press of MIT and J. Wiley & Son.
- Mynt, M. (1954), 'The Gains from International Trade and the Backward Countries', *Review of Economic Studies*, 22 (2): 234–51.
- Palma, J.G. (1978), 'Dependency, a Formal Theory of Underdevelopment or a Methodology for the Analysis of Concrete Situations of Underdevelopment?', *World Development*, 6 (7/8).
- Palma, J.G. (1981), 'Dependency and Development: A Critical Overview', in D. Seers (ed.), *Dependency Theory: A Critical Reassessment*, London: Frances Pinter.
- Pinto, A. (1965), 'La concentración del progreso técnico y de sus frutos en el desarrollo', *El Trimestre Económico*, 25: 3–69.
- Pinto, A. (1973), *Inflación: raíces estructurales*, Mexico: Fondo de Cultura Económico.
- Pinto, A. (1974), 'Heterogeneidad estructural y el modelo de desarrollo reciente', in J. Serra (ed.), *Desarrollo Latinoamericano, ensayos Críticos*, Mexico: Fondo de Cultura Económico.
- Pinto, A. and Knakel, J. (1973), 'The Centre-Periphery System Twenty Years Later', *Social and Economic Studies*, March: 34–89.
- Prebisch, R. (1980), 'Prólogo', in Rodríguez (1980).
- Rodríguez, O. (1980), *La teoría del subdesarrollo de la CEPAL*, Mexico: Siglo XXI Editores.
- Rodríguez, O. (2006), *El estructuralismo latinoamericano*, Mexico: Siglo XXI Editores.
- Taylor, L. (1983), *Structuralist Macroeconomics*, New York: Basic Books.
- UNESCO (1949), *Postwar Price Relations Between Under-developed and Industrialized Countries*, New York: UN.
- Viner, J. (1953), *International Trade and Economic Development*, Oxford: Clarendon Press.

11 Marxism and development

*Bob Sutcliffe*¹

Marx's first thoughts

Just as Darwin discovered the law of development of organic nature, so Marx discovered the law of development of human history . . . [that] the degree of economic development attained by a given people or during a given epoch form[s] the foundation upon which the state institutions, the legal conceptions, art, and even the ideas on religion, of the people concerned have been evolved, and in the light of which they must, therefore, be explained, instead of vice versa, as had hitherto been the case. (Engels, (1975 [1883]), p. 467)

Engels's eulogy, delivered at Marx's burial in 1883, is an assertion of Marx's pre-eminent role as a theorist of development in general and of the fundamental importance of economic development for Marxism. This chapter briefly outlines Marx's own ideas on the process and the ways in which later Marxists have built on and adapted these ideas.

Marx viewed human history as a giant spiral tracing the development of the productivity of labour (the forces of production) in relation to the changing social structure within which production took place (the social relations of production). The forces of production tend to grow through history,² although at varying speeds depending on whether the social relations create a favourable or unfavourable climate for material progress. At key moments the forces of production find themselves held back by the form of society and this creates pressure for revolutionary transition from one social system to another, for instance from feudalism to capitalism, which was to play a pivotal role in the development of human history.

Being a system driven by the pursuit of profit in competitive conditions, capitalism would impel a sharp acceleration in the development of the productive forces to such an extent that the universal elimination of want and of involuntary labour could become possible. But capitalism was also a uniquely unequal system, polarizing people into a minority of property owners and a majority of propertyless proletarians. Under capitalism the elimination of want was potential, only realizable after a transition to a fully socialist society. In that way Marx envisioned human society advancing along the axis of scientific and material progress while at the same time following a circular movement from primitive communism, through

various forms of class society and ultimately to a new communism and equality which would be combined with an advanced state of development of the forces of production.³

Marx regarded capitalism as a system which is abhorrent because it rests on exploitation and generates inequality but historically progressive because it brings about an unprecedented development of the productive forces and creates its own 'gravediggers', the propertyless working class.

From his early writings until the publication of the first volume of *Capital* in 1867, Marx had three great expectations. The first ('repetition') was that the rapid capitalist industrialization which he observed in Britain would soon be repeated in other parts of the world. 'The country that is more developed industrially', he wrote, 'only shows, to the less developed, the image of its own future' (Marx, 1975 [1867]).

The second expectation ('universalization') was that the spread of capitalist growth would lead not to independent capitalist countries but to a single, unified interdependent system. In the *Communist Manifesto* Marx and Engels expounded a famous vision of the way capitalism would pervade the globe:

The bourgeoisie has, through its exploitation of the world market, given a cosmopolitan character to production and consumption in every country . . . All old-established national industries . . . are dislodged by new industries . . . that no longer work up indigenous raw material, but raw material drawn from the remotest zones; industries whose products are consumed, not only at home, but in every quarter of the globe . . . In place of the old local and national seclusion and self-sufficiency, we have intercourse in every direction, universal interdependence of nations. (Marx and Engels, 1975 [1848])

The third expectation ('utopia') was that a revolutionary proletariat would 'expropriate the expropriators' and install a society of freedom, both freedom from want and freedom for humans to realize their capacities. In this utopia⁴ the existing division of labour would end, multifaceted work would 'become not only a means of life but life's prime want' and 'society [could] inscribe on its banners: from each according to his ability, to each according to his needs!' (Marx, 1975 [1875]).

Second thoughts?

Marx's favourite motto was '*de omnibus dubitandum*' and his later writings often hint at some second thoughts about all three of his development expectations. This was not only because events were moving slower than he had foreseen; his theoretical work, too, began to suggest possible contradictions with his earlier predictions. The urgent and universalist tone which suffused earlier writings gave way to more complex treatments of the forces leading to monopoly and capitalist concentration, and to economic crisis

which might slow or halt capitalist growth before it had created the productive basis for communism.

The main pressure to rethink his expectations came from problems in applying Marxist ideas to contemporary politics. Among those were his attitudes to British imperialism in India, the question of national liberation in general and prospects of a transition to socialism in Russia. Marx had initially believed that:

England . . . in causing a social revolution in Hindoostan, was actuated only by the vilest interests, and was stupid in her manner of enforcing them. But that is not the question. The question is, can mankind fulfil its destiny without a fundamental revolution in the social state of Asia? If not, whatever may have been the crimes of England she was the unconscious tool of history in bringing about that revolution. (Marx, 1969 [1853a])

He confidently predicted that '[t]he millocracy [industrial capitalists] have discovered that the transformation of India into a reproductive country has become of vital importance to them and that, to that end . . . [t]hey intend now drawing a net of railroads over India. And they will do it' (Marx, 1969 [1853b]).

In later years Marx came to give more weight to the crimes and less to the hope of economic transformation, becoming more supportive of the anti-colonial struggle. By 1881, two years before his death, both the tone and the content had shifted:

What the British take . . . from them (the Indians) without any equivalent . . . amounts to more than the total sum of the income of the 60 million of agricultural and industrial labourers of India. This is a bleeding process with a vengeance. (Marx, 1975 [1881])

There was a parallel evolution in Marx and Engels's attitude towards other nationalist movements which they had once opposed. They supported Irish self-rule because the failure to settle the Irish question was threatening working-class unity in Britain, the country where they had high hopes for the development of socialism: 'the national emancipation of Ireland is no question of abstract justice or humanitarian sentiment but the first condition of [English workers'] own social emancipation' (Marx, 1975 [1870]). And their support for Polish national liberation was premised on the belief that it would weaken Tsarist Russia, the regime they regarded as the main bastion of reaction in Europe.

Nationalism, then, was supported in order to neutralize a cause of fissure in the proletarian movement or to weaken a particular section of the international ruling class, but not because of any general belief in the necessity of national capitalist development strategies. Marx was a fierce critic of the

writings of Friedrich List (1856), advocate of a nationalist and protectionist development strategy for Germany and the United States (see Cowen and Shenton, 1996, pp. 154–69), and never abandoned the idea that development should be universal.

In 1881 the Russian revolutionary Vera Zasulik sought Marx's guidance on the debate between Russian Marxists advocating capitalist development and the Narodniks who believed that capitalism could not develop Russia and who therefore argued for a transition to socialism based on existing peasant communes. The question clearly perplexed Marx and his reaction was not to reassert his earlier opinions; after serious study of the question he penned no less than five drafts of his reply to Zasulik without reaching a definitive position (Shanin, 1983).

Some have seen these intimations of diminished expectations as fitting into a coherent whole with alongside earlier apparently more optimistic ideas (for instance, Melotti, 1977); others have seen Marx edging towards radically different positions (in different ways, Shanin, 1983; Booth, 1985; Lim, 1992). Marx was certainly prepared to re-examine the three original expectations in the light of historical events and to espouse more flexible political tactics. While his thinking evidently evolved, there is no convincing evidence that he fundamentally changed the idea of the ambiguous progressiveness of capitalism, the opposition to national paths to development or the nature of the socialist objective of development.⁵

Nonetheless, in examining the situation in Russia, Marx had been obliged to face the possibility that capitalism might not accomplish the development of the whole world. The implication of that possibility was that perhaps something other than capitalism would have to shoulder the task of developing the productive forces – a question later followers would have to confront.

Marx's followers: development and imperialism

Fifteen years after Marx's death Lenin still argued against the Narodniks that capitalism in Russia, although brutal and truncated, was historically progressive, implying that the revolutionary impulse would come from the working class (Lenin, 1977 [1899]). Trotsky's theory of combined and uneven development was a complementary way of seeing Russian peculiarities in the context of Marx's expectations. History, he argued, did not proceed as an exact series of simultaneous transformations or even repetitions in backward countries (Trotsky, 1969 [1906], 1977 [1930]). The latter could advance unevenly in leaps; separate steps in the journey of development in the more advanced countries might be combined together in more backward ones resulting in 'an amalgam of archaic with more contemporary forms' (Trotsky, 1977 [1930], p. 27). Trotsky used this idea to explain

both why technologically backward Russia could be politically advanced, and also why the revolution was necessarily international. An economically backward revolutionary nation could take advantage of the forces of production in the more advanced nations.⁶

The central question confronted by Marxists in the generation which followed Marx was imperialism (for a survey, see Brewer, 1990). By the first years of the twentieth century nationalist and protectionist forms of development, exactly the kind of repetition which List had supported and Marx opposed, had produced a small group of leading countries contending for world hegemony, and ruling over rival empires. This was what Lenin called *Imperialism, the Highest Stage of Capitalism* (Lenin, [1916]), the title of a book promoting the idea that World War I was an inter-capitalist struggle in which the working classes should oppose their own bourgeoisies, turning the inter-imperialist war into a series of revolutionary civil wars. *Imperialism* reached the conclusion that in an overall sense this ‘monopoly stage’ of capitalism could no longer be considered progressive – not because economic development in all countries would cease, but because competition and war between the leading imperialist powers would destroy more than capitalism could create. Permanent inter-capitalist fratricide fatally wounded Marx’s vision of universalization under capitalist relations. This analysis would be a major part of the theoretical background to the political strategy which led to the 1917 Bolshevik Revolution; in ditching the universalization expectation, Lenin transformed the nature of the utopian one.

Lenin’s book and that of his fellow Bolshevik Nicolai Bukharin (1973 [1915]) were influenced by the Social Democrat Rudolf Hilferding whose remarkable *Finance Capital* was published in 1910 (Hilferding, 1981 [1910]).

Building on Marx’s later writings, Hilferding furnished a detailed analysis of the new monopoly stage of capitalism. Finance capital was the bloc formed in all leading countries between industrial, commercial and banking capital – a ‘holy trinity’, to which the state became the slavish servant. Hilferding argued that the epoch of finance capital meant that Marx’s repetition expectation had only been realized in a limited number of countries and that to some extent it had been replaced by new obstacles to the development of weaker countries. He sounded a whole series of pre-echoes of views which later became commonplace:

As long as the export of capital served primarily for the construction of a transport system and the development of consumer goods industries in a backward country, it contributed to the economic development, in a capitalist form, of that country. Even so . . . [t]he bulk of the profit flowed abroad . . . [which] slows down enormously the pace of accumulation, and hence the further development of capitalism, in the debtor country. In large economic territories . . . a

national assimilation of foreign capital soon occurred . . . In the small economic territories, however, this assimilation was more difficult to achieve, because an indigenous capitalist class emerged much more slowly and with greater difficulty.

Such emancipation became quite impossible when the character of capital exports changed, and the capitalist class in the large economic territories became less concerned with establishing consumer goods industries in foreign countries than with acquiring control over raw materials for their ever growing producers' goods industries . . . [The] capitalist development [of the weaker European countries], and along with it their political and financial development, was stunted at the outset. As economic tributaries of foreign capital, they also became second-class states, dependent on the protection of the great powers. (Hilferding, 1981 [1910], pp. 329–30)

Rosa Luxemburg, another theorist of imperialism of this epoch, also saw the export of capital as prejudicial to peripheral countries (such as Egypt and South Africa), especially to their poorer classes who were usually required to repay the debts incurred and wasted by their rulers (Luxemburg, 1951 [1913]). But her theory of imperialism was only remotely connected with those of Hilferding, Bukharin and Lenin. Arguing that capitalism suffered from a permanent shortage of demand (underconsumptionism), Luxemburg concluded that it was forced to avert collapse by absorbing non-capitalist areas and activities. Imperialism had nothing to do with monopoly or with nations; it was a systemic imperialism of capitalism as a mode of production, rapaciously seeking its surplus value from other modes of production. But this process – really a version of Marx's primary or primitive accumulation – could not continue indefinitely since once the non-capitalist world was completely absorbed then the system would collapse.

While she did not share Lenin's view that capitalism had changed from a progressive to a retrogressive system, Luxemburg did for different reasons share his opinion that human society was approaching a precipice in which all the historical development of the productive forces would be threatened and the choice was between 'socialism or barbarism'. In this they both differed from many conservative socialists who continued to believe that capitalism, left largely to itself, would develop the productive forces and the working class until socialism became both feasible and inevitable.

Somewhere between the two currents stood Karl Kautsky, who argued, to Lenin's fury, that the epoch of conflict of the great powers would give way to a period of cooperation between them (Kautsky, 1970 [1914]). This 'ultraimperialism' would in many ways be worse than imperialism, especially for the less-developed areas of the world which would be collectively exploited by the ultraimperialist alliance. From a different viewpoint to that of Hilferding, Kautsky, too, pre-echoes the way many Marxists and Left radicals were to look at the world half a century later.

Suddenly, at the height of these debates about imperialism, and in conditions where Marx's expectations about the development of capitalism had not been fulfilled, Marxists found themselves with the responsibility of managing an economy in desperate need of development.

A non-capitalist road?

The new Bolshevik rulers of Russia took power still believing that the transition to socialism required a high prior development of the forces of production and must be conducted at a global level. Once the hope of other European revolutions was betrayed, the new communist state had to search for a means of survival and, if possible, progress. A short period of 'war communism', characterized by almost total state control and the breakdown of regular exchange, gave way in 1921 to the less ambitious and stabilizing New Economic Policy (NEP) under which a large measure of market autonomy was restored.

Between the introduction of the NEP and Stalin's seizure of complete power in 1928, there was a brief window in which questions of development strategy were seriously debated among Marxists. The leading protagonists were Bukharin, who increasingly leaned towards the position that the development of a capitalist agriculture was a necessary precondition for eventual industrialization and who therefore saw the more market-friendly NEP as a long-term necessity, and Preobrazhensky, more sympathetic to the Left opposition, who argued for a more rapid pace of industrialization, financed by squeezing a surplus out of agriculture. In a debate which has not lost its relevance,⁷ both of them were searching for a way to achieve what Marx had expected of capitalism – the creation of the material conditions for socialism; they differed about whether this would occur by imitating capitalist development or by following a novel non-capitalist route. (A debate between Marxists about similar issues took place during the early years of the Cuban Revolution.)⁸

Also during the 1920s G.A. Fel'dman designed two sector models, based on Volume 2 of Marx's *Capital*, as a method of planning a socialist economy (Ellman, 1987a). His ideas were partially incorporated into Soviet planning methods and later aroused interest outside the USSR, being influential on the early Indian planners, especially P.C. Mahalanobis, and other Marxist writers on development (Ehrlich, 1978; Chakravarty, 1987; Sen, 1987). Fel'dman was politically purged, and Bukharin and Preobrazhensky were killed when Stalin imposed 'Socialism in one country', the definitive abandonment of a universalist perspective on development.

Soviet industrialization survived the trauma of forced agricultural collectivization, the world economic crisis of the 1930s and three years of Nazi invasion. A Soviet economic model established itself, consisting of highly

centralized planning, virtual autarchy, high rates of investment, concentration on producer goods and heavy industry in order to build a strong industrial productive base and maximize output and consumption in the long run (Bardhan, 1986). The country emerged from World War II with an enhanced industrial and technological capacity. Soviet planning acquired a positive reputation just at a time when colonialism was collapsing and the development of poor countries was on the international agenda.⁹ Both India and China in different ways adopted aspects of the Soviet model, although it failed to transplant successfully. Nonetheless, the apparent existence of a road to industrialization which was not capitalist, was to have considerable impact on the evolution of Marxist ideas about development under capitalism.

Marxism and the Third World: polarization or convergence?

In the decades following World War II, against the chorus of optimistic modernizing developmentalism emanating from official sources in the West, a growing number of Marxists began to argue that capitalism was no longer capable of producing economic development in the poorer parts of the world. Instead it would create growing polarization between the developed and underdeveloped countries.

Foretastes of this idea of imperialism for a world after decolonization had been present in Marxist writings, including even those of Marx himself, for nearly a century. Lenin insisted that, though still progressive, capitalism in Russia was nonetheless incomplete. Hilferding came close to producing a theory of polarization. In the documents of the Third International this idea also appears at the end of the 1920s (Palma, 1978)¹⁰ and even earlier it had a strong presence among Chinese communists. But after the 1950s it was more emphatically asserted by influential Marxist and radical thinkers. It became enormously influential among mass movements and radical intellectuals throughout the world before strong attacks were directed against it by other Marxists. Its legacy is still very much alive in widespread anti-globalization sentiment.

Elements of theories of inevitable polarization were already circulating among Latin American intellectuals when Paul Baran in the 1950s presented an explicitly Marxist version of it, concluding that 'the capitalist system, once a mighty engine of economic development, has turned into a no less formidable hurdle to economic advancement' (Baran, 1973 [1957], p. 402; also see Baran, 1952). The cause was the onset of monopoly capital, a new stage of the system, characterized a general tendency in the major centres of capitalism to underconsumption and crisis, held at bay only by state spending, militarism and the exploitation of ethnic minorities and economically backward countries.¹¹

Other theorists of polarization, by contrast, saw it as a process which had lasted through the four centuries of existence of a worldwide market, through which a privileged group of countries in the centre could transfer resources from the dominated countries of the periphery through plunder, unequal trade and later investment and indebtedness. Particularly influential were the writings of Andre Gunder Frank which began as an attack on modernization theories exemplified by W.W. Rostow and on the anti-revolutionary perspectives of Latin American communist parties. Frank transformed the meaning of the word 'underdevelopment' from a pre-developmental state into a consequence of worldwide capitalist development. His purpose was to anatomize what he called, in a memorable phrase, 'the development of underdevelopment' during centuries of capitalist history (Frank, 1966, 1991). His name became associated with dependency theory, whose influence penetrated several disciplines – economics, sociology and international relations in particular (see Kay, 1989; Larrain, 1989). An overlapping set of ideas was the world-system theory of Immanuel Wallerstein, influenced by the long-term historical outlook of Fernand Braudel (Wallerstein, 1979, 1983). Samir Amin derived polarization from an analysis of world-scale capital accumulation (Amin, 1974). Proponents of these theories differed considerably over the extent to which development was held back by involvement in the capitalist economy. To some it meant simply impoverishment, to others a more complex and variable form of dependent development (see Evans, 1979; Cardoso and Faletto, 1979). Most of them believed that development of the poorer countries would not be possible without some clear limit to involvement in the unequalizing capitalist world market, an idea encapsulated in the title of Samir Amir's book, *Delinking* (Amin, 1990). Many advocated protectionism, citing Friedrich List and Alexander Hamilton as positive historic precedents. Others, including Baran, saw the way out as repeating Soviet-style industrialization policies.

Not all the advocates of dependency and world-systems theory saw themselves as Marxists in the way Baran had done;¹² but most were strongly influenced by Marxism and have often been labelled 'neo-Marxists' (by Hirschman, 1981 and Brewer, 1990, among others). Like Marx, they have analysed the world in a long historical perspective, put capitalism in the centre of their analysis, found some of the causes of the process of underdevelopment in Marx's own analysis (for instance, the plunder of the wealth of poorer regions which was one element of Marx's primary accumulation of capital), assigned some role to classes (especially the weakness of the dependent bourgeoisie), and expounded a theory of polarization between nations and continents which was arguably a transfigured version of Marx's idea that capitalism simultaneously created wealth and poverty.

But much polarization theory stressed the divergence between countries rather than classes.

While Marx saw capitalism as being progressive in spite of its barbarities, most polarization theorists have not. Lenin for one reason, and later Baran for others, saw the epoch of capitalism which they wrote about as having ceased to be progressive. But many dependency and world-systems theorists regarded capitalism as never having been progressive. Dependency theorists have been criticized by other Marxists for regarding capitalism as an unchanging system throughout its history. Such critics contended that dependency theory failed to recognize that it is not the market and exchange which are the essence of capitalism, but productive capital producing surplus value by exploiting free labour. This leads to the erroneous location of the beginning of capitalism's great polarization of the world in the sixteenth century with the emergence of worldwide markets. Hence they ascribe the process of underdevelopment more to plunder and unequal exchange rather than to more essential features of the capitalism mode of production, and also as a result exaggerate the role of nation and underestimate the role of class in the generation of and the fight against world inequalities.¹³

Most polarization theories, Marxist or not, assumed that the world was very different from the one which Marx had foreseen. Some critics have taken issue with this assumption. 'Post-imperialist' historians have argued that Marx's universalization expectation, the fusion of capitalist countries into a single global system, is already a reality (Sklar, 1976; Becker et al., 1987); their focus is on the emergence of a single global capitalist class. In a more recent, widely discussed global hypothesis, Hardt and Negri assert that it is a world non-ruling class, the 'multitude', which is the most coherent offspring of globalization and the decline of states' authorities. Their decidedly global concept of development is implicit from their main political demands: for the totally free movement of human beings across borders and for a global guaranteed basic wage and access to welfare provisions (Hardt and Negri, 2000).

The 'return to Marx' proposal which has been the most influential, in part because it was a frontal attack on the polarization theorists, made with the same ringing defiance as they had attacked modernization and the Latin American communist parties, was that of Bill Warren, in his book provocatively titled *Imperialism, Pioneer of Capitalism* (Warren, 1980; also see Warren, 1973). He argued that prospects for capitalist development were in fact good, that much of it had taken place since World War II, that colonialism had indeed broken obstacles to progressive social change as Marx had originally predicted, that the obstacles to capitalist development are not those involving relations with developed countries but those to be

found 'in the internal contradictions of the Third World itself', that the policies of the developed countries in general foster rather than stifle industrialization in the underdeveloped ones, and that 'the ties of "*dependence*" (or subordination) binding the Third World and the imperialist world have been and are being markedly loosened with the rise of indigenous capitalisms'.¹⁴ In other words, he was arguing that Marx's first thoughts remained valid and that Marxist thinking about development from Lenin onwards was a saga of errors.

Unlike some other critiques, Warren's attack on dependency was in considerable part an empirical one. He stressed that the economic and social performance of the Third World was not nearly as bad as polarization theorists made out. Although a number of seemingly impartial commentaries have accepted these conclusions (for example Booth, 1985 and Brewer, 1990), it is worth mentioning that from 1950 there was a clear divergence between developed and underdeveloped countries in aggregate until as recently as the 1990s. The average gross domestic product (GDP) per head of Africa, Latin America and Asia (excluding Japan) taken together fell as a percentage of the North (the USA, Canada, the EU and Japan) in every year between 1950 and 1990. If China is excluded, it continues to fall up to at least 2001 and possibly beyond (as calculated from Maddison, 2003).

Nonetheless, if the empirical evidence which Warren relied upon in the 1970s seemed less than convincing, by the final years of the twentieth century the rapid development of a number of Asian countries seemed to give solid support to his position, although others pointed out that none of the Asian success stories were based on free market capitalism but that all of them had depended on vigorous state intervention and protectionism. Nonetheless three decades of breakneck development in China and other parts of Asia is enough to refute the idea of continuous polarization between developed and underdeveloped countries as a global generalization; equally, the continued economic decline of Africa and parts of Latin America refutes the opposite hypothesis (Leys and Saul, 1999).

The years since 1980 have been years of extremely sharp divergence not so much between developed and underdeveloped countries, but between different groups of underdeveloped countries. While the GDP per head of China (measured at purchasing power parity) has risen by 667 per cent during the years 1980 to 2004, that of Latin America has risen by 12 per cent and that of Africa has fallen by 6 per cent (World Bank, 2005). Such a difference, over such a time, surely indicates a more complex global reality than either polarization or convergence theories assume. The dichotomy, which has ended in what has been variously described as an impasse (Booth, 1985) or mutual check-mate (Munck, 1999), needs to be transcended.

Not only are there contradictory development tendencies in what was called the Third World but also the extremes are extraordinarily far apart. At one extreme is Southern Africa where not only is poverty growing, but also a high proportion of the population is infected with a fatal disease which is changing the nature of society and which has reduced life expectancy by decades. At the other extreme is China, the location of the most important surge of capitalist industrialization which has happened in history, presided over, ironically, by those who, without apparent embarrassment, style themselves as Marxists. The overall size of China's GDP rose from 13 per cent of that of the USA in 1978, to 62 per cent in 2004 and at this rate will overtake it in a very few years (World Bank, 2005). This momentous shift in the centre of gravity of world capitalist accumulation creates echoes of the earlier Marxist propositions and debates about development. The advance of China suggests that the centre of capitalist accumulation has geographically shifted from the long-developed countries. Will China (along with other Asian countries) reach the economic level of and challenge the hegemony of the USA? Will it become an imperialist power? Will its thirst for raw materials force it to develop parts of Africa? Or will new forms of polarization occur? And what will be the role in this story of the Chinese working class? These are the questions which Marx asked about nineteenth-century capitalism. Marxists must try to give new answers to them today.

Utopia, production and redistribution

Since the 1980s the influence of Marxism in development has declined. The neoliberal revival and the collapse of actually existing socialism have shifted the global political balance in favour of capitalism's friends. But also, the long debate on imperialism had not prepared Marxism well to make major creative contributions to a number of neglected questions which have come to the fore. Major debates were, therefore, spearheaded by people of other heterodox opinions and currents, often directing their fire not only against conventional development thinking but against Marxism as well.¹⁵

First, feminists challenged Marxists by insisting that women's emancipation is a task which cannot be reduced to class and development in general. It is a central part of the struggle for and the realization of socialist utopia (for a survey of arguments see Parport et al., 2000).

Second, majority opinion in environmental science is that probably the universalization of development in its most widely used meaning is physically not attainable. A number of writers, however, have begun to search for Marxist answers to this and other environmental quandaries (see Martinez Alier, 1991; O'Connor, 1998; Foster, 2000; Löwy, 2002), but it remains a minority pursuit.

A third issue, which partly embraces the previous two, is the nature of the objective of development. Polarization and convergence theories shared an implicit conception that development meant roughly what had been attained in developed countries. Convergence theorists forecast that most countries would reach the destination; polarizationists complained that they will not. Neither side incorporated a thorough critique of the economic and social nature of the destination itself. Booth criticized both for their 'system teleology'; but perhaps the problem is more a shared failure to question the nature of the telos. The discussion of 'human development', launched by the UNDP in 1990, based on A.K. Sen's notion of 'development as freedom', was one influential but limited attempt to do this. More fundamentalist heterodox critics have scorned all conventional (including Marxist) images of the destination of development as dystopias. From post-development or even anti-development perspectives they have rejected development as an aim and have tried to outline a more modest model which often stresses small-scale communities, the maintenance of traditional cultures, a balance with nature, and so on.

So feminists, environmentalists, postmodernists and other radical critics of social and economic orthodoxy have, sometimes with validity, criticized Marxist conceptions of development as no less male-centric, Eurocentric or unsustainable than orthodoxy itself. They have forced some self-critical rethinking about the limitations of Marxist approaches to development. Yet in a sense what all these currents of thought do is to re-pose a problem central to Marx's original thinking about development: the definition of utopia.

There are serious dangers involved in concluding from the valid parts of these criticisms that the whole concept of development, in its orthodox or Marxist version, should be thrown out like old bathwater. The baby which must be saved is Marx's fundamental insight, picked out by Engels in his eulogy, that utopia must rest on an appropriate global material, economic and productive foundation. There are some elements of the often reviled, economistic modernization project which, purged of their unequal, unsustainable and imperialist form, must form a part of the journey to social emancipation. Nonetheless, human productivity is now so advanced that the forces of production are more than enough to produce all reasonable human needs if the composition and distribution of their product was different. Yet since distribution is so unequal, these forces are in fact used on a huge scale to produce unreasonable and destructive 'needs' (what some have referred to as 'overdevelopment'). If the question of development is posed, in the way Marx posed it, as how to translate capitalist productivity into socialist utopia, then the main focus of development on a world scale must now be not so much on growth, but increasingly on distribution.

Notes

1. My thanks to Andrew Glyn and Arthur MacEwan for comments on a draft of this chapter.
2. This is what G.A. Cohen calls Marx's 'development thesis': see Cohen (1978), especially Chapters 6 and 7.
3. For discussions of the various senses in which Marx used the term 'development' see Cohen (1978) and Cowen and Shenton (1996), pp. 117–19.
4. Any reader who doubts whether Marxism can be described as utopian (in the positive sense) should read Geras (2000).
5. An exposition and discussion of the issues dealt with in this and the previous section can be found in Patnaik (2005).
6. For analyses of combined and uneven development, see Elster (1986) and Löwy (1981).
7. See Ehrlich (1950, 1960), Cohen (1973), Day (1975, pp. 196–219), Bukharin (1979), Filtzer (1980), Haynes (1985), Ellman (1987b).
8. The leading participants in these debates were Che Guevara (then Minister of Industry), Carlos Rafael Rodriguez from the wing of the old Cuban Communist Party which had switched its support to Castro, and a number of foreign Marxist theorists including Ernest Mandel and Charles Bettelheim. They once again raised familiar issues: the role of the state and the market, the balance between agriculture and industry, and (perhaps the most originally discussed question) the balance between moral and material incentives. These debates remain interesting although they quickly slipped into history in Cuba as it became increasingly dependent on Soviet aid and followed Moscow's line in international policy. See Brundenius (1984) and Martinez Alier and Martinez Alier (1972, Chapter 6).
9. On the Soviet economy and planning methods see Nove (1992), Davies et al. (1993), Davies (1998), Gregory (2001) and Allen (2003).
10. The idea of polarization was encouraged in the 1928 Congress of the International partly as a cynical manoeuvre to isolate Bukharin and his supporters who regarded further capitalist development as possible and desirable. It disappeared from official communism after the USSR's alliance with the liberal democracies in 1941.
11. This thesis appeared in rudimentary form in Baran's (1973 [1957]) *The Political Economy of Growth* and was later much elaborated in Paul Baran and Paul Sweezy (1966), one of only a handful of Marxist analyses of the whole capitalist system written in the last 50 years.
12. Among the most explicitly Marxist of development theorists were Marini (1991) and Dos Santos (1970). G. Kay (1975), although a critic of dependency theory, developed an explicitly Marxist theory which relates the historical pattern of capital accumulation to the restriction of development, especially in Africa.
13. For two examples of this kind of critique see Laclau (1971) and Brenner (1977). These critiques raised the question of how the capitalist mode of production was to be defined; it was related to a much larger development-related discussion of the relevance of modes of production and their interrelations. For a summary of this and references to other authors see Foster-Carter (1978) and Brewer (1990, pp. 226–36).
14. Warren (1980, pp. 9–10); this summary follows that in Booth (1985, p. 766).
15. On the subject of the decline of Marxist development theory see Bernstein (2005) which analyses the rise and decline of Marxism in the world of academic and para-academic development studies, and Leys (1996) who sees the decline of Marxist development theory as part of the disappearance of development theory in general. He concludes that there is an 'urgent need to revive development theory, not as a branch of policy-oriented social science within the parameters of an unquestioned capitalist world order, but as a field of critical enquiry about the contemporary dynamics of that order itself, with imperative policy implications for the survival of civilized and decent life, and not just in the ex-colonial countries'. The question of new directions in Marxist and critical development thinking is also discussed in Munck and O'Hearn (1999) and Schuurman (1993).

References

Note: the references are all to printed versions. Dates in square brackets refer to the original date of publication. Some of the reprinted sources do not have dates. The majority of the references by Marx, Engels and Marxist authors before 1950, and a few afterwards, are available online at the indispensable Marxists Internet Archive (www.marxists.org).

- Allen, R.C. (2003), *From Farm to Factory*, Princeton, NJ: Princeton University Press.
- Amin, S. (1974), *Accumulation on a World Scale*, New York: Monthly Review Press.
- Amin, S. (1990), *Delinking*, London: Zed Books.
- Baran, P.A. (1952), 'On the Political Economy of Backwardness', *Manchester School of Economics and Social Studies*, 20, January, pp. 66–84.
- Baran, P.A. (1973 [1957]), *The Political Economy of Growth*, Harmondsworth: Penguin Books, p. 402; originally published by Monthly Review Press.
- Baran, P.A. and P.M. Sweezy (1966), *Monopoly Capital: An Essay on the American Economic and Social Order*, New York: Monthly Review Press.
- Bardhan, P. (1986), 'Marxist Ideas in Development Economics: An Evaluation', in John Roemer (ed.), *Analytical Marxism*, Cambridge: Cambridge University Press, pp. 64–77 and Paris: Editions de la Maison des Sciences de l'Homme, pp. 74–5.
- Becker, D.G., J. Freiden, S.P. Schatz and R. Sklar (eds) (1987), *Postimperialism: International Capitalism and Development in the Late Twentieth Century*, Boulder, CO: Lynne Rienner.
- Bernstein, H. (2005), 'Development Studies and the Marxists', in U. Kothari (ed.), *A Radical History of Development Studies: Individuals, Institutions and Ideologies*, London and New York: Zed Books, pp. 111–37.
- Booth, D. (1985), 'Marxism and Development Sociology: Interpreting the Impasse', *World Development*, 13 (7): 761–87.
- Brenner, R. (1977), 'The Origins of Capitalist Development: A Critique of Neo-Smithian Marxism', *New Left Review*, 104: 25–92.
- Brewer, A. (1990), *Marxist Theories of Imperialism: A Critical Survey*, London and New York: Routledge.
- Brundenius, C. (1984), *Revolutionary Cuba: The Challenge of Economic Growth with Equity*, Boulder, CO: Lynn Rienner.
- Bukharin, N. (1973 [1915]), *Imperialism and World Economy*, New York: Monthly Review Press (n.d.).
- Bukharin, N. (1979), 'Notes of an economist', *Economy and Society*, 8 (4): 473–500.
- Cardoso, F.E. and E. Faletto (1979), *Dependency and Development in Latin America*, Berkeley, CA and London: University of California Press.
- Chakravarty, S. (1987), 'Mahalanobis, Prasanta Chandra', in John Eatwell, Murray Millgate and Peter Newman (eds), *The New Palgrave Dictionary of Economics*, London and Basingstoke: Macmillan, Vol. 3, pp. 276–7.
- Cohen, S.F. (1973), *Bukharin and the Bolshevik Revolution: A Political Biography 1888–1938*, New York: Knopf.
- Cohen, G.A. (1978), *Karl Marx's Theory of History: A Defence*, Oxford: Clarendon Press.
- Cowen, M.P. and R.W. Shenton (1996), *Doctrines of Development*, London and New York: Routledge.
- Davies, R.W. (1998), *Soviet Economic Development from Lenin to Khrushchev*, Cambridge: Cambridge University Press.
- Davies, R.W., Mark Harrison and S.G. Wheatcroft (eds) (1993), *The Economic Transformation of the Soviet Union 1913–1945*, Cambridge: Cambridge University Press.
- Day, R. (1975), 'Preobrazhensky and the Theory of the Transition Period', *Soviet Studies*, 27 (2): 196–219.
- Dos Santos, T. (1970), 'The Structure of Dependence', *American Economic Review*, 60, May: 231–6.
- Ehrlich, A. (1950), 'Preobrazhenski and the Economics of Soviet Industrialization', *Quarterly Journal of Economics*, 64 (1): 57–88.
- Ehrlich, A. (1960), *The Soviet Industrialization Debate 1924–1928*, Cambridge, MA.

- Ehrlich, A. (1978), 'Dobb and the Marx–Fel'dman Model: A Problem in Soviet Economic Strategy', *Cambridge Journal of Economics*, 2 (2): 203–14.
- Ellman, M. (1987a), 'Fel'dman, Gigorii Alexandrovich', in John Eatwell, Murray Millgate and Peter Newman (eds), *The New Palgrave Dictionary of Economics*, London and Basingstoke: Macmillan, Vol. 2, pp. 299–300.
- Ellman, M. (1987b), 'Preobrazhensky, Evgeni Alexeyevich', in John Eatwell, Murray Millgate and Peter Newman (eds), *The New Palgrave Dictionary of Economics*, London and Basingstoke: Macmillan, Vol. 3, pp. 945–6.
- Elster, J. (1986), 'The Theory of Combined and Uneven Development: A Critique', in John Roemer (ed.), *Analytical Marxism*, Paris and Cambridge: Maison de Sciences de l'homme, pp. 54–63.
- Engels, F. (1975[1883]), 'Draft of a speech at Marx's graveside', in Marx and Engels, *Collected Works*, London: Lawrence and Wishart, Vol 24, p. 467, also available at <http://www.marxists.org/archive/marx/works/1883/death/justice.htm>.
- Evans, P.B. (1979), *Dependent Development: The Alliance of Multinational, State, and Local Capital in Brazil*, Princeton, NJ: Princeton University Press.
- Filtzer, D.A. (ed.) (1980), *The Crisis of Soviet Industrialization*, London: Macmillan.
- Foster, J.B. (2000), *Marx's Ecology: Materialism and Nature*, New York: Monthly Review Press.
- Foster-Carter, A. (1978), 'The Modes of Production Controversy', *New Left Review*, 107: 47–77.
- Frank, A.G. (1966), 'The Development of Underdevelopment', *Monthly Review*, September, pp. 17–31.
- Frank, A.G. (1991), 'The Underdevelopment of Development', *Scandinavian Journal of Development Alternatives*, 10 (3).
- Geras, N. (2000), 'Minimum Utopia: Ten Theses', in Leo Panitch and Colin Leys (eds), *Socialist Register 2000*, London: Merlin, pp. 41–52.
- Gregory, P. (2001), *Behind the Façade of Stalin's Command Economy: Evidence from the Soviet State and Party Archives*, Stanford, CA: Hoover Books online.
- Hardt, M. and A. Negri (2000), *Empire*, Cambridge, MA and London: Harvard University Press.
- Haynes, M.J. (1985), *Nikolai Bukharin and the Transition from Capitalism to Socialism*, London: Croom Helm.
- Hilferding, R. (1981 [1910]), (ed. T. Bottomore) *Finance Capital: A Study of the Latest Phase of Capitalist Development*, London, Boston, MA and Henley: Routledge & Kegan Paul.
- Hirschmann, A.O. (1981), 'The Rise and Decline of Development Economics', in A.O. Hirschmann, *Essays in Trespassing: Economics to Politics and Beyond*, Cambridge: Cambridge University Press, pp. 1–24.
- Kautsky, K. (1970 [1914]), 'Ultraimperialism', *New Left Review*, 59: 41–6.
- Kay, C. (1989), *Latin American Theories of Development and Underdevelopment*, London: Routledge.
- Kay, G. (1975), *Development and Underdevelopment: A Marxist Analysis*, London: Macmillan.
- Laclau, E. (1971), 'Feudalism and Capitalism in Latin America', *New Left Review*, 67: 19–38.
- Larrain, J. (1989), *Theories of Development: Capitalism, Colonialism and Dependency*, Oxford: Polity Press.
- Lenin, V.I. (1977 [1899]), *The Development of Capitalism in Russia: The Process of the Formation of a Home Market for Large-Scale Industry*, Moscow: Progress Publishers.
- Lenin, V.I. ([1916]), *Imperialism, the Highest Stage of Capitalism*, in V.I. Lenin, *Collected Works*, Vol. 22, Moscow: Progress Publishers, pp. 185–304, also available at <http://www.marxists.org/archive/lenin/works/1916/imp-hsc/index.htm>.
- Leys, C. (1996), 'The Rise and Fall of Development Theory', in Colin Leys, *The Rise and Fall of Development Theory*, Nairobi, Bloomington and Indianapolis, Oxford: EAEP, Indiana University Press, James Currey.
- Leys, C. and J. Saul (1999), 'Sub-Saharan Africa in Global Capitalism', *Monthly Review*, July–August: 13–30.

- Lim, J.-J. (1992), 'Marx's Theory of Imperialism and the Irish National Question', *Science and Society*, 56 (2): 168–9.
- List, F. (1856), *The National System of Political Economy*, Philadelphia, PA: J.P. Lipincott and Co.
- Löwy, M. (1981), *The Politics of Combined and Uneven Development: The Theory of Permanent Revolution*, London: Verso.
- Löwy, M. (2002), 'Marx to Ecosocialism', *Capitalism Nature Socialism*, 13 (1): 121–33.
- Luxemburg, R. (1951 [1913]), *The Accumulation of Capital*, London: Routledge & Kegan Paul.
- Maddison, A. (2003), *World Historical Statistics*, CD version, Paris: OECD.
- Marini, R.M. (1991), *Dialectica de la Dependencia*, Mexico City: Era.
- Martinez Alier, J. (1991), 'Ecology and the Poor: A Neglected Dimension of Latin American History', *Journal of Latin American Studies*, 23 (3): 621–39.
- Martinez Alier, J. and V. Martinez Alier (1972), *Cuba: economía y sociedad*, Paris: Ruedo Iberico.
- Marx, K. (1969 [1853a]), 'The Future of British Rule in India', in Shlomo Avineri (ed.), *Karl Marx on Colonialism and Modernization*, New York: Anchor Books, p. 94 (originally published in the *New York Daily Tribune*, 25 June).
- Marx, K. (1969 [1853b]), 'The Future of British Rule in India', in Shlomo Avineri (ed.), *Karl Marx on Colonialism and Modernization*, New York: Anchor Books, p. 134 (originally published in the *New York Daily Tribune*, 8 August).
- Marx, K. (1975 [1867]), Preface to *Capital, Vol. I*, (1st German edn.), in Marx and Engels, *Collected Works*, Vol 35, London: Lawrence and Wishart, p. 7, also available at <http://www.marxists.org/archive/marx/works/1867-c1/p1.htm>.
- Marx, K. (1975 [1870]), 'Letter to Sigfrid Meyer and August Vogt', in Marx and Engels, *Collected Works*, Vol 43, London: Lawrence and Wishart, p. 148, also available at <http://www.marxists.org/archive/marx/works/CW/Volume43/index.htm>.
- Marx, K. (1975 [1875]), *Critique of the Gotha Programme*, in Marx and Engels, *Collected Works*, Vol 24, London: Lawrence and Wishart, pp. 75–99, also available at <http://www.marxists.org/archive/marx/works/CW/Volume24/index.htm>.
- Marx, K. (1968 [1881]), Letter to Nikolai Danielson, *Marx–Engels Correspondence*, Moscow: International Publishers.
- Marx, K. (1975 [1881]), 'Letter to Nikolai Danielson', in Marx and Engels, *Collected Works*, Vol 46, London: Lawrence and Wishart, p. 60, also available at <http://www.marxists.org/archive/marx/works/CW/volume46/index.htm>.
- Marx, K. and F. Engels (1975 [1848]), *The Manifesto of the Communist Party* in Marx and Engels, *Collected Works*, Vol 6, London: Lawrence and Wishart, pp. 477–517, also available at <http://www.marxists.org/archive/marx/works/1848/communist-manifesto/index.htm>.
- Melotti, U. (1977), *Karl Marx and the Third World*, London and Basingstoke: Macmillan.
- Munck, R. (1999), 'Deconstructing Development Discourses: Of Impasses, Alternatives and Politics', in R. Munck and D. O'Hearn, *Critical Development Theory: Contributions to a New Paradigm*, London: Zed Books, pp. 196–210.
- Munck, R. and D. O'Hearn (1999), *Critical Development Theory: Contributions to a New Paradigm*, London: Zed Books.
- Nove, A. (1992), *An Economic History of the USSR*, Harmondsworth: Penguin Books.
- O'Connor, J. (1998), *Natural Causes: Essays in Ecological Marxism*, New York: Guilford.
- Palma, G. (1978), 'Dependency: A Formal Theory of Underdevelopment or a Methodology for the Analysis of Concrete Situations of Underdevelopment', *World Development*, 6: 881–924.
- Parport, J.L., M.P. Connelly and V.E. Barriteau (eds) (2000), *Theoretical Perspectives on Gender and Development*, Ottawa: ICRC.
- Patnaik, P. (2005), 'Marx as a Development Economist', in Jomo K.S. (ed.), *The Pioneers of Development Economics: Great Economists of Development*, New Delhi and London: Tulita Books and Zed Books, pp. 1–9.
- Schuurman, F.J. (ed.) (1993), *Beyond the Impasse: New Directions in Development Theory*, London: Zed Books.

- Sen, A. (1987), 'Dobb, Maurice', in John Eatwell, Murray Millgate and Peter Newman (eds), *The New Palgrave Dictionary of Economics*, London and Basingstoke: Macmillan, Vol. 1, pp. 910–12.
- Shanin, T. (presenter) (1983), *Late Marx and the Russian Road; Marx and 'the Peripheries of Capitalism'*, London, Melbourne and Henley: Routledge & Kegan Paul, History Workshop Series.
- Sklar, R. (1976), 'Postimperialism, a Class Analysis of Multinational Corporate Expansion', *Comparative Politics*, October, pp. 75–92.
- Trotsky, L. (1969 [1906]), *Results and Prospects*, in L. Trotsky, *The Permanent Revolution*, New York: Pathfinder Press.
- Trotsky, L. (1977 [1930]), *The History of the Russian Revolution*, London: Pluto Press.
- Wallerstein, I. (1979), *The Capitalist World-Economy*, Cambridge: Cambridge University Press.
- Wallerstein, I. (1983), *Historical Capitalism*, London: Verso.
- Warren, B. (1973), 'Imperialism and Capitalist Industrialization', *New Left Review*, 81, September–October: 3–44.
- Warren, B. (1980), *Imperialism, Pioneer of Capitalism*, ed. J. Sender, London: Verso.
- World Bank (2005), *World Development Indicators 2005*, online version.

12 Institutional development economics

Kenneth P. Jameson¹

Introduction

Institutionalism was integral to development economics in its post-World War II formative years. By 1977 Philip Klein (p. 785) could write: ‘in the field of development economics the victory (of institutionalism) has been so complete that many economists fail to realize it’. Development economics seems to be at a similar juncture today. The *World Development Report 2002* (World Bank, 2001) was entitled ‘*Building Institutions for Markets*,’ and the 1999 IMF ‘Conference on Second Generation Reforms’ focused on institutions and development. Meier (2005, p. 183) summed it up well:

Policy recommendations have evolved in response to the lessons of experience and the evolution in the analytics of development. Development economists in the 1970s advised developing countries to ‘get prices right’. Then in the 1980s and 1990s, they said ‘get macro policies right’. Now they say ‘get institutions right’.

However, such outcomes are never so simple. The earlier ‘victory’ of institutionalism was short-lived, even pyrrhic. By 1981, one of its foremost practitioners, Albert Hirschman (1981, p. 1) was writing of ‘the rise and decline of development economics’. He wrote wistfully that ‘(t)he forties and especially the fifties (that) saw a remarkable outpouring of fundamental ideas and models which were to dominate the new field and to generate controversies that contributed much to its liveliness’. This suggests the need for a closer examination of the relation of development economics and institutionalist economics before declaring that institutionalism has triumphed.

Both institutionalism and development are holistic, attempting to bring all major elements of the economy and society into their analysis. Looked at over time, they are like the two strands of the double helix, functioning separately at some points, while winding about each other and linking or bonding at other points into institutionalist development economics.

In some cases, the link is through the historical situation that draws the two into contact. For example, Adam Smith was writing at a time of significant economic and institutional change, and of necessity holistically linked development and institutionalism. His treatment of the ‘Causes of

Prosperity of the New Colonies' analyzed the prosperity, that is, development, of the English colonies compared particularly to the Spanish colonies. His understanding of development was heavily institutionalist:

The colonists carry out with them a knowledge of agriculture and of other useful arts, superior to what can grow up of its own accord in the course of many centuries among savage and barbarous nations. They carry out with them too the habit of subordination, some notion of the regular government which takes place in their own country, of the system of laws which supports it, and of a regular administration of justice; and they naturally establish something of the same kind in the new settlement. (Smith, 1976 [1776], Volume 2, Book IV, Chapter VII, Part II, p. 76)

A detailed description of the development process in the various colonies followed: the education and health of children, the abundance and inheritance of land, the sources and uses of government revenues, and the type of production undertaken. Writing at that time of change caused Smith to link development and institutionalism as described by Sobel (1979, p. 350): 'the ideas of the great economic system-builders should be interpreted in the context of the problems, knowledge, and intellectual method of the time in which they wrote. Assumptions about reality introduced into their analysis were necessarily drawn from the institutional framework of the time.'

This chapter addresses two issues. First, it traces the double helix of development and institutionalist economics during the twentieth century. Second, it examines the points where links of institutionalism and of development have been created, resulting in today's dynamic institutionalist development economics.

We first trace the institutional analysis formulated before World War II whose constructs became available to later development economists. The most common link across the double helix resulted from the effort to apply development and institutionalist economics to a common set of economic and social phenomena. Then we describe the economic reality that challenged economists after World War II and stimulated institutionalist development economics as described by Hirschman. The cross-fertilization of ideas and concepts across the two frameworks was important in the early post-war years.² After that we examine institutionalism and development when their links and momentum were strongest, as well as the subsequent 'decline' of both. We then trespass into the realm of 'new institutionalism and development' and analyze the new bonds forming in response to the recent experience of development. Following that we examine the various versions of institutionalist development economics today. Hirschman attributed the decline of development economics to the error of Western economists who perceived developing countries 'to have only *interests* and

no passions' (1981, p. 24). The continued viability of institutionalist development economics depends on avoiding that same error in the future.

Let us turn now to the ideas of the institutionalists in the formative period of their greatest vibrancy, focusing on those that would later provide links to development economics.

The original institutionalist strand

The holism of Smith had long been lost by the beginning of the twentieth century. Early institutionalism, led by Thorstein Veblen, was a reaction to the static equilibrium analysis of the neoclassical school. He focused on long-term change and development, with technology as the driving force for change and cultural rigidities the wall holding back the forces of change. Two of institutionalism's fundamental characteristics were rooted in Veblen's analysis. The first was the dynamism of technology embodied in the corporation. Technical changes fostered by the engineering mentality, as opposed to the banking mentality, were constantly altering the economy and preventing it from staying at rest. This movement was consistent with the human instinct toward more effectively satisfying human needs. However, technological innovations confronted the second key category, habits or ceremonial judgments. The changes technology sets in motion cause an opposite reaction when they clash with the habits of individuals and groups in the society who resist change and instinctually react against the challenge to existing social relations. As a result, economics must be evolutionary, providing an understanding of institutions and cumulative change, the type of change that later became the subject of development economics.

Veblen's analysis provided the starting point for three later frameworks that crystallized much of institutionalist thinking (Fusfeld, 2000). The first was John R. Commons and the Wisconsin School who, consistent with the Progressive movement in Wisconsin, looked to government to guide the process of change and to find institutional mechanisms to bridge the fissures that social change invariably caused. This confidence in government involvement in social change provided support for the attitudes toward development that became prevalent at the end of World War II. In addition, the emphasis on institutions, such as the legal system, provided the central defining characteristic of the school.

The second strain originated with Clarence Ayres and the Texas School, who emphasized both technology and the cultural elements that led to resistance to technologically generated change. Ayres concentrated on detailing the factors that could lead to resistance, to the ceremonial features of a society such as its 'mystical' rites and ceremonies and its system of convention. Following World War II, many societies underwent radical economic and social change, which often sparked resistance and reaction

to the change. Once again, the institutionalist analysis was attractive to those working in development economics.

Wesley Clair Mitchell and the Columbia School constituted the final strain. They fostered an empirical methodology and gave primacy to induction over deduction, to empirical institutional investigation over formal model building. They paralleled development economists working out of development experiences and basing their insights and prescriptions on experience, rather than on a given theoretical structure.

The result was that the framework described as early as 1927 as institutionalist economics ‘consisted of aspects of Veblen’s analysis of institutions combined with a much greater emphasis on legal institutions, pragmatic social reformism, and a strong empirical . . . view of proper methodology’ (Rutherford, 2000, p. 292). This was a period of vibrancy and of heightened importance for institutional economics. One effect was that the institutionalist themes that had initially appeared in Veblen and had been elaborated on by Commons, Ayres and Mitchell and their fellow institutionalists provided an alternative to orthodox economics with its liberal, static and individualistic starting points. The economists committed to addressing the problems of development that came to the fore at the end of World War II required such a framework.

The development economics strand

The ‘problems, knowledge, and institutions’ of greatest importance in the formative years of development economics were those of Central and Eastern Europe (Warner and Jameson, 2004). Latin America already had 150 years of independent development; Africa, the Caribbean and important parts of Asia were still colonies. On the other hand, by the 1930s the growing economic disparities between Eastern and Western Europe, first documented by Colin Clark, had stimulated research and thinking about development in Central and Eastern Europe. The end of World War II brought those issues to the forefront; all sides were in agreement that the economic and social institutions of Central and Eastern Europe required transformation. At the same time, the war’s end provided an influx to the West of a group of economists who would contribute many of the ideas and models of the new discipline. They knew and had practiced neoclassical economics. However, its inadequacies for dealing with post-war development led them to take an approach quite consistent with the then important US institutional economics.

For example, Paul Rosenstein-Rodan was born in Poland and began developing his ‘big push’ theory of industrialization in the early 1940s. He believed the state would be central, along with investment that incorporated modern technology. But institutional change would be central to its

success: 'An institutional framework different from the present one is clearly necessary for the successful carrying out of industrialization in international depressed areas' (1943, p. 204). Albert Hirschman's early education was in Berlin and he received his degree in Trieste. He joined the debate on industrialization with his theory of unbalanced growth. In both cases the implicit concern was how to harness the technological dynamism of industry in service of development, the same theme as Ayres's. Both authors became prominent in development economics through their work at the World Bank and in Colombia. Their non-orthodox analysis linked them with institutionalism through their common categories of analysis. In addition, some of their writings, Hirschman's for example, show familiarity with institutionalist writers such as Veblen.

Less well-known development economists such as K. Mandelbaum, Antonin Basch and Ladislav Feierabend had similar views; all looked to the state to guide the movement from agrarian-based to industrial societies. This was in part a reaction to the conservative peasant-dominated governments of Eastern Europe before the end of the war, though it also reflected their sense of the success of projects such as the Tennessee Valley Authority (TVA).

For them, the state had a central role in changing Eastern European attitudes as well, particularly among the peasants. Further land reform and investment in agriculture had to be undertaken, and in this fashion the resistance to modernization would be broken and the technological advancements would foster development. Excess labor could be mobilized into the industrial sector. Rosenstein-Rodan suggested that removing this disguised unemployment would actually increase agricultural output. In the process the institutions of modern society, more related to Western Europe than to the feudal past of the Eastern region, would appear, consolidate and facilitate development. However, these were not simply market institutions. Indeed, Hirschman showed that free trade could end up hurting the poorer country.

There were other themes in Eastern European development: regional development, infrastructure growth, foreign investment. However, the central ideas and approaches conformed quite closely to those that had come to the forefront in the institutionalist golden age of the 1930s and 1940s. This laid the base for the bonds and links of institutionalism and development until the orthodox resurgence in the 1980s.

Institutionalist development economics: dynamism gained, optimism lost

Institutionalist development economics enjoyed growing interest and prestige until the 1980s because of its contribution to understanding the complexities of post-World War II economies and societies. The importance of development grew as decolonization proceeded and resources and infra-

structure were mustered to confront its challenges. Institutionalism's field of application increased as well, for the new countries that became objects of study presented a welter of different institutional contexts. This stimulated work that could be directly termed institutionalist development economics. Hirschman continued with his insightful observation of development processes, and how hidden rationalities and inverted sequences could affect the process of social change that he still viewed with a 'bias for hope'. Gunnar Myrdal (1968) made significant contributions in his evidence-intensive and magisterial *Asian Drama*, as well as his theoretical constructs of circular and cumulative causation and spread and backwash effects. Myrdal had been trained as a neoclassical macroeconomist and he was critical of institutional economics early in his career. However, his work on *Asian Drama* (Myrdal, 1968) and as Secretary-General of the Economic Commission for Europe led him to understand the complexity of development and the shortcomings of conventional economics (deGregori and Shepherd, 1994, p. 111). His methodology and concerns became central to the institutionalist development project. In a later and optimistic edition of his 1958 book, Hirschman (1978 [1958], p. 187) noted the similarity of his work to Myrdal's cumulative causation:

I now find that Gunnar Myrdal has addressed himself to similar problems . . . and has had recourse to the same conceptual tools that are employed here: his 'backwash' and 'spread' effects correspond exactly to my 'polarization' and 'trickling down' effects. Nevertheless . . . Myrdal's analysis strikes me as excessively dismal.

The corpus of institutionalist development economics literature grew tremendously in the decades before the 1980s. Virtually every issue of the institutionalist *Journal of Economic Issues* had one or more articles on development issues. The advances in understanding were significant and the contributions numerous. As a result, by the late 1970s authors were beginning to look back over the period and to summarize the work that had grown up and the relation of development and institutionalism. Myrdal (1974), Klein (1977) and Street (1987) all contributed survey articles. In addition, links between institutionalism and other analytical approaches to development, such as Latin American structuralism or dependency theory, were specifically noted and developed (Street and James, 1982). Sunkel (1989, p. 525) found similarities and, more importantly, complementarities among the approaches:

Institutionalism is also much stronger as regards the theoretical and conceptual groundings of its approach and its theory of socioeconomic change as a distinct and positive alternative to neoclassical orthodoxy . . . Structuralism is particularly strong in its conceptual approach and historical interpretation of Latin American under-development and dependency.

For the most part, the surveys documented the application and applicability of the central institutional constructs introduced above: technology and its role in driving the evolution of economies; the universality of technology; the role of habits in affecting individual behavior; the cumulative nature of change and the inadequacy of equilibrium constructs to capture developmental change; and the importance of induction with specific institutional detail rather than global deductive modeling. The articles exhibited great optimism about the value of cross-fertilization among institutionalism, structuralism and dependency theory in creating a more vibrant institutionalist development economics.

By the mid-1980s that optimism was gone, as reflected in Hirschman's observations. The reason was cogently captured by Bolin (1984, p. 643):

One need only pick up the daily newspaper to be aware of the conspicuous failure of prevailing policies of capital infusion and technology transfer to achieve development in less developed countries. The United States cannot point to a single success story among Third World countries to demonstrate the effectiveness of its development strategies since World War II.

His answer was to look to psychology and constructs such a 'need-achievement', to develop a 'Development Potential Index', and consciously to 'instill' the required development attitudes in developing countries. This was a far cry from traditional institutionalism, even though Bolin claimed an institutionalist perspective on development. This flailing about was symptomatic of institutionalist development economics' loss of dynamism in the face of the intransigent problems of development. Nonetheless, Sen (1983) disagreed strongly with Hirschman on the decline of development economics. He cited four of its theoretical themes that history seemed to vindicate. Although he did not defend development economics' link to institutional analysis, in the same article he documented 'fast growth and slow social change' as notable contributors to famines. These are constructs that institutionalism has long utilized. Sen and Hirschman might differ on whether institutionalist development economics had gone into decline; however, they could not deny that orthodox, market based approaches were in the ascendancy at that time. 'Get prices right' took center stage as the reductionist solution to the problems of development. The strands of the helix had separated.

New institutionalism and the market solution

Mirowski (2001) has pointed out that the protean nature of orthodox, market-based economics is its strong point. By the end of the 1970s, the frustration with the development process created a vacuum that orthodox economists were quite willing to fill. They advocated expansion of market

relations through reducing the role of the state, privatizing state enterprises, removing price and interest rate controls, and attacking interference in market processes, particularly in the labor market. Improvements in resource allocation efficiency and removal of the dead hand of government would provide the needed impulse to growth and development. Chile under General Pinochet became their laboratory and exemplar. One irony was that first the East Asian countries and then the Newly Industrialized Countries (NICs) were becoming development success stories at this time by following strong state policies that avowedly used government policy to guide the market outcomes (Chang, 2002). Another irony was that the strongest exponents of the 'diminish government' program were the government-charted international financial institutions, such as the World Bank, the International Monetary Fund (IMF) and the (IDB).

The debate was not completely one-sided. For example, Dietz (1994) provided an institutionalist critique of the neoliberal policy package and suggested a six-point institutionalist alternative. Nonetheless, development economics came to be dominated by 'orthodox economics' because of its claim of 'monoeconomics' and its assertion of mutual benefit between underdeveloped and industrial countries (Hirschman, 1981). Development economics had rejected the monoeconomics claim. So had traditional institutionalism, which in addition was skeptical about assuming mutual benefit under existing institutional structures. Nonetheless, economic policy in developing countries was soon dominated by a set of market-oriented prescriptions advocated by the international financial institutions and the aid agencies. Their most representative rendition was in the 'Washington Consensus' summarized by John Williamson (1990).

At first glance, institutionalism should find a place within orthodox development economics through the growing interest in the New Institutional Economics (NIE) of North (1990) and Williamson (1985). Closer examination indicated fundamental differences from 'old' institutionalist development economics. The NIE was built on individualistic neoclassical foundations and its concern was how rational action would generate the market institutions with their unique capacity to minimize transaction costs. The starting point was Williamson's 'in the beginning there were markets' (Hodgson, 1998, p. 182) and the analysis was based on methodological individualism (p. 176). This is very different from traditional institutionalism where individuals interact in institutions and individual purposes are molded by socio-economic conditions. Such evolutionary economics does not necessarily expand market relations, and markets are only one possible mechanism for economic interaction. So from a development perspective, early NIE reduced to a new version of the theory of underdevelopment:

countries were not developed because they lacked market institutions, their transactions costs were too high, and rent-seeking behavior was prevalent. Policy was reduced to getting prices and policy right through market institutions. Hirschman (1981, p. 21) describes this period as a silencing of development economists who retreated from the belief that 'all good things go together' to holding that 'good economics is good for people'. So while work continued in institutionalist development economics, often under the rubric of 'area studies', it was marginalized from the mainstream and from the international policy sphere.

On the other hand, the NIE did provide another potential explanatory avenue when the problems of development did not succumb to market solutions. Development economists could consider institutions, just as they turned to orthodox economics in the 1970s. For example, Nugent (1998, p. 8) found that the 'institutional content' of the *Journal of Development Economics* increased from 15 percent in the 1970s to 27 percent later. In the *Handbook of Development Economics*, its presence rose from 7.5 percent in 1988's first volume to 36 percent in Volume 3 in 1995. The links of institutionalism and development economics were being reconstituted.

Institutionalist development economics today

Economic development in 2008 continues to be complex and problematic. No one has found a simple formula for development. There is growing evidence that the contemporary experience is chastening the market fundamentalists, validating a more expansive view of development processes. This entails accepting the centrality of a broad range of institutions, not just those of markets and private property, and the centrality of institutional interaction with human agency and habit. The experience of China and India is a challenge to the market orthodoxy, because they have found a balance of state and market, of control and freedom, which is accelerating their growth and engendering doubts about that simple reductionist market approach to development.

The complexity of the development process has already moved institutionalism back to the center of efforts to understand development and to find successful policies. Links between the strands of development and of institutionalism are being re-established at numerous points. Their specific approaches differ from old institutionalism, but the themes of innovation, habit and evolution in a historically determined institutional context have their roots there.

Researchers using traditional intuitionist approaches are shedding light on development experiences and possibilities. Again today virtually every issue of the *Journal of Economic Issues* contains a development article. For example, in their study of Nepal and the effort to institute a

financial sector ‘institutional’ reform, Adams and Brunner (2003) show how Ayresian institutionalist categories can account for why the project ran into severe problems and fell far behind plans. They show that it would have been more successful had it accepted Ayres’s insight that ‘what happens to any society is determined jointly by the forward urging of its technology and the backward pressure of its ceremonial system’ (p. 364).³ Ho and Schneider (2002) utilized Myrdal’s backwash and spread effects and cumulative causation to analyze South Africa’s path since the end of apartheid and to suggest alternative policies. Stettler (2002) drew on Commons’s treatment of the emergence of rules and order to analyze disorder in the Maasai territory after private property rights were instituted. These studies update the analytical framework and imply that many of the issues faced in development have not changed fundamentally since institutionalism’s heyday.

Chang (2003) provides another link by updating the institutionalist critique of mainstream development economics. He documents the ‘unholy alliance’ of neoclassical economics’ treatment of market failure with the Austrian–libertarian belief that any intervention must be suppressed. He then offers an institutionalist political economy, empirically or inductively based, that places institutions and political decisions at the core of the economy and leads to suggestions for alternative development policies (Chang and Gabel, 2004).

There is also movement in NIE. It has moved away from its naive concentration on ‘market institutions’ to a much richer understanding of the entire institutional framework of economic development. Hoff’s (2003, p. 215) summary of historical evidence suggests: ‘(t)he historical findings reported here are sharply at variance with (NIE) views of economic development, (i.e.) the functionalist view that institutions are endogenous, flexible, and efficient’. As Hodgson (1998, p. 177) wrote: ‘Accordingly developments in the “new” institutionalism show some signs of yielding some ground to the “old”, or at least creating the possibility of a fruitful dialog between the two approaches.’ This can be seen in the evolution of Williamson’s and North’s writings. For example, Williamson (2000) now writes of four levels of the ‘economics of institutions’. Though he pushes embeddedness, traditions and norms to a nearly irrelevant long-run, and envisions ‘getting the institutional environment right (for markets)’, he admits the wider importance of institutions. Indeed, he even cites Commons, Myrdal and Mitchell as forefathers of NIE. Writing of North, Hoff (2003, p. 224) notes that he adopted the functionalist view in his early work ‘but abandoned it in his later work’. For example, North now calls for ‘a clear understanding of the new institutional economics’, development of ‘a body of political-economic theory’, and ‘a better understanding of the

social norms and informal constraints' that affect performance (North, 2000, p. 491).

An additional link of an evolutionary nature originates with the 'historical and comparative institutional analysis' (HCIA) school. Its premise is that there are multiple institutional equilibria possible; as a result the outcome is path-dependent and a result of a particular historical evolution. The actual institutions depend on historical conditions and cultural beliefs; market institutions are only one possibility. The HCIA case studies document many other institutional forms that have developed as societies confront economic issues (Greif, 2000).

Next, mathematization was traditionally the special purview of orthodox economics. Today, increased computer power and mathematical sophistication have made it possible to model the very complex processes that underlie development (Colander, 1996). The best contemporary examples are the computable general equilibrium models (CGEs) and models based on social accounting matrixes (SAMs) that allow specific modeling of many of the concerns of institutionalists, for example income inequality, technological differentiation or ethnic complexity. System dynamics models that can capture elements of the evolution of economic systems provide another example. The mathematically oriented 'microfoundations of institutional arrangements' (Bardhan, 2000) provide another linkage between strands. Bardhan's (2005) methodology is quite mainstream, with complex micro-based modeling of a variety of phenomena. On the other hand, his purview is far broader than the mainstream fixation on the market. He examines the interaction of the private sphere, the government and the community, and he brings acute empirical scrutiny to experiences of collective action, distributive conflicts, accountability and cooperative action.

The final link is epistemological convergence and it is here that the two may form unbreakable bonds. Geoffrey Hodgson (2004, 2005) returns to the origins of institutionalism and highlights the approach to knowledge that existed at that time: its psychology was based in instinct theory; it was evolutionary, especially with Veblen; and the philosophical guide was the pragmatism of John Dewey. The post-World War II attack on institutionalism substituted behaviorism, an ahistorical mechanical model, and logical positivism. Hodgson (2004) makes a strong case that psychology has moved back toward an instinct theory, that an evolutionary perspective and 'emergentism' are becoming accepted as better explanations of complex social systems, and that logical positivism is being replaced by pragmatism. Each of these has taken on modern characteristics growing out of methodological and empirical advances in recent decades. Nonetheless, the core psychological and philosophical bases of old

institutionalism have returned to prominence in social science and this will be reflected in our understanding of development.

Conclusion

All of these factors converge to suggest that the future of development and institutionalism are inextricably intertwined, that these two strands of the double helix of institutionalist development economics will continue to be linked. This is not surprising since such was the case from the first writings of development economists. It is not clear that there will be a unified institutionalist development economics. However, the links and bonds between institutionalism and development will only grow stronger, which can only improve our understanding of the development process. It will allow us to avoid forgetting that ‘interests’ can propel countries to higher economic status, to development, while the ‘passions’ they embody will insure that no country can simply ‘lumber through the various stages of development single-mindedly’ (Hirschman, 1981, p. 24).

Notes

1. My thanks to Chuck Wilber and Jim Weaver for their comments and suggestions on an earlier draft and to the editors of the volume for their helpful comments.
2. Some institutionalists allow application of the term only to those whose work builds specifically on Veblen, Commons, Ayres, and so on, the forefathers of institutionalism. My use is broader and includes development economists whose concepts are closely related to traditional institutionalism’s. I find support from the inclusion of these development economists in institutional compendia such as Hodgson et al. (1994).
3. Some of the traditional institutionalist themes may need to be re-examined in light of contemporary development problems. For example, in many areas, it has become clear that the ‘developmental state’ that was to be the organizer of development has failed. Ecuador, Bolivia, Papua New Guinea, and many African countries come immediately to mind.

References

- Adams, John and Hans-Peter Brunner (2003), ‘Technology and Institutions in the Process of Economic Reform: Achieving Growth with Poverty Reduction in South Asia’, *Journal of Economic Issues*, 37 (2): 363–9.
- Bardhan, Pranab (2000), ‘Distributive Conflicts, Collective Action, and Institutional Economics’, in Gerald Meier and Joseph Stiglitz (eds), *Frontiers of Development Economics: The Future in Perspective*, New York: Oxford University Press, pp. 269–90.
- Bardhan, Pranab (2005), *Scarcity, Conflicts, and Cooperation: Essays in the Political and Institutional Economics of Development*, Cambridge, MA: MIT Press.
- Bolin, Meb (1984), ‘An Institutional Perspective on Economic Development’, *Journal of Economic Issues*, 18 (2): 643–50.
- Chang, Ha-Joon (2002), *Kicking Away the Ladder: Development Strategy in Historical Perspective*, London: Anthem Press.
- Chang, Ha-Joon (2003), ‘The Market, the State, and Institutions in Economic Development’, in Ha-Joon Chang (ed.), *Rethinking Development Economics*, London: Anthem Press, pp. 41–60.
- Chang, Ha-Joon and Ilene Grabel (2004), *Reclaiming Development: An Alternative Economic Policy Manual*, London: Zed Books.
- Colander, David (1996), ‘New Institutionalism, Old Institutionalism, and Distribution Theory’, *Journal of Economic Issues*, 30 (2): 433–42.

- DeGregori, Thomas and Deborah Shepherd (1994), 'Myrdal, Gunnar', in Geoffrey Hodgson, Warren Samuels and Mark Tool (eds), *The Elgar Companion to Institutional and Evolutionary Economics*, Aldershot, UK and Brookfield, US: Edward Elgar Publishing, pp. 107–12.
- Dietz, James (1994), 'Development Policy', in Geoffrey Hodgson, Warren Samuels and Mark Tool (eds), *The Elgar Companion to Institutional and Evolutionary Economics*, Aldershot UK and Brookfield, US: Edward Elgar Publishing, pp. 139–45.
- Fusfeld, Daniel (2000), 'A Manifesto for Institutional Economics', *Journal of Economic Issues*, 34 (2): 257–66.
- Greif, Avner (2000), 'The Fundamental Problem of Exchange: A Research Agenda in Historical Institutional Analysis', *European Review of Economic History*, 4: 251–84.
- Hirschman, Albert (1978 [1958]), *The Strategy of Economic Development*, New York: Norton Library.
- Hirschman, Albert (1981), 'The Rise and Fall of Development Economics' Essay 1 in *Essays in Trespassing: Economics to Politics and Beyond*, Cambridge: Cambridge University Press, pp. 1–24.
- Ho, P. Sai-wing and Geoffrey Schneider (2002), 'Africa Drama: Myrdal and Progressive Institutional Change in South Africa', *Journal of Economic Issues*, 36 (2): 507–15.
- Hodgson, Geoffrey (1998), 'The Approach of Institutional Economics', *Journal of Economic Literature*, 36 (1) 166–92.
- Hodgson, Geoffrey (2004), *The Evolution of Institutional Economics: Agency, Structure and Darwinism in American Institutionalism*, London: Routledge.
- Hodgson, Geoffrey (2005), 'Institutions and Economic Development: Constraining, Enabling and Reconstituting', in Gary Dymksi and Silvana De Paula (eds), *Reimagining Growth: Towards a Renewal of Development Theory*, London and New York: Zed Books, pp. 85–98.
- Hodgson, Geoffrey, Warren Samuels and Marc Tool (eds) (1994), *The Elgar Companion to Institutional and Evolutionary Economics*, Aldershot, UK and Brookfield, US: Edward Elgar Publishing.
- Hoff, Karla (2003), 'Paths of Institutional Development: A View from Economic History', *World Bank Research Observer*, 18 (2): 205–26.
- Klein, Philip (1977), 'An Institutional View of Development Economics', *Journal of Economic Issues*, 11 (4): 785–808.
- Meier, Gerald (2005), *Biography of a Subject: An Evolution of Development Economics*, Oxford: Oxford University Press.
- Mirowski, Philip (2001), 'Economism from Affront to Façade or the Contours of American Economic Orthodoxy 1929–1989', in Amitava K. Dutt and Kenneth P. Jameson (eds), *Crossing the Mainstream: Ethical and Methodological Issues in Economics*, Notre Dame, IN.: University of Notre Dame Press, pp. 225–46.
- Myrdal, Gunnar (1968), *Asian Drama: An Inquiry into the Poverty of Nations*, New York: Pantheon.
- Myrdal, Gunnar (1974), 'What is Development?' *Journal of Economic Issues*, 8 (4): 729–36.
- North, Douglass (1990), *Institutions, Institutional Change, and Economic Performance*, Cambridge: Cambridge University Press.
- North, Douglass (2000), 'Needed: A Theory of Change', in Gerald Meier and Joseph Stiglitz (eds), *Frontiers of Development Economics: The Future in Perspective*, New York: Oxford University Press, p. 491.
- North, Douglass (2005), *Understanding the Process of Economic Change*, Princeton, NJ: Princeton University Press.
- Nugent, Jeffrey (1998), 'Institutions, Markets, and Development Outcomes', in Robert Picciotto and Eduardo Wiesner (eds), *Evaluation and Development: The Institutional Dimension*, New Brunswick, NJ: Transaction Publishers, pp. 7–23.
- Rosenstein-Rodan, P.N. (1943), 'Problems of Industrialisation of Eastern and South-Eastern Europe', *Economic Journal*, 53 (210/211): 202–11.
- Rutherford, Malcolm (2000), 'Institutionalism Between the Wars', *Journal of Economic Issues*, 34 (2): 291–304.

- Sen, Amartya (1983), 'Development: Which Way Now?' *Economic Journal*, 93 (272): 745–62.
- Smith, Adam (1976 [1776]), *An Inquiry into the Nature and Causes of the Wealth of Nations*, ed. Edwin Cannan, Chicago, IL: University of Chicago Press.
- Sobel, Irvin (1979), 'Adam Smith: What Kind of Institutionalist Was He?' *Journal of Economic Issues*, 13 (2): 347–69.
- Stettler, Michael (2002), 'John Commons on Institutional Disorder and an Application Relating to Property Rights in Maasai Territory', *Journal of Economic Issues*, 36 (2): 517–25.
- Street, James (1987), 'The Institutionalist Theory of Economic Development', *Journal of Economic Issues*, 21 (4): 1861–86.
- Street, James and Dilmus James (1982), 'Institutionalism, Structuralism, and Dependency in Latin America', *Journal of Economic Issues*, 16 (3): 673–90.
- Sunkel, Osvaldo (1989), 'Structuralism, Dependency and Institutionalism: An Exploration of Common Ground and Disparities', *Journal of Economic Issues*, 23 (2): 519–44.
- Warner, James M. and Kenneth P. Jameson (2004), 'The Role of Eastern Europe in Development Economics History', *History of Economics Review*, 29: 65–87.
- Williamson, John (ed.) (1990), *Latin American Adjustment: How Much Has Happened?* Washington, DC: Institute for International Economics.
- Williamson, Oliver (1985), *The Economic Institutions of Capitalism*, New York: Free Press.
- Williamson, Oliver (2000), 'The New Institutional Economics: Taking Stock, Looking Ahead', *Journal of Economic Literature*, 38 (3): 595–613.
- World Bank (2001), *World Development Report 2002: Building Institutions for Markets*, Oxford: Oxford University Press.

13 Neoclassical development economics

Salim Rashid

Introduction

Neoclassical approaches to economic development (hereafter Necdev) focus upon markets as the primary means to enrich poor economies. As such its primary analytical tools are optimization and equilibrium. These are neutral tools and yet Necdev is also associated with some specific policies. These are sometimes summarized in the phrase: ‘Stabilize (balance the budget), Liberalize (remove tariffs) and Privatize (minimize state control and ownership)’, which are collectively termed the ‘Washington Consensus’. This is a rather concrete and hence workable approach to getting some 2 billion poor people out of their misery. Earlier attempts at economic development envisioned much more comprehensive goals and even talked of ‘transforming’ societies. Eschewing such grandiose visions, Necdev argued that it is better to aim lower but achieve our goals. There is an inherent tension between the analytics of Necdev and its policy implications and this is a theme addressed later. Exposing the faults of others is a contribution, and Necdev as policy has been invaluable in this role. Necdev has acutely displayed the reasons for failure when greed and markets have been ignored. Yet Necdev itself has had limited direct success in creating economic development. To understand why Necdev is the most influential story in this field today we need to set out the intellectual roots of Necdev, observe its limitations, and then trace the recent history of development economics.

Roots of neoclassical development economics

Conventional wisdom has it that ‘scientific’ economics began in the late eighteenth century and it is notable that our modern policies for economic development mirror those found in the nineteenth century. In the *Wealth of Nations*¹ Adam Smith (1776 [1881]) popularized three axioms which could then be directly applied to the problem of maximizing economic growth:

1. All individuals desire to maximize wealth. (Greed)
2. All individuals know better than government what will maximize their wealth. (Knowledge)
3. National wealth is the sum of individual wealth. (Additivity)

These premises were widely known and used before Adam Smith, but they were taken as empirical rules rather than as axioms. The great advantage of turning the empirical rules into axioms is that it enables us to bypass history and culture and get right at advocating economic policies to maximize growth. If history and culture could be thus ignored, then a powerful intellectual breakthrough would have been achieved. The entire modern debate hinges on the extent to which this presumed axiomatic simplification can be taken as valid.

Axioms 1 and 2, in conjunction, prove that individual wealth is maximized when government leaves individuals alone. Axiom 3 then says that maximizing individual wealth suffices to maximize national wealth. If we describe the axioms loosely as greed, knowledge and additivity, then greedy and knowledgeable individuals surely do not need government in order to maximize their wealth and additivity suffices to assure us that, since the aggregate is the simple sum of the individuals, the aggregate also does not need government to maximize economic growth. Modern readers, familiar with externalities in the form of, say, pollution, will probably be most curious about the validity of axiom 3, but in most less-developed countries LDCs axioms 1 and 2 are also worth questioning.²

What Necdev has done is to extend the scope of greed to include history and culture, and to project the depth of our knowledge into the indefinite future. Necdev has also added precision into this subject by insisting upon the need for mathematics; as additivity is mathematically convenient, Necdev silently accepts additivity unless forced to do otherwise. Necdev believes intensely in markets. But by paying scant attention to the social preconditions that allow markets to function well, it implicitly makes the market an autonomous part of society. Can this work? What are the salient facts about the practical value of Necdev policies?

Limitations

The most important fact for students of economic development is that several East Asian economies have grown for almost 50 years at rates unheard of in human history. Speaking crudely we can state the following. For centuries mankind struggled with growth rates of 0.5 per cent. This increased to 1 or even 2 per cent in the eighteenth and nineteenth centuries. It increased to 3 per cent in many places in the early twentieth century. But in the latter half of the twentieth century Japan, South Korea, Taiwan and China achieved growth rates of 7 per cent or more for decades. These countries did not follow Necdev policies, their policy-makers provide virtually no economists of note and, as we look more into details, their policy templates look specific to themselves.³ If economic growth is what we want, why should we look at Necdev? The

argument is not about markets versus planning, but rather about the extent to which markets will be used and the limits within which they will be free.

A look at one weapon in the economist's armory – comparative advantage – gives an important indication why *Necdev* is often felt to be a distraction. Comparative advantage assures us that Europe and China will find trade to be mutually beneficial even if Europe is better at doing everything. True. But the question we want answered is how Europe came to be better at doing everything. Are they better human beings (the implicit and explicit answer of many)? If not, then let us ask what it is in their history that led them to be where they are. The larger question needs to be appropriately framed, before the optimizing activities of greedy individuals can help in deciding appropriate economic policies.

To get to the larger framework, let us ask, 'How is economic development distinct from just economic growth?' If we take economic development (ED) to have begun with President Truman's speech in 1947⁴ and not with vague concerns about 'development', or with the first use of a phrase, then there are three reasons for separating ED from growth:

1. Understanding the possibility and desirability of speedy growth.
2. Emphasizing the need for redistribution to the have nots.
3. Accepting that such economic progress can require global, discrete changes.

The first goal is both positive and normative, the second is redistributive and the third is interventionist. ED is one those rare subjects where theory and policy are intertwined because the subject is defined by the desire to see change. This is why other fields have often referred to ED as 'modernization theory'.

An early attempt to demonstrate the virtues of *Necdev* policies came with the occupation of the Philippines by the USA. The policy of the Philippine Commission exhibits several of the features later seen in *Necdev* policies. Filipinos were to be given some local autonomy, yet kept under tutelage; they were to be 'educated', but with no consciousness that the proposed course was so comprehensive that it would feel like indoctrination to those opposed to American ideas: 'The character of the people contains many discouraging defects which can only be cured by careful tutelage and widespread education'.⁵ As for economic policies, these consisted of opening up Filipino land to US investors, reducing tariffs and building roads and harbours. It did not occur to the members of the Commission, well meaning though they were, that something more and something other than American investment was needed to pull the Philippine economy out

of its poverty. Indeed the policy has considerable similarities with those proposed for Ireland in the 1820s.

ED wants to get things done, it focuses upon policies and Necdev as policy answers the call by providing us with specific policies. Right or wrong, this does further the argument. An alternative interpretation of Necdev represents it only as a method: 'You tell me what people want and I will tell you the optimal way to get it.' As method, Necdev is simply constrained optimization. Even though many famous names have espoused the claim that Necdev is only a method, this loose usage creates confusion. Constrained optimization is a mathematical problem and should be left primarily to mathematicians; what such optimization can tell us depends wholly upon what is being optimized and which constraints are taken to be binding. The hard problem is to decide upon the objective function and the constraints, and Necdev as method is of no help in settling the hard problem.

The use of optimization and equilibrium as the preferred tools of Necdev has produced a fissure in the idea of economics as science. Since markets are incomplete in most LDCs we have to find ways in which optimizing agents adapt to this absence. This makes game theory a natural tool to examine problems of development. However, game-theoretic models allow us to talk about such issues as power and asymmetrical relations, and models including such notions frequently suggest market interventions to be beneficial – for example, they can provide an economic rationale for usury laws because the borrowers willing to pay the highest interest rates may be the least desirable ones. Such embarrassing outcomes do not greatly affect the impact of Necdev because every such outcome is the result of a specific construction – it has the flavour of a 'Just So' story. By wiggling the model one can typically show that a different intervention, or maybe none at all, was desirable. And one can always argue that such incomplete markets are a transitory phenomenon – economic growth will soon erase such oddities.

The interpretation of Necdev as method has deepened the hold of Necdev as policy by encouraging the view that history and culture are simply the results of the self-interest of earlier generations. If institutions and culture really could be shown to be simply ossified self-interest, then this school of thought might have a hope of illuminating ED. But we are hard-pressed to find compelling empirical support for such a view; instead, we are faced with a plethora of models purporting to 'explain' the caste system, racial bias, ethnic conflict and the like. Which explanation can we count on?

The axiomatic systems of 'economic science' build upon the following assumptions:

1. Uniformity of individuals and firms.
2. Selfish optimality for all.
3. Individualistic constraints for all.
4. Unlimited decision-making skills.
5. Instant or rapid equilibrium.

By varying the extent to which we dispense with any one or more of these assumptions, we can produce alternative models for any given set of facts almost at will. Using Necdev as method, multiple theories, each of which ‘explain’ economic phenomenon, are, so to speak, always available. How does it help to have 17 different game-theoretic models for the caste system, unless our goal is to maximize academic publications?⁶ The fog created by Necdev as method allows Necdev as policy to appear as the only viable ‘scientific’ method to deal with problems of economic development.

Recent history of development economics and the role of neoclassical development economics

The modern understanding of ED usefully dates from Truman’s ‘Five Points’ speech.⁷ But Eastern Europe’s economic condition had aroused the interest of economists such as Rosenstein-Rodan (1943), and there were also colonial reports from the 1920s which used the phrase ‘economic development’ and argued for policies to achieve it. The overwhelming impression one gets from these early accounts is of lost potential – surplus labour and wasted resources abound in such underdeveloped countries. During the 1940s this anger at avoidable waste – already aroused by the Great Depression – coupled with a widespread belief in the limited good to be achieved by the market, led many to hope for little from the market mechanism. By contrast, the Soviet economy was believed to be growing by leaps and bounds, and even the free market economies provided astounding examples of the success of large coordinated planning, such as the Tennessee Valley Authority (TVA) within the USA and the Marshall Plan for Europe. ED was born with the large hopes aroused by planning and direction from above.

But even as large problems and large solutions were being addressed, economists arose to insist on the need for small steps that were necessary prerequisites. Indeed, in the 1950s, the argument for planning was based on the idea that the poor of the world were either not materialistic enough or not knowledgeable enough to participate properly in the market. But the same argument should have led the planners to realize that to the extent that they succeeded in creating material prosperity, they were also creating the people and the economic structures that would falsify the planners’ own assumptions. Sometimes the planning model was made to realize this with

a jolt, as in V.K.R.V. Rao's (1952) demonstration that the predominantly agrarian structure of most LDCs made the application of Keynesian remedies futile.

The profession expressed unease about what to do, as someone who knows something very important is happening, feels the urge to do something, but has no clear conception what. This led to some frustrating episodes, such as the attempt to make the Lewis model some kind of paradigm of development, thus providing neoclassicals with the chance to show that nothing very unusual was at stake; or in Walt Rostow's (1960) stages of growth, a book which provided us with some memorable phrases for labelling different levels of development, such as the 'take-off', but which was of no help in solving any real problem. The confusion is apparent in the occasional remarks made by Rostow about the practical steps that need to be taken. Rostow felt that the wealth in LDCs that is:⁸

largely concentrated in the hands of those who own land, must be shifted into the hands of those who will spend it on roads and railroads, schools and factories rather than on country houses and servants, personal ornaments and temples. (Rostow, 1960, p. 19)

If we have a regime of private property, who is to 'shift' all this income from the landlords and how will they do it? No wonder that those who were more clearheaded about such events, such as the Mexican economist Edmundo Flores, welcomed the admission on Rostow's part and went on to argue that the only practicable action implied was redistributive land reform.

The literature on land reform from 1950 to 1975 can be characterized by the words: 'Think revolution but do not say the word.' While the economists hummed and hawed about the virtues of the market, policymakers decided that communism was such a great threat that land reform had to be undertaken as triage. It is surely ironic that the most sustained economic activity undertaken by the United States was almost certainly the effecting of land reform in Japan, Taiwan, and South Korea – as well as the attempts made in many other countries through the United States Agency for International Development (USAID). Development economics was born with a split conscience.

In principle, the planners had much logic on their side. The welfare theorems of modern economics show how planning can be superior – Pigou, Hotelling, Arrow were all superior theorists – but they could not advise us how to find the bureaucracy that would implement their plans. Planning seemed inevitable, but how? Despite the general uncertainty about what to do, there was a steady but determined trickle of market-based criticism of the plan models. Jacob Viner questioned whether there really was surplus labour – in the sense of labour with zero marginal product. Was it really

impossible to plough, hoe, weed or fertilize a little more? Hla Myint argued that the dramatic increases in rice production in the early twentieth century in Burma showed that peasants could respond quite well to profits, while Bauer demonstrated the inefficiency, if not the counterproductive nature, of marketing boards in West Africa.

The intellectual reaction solidified with Ronald Coase's demonstration that the existence of externalities alone need not provide grounds for intervention; if people were greedy and could negotiate costlessly, they could also reach decentralized solutions. The important point to be emphasized is that there were no conceptual errors in what was said by the development planners – only vast faith in the ability of experts to spot, and of bureaucrats to 'fix', the failures of the market. We see one example of such faith in Lenin's claim that modern industry had made running a factory just a bureaucratic exercise; nor should one ascribe this naivety to the Communists alone, since Max Weber is equally mechanical in his interpretation of the function of bureaucracy.

It was the evolution of facts that changed the intellectual scene. Two dominant pessimistic 'facts' about population and trade had supported the planning mentality of early ED. Population pessimism, really a revived Malthusianism, believed that people would outrun food at such speed that it needed all our effort just to feed the world – there really was not time to worry about growing rich.⁹ Trade pessimism claimed that the facts showed LDC products to be at a continual disadvantage in competition with developed-country (DC) products – the Singer–Prebisch thesis argued that the terms of trade were constantly turning against the LDCs so growth through trade was really like running up a slippery slope. Hence import substitution industrialization became embedded in Latin America for about two decades. Time did not support either pessimistic expectation: commodity prices did not continue to go down *vis-à-vis* manufacturing prices, export-led growth gave startling success to some East Asian economies, and countries like Bangladesh showed dramatic decreases in fertility while still being quite poor.

Ian Little demonstrated the failure of detailed planning in several protective trade regimes and Deepak Lal popularized the practical failure with an obituary wish: 'the demise of development economics is likely to be conducive to the health of both the economics and the economies of developing countries'.¹⁰ The tone of some of the discussion is inexplicable if one forgets that what was proposed for the LDCs was also meant to indicate weaknesses in DCs, and later the point was turned inwards with force when the success of privatization in the UK was held to indicate its success in LDCs.¹¹ In a later article Lal (1985) tried to provide a moderate defence of his extreme language. It amounted to two statements:

firstly, that the market distribution of income and wealth was not obviously worse than that generated by an economist's welfare function; and secondly, that the existence of market failure did not prove the need for intervention – bureaucratic failures might be greater. True. But Lal should also have then pointed out the purely pragmatic turn of the debate. Neither side had a clear case; what can be supported should thus depend upon one's assessment of the relative strengths of the market versus the bureaucracy – but this is just the sort of question that only history and culture can illuminate. So a proper resolution of the debate would have required some iconoclasm on the part of those who struggled with the questions affecting development.

Such open-mindedness conflicted with other parts of Necdev, indeed of the entire mindset of those focused upon economic development. Both the planners and the mathematically inclined members of Necdev felt that uniformity – empirical regularities – were an essential support for their programme. Simon Kuznets, then Syrquin and Chenery, expended admirable effort in finding such regularities. While Kuznets was modest about his findings, Syrquin and Chenery promised more: 'Although development experiences may vary over time and across countries, there is sufficient uniformity within them for the main features of structural transformation to emerge as clear and consistent patterns of modern economic growth.'¹² When such regularities are meant to support policies, however, they turn out to be vacuous or trivial.¹³ Those uniformities that do exist are of a different nature than those the quantitative economists pin their hopes upon. They are the uniformity of nationalism, of social values as a binding element of cultural values upholding respect for property – but these are just the factors not amenable to logic-chopping and model-building. Few are so clear as Bhagwati on the idiosyncratic role to be played by theory:¹⁴ 'It is perhaps true that the only valid generalization in development economics is that no generalization is possible.'

The old school of ED planners initially questioned whether greed was sufficiently prevalent to allow markets to work, then they looked to externalities as justification for intervention; nowadays the attention has shifted to informational problems. Throughout, the most important question they posed was whether only small steps are valid, in which case the market could be left to grope towards the solution, or whether large ones are also needed. The planners showed repeatedly, and correctly, that there can be market failures and coordinated action was called for – but they could not tell us who was to direct the large step that needed to be taken. Practically, they handed the torch to the bureaucracy, without showing any appreciation for what a bureaucracy could actually do, or for the culture created by a bureaucratic economy.

Necdev will say, in asides, that culture is very important, but as we can neither understand it and nor can we hope to mould it, we had best mind our own business. If only such a virtuous possibility were open to us. But such an attitude fails to understand that even the bare-bones market in which transactions occur is still a culture, albeit a very desiccated one and hence incapable of supporting humans for very long. We have no choice but to engage with 'culture', in the wider sense, and we cannot presume that social values will be explicable by self-interest alone or be amenable to our mathematics. When Necdev economists rushed into the former communist countries – as history and culture can be ignored, they had every right to do so – with the cry 'get the prices right', they soon found this to be inadequate. So they modified it to 'get the institutions right', and even this was found wanting; finally we hear 'get the culture right'. But at this point our market-makers are designing institutions and cultures, so they are also social planners and the main point has been conceded.¹⁵ Of course culture matters. The paradox of capitalism is that it is a system ostensibly based on self-interest, yet wholly dependent on non-economic virtues for its foundations and success.¹⁶ Once Necdev accepts that non-economic factors need to be considered, not only alongside economic ones, but also given primary importance on many occasions, it will find a positive and valuable place in the vital transforming challenge that is economic development.¹⁷

Notes

1. Wesley Mitchell first provided this compact and lucid characterization in *Types of Economic Theory*, pp. 61–65.
2. Smith was generally unaware of the actual economic conditions or problems facing Scotland.
3. In Latin America, the resurgence of Necdev is known as neoliberalism. For an evaluation of its impact, see Amman and Baer (2002).
4. Rist (2000).
5. May (1980), 15,142.
6. Lucas describes neoclassical economics as the 'use of mathematically explicit descriptions of agents' preferences and the technology available to them' (p. 123). The objective evades the two principal points needed for economic policy. First, what do we do with the multiplicity of mathematical models and solutions that can be formulated? Secondly, it does not even ask whether there are large questions that need to be handled even if they cannot be mathematically modelled.
7. Address by President Harry S. Truman on 12 March 1947 before a joint session of congress, which called for economic and military support for Greece and Turkey, so as to prevent these countries from falling into the Soviet sphere of influence. This speech laid down the so-called Truman Doctrine. The complete text of the speech is available at the website of Avalon Project at Yale Law School: <http://www.yale.edu/lawweb/avalon/trudoc.htm>.
8. Rostow (1960), p. 19.
9. Paddock and Paddock (1967) provides a convenient example of the hysteria that can be generated.
10. Lal (1983), p. 130.

11. At the conclusion of a calm article on the role of services, Spengler (1967) turns with vehemence to the prevalence of planning in the US economy.
12. Syrquin and Chenery (1989), article summary, p. 145. An even grander achievement is promised by Lucas: 'if we look at the right evidence, organized in the right way, we can get very close to a coherent and reliable view of the changes in the wealth of nations that have occurred in the last two centuries and those that are likely to occur in the present one'. This goal is achieved by actually subsuming history satisfactorily within the fold of Necdev: 'I do not think we can understand the contemporary world without understanding the events that have given rise to it.' (Lucas, 2004, p. 5).
13. See Levine and Renelt (1992).
14. Bhagwati (1984, p. 23).
15. What are called the 'Economics of Transition' economies can be better understood as the economic development problems of those countries that have a history of Marxist socialism in their recent past.
16. A valuable instance of learning from experience is DeLong (2004).
17. One needs to ask how far the economic plight of the Middle East is due to the fact that it was deliberately carved up to serve Western purposes. In the words of the India Office in London in 1916: 'What we want is not a United Arabia: but a weak and disunited Arabia, split up into little principalities so far as possible under our suzerainty – but incapable of coordinated action against us' (Black, 1966, pp. 196–7).

Bibliography

- Amann, Edmund and Werner Baer (2002), 'Neoliberalism and its Consequences in Brazil', *Journal of Latin American Studies*, 34 (4): 945–59.
- Bauer, Peter T. and Basil S. Yamey, *The Economics of Under-Development Countries*, Chicago, IL: University of Chicago Press.
- Bhagwati, Jagdish (1966), *The Economics of Underdeveloped Countries*, New York: McGraw Hill.
- Bhagwati, Jagdish (1984), 'Development Economics: What Have We Learned?', *Asian Development Review*, 2 (1).
- Black, C.E. (1966), *The Dynamics of Modernization: A Study in Comparative History*, New York: Harper & Row Publishers.
- Black, Edwin (2003), *Banking on Baghdad: Inside Iraq's 7,000-Year History of War, Profit and Conflict*, Hoboken, NJ: John Wiley & Sons.
- Chakravarty, Sukhamoy (1987), 'The State of Development Economics', *The Manchester School of Economic & Social Studies*, 55 (2): 125–43.
- Cohen, Daniel, *The Wealth of the World and the Poverty of Nations*, trans. Jacqueline Lindenfield, Cambridge, MA: MIT Press.
- DeLong, J. Bradford (2004), 'Should we still support Untrammelled International Capital Mobility? Or are Capital Controls Less Evil Than We Once Believed?', *The Economist's Voice*, 1 (1): Article 1.
- Harris, Nigel (1987), *The End of the Third World: Newly Industrializing Countries and the Decline of an Ideology*, Harmondsworth: Penguin Books.
- Hirschman, Albert O. (1958), *The Strategy of Economic Development*, New Haven, CT: Yale University Press.
- Kuznets, Simon (1973), 'Development and Underdevelopment: Historical Interpretations', *American Economic Review*, 63 (3): 247–58.
- Lal, Deepak (1983), *The Poverty of 'Development Economics'*, London: Institute of Economic Affairs.
- Lal, Deepak (1985), 'The Misconceptions of "Development Economics"', *Finance and Development*, June: 10–13.
- Levine, Ross and David Renelt (1992), 'A Sensitivity Analysis of Cross-Country Growth Regressions', *American Economic Review*, September: 942–63.
- Lucas, Robert E. (2004), 'The Industrial Revolution: Past and Future', *The Region*, May: 5–20.

- May, Glenn Anthony (1980), *Social Engineering in the Philippines: The Aims, Execution and Impact of American Colonial Policy 1900–1913*, Westport, CT: Greenwood Press.
- Nurkse, Ragnar (1953), *Problems of Capital Formation In Underdeveloped Countries*, New York: Oxford University Press.
- Paddock, William and Paul Paddock (1967), *Famine – 1975! America's Decision: Who Will Survive?* Boston, MA: Little, Brown & Company.
- Rao, V.K.R.V. (1952), 'Investment, Income, and the Multiplier in an Underdeveloped Economy', *Indian Economic Review*, 1 (1): 55–67.
- Rist, Gilbert (2000), *The History of Development*, London: Zed.
- Rosenstein-Rodan, Paul (1943), 'Problems of Industrialization in Eastern and South-Eastern Europe', *Economic Journal*, 53 (210/211): 202–11.
- Rostow, W.W. (1960), *The Stages of Economic Growth: A Non-Communist Manifesto*, Cambridge: Cambridge University Press.
- Sen, Amartya (1983), 'Development: Which Way Now?' *Economic Journal*, 93 (December): 745–62.
- Smith, Adam (1776 [1881]), *An Inquiry into the Nature and Causes of the Wealth of Nations, Vols I and II*, in R.H. Campbell and A.S. Skinner (eds), *Glasgow Edition of the Works and Correspondence of Adam Smith*, Indianapolis, IN: Liberty Fund.
- Smout, T.C. (1987), 'Where had the Scottish Economy Got to by the Third Quarter of the Eighteenth Century?' in Istvan Hont and Michael Ignatieff, *Wealth and Virtue: The Shaping of Political Economy in the Scottish Enlightenment*, London and New York: Cambridge University Press, pp. 45–72.
- Spengler, J.J. (1967), 'Services and the Future of the American Economy', *South Atlantic Quarterly*, Winter, 104–15.
- Syrquin, Moshe and Hollis B. Chenery (1989), 'Three Decades of Industrialization', *World Bank Economic Review*, 3 (2): 145–81.

PART III

MACROECONOMICS OF GROWTH AND DEVELOPMENT

14 General long-run approaches to growth and development

*Amitava Krishna Dutt*¹

Introduction

Although the level of economic development is by no means synonymous with the level of per capita real income and product, most development economists take the position that the level of development is quite closely correlated with them.² Given this, a country is said to be developed if it has a high level of per capita real income or product, which requires that it has experienced a significant growth in per capita real income and product over a long period of time. To understand why development occurs and why some countries remain less developed is therefore to understand why growth occurs in some countries and why it does not in others. Development economists would like nothing better to have a unified approach and, better still, a single model which answers the question of what causes growth and development.

The search for such an approach and model has proved elusive. This quest is also arguably misguided, because there is no reason why all countries which are poor are poor for the same reason or can be characterized in the same way, or why all countries which grow experience growth for the same reason. Moreover, it is reasonable to expect that a process which is as complex as development depends on a variety of factors, and not everyone will agree about their relative importance even for a particular part of the world. Indeed, the analysis of long-run approaches to growth and development has spawned a number of different approaches and models. It is sometimes argued that there are two broad traditions of growth and development which represent different visions.³ It has also been argued that the growth process can be examined in terms of different analytical theories and models, such as classical-Marxian models, neo-classical models, new growth theory models and models focusing on aggregate demand. Finally, growth and development can be said to depend on, and be constrained by, a number of different factors, such as saving, population growth, education, geography, entrepreneurship and international factors. Some visions and theories may stress some factors more than others, but the correlation between factors and theories is not perfect.

The purpose of this chapter is to provide a broad overview of general approaches to long-run growth and development. Its focus will be on alternative theories of growth and development in an aggregative macroeconomic closed-economy framework, considering in turn, classical-Marxian, neoclassical and aggregate demand-determined models and their relevance for stagnation in less-developed countries (LDCs). The discussion of these models will be used to shed light on different factors determining growth and stagnation which are reviewed in the final section.

Classical-Marxian growth theory

Early theories of growth, developed by the classical economists such as Adam Smith and David Ricardo and continuing in the approach of Marx, followed what can be called the surplus approach. According to this approach capitalists use their capital stock and hired labor to produce output. After workers are paid their wages, which are held down by population dynamics (with population increasing when the wage increases above its subsistence level) or by the existence of a reserve army of the unemployed, capitalists receive the surplus production as profits. Capitalists save out of their profits and invest aggressively in order to survive their competitive struggle with other capitalists, and this investment adds to the expansion of capital stock and hence to the growth of production. Assuming fixed labor-output and capital-output ratios given by b and a , with a fixed stock of capital given at K as a result of past investment, production and employment, respectively, are given by $Y = K/a$ and $L = bY$, total profit is $\Pi = Y - wL$, where w is the fixed real wage, so that total saving is given by $S = s_c \Pi$, where the fraction s_c is the exogenously-given saving rate of capitalists and S is aggregate saving, and where workers are assumed not to save. These assumptions imply that the rate of profit, r , and the rate of growth of capital, $g = (dK/dt)/K$ (assuming away depreciation, for simplicity), are given by:

$$r = (1 - wb)/a \quad (14.1)$$

and

$$g = s_c(1 - wb)/a \quad (14.2)$$

Since the output-capital ratio is fixed, total production also grows at rate g and, with a given rate of population growth, this determines the rate of growth of per capita income and production. Growth in this model is unconstrained by the availability of the supply of labor either because labor is always assumed to be unlimited supply, growing at some exogenously-fixed

rate $n \geq g$, or because labor supply growth is endogenous. If the growth of output and labor demand increases too rapidly, the adjustments in n due to endogenous labor supply (which can include immigration), and a due to endogenous technological change, can relax the labor supply constraint if and when it emerges. This model therefore views growth to be constrained by the availability of capital, which is accumulated through capitalist saving. An increase in the rate of growth is brought about by an increase in the saving rate of capitalists, s_c , by a reduction in the real wage, w (that is, a weakening in the bargaining power of workers),⁴ and by technological improvements (a fall in b and a rise in a).

This approach found resonance in early development economics. The idea that growth in LDCs is constrained by the availability of capital and requires increases in the rate of capital formation is found in early theories of development. For instance, it is found in the so-called Harrod–Domar equation which is $g = sa$, where s is the overall saving rate of the economy, which follows from equation (14.2), by suppressing income distributional considerations and noting that $s = s_c(1 - wb)$.⁵ Growth in LDCs is low because of low saving rates (the poor countries cannot save much, so they remain poor, as an often-discussed vicious circle story has it), and because of low productivity of capital due to inefficiency, inappropriate technology or backward technology. The policy implication for this approach is that the saving rate of the economy had to be expanded (often through government planning and intervention in the economy), and the capital-output ratio had to be reduced by increases in efficiency, the use of more labor-intensive techniques, or technological change.⁶ The approach is also closely related to Lewis's (1954) dual-economy model with surplus labor, which justifies the fixed real wage by the assumption of unlimited supplies of labor or disguised unemployed in a non-capitalist subsistence sector. Lewis also emphasized the importance of raising saving, as well as the problem of a reduction in profit and saving due to increases in w brought about by the disappearance of surplus labor, or by increases in the terms of trade of subsistence agricultural products which squeeze profits in the capitalist industrial sector because it requires the payment of a higher wage in terms of industrial goods (although this takes us beyond our one-sector framework). Lewis's model, however, was optimistic about growth prospects in LDCs because of the existence of surplus labor. Finally, neo-Marxist development economists like Baran (1957) analyzed the growth prospects of LDCs in terms of the surplus approach, arguing that domestic saving and investment is reduced by the dissipation of the surplus due to high consumption by LDC capitalists due to international demonstration effects, other unproductive spending and investment by them, and to the leakage of surplus abroad (in an open economy context). They typically drew a revolutionary implication

from this approach, arguing for the necessity of a socialist revolution to increase saving and investment and increase growth.

The implications of this capital-based approach have attracted a fair amount of criticism. Many LDCs which increased their saving significantly failed to record significant growth, experiencing, instead, increases in their capital–output ratio. It was argued that the capital-based approach, especially in its planning version, failed to take into account the importance of technological change, efficiency and incentives. From a different perspective it was argued that increases in saving did not necessarily promote investment if there were insufficient incentives due to the lack of aggregate demand (itself brought about by a rise in saving and a fall in consumption demand). The characteristic of the model which, however, was most responsible for its displacement was the assumption of surplus labor which, paradoxically, was an assumption which is appropriate for many labor-abundant LDCs. The neoclassical approach which displaced it, although mostly in applications to developed economies, assumed that labor is fully employed.

Old and new neoclassical growth theory

Solow's (1956) model, the paradigmatic neoclassical growth model, assumes that labor is fully employed, and also that capital and labor can be substituted in production in the smooth production function:

$$Y = F(K, L),$$

which exhibits diminishing returns to factors and constant returns to scale. The constant returns to scale assumption implies that the production function can be written in intensive form as:

$$y = f(k) \tag{14.3}$$

where $y = Y/L$ and $k = K/L$, per worker or per capita output and capital per worker. The wage and the rental adjust to maintain the full employment of labor and capital. A fraction s of total income is assumed to be saved (no distinction is made between saving from wages and profits) and automatically invested (as in the classical-Marxian model), so that $S = sY = I$. The fixed saving-rate assumption has been modified in subsequent models by the assumption of intertemporal substitution by consumers, either in an infinite horizon or overlapping–generations framework, without any fundamental changes in results about the determinants on long-run growth. Capital accumulates due to investment, so that $dK/dt = I$ (assuming away depreciation) and labor supply and employment (because labor is fully

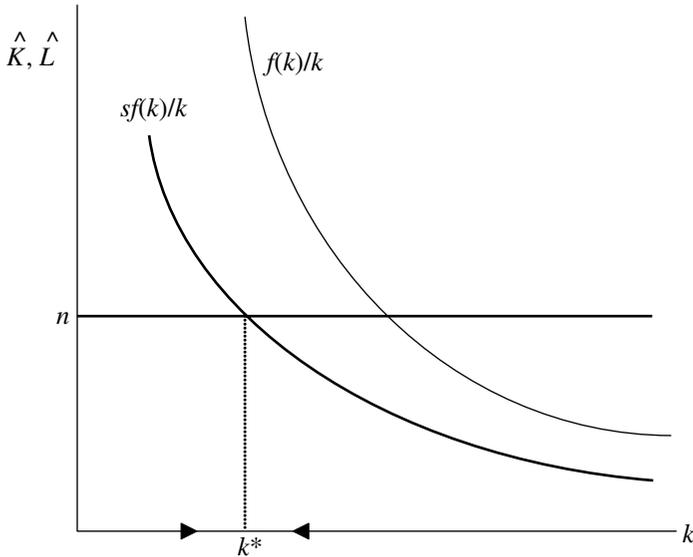


Figure 14.1 Neoclassical growth model

employed) grows at the exogenously fixed rate n . The dynamics of the model can be examined in terms of the rate of change of k . We have:

$$\hat{k} = \hat{K} - \hat{L} = I/K - n = sY/K - n = sy/k - n = sf(k)/k - n.$$

The dynamics are shown in Figure 14.1, where the $f(k)/k$ line which shows the average product of capital, is negatively sloped due to diminishing returns, and is assumed to have the two axes as asymptotes (obeying the so-called Inada conditions which require both capital and labor to be essential for production). The capital growth curve is shown by the line $sf(k)/k$. Starting from any initial level, k will change over time to the steady-state position shown by k^* , where capital grows at the same rate as labor supply, so that k is constant, so that per capita income, y , also becomes constant.

Growth can be accommodated in this model by measuring labor in efficiency units and assuming that the efficiency factor of labor, E , grows at a given rate, say λ . We can reinterpret y as output per worker in efficiency units, or Y/EL , and $k = K/EL$. At steady state, k and y will attain their equilibrium values, but output per worker, Y/L , will grow at the rate λ to keep y constant. The model implies that a rise in the saving rate, s , will shift up the $sf(k)/k$ in Figure 14.1 and increase the steady-state value of k , and hence y . There is thus a level effect on per capita output (for a given efficiency of

labor), but the steady-state growth rate of per capita output is unchanged at λ . Thus, in this model the rate of growth is given by the exogenously given rate of growth of labor's efficiency factor. In most extensions of the Solow model which endogenize the rate of technological change, for instance, due to learning by doing, or through research activity, the growth rate is still determined by exogenous parameters, and is not affected by the rates of saving and investment.⁷

Empirical studies, usually using cross-country data, found that long-run average growth is positively affected by the rates of saving and investment, apparently contradicting the implication of the Solow model that an increase in the saving rate does not increase the economy's steady-state growth rate. The model also implies that if two countries with different initial levels of income have identical parameters and technology, the richer country with a higher capital-labor ratio and per capita output initially will be closer to the common steady state for the countries, and grow at a slower rate. This seemed to contradict the empirical finding of divergence of income between rich and poor countries (although the model can be made consistent with this observation by assuming different parameter values for the two countries, for instance, a higher saving rate for the rich country). The discrepancies between the model and empirical result, as well as the unease with explaining growth in terms of exogenous technological factors, led to the development of new or endogenous growth theories by Romer (1986), Lucas (1988) and others.

This approach, which continues to follow the neoclassical tradition of assuming full employment, but departs from the assumption of diminishing returns to capital, the produced means of production, has now come to dominate growth theory. The essence of the approach can be shown with the simple AK model with the production function:

$$Y = AK$$

where A is the productivity of capital. This production function states that output increases proportionately with capital, without exhibiting diminishing returns, and is not affected by the amount of labor employed. Continuing with all of the other assumptions of the Solow model, we obtain:

$$\hat{k} = sA - n.$$

Assuming that $sA > n$, starting from any initial value of k , the economy grows at the rate given by $\hat{y} = \hat{k} = sA - n$. An increase in the rate of saving (and investment), s , implies a permanent increase in the rate of growth of

the economy. Since capital, the produced means of production, is not subject to diminishing returns, an increase in the rate of capital accumulation raises the rate of growth of the economy in the long run. In fact, the absence of diminishing returns to the produced means of production is not necessary for this result; all that is necessary is that there is a lower bound to its average product which exceeds the rate of growth of labor supply.

Several interpretations of the AK production function can be given, following key contributions to the new growth literature.⁸ Romer (1986) takes investment to increase not only capital in the usual sense of private capital goods, but also the stock of knowledge, which is a public good and which increases the efficiency factor of labor, the two effects together implying non-diminishing returns to capital. Lucas (1988) can be interpreted as referring broadly to investment so as to include both physical and human capital; the growth of human capital increases the efficiency of labor, implying non-diminishing returns to capital. Yet other models countered diminishing returns to produced means of production by allowing increasing product variety and increasing returns within firms which produce these products (as distinct from externalities and public goods). The models with this form of increasing returns have to introduce imperfect competition to limit the size of firms, and this market power allows the consideration of innovative activity driven by profit-maximizing producers of knowledge who obtain temporary profits from new products (see Aghion and Howitt, 1998). New innovations, in turn, increase the stock of knowledge on which future innovators can draw, but make obsolete products developed by earlier innovators.

New growth theory has apparently made the neoclassical model more consistent with the data. Growth depends on the rates of saving and investment, and the absence of diminishing returns gives poor countries no growth advantages over richer ones, and in fact, increasing returns to capital can do just the reverse. Moreover, it has the theoretical advantage of making the long-run rate of growth depend on the determinants of technological change, such as government policy, spending on technological change, and the patent system. However, both types of models assume that labor is fully employed along the growth path, which does not appear to reflect well the empirical reality of unemployed and underemployed labor in many LDCs. To this extent they may be in fact taking a step back from the classical approach.

There has been some criticism of the standard neoclassical growth theory without market imperfections and with the full employment of resources, but from those who remain neoclassical in the sense of using optimizing underpinnings of behavior but introducing market imperfections systematically into the analysis. Banerjee and Duflo (2005)

summarize evidence which suggests that there are many kinds of market imperfections in LDCs, including those in credit and insurance markets due to imperfect information, externalities, and incomplete contracts within and between generations. These distortions can result in across-the-board inefficiency, but also differences in efficiency across firms. Given such differences in efficiency across firms, Banerjee and Duflo (2005) argue that it is incorrect to use the aggregate production function (which assumes that efficiency differences across agents are removed by market forces), and advocate the use of disaggregated models which take into account various kinds of market imperfections.

Growth and aggregate demand

It may be supposed that the existence of surplus labor in LDCs – a feature accepted by most early development economists – made aggregate demand an important issue for them, but this was not the case. In fact, it was argued that the Keynesian approach was irrelevant for LDCs, which were characterized as subsistence economies, unlike capitalist economies in which hired labor is used for production for the purpose of making profits and where savers and investors were different individuals and institutions (see Rao, 1952 [1958]; Das-Gupta, 1954). Moreover, it was argued that supply constraints due to shortages of wage goods (consisting of agricultural products), capital goods, working capital and skilled labor, and foreign exchange and government controls, rather than demand constraints, limit production and growth in LDCs (see Rao, 1952 [1958]; Kalecki, 1976). Demand constraints were also argued to be irrelevant because of high population levels (which meant that there were many consumers) and low levels of consumption which left many consumption ‘needs’ unmet (Das-Gupta, 1954).⁹ Note that the irrelevance of aggregate demand was not related to labor shortages and full employment as is implicitly the case in neoclassical models.

In the 1970s and 1980s, partly perhaps as a result of changes in the structures of many LDCs which made many of them ‘semi-industrialized’ in the sense of having large capitalist sectors within subsistence economies, and partly because of the perceived failures of earlier development theories to deal with the problems of demand deficiencies, theories in which aggregate demand issues took center stage – drawing on the analysis of not only Keynes (1936), but also Marx (1867) and Kalecki’s (1971) work on advanced capitalist economies – emerged in the development literature (see, for instance, Rakshit, 1982, 1989; Taylor, 1983; Dutt, 1984). Some, in fact, argued that the fragmented nature of commodity markets and credit markets (which reduce the ability of credit markets and within-country trade in goods to overcome demand deficiencies), and the importance of

assets such as land, gold and other precious metals and food stocks which allow potential investors in LDCs to divert their assets to unproductive channels in the face of economic uncertainty, arguably make aggregate demand issues highly relevant for LDCs.

A simple model of growth determined by aggregate demand can be represented as follows. Planned investment as a ratio of capital stock depends positively on the rate of profit (which increases both the availability of finance and profit expectations, and possibly also the rate of capacity utilization, which shows the degree of buoyancy of markets), and saving, as in the classical Marxian approach, is a fraction s_c of profits while workers do not save, imply that $S/K = s_c r$. The saving and investment functions are shown in Figure 14.2 where the standard stability condition that saving adjusts more to changes in the profit rate than does investment, is assumed. We return to the assumption of fixed coefficients of production, since factor substitution does not have a major role to play in this approach. Labor is assumed to be unlimited supply. One interpretation of the model assumes that the economy fully utilizes its capital, and planned investment and saving are brought into equality through changes in the price level. Thus, if planned investment exceeds saving, there is an excess demand for goods which increases the price level, and with the money wage constant or at least not adjusting fully to the price change, r increases, taking the economy to saving–investment equality (Robinson, 1962). Price flexibility,

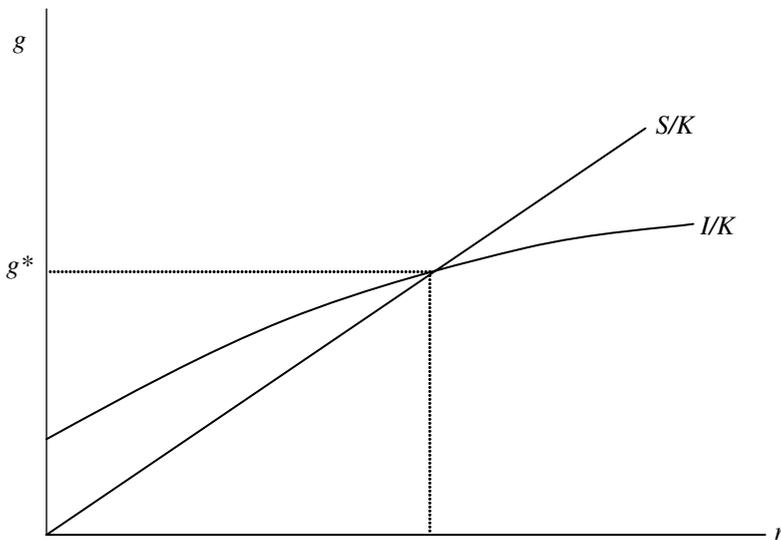


Figure 14.2 Growth model with aggregate demand

as assumed in this interpretation, has often been argued to be inconsistent with the oligopolistic structure of many LDC industrial sectors, which are sometimes seen as having excess capacity due to the lack of demand. This suggests a second interpretation of the model, in which firms set the price as a mark-up on wage costs, so that:

$$P = (1 + z) bW, \quad (14.4)$$

where z is exogenously given, dependent on the degree of industrial concentration and on the relative bargaining position of workers à la Kalecki, and where firms who normally maintain excess capacity and adjust their production according to demand, make investment plans depending on both the rate of profit and the rate of capacity utilization. When investment exceeds saving, firms increase output and raise capacity utilization, which increases the rate of profit, which is given by:

$$r = (z/(1 + z)) Y/K, \quad (14.5)$$

resulting in saving–investment equality and the clearing of the goods market. Both interpretations of the model imply that aggregate demand is the main force behind growth. If there is an increase in autonomous investment by firms, the rate of capital accumulation increases by increasing saving either by squeezing real wages and increasing forced saving, or by increasing output. Thus, business psychology and animal spirits have an important influence on growth. Increasing the saving rate, however, pushes up the saving function and reduces the rate of growth by depressing consumption demand. In the second interpretation of the model an increase in the mark-up, z , reduces the real wage as shown by equation (14.4), and also pushes the investment function downward because for a given profit rate a higher mark-up implies a lower rate of capacity utilization (see equation 14.5) and lower investment, and therefore reduces the rate of accumulation and growth. Thus, an improvement in income distribution raises the rate of growth, an effect to which we return below.

It should be noted that models of long-run development in LDCs which stress aggregate demand issues do not neglect all other constraints. Many of them prominently feature the agricultural sector (often with non-capitalist forms of production organization), and introduce foreign exchange constraints by introducing export functions and exogenously-given capital inflows, full capacity constraints, and fiscal constraints with upper limits to public sector borrowing limits (see Taylor, 1983, 1991). Sufficiently high levels of aggregate demand can theoretically make the economy ‘hit’ some of these constraints. But their incorporation does not

make aggregate demand issues irrelevant (see Dutt, 1997). First, many of the constraints may not usually bind, so that at most times aggregate demand factors will determine the growth path of the economy. Second, many of the constraints may be affected by changes in aggregate demand or factors that affect aggregate demand. For instance, changes in government investment expenditure will affect agricultural growth (as well as that of manufacturing) by affecting the amount of infrastructure, thereby easing the agricultural constraint. Moreover, changes in growth determined by aggregate demand can imply faster technological change which has a positive effect on exports and foreign capital inflows, both of which can ease the foreign exchange constraint. Third, the uncertainty generated by the instability of the growth process caused by the existence of a multiplicity of constraints may make aggregate demand more relevant for the growth process (Bagchi, 1988). What they do imply, however, is that it is not appropriate to relegate aggregate demand issues to the short run and to assume that short- and long-run behavior of the economy are unrelated. It is quite possible that sharply contractionary fiscal and monetary policies may reduce output in the short run and negatively affect investment, infrastructural investment and technological change, and slow down the long-run growth of the economy.

Low-level equilibrium traps

The theories discussed in the preceding sections all determine the rate of growth of the economy. They can be used to explain low levels, or the lack, of growth in countries in terms of some parameters of the model. For instance, in the new growth theory approach, countries will experience persistent underdevelopment or low rates of growth if they have a low saving rate, a low level of productivity and a high rate of growth of labor supply or population. If these parameters could be changed, the countries would be able to grow faster. The models help us to understand what parameters affect growth and development and how. However, they must be supplemented with some explanation of why the parameters take the values that they do. An alternative approach to understanding stagnation is to develop models which produce low-growth equilibria endogenously in the sense that the long-run growth rate of the economy – whether it is high or low – depends on the initial state of the economy. The approach therefore explains why economies with a similar structure can end up growing or stagnating, depending on where they start from and their structure. Models representing this approach are called models with low-level equilibrium traps or poverty traps.

These models can be interpreted as modifying the models discussed earlier, for instance, by endogenizing some of their parameters, or by

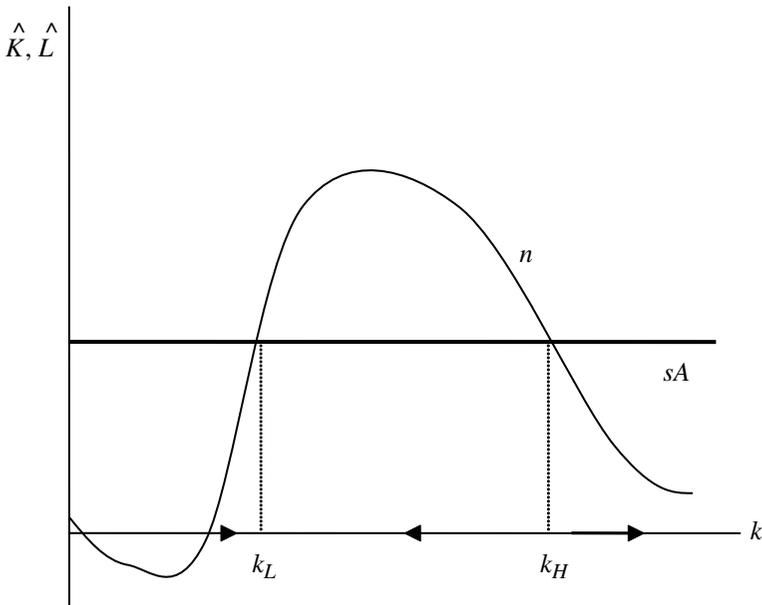


Figure 14.3 *Model with a low-level equilibrium trap*

introducing additional features into them. For instance, the new growth theory model can be modified to make the rate of growth of population depend on the level of per capita income. Under the assumption that the rate of population growth is low (because of high death rates) at low levels of income (and hence, capital-labor ratios), rises with increases in income (because death rates fall), and then falls (because birth rates fall with increases in income), the modified new growth theory model can be illustrated as in Figure 14.3. The low-level equilibrium trap is shown by k_L , while the critical minimum level of k is k_H . If the economy starts at an initial state of $k < k_H$, it will move towards the equilibrium at k_L , at which the rate of per capita income growth is zero: increases in growth increase population, and this serves to reduce the capital-labor ratio. However, if it starts above k_H , it will enjoy sustained growth because growth reduces the rate of population growth below the rate of capital formation. The economy can break out of the low-level trap if it can increase its saving rate significantly, pushing up the sA line sufficiently to make k_H lower than k . The model has the attractive property that any increase in the saving rate will not be able to achieve this result; a critical minimum increase is required.

Many other examples of neoclassical models – with similar multiple equilibria with a low-level equilibrium trap and a critical minimum level

above which the economy can experience sustained growth or approach a high-growth equilibrium – are available, capturing other relevant aspects of the development process. The presence of increasing returns at higher levels of k can make the capital growth line slope upwards beyond a certain scale at which strong external economies kick in and thereby produce a U-shaped capital growth curve (exhibiting initially diminishing and subsequently increasing returns to capital). With a constant n the model may then have multiple equilibria with a low-level trap. Increases in consumption from very low levels when income increases can also cause the capital growth curve to slope downwards for a stretch, followed by a positive slope due to rising saving rates. Other examples include credit and insurance market imperfections due to imperfect information, low productivity caused by low levels of nutrition and human capital, and institutional and organizational factors, such as corruption, incomplete property rights and kinship ties (see Azariadis and Stachurski, 2005). These mechanisms produce low-level poverty traps: if per capita income is initially below a critical minimum level the economy will converge to a poverty trap, while if the economy happens to attain a level beyond that critical minimum, it sets off into self-sustained growth (or to a higher equilibrium).

Models rooted in the classical or Keynesian tradition can also produce multiple equilibria and low-level equilibrium traps. Ros (2000) combines features of the classical approach of Lewis (1954) with increasing returns to show how the interaction of surplus labor and external economies can produce multiple equilibria and a low-level equilibrium trap. A simple extension of the model of growth and aggregate demand, which assumes that the investment function is non-linear, also implies multiple equilibria. For example, an investment function that is upward-rising but S-shaped, capturing the idea that at low levels of profit increases in profit have a small effect on desired investment but that as profits increase the response is stronger, can imply a low-level trap and a high growth equilibrium. The low growth equilibrium results from pessimistic self-fulfilling expectations.

Constraints on growth and development

The models discussed in the previous sections emphasize different determinants of, and constraints on, long-run growth and development. The determinants of growth emphasized in these models include saving rates and investment parameters, technological parameters, the rate of growth of labor supply and income distribution. The different models imply that these determinants need not affect economic growth in the same way in all economies. For instance, efforts to increase the saving rate will increase the rate of growth in economies that are saving-constrained and where diminishing returns to capital is not strong, but may have an adverse effect on

growth in demand-constrained economies, or no long-run effect with strong diminishing returns to capital in supply-constrained economies. To illustrate such issues further we discuss briefly the role of two determinants of growth – technological factors and income distribution – in some of the models discussed above.

New growth theory stresses various sources of technological change, including learning by doing, education and human capital formation, and profit-seeking innovative activity, and shows that faster technological change speeds up growth. A common implication of these models is that given the public goods nature of technological change (see Romer, 1986), there will be underinvestment in the creation of knowledge in the free market economy compared to what is socially optimal; the implication is that government intervention can raise the rate of growth by speeding up technological change. A few models, however, suggest that it is also possible for the free-enterprise economy to overinvest in research and development, because the private research firm does not take into account its business-stealing effect, that is, the fact it takes business away from firms which profited from an invention made obsolete by its invention (see Aghion and Howitt, 1998).

In demand-led models, technological change is often seen as being driven by demand growth. Aggregate demand growth, by speeding up economic growth, can increase the rate of productivity growth due to learning by doing or due to the fact that technological change, especially technological diffusion of a labor-saving variety, can be speeded up due to the emergence of labor scarcity. On the consequences of technological change this approach implies a more ambiguous role to technological change than in the neoclassical and new growth theory approaches. For instance, it is possible that faster technological change can reduce the rate of growth of labor demand, and thereby increase unemployment, rather than increase growth, and higher unemployment may even have the effect of reducing the share of wages in income and reducing aggregate demand and growth. On the other hand, technological change can have the effect of increasing investment demand because firms install new machinery to take advantage of newer production methods, and also increasing consumption demand due to product innovation, thereby increasing aggregate demand and economic growth. Thus, it is possible for there to be a two-sided synergistic relation between capital accumulation and technological change.

Technological change played a central role in Schumpeter's (1911 [1934]) theory of economic growth and development. In this theory, technological progress, involving – for instance – new production techniques, new products, new managerial methods and new sources of supply of inputs, occurred because of the efforts of profit-seeking entrepreneurs who created

new technology, who earned profits from temporary monopoly power, and whose creations would be destroyed by new innovations in the future. Schumpeter's ideas have had an influence on both new growth theory (see Aghion and Howitt, 1998) and demand-based approaches, and have also spawned other evolutionary approaches to growth (see Nelson and Winter, 1982).

It may be argued that these approaches to technological change have limited applicability to LDCs in which technological change is more a matter of importing technology than creating it. However, since the distinction between technology creation, on the one hand, and diffusion and transfer, on the other, is not as great as thought earlier (see Bell and Pavitt, 1993), and since the effects of technological change raise similar issues in advanced countries and LDCs, they have great relevance for LDCs.

On the role of income distribution, it may be supposed that since saving positively affects economic growth in new growth models, faster growth would result from increasing inequality, since the rich do most of the saving. If there are investment indivisibilities – for instance, large set-up costs for new industries – and if there are credit market imperfections, concentration of wealth among a few can, indeed, increase investment and the rate of growth. However, a number of models suggest that a more equal income distribution can increase growth (see Aghion and Howitt, 1998). For instance, if there are no credit markets (or if these markets are imperfect), investors with higher endowments of wealth will invest more than those with less. If there are diminishing returns to capital, a misallocation of capital and a low level of aggregate output will result, which can result in a lower rate of technological change and growth (if aggregate output affects technological change due to learning by doing). Other models, such as those which consider cooperation between the rich and poor and in which inequality leads to less cooperation, shirking and free-riding by the poor, or in which inequality leads median voters to push for growth-reducing redistribution, produce similar results.

Income distribution affects growth in classical-Marxian models because of the assumption that the propensity to save out of profit is greater than that to save out of wage income. In the classical-Marxian model a rise in the profit share increases saving and capital accumulation, and therefore the rate of growth; thus inequality helps growth. However, in demand-led growth models, a rise in the share of profits due to, for instance, a rise in the price-cost mark-up charged by firms, has the effect of making the distribution of income less equal and reducing the rate of growth of capital stock and output. This occurs because the shift in income distribution towards profits increases saving, thereby reducing consumption and aggregate demand, reducing capacity utilization and hence the rate of investment and

growth (see Dutt, 1984, 1990). Such models, however, need not necessarily have this implication. For instance, if investment depends both on capacity utilization and the profit mark-up (since both are likely to affect profit expectations), a rise in the mark-up can increase both the profit share and the rate of investment and growth because of the direct effect of the mark-up on investment (Bhaduri and Marglin, 1990).

Space limitations prevent me from discussing other determinants of growth and development. In conclusion, I note two types of omissions. First, I have focused on models and on economic constraints which are often endogenous to the process of growth and development and hence sometimes called proximate determinants, and not discussed what some have called fundamental factors. Included in the list of such factors are institutional and geographical variables. Although this distinction is usually made in the context of the estimation of growth equations to avoid the problem of simultaneity, it is also relevant for the understanding of growth processes and for formulating appropriate policies for growth and development. For instance, if saving and investment rates – proximate determinants in some of the models – need to be increased to increase growth, it may be necessary to change institutions which strengthen property rights or cultural norms among entrepreneurs which encourage long-term investment that may be considered more fundamental factors. We have not examined institutional and cultural features here.¹⁰ Second, the analysis has been conducted for the most part using a one-sector, closed-economy framework. This implies that I have not examined growth problems related to sectoral issues, such as agricultural and environmental constraints, or open-economy considerations, such as foreign exchange constraints and the relation between trade, technological change and growth.¹¹ The models discussed here, however, can be suitably modified to examine these omitted factors.

Notes

1. I am grateful to Jaime Ros and Lance Taylor for their helpful comments on earlier drafts of this chapter.
2. Some would argue that development involves much more than the goal of economic growth as measured by the level of real per capita income, such as improvements in income distribution, and the fulfillment of the basic needs and capabilities of people. Some would also argue that development involves other changes in the economy, such as those in the sectoral composition of output and trade.
3. For instance, Chakravarty (1980) distinguishes between the Mill–Marshall and Marx–Schumpeter traditions, which – from today’s perspective – can be called the orthodox and heterodox visions. See Taylor (2004, Chapter 11) for a more detailed historical discussion of alternative views.
4. Although the classical economists, especially Marx, often assumed the real wage to be given by subsistence requirements, they interpreted subsistence broadly to include historical, moral and political factors, which makes it possible for the real wage to be changed by the state of what Marx called the class struggle.

5. This model is closer to that of Domar (1946) than to that of Harrod (1939), who assumed an independent investment function in the form of an accelerator, thereby bringing in effective demand into the model, an issue which is discussed below. In this model, all saving is automatically invested.
6. Higher labor intensity implies that b rises when a falls. It is implicitly being assumed that the rise in b does not offset the growth enhancing effect of a fall in a which, as can be seen from equation (14.1) implies that capitalist firms only adopt technological changes if their rate of profit rises. Or it is assumed that there is no difference in the saving rate of workers and capitalists, so that a rise in the wage share does not reduce the overall saving rate, as assumed in this model.
7. Some extensions of the Solow model in which the share of saving allocated to education and research and development affects the rate of technological change allow long-run growth to be endogenously determined. See, for instance, Uzawa (1965).
8. For a fuller discussion, see Chapter 15 in this volume on new growth theory.
9. These arguments are, of course, erroneous because they ignore effective demand issues.
10. There is no presumption that these factors are necessarily more fundamental than other more narrowly economic factors, and may in fact respond to changes in the level of development relatively quickly. They are discussed elsewhere in this *Handbook*, especially in Chapter 61 on institutions and Chapter 62 on culture.
11. These issues are discussed elsewhere in this *Handbook*, especially in Chapter 17 on sectoral interactions, Chapter 18 on open economy issues, Chapter 28 on the environment and Chapter 36 on trade.

References

- Aghion, Philippe and Peter Howitt (1998), *Endogenous Growth Theory*, Cambridge, MA: MIT Press.
- Azariadis, Costas and John Stachurski (2005), 'Poverty Traps', in P. Aghion and S. Durlauf (eds), *Handbook of Economic Growth*, Vol. 1A, Amsterdam: Elsevier, pp. 295–384.
- Bagchi, Amiya K. (1988), 'Problems of Effective Demand and Contradictions of Planning in India', in A.K. Bagchi (ed.), *The Economy, Society and Polity. Essays in the Political Economy of Indian Planning*, Calcutta: Oxford University Press, pp. 227–66
- Banerjee, Abhijit and Esther Duflo (2005), 'Growth Theory through the Lens of Development Economics', in P. Aghion and S. Durlauf (eds), *Handbook of Economic Growth*, Vol. 1A, Amsterdam: Elsevier, pp. 473–552.
- Baran, Paul (1957), *The Political Economy of Growth*, New York: Monthly Review Press.
- Bell, Martin and Keith Pavitt (1993), 'Accumulating Technological Capability in Developing Countries', *Proceedings of the World Bank Annual Conference on Development Economics*, Washington, DC: World Bank, pp. 257–81.
- Bhaduri, Amit and Stephen A. Marglin (1990), 'Unemployment and the real wage: the economic basis of contesting political ideologies', *Cambridge Journal of Economics*, 14 (4): 375–93.
- Chakravarty, Sukhamoy (1980), *Alternative Approaches to the Theory of Economic Growth*, Delhi: Orient Longman.
- Das-Gupta, A.K. (1954), 'Keynesian Economics and Under-Developed Countries', *Economic Weekly*, 26 January; reprinted in A.K. Das-Gupta, *Planning and Economic Growth*, London: George Allen & Unwin.
- Domar, Evsey D. (1946), 'Capital Expansion, Rate of Growth, and Employment', *Econometrica*, 14: 137–47.
- Dutt, Amitava Krishna (1984), 'Stagnation, Income Distribution and Monopoly Power', *Cambridge Journal of Economics*, 8 (1): 25–40.
- Dutt, Amitava Krishna (1990), *Growth, Distribution and Uneven Development*, Cambridge: Cambridge University Press.
- Dutt, Amitava Krishna (1997), 'The Relevance of Keynesian Policies in Semi-Industrialized Economies: Theoretical Issues and an Empirical Illustration', in Philip Arestis and

- Malcolm Sawyer (eds), *The Relevance of Keynesian Economic Policies Today*, London and New York: Macmillan and St Martin's Press, pp. 163–85.
- Harrod, Roy F. (1939), 'An Essay in Dynamic Theory', *Economic Journal*, 49: 14–33.
- Kalecki, Michal (1971), *Selected Essays on the Dynamics of the Capitalist Economy*, Cambridge: Cambridge University Press.
- Kalecki, Michal (1976), *Essays on Development Economics*, Brighton: Harvester.
- Keynes, John Maynard (1936), *The General Theory of Employment, Interest and Money*, London: Macmillan.
- Lewis, W. Arthur (1954), 'Economic Development with Unlimited Supplies of Labour', *The Manchester School*, May, 22: 139–92.
- Lucas, Robert E. (1988), 'On the Mechanics of Economic Development', *Journal of Monetary Economics*, 22: 3–42.
- Marx, Karl (1867), *Das Kapital, Band 1*, translated (1976), *Capital*, Vol I, Harmondsworth: Penguin.
- Nelson, Richard R. and Sidney G. Winter (1982), *An Evolutionary Theory of Economic Change*, Cambridge, MA: Harvard University Press.
- Rakshit, Mihir K. (1982), *The Labour Surplus Economy*, Delhi: Macmillan.
- Rakshit, Mihir K. (ed.) (1989), *Studies in the Macroeconomics of Developing Countries*, Delhi: Oxford University Press.
- Rao, V.K.R.V. (1952), 'Investment Income and the Multiplier in an Underdeveloped Economy', *Indian Economic Review*, February; reprinted in A.N. Agarwala and S.P. Singh (eds) (1958), *The Economics of Underdevelopment*, Bombay: Oxford University Press, pp. 205–18.
- Robinson, Joan (1962), *Essays in the Theory of Economic Growth*, London: Macmillan.
- Romer, Paul M. (1986) 'Increasing Returns and Long-Run Growth', *Journal of Political Economy*, 94: 1102–37.
- Ros, Jaime (2000), *Development Theory and the Economics of Growth*, Ann Arbor, MI: University of Michigan Press.
- Schumpeter, Joseph (1911), *The Theory of Economic Development: An Inquiry into Profits, Capital, Credit, Interest and the Business Cycle*, English trans. (1934), Cambridge, MA: Harvard University Press.
- Solow, Robert M. (1956), 'A contribution to the Theory of Economic Growth', *Quarterly Journal of Economics*, 70: 65–94.
- Taylor, Lance (1983), *Structuralist Macroeconomics*, New York: Basic Books.
- Taylor, Lance (1991), *Income Distribution, Inflation and Growth*, Cambridge, MA: MIT Press.
- Taylor, Lance (2004), *Reconstructing Macroeconomics: Structuralist Proposals and Critiques of the Mainstream*, Cambridge, MA: MIT Press.
- Uzawa, Hirofumi (1965), 'Optimum Technical Change in an Aggregative Model of Growth', *International Economic Review*, 6 (1): 18–31.

15 New growth theory and development economics

Heinz D. Kurz and Neri Salvadori

Introduction

After a period of relative calm in the aftermath of the so-called Cambridge–Cambridge controversies in the theory of capital (for an overview, see Kurz and Salvadori, 1995, Chapter 14) which had dealt a serious blow to the long-period version of neoclassical theory in general and the Solovian growth model based on an aggregate production function in particular (Solow, 1956), since the mid-1980s, growth economics has become again one of the most vibrant areas of research. The revival was spurred by theoretical and empirical contributions. New modelling techniques imported from other areas in economics were used in order to ‘endogenize’ technological progress within a macroeconomic intertemporal general equilibrium framework and thus overcome a major shortcoming of Solow’s model – the treatment of technological change as exogenous. This is also the reason why the new class of models are frequently dubbed ‘endogenous’. The construction of new data sets for a large number of countries, in particular the Penn World Table (Summers and Heston, 1991; Heston et al., 2002), and Maddison (2001), has led to a revived interest in empirical studies which in turn have thrown up new problems for growth theory. The existing literature is huge and still rapidly growing. For summary accounts of the present state of the art in this area of research see, in particular, Aghion and Howitt (1998), Jones (2002), Barro and Sala-i-Martin (2003), Helpman (2004) and Aghion and Durlauf (2005).

While neoclassical growth theory had first to overcome the straitjacket of the exogeneity of the long-term growth rate in Solow, in alternative approaches this was not necessary. There the growth rate has always been considered as endogenous, shaped by the behaviour of agents, the distribution of income, social institutions and so on. The very stress classical authors in the tradition of Adam Smith laid on the unintended social consequences of purposeful activities of individuals or groups of people acting within a highly complex system of an ever deeper division of social labour, characterized by scarce natural resources, technical innovations and changing social relations, is incompatible with the idea of an exogenously given

growth rate. This impression is already corroborated when we look at what may be called their linear growth 'models' (see Kurz and Salvadori, 1998, 1999, 2006; Salvadori, 2003). Marx's scheme of extended reproduction constitutes a two-sector, the von Neumann model an n -sector and the so-called Harrod–Domar model a one-sector model of endogenous growth. An important difference with the so-called new growth models is that the latter are intensive models, concerned with explaining the growth of income per capita, not extensive ones. However, intensive models are also encountered in these non-neoclassical traditions, be they classical, Marxist, Keynesian or evolutionary. For summary accounts of this kind of literature that is surprisingly largely ignored in Aghion and Durlauf (2005), see Dutt (1990), Foley and Michl (1999), Salvadori (2003), Nelson (2005) and Bhaduri (2007).

Since sustained growth in income per capita is arguably the most important determinant of living standards and since other measures of living standards, such as life expectancy and the Human Development Index (HDI), typically, though not always, move together with income per capita, to understand the causes of economic growth is of direct relevance to development economics.

There is also the following phenomenon that emphasizes the close connection between economic growth and development. For thousands of years incomes per capita, measured in some broad way, were not all that different throughout the world, and whenever they rose somewhat in some areas due to technological innovations this rise discharged itself first and foremost in a Malthusian way on a growing population which, in turn, due to diminishing returns to land, tended to annihilate the increase in income per capita. Hence, while there were lasting increases in population in some areas of the world, there were hardly any in income per capita. It was only with the Industrial Revolution that sustained growth of GDP per capita became a normal fact of life in Europe and the Western offshoots, with large parts of the rest of the world at first remaining stagnant. This led to what the historian Kenneth Pomeranz (2000) called the 'Great Divergence', the spreading out especially in living standards. Seen from a long-term historical perspective, the problems of economic growth and those of development are thus two sides of a single coin, with the latter being perceived as a problem only after the former had made an appearance which by all historical standards can only be called impressive.

We deal first with neoclassical growth models, followed by a brief discussion of approaches trying to come to grips with the 'fundamental' causes of growth and development, in which the role of economic, political and social institutions is emphasized. We conclude with a discussion of Keynesian and evolutionary contributions.

Neoclassical theory

At the forefront of research in contemporary growth economics are the proximate and fundamental causes of technological progress, which, together with human capital formation, is seen as the prime mover of the system and a main source of rising living standards. The main reference point of this literature is still the Solow model, and while the model at first has been rejected by major advocates of new growth theory, it recently had a comeback in an augmented form. Interestingly, it was not rejected because of a lack of solid micro foundations of its technology and production side. This is somewhat surprising because one criticism levelled at it was that its treatment of savings behaviour lacked such foundations. Yet replacing a Keynesian savings function by the assumption of an immortal representative agent maximizing his or her intertemporal utility can hardly be said to meet the criterion. The main reasons for its rejection were rather the following: (1) it takes as given the behaviour of the variable meant to depict the main driving force of growth: technological knowledge; (2) it implies that decisions of agents to save more or less have no impact on the steady-state rate of growth; (3) it has little to offer in terms of policy advice for long-run growth; and (4) it predicts the convergence of levels of income per capita on a world scale. According to the logic of the model, countries that exhibit similar structural parameters (savings rate, population growth) should in the long run have similar levels of income per capita. This is brought about by poor countries with a lower capital-to-labour ratio growing faster than rich countries. Alas, the propositions of the model as to convergence have not generally been confirmed empirically. While there are 'convergence clubs' (William Baumol), for example the Organisation for Economic Co-operation and Development (OECD) economies, there is no universal catching up of the less-developed countries. In addition to remarkable success stories (for example the so-called Asian Tiger economies) there are a number of disaster stories (for example sub-Saharan Africa). Cross-country studies indicate by and large a strong negative correlation between population growth and income per capita growth and a strong correlation between the latter and the share of investment (alias savings) in gross domestic product (GDP). Another implication of Solow's model has not been corroborated by empirical studies, namely, that poor countries exhibit higher rates of return, because capital is relatively more scarce. If this were to be true, one would expect massive flows of capital from rich to poor countries, which, however, have not been observed across poor countries as a whole. The implicit assumption of Solow's model that advances in technical knowledge are both a free and a public good, cannot be sustained. There exist, and possibly persist, technological or 'idea gaps' (Paul Romer) between developed and developing countries. However, a

poor economy may benefit in terms of economic growth from its relative backwardness, and there is some evidence that its openness helps in this regard. Economies that follow an isolationist strategy (for example North Korea) are cut off from the flow of ideas and in the medium term suffer stagnation or even negative growth.

Let us now have a closer look at some of the new growth models. A central feature of them is that they abandon the conventional marginalist assumption of diminishing returns to capital accumulation. This is done by broadening the concept of capital to encompass physical capital, human capital and 'ideas' or 'knowledge', and by invoking positive externalities with regard to the accumulation of capital. The first generation of models attempted to integrate a range of growth mechanisms in a neoclassical macroeconomic framework. The most important mechanisms concern the creation of new technical knowledge in research and development (R&D) departments of firms (Romer, 1986) and the formation of human capital in education processes (Lucas, 1988). These two mechanisms swiftly got accepted as the main engines of growth. Both mechanisms rely on positive externalities which counteract any tendency of the marginal product of capital (and thus the rate of profit) to fall. Romer stipulated a 'research technology' that is concave and homogeneous of degree one:

$$\dot{k}_i = G(I_i, k_i)$$

where I_i is an amount of forgone consumption in research by firm i and k_i is the firm's current stock of knowledge. The production function of the consumption good relative to firm i is:

$$Y_i = F(k_i, K, \mathbf{x}_i)$$

where K is the accumulated 'stock of knowledge' in the economy as a whole and \mathbf{x}_i is the vector of all inputs different from knowledge. The function is taken to be homogeneous of degree one in k_i and \mathbf{x}_i and homogeneous of a degree greater than one in k_i and K . Romer assumes that factors other than knowledge are in fixed supply. This implies that knowledge is the only capital good utilized in the production of the consumption good. Spillovers from private research and development activities increase the public stock of knowledge K . A positive externality is taken to be responsible for per capita income growth. Different from the Solow model, agents via their behaviour do have an impact on the long-term growth rate.

Similarly in the model of human capital formation by Lucas (1988) in which agents are assumed to have a choice between two ways of spending their (non-leisure) time: to contribute to current production or to

accumulate human capital. Lucas's conceptualization of the process by means of which human capital is built up is the following:

$$\dot{h} = \nu h(1 - u)$$

where ν is a positive constant. With the accumulation of human capital there is said to be associated an externality: the more human capital society as a whole has accumulated, the more productive each single member will be. This is reflected in the following macroeconomic production function:

$$Y = AK^\beta(uhN)^{1-\beta}h^{*\gamma}$$

where the labour input consists of the number of workers, N , times the fraction of time spent working, u , times h which gives the labour input in efficiency units. Finally, there is the term h^* . This is designed to represent the externality. The single agent takes h^* as a parameter in his or her optimizing by choice of c and u . However, for society as a whole the accumulation of human capital increases output both directly and indirectly, that is, through the externality.

In both models we are confronted with a variant of a public good problem: the individual optimizing agent faces constant returns to scale, yet for society as a whole returns are taken to be increasing.

The lessons to be drawn for developing countries are straightforward: it is not so much a lack of physical capital relative to population (as in the so-called 'Harrod-Domar' model) or relative to the labour force (as in the Solow model) that accounts for low levels of income per capita, but a lack of human capital and technical knowledge. In order to catch up, a less-developed country is well advised to invest in its education system and infrastructure and to try to get closer to the frontier of technological knowledge by providing incentives to domestic firms to imitate and innovate and by encouraging foreign direct investments of technologically advanced firms.

Given the stress laid on knowledge and human capital in these models, it comes as a surprise that hardly any attempt was made to clarify whether and how these magnitudes can be measured. Obviously, if and only if they are cardinally measurable can anything be said about returns to scale, marginal and average products, growth rates, and so on (see Kurz, 1997 [2003]; Steedman, 2003). One aspect of the problem can be highlighted with reference to Lucas's model, who for simplicity assumed that all workers are possessed of the same amount of human capital. Yet if this is the case, wherein could the externality consist? Similarly in Romer's model, there is the problem of multiple counting of the same particles of knowledge in

building up an aggregate measure of the social stock of knowledge. In short, while the models are suggestive, they lack conceptual clarity.

The simplest and for a while most popular new growth model was the linear or AK model according to which:

$$Y = C^\alpha H^\beta = AK$$

where K represents a broad measure of aggregate capital, consisting of physical capital C and human capital H , and A is a given and constant productivity parameter. The net rate of profit is exogenously given and equals $A - \delta$, where δ is the overall rate of depreciation. In long-run competitive equilibrium the two kinds of capital receive the same rate of return. As capital accumulates, total output expands proportionally, and with a constant population income per capita grows with the same rate as total output. In this model savings (alias investment) assume centre stage. The higher the savings (alias investment) rate, the higher is the growth rate of income per capita. The lesson to be drawn from an economic policy point of view is simple: The process of development is speeded up by whichever policy leads to an increase in savings in physical and human capital.

There is a close similarity between this model and the Harrod–Domar model, because in both models diminishing returns to capital are absent by construction. The important difference is that whereas the Harrod–Domar model assumes a given and constant input proportion of labour and capital and thus allows for labour unemployment, the AK model has effectively replaced the concept of labour by that of human capital. What in the Solow model (as well as in the Harrod–Domar model) was a non-accumulable factor of production, labour, has now become an accumulable one. While Solow had subsumed land (and, more generally, scarce natural resources) under capital, the AK model can be said to have gone to the extreme by also subsuming labour under it. With only a single accumulable factor contemplated, the possibility of perpetual growth should come as no surprise (see Kurz and Salvadori, 1998). Of William Petty’s famous 1662 dictum ‘that Labour is the father and active principle of Wealth, as Lands are the Mother’, nothing is left in this class of models: the parent of wealth is Capital.

The class of ‘horizontal innovation models’ was started by Romer (1990) who combines: (1) the endogenous production of new ‘industrial designs’ as in Romer (1986); with (2) the formalization of the role of human capital in economic growth as in Lucas (1988); and (3) a product-diversity specification of physical capital which he derives from the model of monopolistic competition with regard to consumption goods of Dixit and

Stiglitz (1977). Romer's argument relies on three premises: (1) technological progress is the prime mover of economic development; (2) such progress is in large part the result of deliberate actions of agents responding to market signals; (3) technical instructions of how to use raw materials and other inputs are fundamentally different from other goods. While the development of an 'industrial design' or of some other economically useful knowledge incurs costs, once this knowledge is available it can be used time and again without generating significant further costs. Hence the cost under consideration is a kind of fixed cost. Differently from human capital, industrial designs are said not to need to be embodied, and differently from ordinary goods they are non-rival and also only partially excludable. 'Growth is driven fundamentally by the accumulation of a partially excludable, nonrival input' (Romer, 1990, p. S74). Knowledge per capita can be accumulated without limit, and because of its incomplete excludability there will be spillovers which drive the process of growth. The presence of non-rival inputs of necessity involves non-convexities. Concavities in techniques are disruptive of the received concept of equilibrium and imply a more 'open-ended' vision of economic development and growth. As Romer's paper shows, it takes some considerable effort to tame the model and subdue it again to the equilibrium method.

Economic development is typically bound up with an expanding variety of intermediate products. Romer tries to capture this fact in terms of the following formalization. The final product is taken to be produced according to the production function:

$$Y(H_Y, L, x) = H_Y^\alpha L^\beta \sum_{i=1}^{\infty} x_i^{1-\alpha-\beta}$$

where H_Y denotes the amount of human capital employed in the final output sector, L the number of workers, and $\sum_{i=1}^{\infty} x_i^{1-\alpha-\beta}$ the employment of intermediate products. Since at a given moment in time there is only a finite number Z of them, $x_i = 0$ for all $i > Z$. Final output is thus seen as an additive separable function of the various intermediate products. It is an ever-growing number of (patented) designs, Z , that propels economic expansion and makes income per capita grow. On the one hand an increasing Z increases the productivity of labour and human capital in the production of final output. On the other hand the non-excludability of knowledge in research involves a direct positive relation between the total stock of designs and knowledge, Z , and the productivity of human capital employed in research. The research technology Romer postulates has the rate of increase of Z as a linear function of Z (and of human capital employed in the research sector). This implies that 'unbounded growth is more like an assumption than a result of the model' (Romer, 1990, p. S84).

Romer's model has a number of disquieting features. We have already pointed out that knowledge and human capital must boldly be assumed to be cardinally measurable. The postulated heterogeneity of intermediate products is more apparent than real. Since they all exhibit the same input proportions in production, they cannot be distinguished and all represent unspecifically given amounts of forgone consumption. Once introduced, an intermediate product will be used forever: it neither depreciates nor becomes economically obsolete. New economic knowledge is never the enemy of existing knowledge. If, say, final output happens to be wheat, then wheat in ancient Egypt was produced by means of digging sticks only, whereas today it is taken to be produced by means of digging sticks and ploughs and oxen and tractors and combine harvesters and so on, all employed simultaneously. This is certainly an extreme conceptualization of a growing capital input diversity which is squarely contradicted by an even casual observation of facts. New capital goods frequently replace old ones in a similar way as new particles of knowledge supersede their ancestors.

This latter fact is taken into account in so-called Schumpeterian growth models with quality-improving innovations championed by Aghion and Howitt (1998). These models revolve around Schumpeter's concept of 'creative destruction'. Accordingly, new technical devices are hardly ever a general good. While innovators may benefit, those that have invested their capital in previous vintages of technical knowledge suffer from them. The double-edged character of innovations raises the problem of the socially optimal rate of technological advancement.

Some broader issues regarding the fundamental causes of growth

Schumpeter, when talking about innovations, the main force of economic development, referred to 'new combinations'. The idea that new knowledge of whichever sort consists of the (re)combination of given pieces of knowledge can be traced far back in natural philosophy. It was referred to by Adam Smith who in a famous passage in *The Wealth of Nations* discussed the combination of existing ideas in order to create new ones in the context of the emergence of a new profession in an ever deeper division of labour, that is:

philosophers or men of speculation, whose trade it is, not to do anything, but to observe everything; and who, upon that account, are often capable of *combining together the powers of the most distant and dissimilar objects*. In the progress of society, philosophy or speculation becomes, like every other employment, the principal or sole trade and occupation of a particular class of citizens. (Smith, 1976 [1776], I.i.9; emphasis added)

Weitzman (1998), in an attempt 'to get inside the black box of innovation' and build up an explicit model of knowledge production, gave the

combinatoric metaphor a more precise form. New ideas, he maintained against Romer and others, are not some exogenously determined function of ‘research effort’ in ‘the spirit of a humdrum conventional relationship between inputs and outputs’. Rather, ‘when research effort is applied, new ideas arise out of existing ideas in some kind of cumulative interactive process’ (p. 332). What happens may be compared to the activities in an agricultural research station in which pairs of existing ‘idea-cultivars’ are combined to bring about new ‘hybrid ideas’ (where the word ‘cultivar’ is an acronym for cultivated variety). With I as the number of idea-cultivars, the corresponding number of different binary combinations that can be got from I is $C_2(I)$, which is given by:

$$C_2(I) = \frac{I!}{(I-2)!2!}$$

For example, with $I = 5$, we have $C_2(5) = 10$; and with $I = 6$, we have $C_2(6) = 15$, and so on.

The important message of Weitzman’s otherwise rather mechanistic argument is that the growth in the number of ideas that results from combining reconfigured existing ideas is remarkable and well exceeds exponential growth. If the entire potential of recombinatory possibilities could always be exploited, then the growth of the number of knowledge particles would over time increase almost without limit. Yet the capacity to process new ideas depends on the resources devoted to the task and the productivity of these resources. According to Weitzman it is sensible to assume that the ultimate constraint on economic expansion is linear. This implies that steady-state growth rates are linearly proportional to aggregate savings – not unlike the situation in the Harrod–Domar model.

Reaching beyond the confines of economics, narrowly defined, there have been studies focusing on what are frequently called the ‘fundamental’ causes of growth. The emphasis is on the role of economic and political institutions; see, for example, Alesina and Rodrik (1994). These studies typically proceed by superimposing upon one of the endogenous growth models some mechanism designed to capture the interaction of social, political and economic institutions and the role of income inequality. Factors such as property rights and their enforcement, ‘social capability’ (Moses Abramovitz), the quality of governance, corruption, religion and so on, are investigated with respect to their impact on the growth and development performance of countries. Economic institutions decide the incentives of economic agents and the constraints they face. Because different groups of society benefit from different economic institutions, there is typically a conflict over the alternatives. Political power which depends on political institutions and the distribution of resources decides the outcome

of the conflict. Political power may, however, be eroded by socio-economic developments. Thus a picture emerges showing the complex dynamics of the interaction of economic, institutional, political and social forces.

From the point of view of development economics, of particular relevance are studies of self-reinforcing mechanisms that perpetuate poverty; see, for example, Azariadis (1996). These mechanisms typically involve an adverse impact on the formation of human and physical capital and the adoption of best-practice technology. They can be traced back to market, institutional and political failures. There is strong empirical evidence that weak law and contract enforcement, an insufficient protection of property rights, confiscatory taxation and a corrupt bureaucracy act as disincentives to enterprise and capital accumulation and entail unproductive rent-seeking behaviour. Despotism is considered incompatible with sustainable economic development, while democracy and good political institutions can be expected to foster it. Then there are studies that investigate the role of largely exogenous factors on growth, such as the geography of a country (natural resources, climate, topography), its 'culture' and ethnic diversity. Here, finally, the received division of labour amongst the (social) sciences is somewhat overcome and economics, sociology, political science and history are brought together in order to come to grips with inherently intricate problems.

Such multidisciplinary approaches are indeed badly needed as the following case also shows. However, the case testifies to some economists' insistence on what they consider 'rigorous' explanations, that is, the explanation of some grand historical fact in terms of an utterly simple model in which some utility-maximizing agents faced with changing budget constraints shape centuries or, as in the present case, even millennia. What is dubbed 'unified growth theory' purports to 'capture the complexity of the process of growth and development over the entire course of human history' (see Aghion and Durlauf, 2005, Chapter 4, p. 174). The approach focuses attention on the take-off from what is called an epoch of stable Malthusian stagnation to a Post-Malthusian Regime of persistent growth in income per capita which occurred in Europe and the Western offshoots at the beginning of the nineteenth century. The main reasoning is simple and goes something like this. Any rise in income per capita above subsistence levels due to small technological improvements in the Malthusian period was swiftly followed by an expansion of population that made income per capita fall again to around its former level. (One wonders how, for example, the Roman Empire could ever get off the ground and where the surplus product it needed in order to do so came from.) It was only the acceleration in technological progress during the Post-Malthusian Regime that broke the vicious circle. The increase in income per capita was no

longer fully channelled into an increase in population, but stimulated the accumulation of human capital in the form of literacy rates, schooling and health. Then, during the second phase of the Industrial Revolution, industrial development was based more and more on increased skills of workers. The human capital formation in turn brought about a demographic transition and paved the way to an era of sustained economic growth. The latter is characterized by a significant increase in the average growth rate of output per capita and a decline in population growth. This broad development is then explained in terms of a model where utility-maximizing fertility behaviour implies that households at low levels of income find it preferable to allocate a large part of their time to raise children and only at higher levels of income change their allocation of time from child rearing to human capital formation and consumption.

Karl Marx in the preface to the first edition of volume I of *Capital* stressed: 'One nation can and should learn from others [but] . . . it can neither clear by bold leaps, nor remove by legal enactments, the obstacles offered by the successive phases of its normal development. But it can shorten and lessen the birth-pangs' (Marx, 1954, p. 20). The conclusion drawn from unified growth theory is similar and adds a new meaning to Marx's reference to birth-pangs: a state of sustained economic growth can only be reached after the forces shaping fertility behaviour have brought about the transition to the Post-Malthusian Regime. The advocates of the 'unified' approach appear to be optimistic that sooner rather than later all economies will have gone through the transition phase to the regime of sustained economic growth worldwide. After an epoch of 'Great Divergence' some economists see even an epoch of 'Great Convergence' ahead for mankind in which the problem of development will vanish.

The idea of unbounded growth of income per capita on a world scale implies that we can escape the limits to growth imposed by environmental constraints. This is a highly controversial claim (see Aghion and Durlauf, 2005, Chapter 28). Industrial pollution, global warming and climatic change, soil erosion, the reduction of biodiversity, and so on, point towards problems that cannot adequately be dealt with in terms of the usual macro-economic growth models. Measuring living standards *vis-à-vis* fundamental changes in the quality of the environment and life becomes a tantalizing task and ought to prevent social scientists from getting complacent about their achievements as to explaining the world.

Alongside an avalanche of theoretical contributions there has been a no less impressive avalanche of empirical ones. The first round of papers was motivated by the observation that there need not be a convergence of per capita income levels worldwide, as predicted by the Solow model. The new growth models were expected to do a better job. Yet Mankiw et al.

(1992) contested this and insisted that the empirical performance of the conventional Solow approach was superior to the ‘endogenous’ growth models. While on the surface this may well be so, Felipe and McCombie (2005) have shown that there is no reason to become complacent, for the ‘good fit’ of the Solow model is simply a reflection of the fact that an accounting identity has been tested. For summary accounts of the empirical findings see, for example, Temple (1999) and Bosworth and Collins (2003).

Non-neoclassical contributions

Aghion and Durlauf (2005) give the impression that the revival of growth economics was largely restricted to the neoclassical camp whose characteristic feature is that the problems at hand are redefined in such a way that they can be subjected to constrained optimization. This neglects contributions coming from other traditions. Here is not the place to provide a comprehensive account of classical, Marxian, Keynesian and evolutionary approaches to the problem of growth and development; see, therefore, Dutt (1990), Foley and Michl (1999), Salvadori (2003), Nelson (2005) and Salvadori and Panico (2006).

Neoclassical models, old and new, typically assume full employment of labour and full capacity utilization and thus follow Solow’s example, who in his 1956 contribution explicitly set aside problems of effective demand and assumed what he called a ‘tight-rope view of economic growth’. This does not mean that there are no such problems, as Solow was to stress time and again and also recently (see Aghion and Durlauf, 2005, p. 5). Despite his warnings, neoclassical growth theorists continue to be concerned almost exclusively with the evolution of potential output and ignore all effective demand failures. Interestingly, the subject index of the *Handbook* just referred to has no entry on capacity or capital utilization. Ignoring the demand side, that is, assuming Say’s law, is justified in terms of the overwhelming importance of long-run growth compared with short-run fluctuations. However, there appears to be a misunderstanding involved here, as the following example can clarify. Assume two identical economies except for the fact that one, due to a better stabilization policy, manages to realize on average, over a succession of booms and slumps, a higher average rate of capacity utilization than the other economy. With Y as actual and Y^* as potential output, s as the savings rate, v as the actual and v^* as the optimal output-to-capital ratio and $u = Y/Y^*$ as the average degree of utilization of productive capacity, we have:

$$g_i = \frac{S}{Y} \frac{Y}{K} = \frac{s}{v} = \frac{S}{Y} \frac{Y^*}{K} \frac{Y}{Y^*} = \frac{s}{v^*} u_i \quad (i = 1, 2)$$

Assume now that $s = 0.2$ and $v^* = 2$, but $u_1 = 0.8$ and $u_2 = 0.7$. Then the first economy would grow at 8 per cent per year, whereas the second would grow at only 7 per cent. This may seem a trifling matter, and in the short run it surely is, but according to the compound (instantaneous) interest formula after about 70 years the first economy would be larger than the second one by the amount of their (common) size at the beginning of our consideration. Hence effective demand matters. Experience also suggests that there is no reason to presume that actual savings can be expected to move sufficiently close around full employment and full capacity savings. Persistently high rates of unemployment in many countries, both developed and less-developed ones, strongly indicate that the problems of growth and development cannot adequately be dealt with in terms of the full employment assumption.

A recurrent tenet of Keynesian models is that different components of effective demand affect the rate of growth differently. It was already Roy Harrod who stressed that government policies have to be used both to stabilize the economy and to achieve higher growth. This theme was taken up by Nicholas Kaldor who discussed the interaction between public debt and interest rates. A monetary policy causing widely fluctuating short-run interest rates is said to raise the long-term rate to levels which may curb accumulation unless the rate of profit is raised too. This, Kaldor maintained, can be accomplished by stimulating effective demand via tax cuts and by fiscal policy.

The relationship between the rate of growth of effective demand and the rate of profits in the simplest framework possible is expressed by the so-called Cambridge equation:

$$r = s_c g$$

where s_c is the propensity to save of capitalists. This relationship, or some variant of it, holds in a large number of cases. In more recent times, taking up suggestions especially by Michal Kalecki and Joseph Steindl, there have been attempts to analyse investment behaviour more carefully. The presence of an 'investment function' in addition to and independently of the savings function is indeed a characteristic feature of Keynesian models. This has led to a class of investment-led growth models, in which growth is typically seen to depend on two main, but interrelated factors: profitability and effective demand. As regards the second factor there is wide agreement and strong empirical evidence that investment responds positively (negatively) to rising (falling) levels of capacity utilization. Profitability in turn is governed by the innovative potential that can be exploited at a given moment of time and by income distribution. Put in a nutshell, the type of investment function typically employed looks as follows:

$$\gamma = \gamma(r, r^e, i, u)$$

where γ is the share of investment, r the current rate of profit as an indicator of the possibilities of internal financing, r^e the expected rate of profit, i the long-term rate of interest, and u the degree of capacity utilization. The characteristic features of these models are essentially three. Firstly, income distribution and growth are simultaneously determined. Secondly, the ‘paradox of thrift’ is not limited to the short run: an increase in the overall propensity to save, other things being equal, reduces both the rate of growth and the rate of profit. This is exactly the opposite of what neoclassical models typically predict. Finally, the rate of growth depends negatively on the real wage rate provided the system is in what is called a profit-led growth regime. However, this need not be the case. There exist constellations of the parameters which give the model an ‘underconsumptionist’ flavour with the growth rate rising together with the real wage rate over a certain range. For a summary account of this class of models, see Commendatore et al. (2003).

Recent economic history shows that the Keynesian approach can be used to interpret reasonably well, for example, the economic development of the United States, which for many years followed a policy of massive budget deficits. The remarkable growth performance of China can also be explained with reference to the leading role of investment and, via the multiplier, of effective demand, whereas an explanation presupposing the full employment of labour is bound to lead astray with respect to an economic system in transition from a dominantly agricultural to an industrial economy, with hundreds of millions of workers from rural areas in search of jobs in cities.

There is also evidence that sluggish economies have less potential to adopt and develop new technology. This brings us to contributions in which technical progress is endogenous. There are several mechanisms discussed in the literature as to how the growth of output affects the growth of labour productivity, especially in the manufacturing sector, from Adam Smith’s concept of the division of labour to Verdoorn’s and Kaldor’s laws. The virtuous circle contemplated in this kind of literature sees profitability positively related to the growth of labour productivity, which is seen to be positively related to the growth in output. High rates of profit in turn will feed high rates of investment growth and thus high rates of output growth.

Another class of contributions comes from evolutionary economics. This is a rapidly growing field and its main concern, following in the footsteps of Joseph A. Schumpeter, is to analyse why and how the economic system incessantly changes from within. The focus of attention is more on development than growth, or rather it is insisted that the economic process

generates rapid qualitative change that cannot be captured in terms of the usual highly aggregate growth models. The origins of the evolutionary approach can be traced back to the classical economists, especially Adam Smith. The approach was taken up in parts by Alfred Marshall and then Schumpeter. The modern discussion was largely shaped by Nelson and Winter (1982). The evolutionary approach centres on a dynamic analysis in which random elements change the population of firms or the technology they use via a selection mechanism on existing variety. Discovery, learning and imitation assume centre stage in the argument which is about population dynamics and the economic effects it entails. For obvious reasons, evolutionary economics rejects such neoclassical concepts as the 'representative agent' or the 'aggregate production function'. Neoclassical growth theory is said to suffer from a detachment between formal and 'appreciative' theory (Nelson, 2005), where the latter is close to empirical studies of the actual behaviour of firms. Important contributions to evolutionary growth economics came from, among others, Giovanni Dosi, Stanley Metcalfe and Gerald Silverberg. For a summary account see Santangelo (2003).

References

- Aghion, P. and S.N. Durlauf (2005), *Handbook of Economic Growth*, 2 vols, Amsterdam: Elsevier.
- Aghion, P. and P. Howitt (1998), *Endogenous Growth Theory*, Cambridge, MA: MIT Press.
- Alesina, A. and D. Rodrik (1994), 'Distributive Politics and Economic Growth', *Quarterly Journal of Economics*, 109: 465–90.
- Azariadis, C. (1996), 'The Economics of Poverty Traps', *Journal of Economic Growth*, 1: 449–86.
- Barro, R. and X. Sala-i-Martin (2003), *Economic Growth*, 2nd edn, Cambridge, MA: MIT Press.
- Bhaduri, A. (2007), *Growth, Distribution and Innovations: Understanding their Interrelations*, London: Routledge.
- Bosworth, B.P. and S.M. Collins (2003), 'The Empirics of Growth: An Update', *Brookings Papers on Economic Activity*, (2): 113–79.
- Commendatore, P., S. D'Acunto, C. Panico and A. Pinto (2003), 'Keynesian Theories of Growth', in N. Salvadori (ed.), *The Theory of Economic Growth: A Classical Perspective*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 103–38.
- Dixit, A.K. and J.E. Stiglitz (1977), 'Monopolistic Competition and Optimum Product Diversity', *American Economic Review*, 67: 297–308.
- Dutt, A.K. (1990), *Growth, Distribution and Uneven Development*, Cambridge and New York: Cambridge University Press.
- Felipe, J. and J.S.L. McCombie (2005), 'Why are Some Countries Richer than Others? A Skeptical View of Mankiw–Romer–Weil's Test of the Neoclassical Growth Model', *Metroeconomica*, 56: 360–92.
- Foley, D.K. and T.R. Michl (1999), *Growth and Distribution*, Cambridge, MA and London: Harvard University Press.
- Helpman, E. (2004), *The Mystery of Economic Growth*, Cambridge, MA: Harvard University Press.
- Heston, A., R. Summers and B. Aten (2002), 'Penn World Table Version 6.1', Center for International Comparison at the University of Pennsylvania (CICUP).
- Jones, C. (2002), *Introduction to Economic Growth*, 2nd edn, New York: W.W. Norton.
- Kurz, H.D. (1997), 'What could the "New" Growth Theory Teach Smith or Ricardo?' Economic Issues Lecture given at the Royal Economic Society Conference, Staffordshire

- University, 24–27 March; published in *Economic Issues*, 2: 1–20; reprinted in H.D. Kurz and N. Salvadori (2003), *Classical Economics and Modern Theory: Studies in Long-period Theory*, London: Routledge, pp. 137–62.
- Kurz, H.D. and N. Salvadori (1995), *Theory of Production: A Long-Period Analysis*, Cambridge: Cambridge University Press.
- Kurz, H.D. and N. Salvadori (1998), ‘“Endogenous” Growth Models and the “Classical” Tradition’, in H.D. Kurz and N. Salvadori, *Understanding ‘Classical’ Economics. Studies in Long-period Theory*, London: Routledge, pp. 66–89.
- Kurz, H.D. and N. Salvadori (1999), ‘Theories of “Endogenous” Growth in Historical Perspective’, in M.R. Sertel (ed.), *Contemporary Economic Issues: Proceedings of the Eleventh World Congress of the International Economic Association, Tunis, Vol. 4: economic Behaviour and Design*, London: Macmillan, pp. 225–61.
- Kurz, H.D. and N. Salvadori (2006), ‘Endogenous Growth in a Stylised “Classical” Model’, in G. Stathakis and G. Vaggi (eds), *Economic Development and Social Change: Historical Roots and Modern Perspectives*, London: Routledge, pp. 106–24.
- Lucas, R.E. (1988), ‘On the Mechanics of Economic Development’, *Journal of Monetary Economics*, 22: 3–42.
- Maddison, A. (2001), *The World Economy in Millenium Perspective*, Paris: OECD.
- Mankiw, G., D. Romer and D. Weil (1992), ‘A Contribution to the Empirics of Economic Growth’, *Quarterly Journal of Economics*, 107: 407–37.
- Marx, K. (1954), *Capital*, Vol. 1, Moscow: Progress Publishers.
- Nelson, R. (2005), *Technology, Institutions, and Economic Growth*, Cambridge, MA and London: Harvard University Press.
- Nelson, R.R. and S.G. Winter (1982), *An Evolutionary Theory of Economic Change*, Cambridge, MA: Harvard University Press.
- Park, M. (2007), ‘Homogeneity Masquerading as Variety: The Case of Horizontal Innovation Models’, *Cambridge Journal of Economics*, 31(3): 379–92.
- Pomeranz, K. (2000), *The Great Divergence: China, Europe and the Making of the Modern World Economy*, Princeton, NJ: Princeton University Press.
- Romer, P.M. (1986), ‘Increasing Returns and Long-Run Growth’, *Journal of Political Economy*, 94: 1002–37.
- Romer, P.M. (1990), ‘Endogenous Technological Change’, *Journal of Political Economy*, 98: S71–S102.
- Salvadori, N. (ed.) (2003), *The Theory of Economic Growth: A Classical Perspective*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.
- Salvadori, N. and C. Panico (eds) (2006), *Classical, Neoclassical and Keynesian Views on Growth and Distribution*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.
- Santangelo, G.D. (2003), ‘The Evolutionary Perspective on Growth’, in N. Salvadori (ed.), *The Theory of Economic Growth: A Classical Perspective*, Cheltenham, UK and Northampton, MA, USA, pp. 204–20.
- Smith, A. (1976), *An Inquiry into the Nature and Causes of the Wealth of Nations*, Vol. II of *The Glasgow Edition of the Works and Correspondence of Adam Smith*, ed. R.H. Campbell, A.S. Skinner and W.B. Todd, Oxford: Oxford University Press; 1st edn 1776.
- Solow, R. (1956), ‘A Contribution to the Theory of Economic Growth’, *Quarterly Journal of Economics*, 70: 65–94.
- Steedman, I. (2003), ‘On “Measuring” Knowledge in New (Endogenous) Growth Theory’, in N. Salvadori (ed.), *Old and New Growth Theories: An Assessment*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 127–33.
- Summers, R. and A. Heston (1991), ‘The Penn World Table (Mark 5): An Expanded Set of International Comparisons, 1950–1988’, *Quarterly Journal of Economics*, 106 (2): 327–68.
- Temple, J. (1999), ‘The New Growth Evidence’, *Journal of Economic Literature*, 37 (1): 112–56.
- Weitzman, M. (1998), ‘Recombinant Growth’, *Quarterly Journal of Economics*, 113: 331–60.

16 Short-run macroeconomic issues in development

Peter J. Montiel

In recent years, the role of short-run macroeconomic factors in development has received substantial attention from the economics profession. A key result has been the recognition that, rather than being essentially orthogonal to an economy's long-run growth path, short-run macroeconomic fluctuations can have significant, first-order impacts on its long-run growth rate. In particular, higher levels of macroeconomic instability seem to be associated with reduced rates of long-run growth.

Unfortunately, recent experience suggests that macroeconomic stability is difficult to achieve under a widely-favored development strategy in which the capital account is opened early in the development process in order to enhance access to foreign savings – what might be referred to as an 'emerging market' approach to development. The experience of countries that have traversed this road suggests that the combination of an open capital account and poorly developed domestic macroeconomic and financial institutions is a recipe for extreme macroeconomic volatility. A key challenge faced by development macroeconomists, therefore, is to identify a set of macroeconomic and financial institutions that are conducive to macroeconomic stability in developing countries that seek to take advantage of the potential benefits of capital account openness.

This chapter will examine the potential outlines of such a set of institutions. It begins with a short overview of the current state of knowledge concerning the link between macroeconomic stability and growth. It then considers the appropriate institutional underpinnings for macroeconomic stability in the areas of fiscal, monetary, exchange rate and financial sector policies, focusing on the role of institutional design in preventing macro policies from themselves serving as a source of macroeconomic shocks. The third section complements this discussion by examining how deficient macroeconomic and financial institutions can undermine macroeconomic stability – and therefore long-run growth – by magnifying the macroeconomic impact of non-policy shocks. The chapter's main arguments are summarized in the final section.

Instability, crises and growth: the evidence

In principle the relationship between short-run macroeconomic volatility and long-run growth could be positive or negative. A positive relationship could arise if the adoption of high-return technologies entails the assumption of increased macroeconomic risk, or if government policies intended to reduce macroeconomic risk tend to misallocate resources, thereby resulting in more stable but lower growth rates. A positive relationship could also arise through a Schumpeterian process of ‘creative destruction’, in which recessions weed out the least efficient firms, so that more frequent recessions are associated with higher rates of productivity growth. On the other hand, if volatility is associated with increased uncertainty and noisier relative prices, higher volatility may result in a misallocation of investment, creating a negative relationship between volatility and growth. Such a relationship could also arise if boom–bust cycles reduce the efficiency of monitoring mechanisms in the financial sector (Gavin and Hausmann, 1998) or induce bank failures that destroy knowledge capital, or if recessions are associated with reduced learning-by-doing.

Across the world, the simple cross-country correlation between volatility and long-run growth appears to be a function of income levels – it is negative for the lowest-income countries, approximately zero for middle-income countries, and may even be positive for advanced countries (Hnatkowska and Loayza, 2004). Hnatkowska and Loayza found that this relationship does not appear to be mediated by standard growth determinants, but it does appear to depend on country characteristics, not only weakening at higher income levels, but also at higher levels of institutional development, and exhibiting an inverted U-shaped relationship with the level of financial development (that is, the effect is strongest at intermediate levels of financial development). The relationship between volatility and growth seems to be driven primarily by ‘crisis’ volatility, rather than by normal cyclical fluctuations.

Other types of evidence are consistent with this finding. For example, Cerra and Saxena (2005) find that recoveries after business cycle troughs are not associated with faster-than-average growth rates in developing countries, as would be required for the level of output to return to its pre-recession trend. This suggests that output losses associated with recessions in developing countries are permanent – that is, recessions are associated with permanent reductions in average growth rates. Moreover, more frequent economic contractions do not tend to be associated with either shallower contractions or faster expansions, so countries that experience more frequent contractions tend to have lower average growth rates. These effects tend to be stronger for middle-income than for low-income developing

countries (that is, recessions have greater permanent costs for the former), but the effect is present in countries of both types.

Macroeconomic policies and macroeconomic stability

The adverse effect of macroeconomic instability on growth implies that the challenge for macroeconomic policies in developing countries is to achieve an optimal trade-off between credibility and flexibility, so that policies can avoid becoming a source of macroeconomic disturbances (credibility) while at the same time being available to be deployed as stabilization instruments in response to exogenous shocks (flexibility). In this section I will consider in turn the roles of fiscal, monetary, exchange rate and financial sector policies from this perspective.

Fiscal policy

Deficient fiscal institutions in developing countries have been associated with frequent episodes of public sector insolvency, heavy reliance on seignorage (resulting in high inflation), and procyclicality in fiscal policy. Possibly the single most important contribution that fiscal policy can make to short-run macroeconomic stability in developing countries is to maintain a safe perception of fiscal solvency, yet the evidence suggests that the perceived solvency of many developing countries tends to be called into question by capital markets at much lower debt-to-gross domestic product (GDP) ratios than several industrial countries have comfortably been able to support, suggesting a lack of credibility in developing-country fiscal institutions.¹ The combination of fiscal rigidities on the expenditure side and highly distortionary tax structures on the revenue side has meant that reliance on seignorage revenues has typically been substantially higher in developing than in industrial countries. The result has too often been high and unstable inflation. A voluminous literature documents the adverse effects of high inflation on growth (Khan and Senhadji, 2001). Finally, there is by now substantial evidence that fiscal policy in developing countries tends to be procyclical (for example, Kaminsky et al., 2004). One interpretation of this phenomenon relies on ‘voracity effects’ (Tornell and Lane, 1998) that create overspending in response to favorable transitory shocks to the public sector’s intertemporal budget constraint (see below). This results in procyclicality as the result of the borrowing constraints that are imposed by creditors during macroeconomic bad times.

Recent work on devising institutional ‘fixes’ to the political economy mechanisms that generate these fiscal problems is motivated by the observation that government programs tend to create benefits that are concentrated either geographically or sectorally, but are typically financed from a common pool of resources. Those who benefit thus fail to internalize the

full costs of the program, since most of those costs are borne by others, especially future generations who will bear the costs of fiscal insolvency.² Budget institutions can affect ‘rules of the game’ within which these agents interact, either by placing constraints on the whole process, or by distributing power and responsibility among the agents. A variety of institutional ‘fixes’ have been proposed (Hausmann, 2004):

1. Quantitative targets that impose numerical constraints on fiscal outcomes, and thus act as precommitment devices (for example, laws that require the budget to be balanced over the business cycle).
2. Delegation of the formulation of ceilings on government borrowing to an autonomous entity that has little incentive to deviate from the medium-term optimum.
3. Modification of the procedural rules that govern the drafting of the budget, its discussion in parliament, and implementation, so as to give agenda-setting powers to agents that can best internalize the cost of funds – for example, the Finance Ministry, as opposed to line ministries or the parliament.
4. Enhanced fiscal transparency, based on the perspective that informational deficiencies tend to worsen agency problems.

There is some evidence that such institutional ‘fixes’ can be effective. Stein et al. (1998), for example, find that budget procedures that include constraints on the deficit, introduce hierarchical elements into the budget process, and are more transparent, produce smaller deficits.

Monetary policy

While for fiscal policy the primary credibility issue concerns public sector solvency, for monetary policy it concerns price-level stability – that is, what is at issue for central banks is anti-inflationary credibility. The link between fiscal solvency and monetary policy that operates through seignorage revenue suggests that the perception that fiscal solvency is achievable without excessive reliance on seignorage revenues is an important requirement for the anti-inflationary credibility of the central bank, and many stabilization programs in high-inflation developing countries during 1980–2000 have foundered in the absence of sufficient fiscal adjustment. The question is what can be done on the purely monetary side of things to enhance the anti-inflationary credibility of the central bank.

Central bank independence In recent years, the focus in developing countries has been to break ‘fiscal dominance’ – essentially the dictation of monetary policy by the finance ministry – by granting formal legal inde-

pendence to central banks and protecting that independence by giving central banks a legal mandate to promote price-level stability, often to the exclusion of other macroeconomic objectives. The purpose is, of course, to divorce monetary policy from the government's financing needs. In principle, effective central bank independence and single-minded pursuit of a price-stability objective could enhance both the central bank's anti-inflationary credibility as well as the perception of the government's solvency, since central bank independence may function as a signaling device: because fiscal insolvency is so costly, only a government confident of its non-inflationary solvency would accede to effective central bank independence. Unfortunately, the jury is still out on the effectiveness of central bank independence in promoting anti-inflationary credibility. What is clear is that it is *de facto*, not *de jure* independence that matters, especially in developing countries (Cukierman et al., 1994).

Inflation targeting As in the case of fiscal policy, the monetary policy regime faces the need to strike an optimal balance between credibility and flexibility. In doing so, the options open to the central bank depend on the economy's degree of integration with world financial markets. The 'impossible trinity' suggests that countries can only choose two from among the three options of perfect financial integration, monetary autonomy and officially determined exchange rates. For developing countries that have opted for financial integration, the choice is therefore between fixing the exchange rate and exercising monetary autonomy.

Most such countries initially tended to retain their soft pegs after opening up their capital accounts, attempting to achieve credibility and preserve flexibility by using the exchange rate as a nominal anchor as well as an instrument for stabilization policy, thus relegating monetary policy to the role of an instrument for managing the stock of international reserves. However, active use of the exchange rate as an instrument of stabilization policy proved difficult under conditions of high financial integration, because large anticipatory capital movements tended to force exchange rate changes, resulting in a rash of currency crises that generated substantial macroeconomic instability. As a result, many middle-income developing countries have altered the choice that they have made from the options offered by the 'impossible trinity', opting for monetary autonomy under floating exchange rates.

While the monetary autonomy that is available under floating exchange rates preserves the flexibility of monetary policy to act as a stabilization instrument, it implies the abandonment of the exchange rate as a nominal anchor, and thus leaves the challenge of securing nominal credibility. Unfortunately, the alternative of using the money supply as a nominal

anchor has not proven very promising anywhere around the world – in industrial or developing countries – because a continuous process of financial reform and innovation has rendered the demand for money highly unstable, loosening the connection between the money supply and the price level.

A recent innovation among several middle-income developing countries (Brazil, Chile, Colombia, the Czech Republic, Hungary, Israel, Korea, Mexico, Peru, Poland, South Africa and Thailand) has been the implementation of an alternative monetary policy regime intended to secure nominal credibility under floating exchange rates: inflation targeting.³ Mishkin (2004) defines this regime as consisting of five parts:

1. A commitment to price stability as the primary goal of monetary policy.
2. The public announcement of a medium-term quantitative target for inflation.
3. An operating procedure in which many variables are monitored to determine the setting of monetary policy instruments.
4. Increased transparency through communication with the public about the plans, objectives and decisions of the central bank.
5. Accountability of the central bank for attaining its inflation objectives.

Overall, the results from the adoption of inflation targeting in developing countries appear to have been relatively positive, despite the handicap that the framework faces in this context (weak institutions, relatively high initial inflation, low initial credibility, more severe macroeconomic consequences associated with exchange rate changes, and more severe external shocks). Fraga et al. (2003), for example, found that inflation dropped after the adoption of inflation targeting in the countries listed above, though not immediately to the levels recorded by industrial-country inflation targeters. However, perhaps because of the factors listed above, emerging-market inflation targeters have on the whole been less successful at hitting their targets than have industrial countries, even though their target ranges have tended to be broader (Ho and McCauley, 2003).

Monetary autonomy under floating rates Inflation targeting represents an attempt to achieve an optimal trade-off between credibility and flexibility in monetary policy. It is compatible with a Taylor rule for monetary policy in which the domestic policy interest rate is adjusted in response to deviations of inflation from its targeted value as well as deviations in real output from its natural level. In the context of developing countries, however, the monetary authorities are likely to have a concern not just with stability of output, but also with that of the exchange rate. This is partly because

exchange rate changes are likely to have a larger impact on the objectives of price stability and full employment (internal balance) that motivate conventional Taylor rules, but also partly because developing countries tend to be more open, and therefore more concerned with the effects of exchange rate changes on the current account of the balance of payments (external balance).

The implication is that the Taylor rules implemented by inflation-targeting developing countries will tend to make the policy interest rate a function not just of the deviation of inflation from its targeted level and of real output from its natural level, but also of the exchange rate from some perceived equilibrium level. The evidence suggests that this is so (see Mohanty and Klau, 2004). This turns out to be a key issue in determining the effective degree of monetary autonomy that developing-country central banks can exercise in countries that are highly integrated with international capital markets. If the weight given to deviations of the exchange rate from its targeted value in developing-country Taylor rules is sufficiently high so that their central banks are effectively unwilling to tolerate deviations of the exchange rate from its targeted value, then such countries will effectively be operating fixed exchange rate regimes and the impossible trinity implies that they will cease to enjoy monetary autonomy, even if their exchange rates are formally floating. This is the key insight of the ‘fear of floating’ literature initiated by Calvo and Reinhart (2002).

Several writers have explored the issue of whether developing countries that officially float their exchange rates effectively enjoy more monetary autonomy than those that fix. Early results for Latin America by Hausmann et al. (1999) suggested that, consistent with ‘fear of floating’, countries with floating exchange rates used domestic interest rates so aggressively to combat exchange rate changes as to leave them with little effective monetary autonomy. However, subsequent work has not tended to uphold this conclusion. Shambaugh (2004), for example, found that monetary independence held for floating-rate developing countries in both the short run and long run.

Overall, then, the evidence suggests that developing countries that have attempted to achieve anti-inflationary credibility while retaining the flexibility to use monetary policy as an instrument of stabilization policy – by adopting floating exchange rates with inflation targeting – have indeed been able to achieve both improved inflation performance as well as enhanced stability through the exercise of effective monetary autonomy.

Exchange rate policy

In the wake of the currency crises of the 1990s, the debate over optimal exchange rate regimes for developing countries has taken on a new life.

Many authors have argued that developing countries that choose to integrate themselves with world financial markets in practice face a choice that is restricted to the extreme ends of the exchange regime spectrum: either 'hard' pegs or floating. In other words, the claim is that intermediate regimes featuring 'fixed but adjustable' exchange rates are off the exchange rate regime menu for financially open developing countries. This is known as the 'bipolar' view of exchange rate regime choice. This view was initially articulated by Eichengreen (1994), but has found many adherents since.

The theory underlying the bipolar view is essentially based on the welfare consequences of extreme macroeconomic instability. The logic is that of second-generation crisis models: when capital mobility is high, a speculative attack on a currency will require a very high domestic interest rate to defend the existing parity. Central banks will rarely find it optimal to sustain such high interest rates, so in the face of a sustained speculative attack they are likely to abandon the fixed parity. Extreme macroeconomic instability results whether the parity is sustained or not – in the form either of very high interest rates or of an abrupt exchange rate change.

However, it is worth noting that second-generation crisis theory does not suggest that any and all fixed-but-adjustable exchange rate regimes will be vulnerable in this way. Rather, such regimes are vulnerable to speculative attack when the economy's 'fundamentals' are weak, meaning that high interest rates are particularly costly and/or an exchange rate depreciation is particularly beneficial. Thus, at bottom the bipolar view boils down to the proposition that in practice, developing countries are unable to manage their affairs so as to avoid the zone of vulnerability in their fundamentals. This being so, they are best advised to avoid vulnerable exchange rate regimes, opting instead for hard pegs or floating exchange rates.

This raises two questions, however. The first is just why it is that developing countries are unable to manage their fundamentals so as to avoid vulnerability. The second is whether, if they are indeed unable to do so, we can be sure that they will avoid macroeconomic instability by opting for one of the polar exchange rate regimes.

Calvo and Mishkin (2003) argue that the absence of strong institutional underpinnings for fiscal solvency, price stability and financial sector stability underlies the inability of many developing countries to escape weak fundamentals. At the same time, however, institutional weaknesses also imply that macroeconomic stability may not be easily securable through a simple choice of exchange rate regime – that is, by opting either for a hard peg or a floating exchange rate. The upshot is that, given the underlying institutional weaknesses, adoption of an exchange rate regime located at the

extreme ends of the regime spectrum may not safeguard developing countries from macro instability.

This leaves the question of whether opting for an extreme exchange rate regime may not help to resolve the economy's institutional weaknesses. The evidence on this issue is mixed. As Calvo and Mishkin point out, the question of whether exchange regime choice can contribute to the strengthening of macroeconomic institutions remains an open one. The obvious implication is that developing countries that are characterized by weak macroeconomic institutions may be ill advised to allow scope for the free movement of capital. I turn to this issue next.

Financial policies

Capital account openness In principle, an open capital account can convey a number of benefits for developing countries. These prospective benefits induced many developing countries to liberalize their capital accounts after the mid-1980s. Unfortunately, the evidence that capital account liberalization has indeed proved to be welfare-enhancing in the context of developing countries is weak at best. The literature on the effects of international financial market integration on long-run growth in developing countries is extensive, but inconclusive.⁴

Why does the evidence seem to be so unkind to the theory? The problem is that in the presence of distortions, foreign savings may be misallocated. There are a large number of potential distortions in developing countries that could generate effects of this type.

In addition to these microeconomic arguments for capital account restrictions, there are a variety of macroeconomic arguments as well. For example, in the presence of fiscal rigidities, the impossible trinity implies that developing countries that maintain fixed exchange rates can only deploy domestic stabilization policies (in the form of monetary policy) if they retain some controls over capital movements. Capital account restrictions may also be attractive for countries that choose to insulate themselves against external financial disturbances (either in the form of changes in industrial-country monetary policies or exogenous variations in country risk premia driven, say, by contagion from crises occurring elsewhere). Finally, capital account restrictions may be intended to affect not the total volume of capital flows, but their maturity composition, since the evidence suggests that a large volume of short-term debt is associated with enhanced vulnerability to currency crises among developing countries (Frankel and Wei, 2004).

The upshot is that, while capital account openness offers the potential to generate substantial benefits, the retention of restrictions on capital

movements may represent a second-best policy in developing countries that exhibit significant domestic microeconomic or macroeconomic institutional weaknesses. The crisis experience of the decade of the 1990s suggests that premature capital account openness in the presence of such weaknesses is fundamentally misguided.

Domestic financial liberalization Until the mid-1980s, policies toward the domestic financial sector in most developing countries were characterized by financial repression. Led by McKinnon (1973) and Shaw (1973), the economics profession increasingly recognized that such policies were inimical to long-run economic growth, primarily because of their adverse effects on the productivity of domestic investment. This led to a wave of domestic financial liberalization in developing countries after the mid-1980s, as described by Williamson and Mahar (1998).

Unfortunately, undertaking financial liberalization without at the same time strengthening the institutional framework in which the financial sector operates has proven to be a recipe for financial instability in developing countries. The problem is that banking as an economic activity is rife with moral hazard problems, and an appropriate institutional environment is required to restrict the scope for moral-hazard lending. Such an environment includes clear and secure property rights, an accessible, efficient and impartial legal system to enforce contracts, appropriate legal protection for creditors, well-specified accounting and disclosure standards, a regulatory system that screens entrants while encouraging competition, adequate capital requirements, and a supervisory system that can effectively monitor the lending practices of domestic financial institutions so as to prevent excessively risky lending. The proliferation of moral hazard lending in the absence of such an environment results in financial-sector balance sheets that are fragile and vulnerable to insolvency in response even to moderate macroeconomic shocks. Under such circumstances, the financial sector can amplify the effects of macroeconomic shocks arising elsewhere, and can itself serve as the source of particularly severe macroeconomic shocks, in the form of banking crises. Caprio and Klingebiel (1997) document the role of an inappropriate institutional environment in generating banking crises under post-liberalization conditions.

The moral of the story is not, of course, that domestic financial liberalization is to be avoided. One of the most robust results from the cross-country growth literature that has mushroomed in recent years is that financial development can have a powerful effect on economic growth (Levine, 2004; Wachtel, 2003). It is, instead, that the pace of liberalization for domestic financial systems that have not already been liberalized should be modulated to reflect the quality of the institutional framework

governing the domestic financial sector, and that improving the quality of this framework deserves high priority in the macroeconomic reform agenda.

External shocks

The effective degree of macroeconomic stability that developing countries can achieve depends not just on the extent to which policies can avoid creating domestic shocks, but also on their ability to preserve stability in the face of non-policy shocks. For developing countries, these consist primarily of external shocks, in the form of fluctuations in the terms of trade and shocks to capital flows. I now consider the role of macroeconomic institutions in preserving domestic macroeconomic stability in the presence of external shocks.

Terms-of-trade shocks The large agricultural sectors and traditional goods market openness that have characterized most developing countries have implied that weather-related phenomena and terms-of-trade fluctuations have been important sources of exogenous macroeconomic shocks for these economies. Recent studies have found that terms-of-trade fluctuations can account for 30–50 percent of the volatility of output among developing countries (Kose, 2002; and Broda, 2004).

Since the terms of trade are exogenous for the vast majority of developing countries, the stability issue that arises in association with such shocks is how to devise domestic macroeconomic institutions that can dampen their effects. As already mentioned, voracity effects on fiscal policy and procyclical capital flows actually tend to amplify these effects. Thus the key challenge on the fiscal side is to design budgetary institutions that can mitigate the common pool problems that give rise to voracity effects as well as to achieve the fiscal flexibility required to allow the use of countercyclical fiscal policy, even if only in the form of automatic stabilizers. Similarly, well-functioning monetary institutions would allow monetary policy to respond to terms-of-trade shocks in a stabilizing manner without jeopardizing the central bank's anti-inflationary credibility.

A key issue in the design of macroeconomic institutions to promote stability in the face of terms-of-trade shocks is the choice of exchange rate regime. Consistent with theory, the evidence suggests that developing countries that are vulnerable to terms of trade shocks achieve a greater degree of output stability under flexible than under fixed exchange rates (Broda, 2004). The implication is that this type of vulnerability should be an important consideration in the choice of optimal exchange rate regime for developing countries.

The upshot is that developing-country vulnerability to terms-of-trade shocks strengthens the case for the institutional reforms in the areas of

fiscal, monetary and exchange rate policies analyzed in the previous section.

'Sudden stops' of capital flows As developing countries have become more integrated financially with the rest of the world, shocks emanating from or magnified by international financial markets have increased in relative importance, and the most severe sources of macroeconomic instability among developing countries in recent years have been shocks of this type. External financial shocks affect the terms on which developing countries can access international capital markets. In 'normal' times this has two components: the world risk-free interest rate and the country risk premium. Fluctuations in the world risk-free interest rate are exogenous to developing countries. The country risk premium, on the other hand, has both exogenous and endogenous components. The exogenous components include the international price of risk as well as the factors that determine the amount of risk that international capital markets associate with a given set of domestic macroeconomic circumstances in debtor countries. The former has clearly fluctuated for reasons that are exogenous to individual developing countries – for example, the significantly reduced international appetite for risk in the wake of the 1998 Russian and Long-Term Capital Management (LTCM) crises. Similarly, country risk premia have at times also fluctuated without apparent changes in the international appetite for risk or in the circumstances of the countries involved – for example, as the result of 'wake-up call' contagion after the 1994 Mexican crisis and the 1997 crisis in Thailand.

The most dramatic and most destabilizing form of external financial shock to developing countries, however, can be interpreted as a very sharp response of the country risk premium to a change in country circumstances – so sharp, in fact, that the country becomes suddenly rationed out of world capital markets. This is known as the 'sudden stop' phenomenon (Calvo, 1998). What stops suddenly is the inflow of foreign capital, caused by sharply heightened perceptions of country risk. Because the contraction in aggregate demand associated with these episodes gives rise to sudden and unexpected real exchange rate depreciations, they are often accompanied by widespread bankruptcies in the real sector and banking crises, and they have consequently been associated with severe output contractions (Calvo and Reinhart, 2000).

Calvo et al. (2004) note that the incidence of sudden stops primarily depends on a developing country's degree of 'real' openness and its extent of domestic liability dollarization. As they point out, these variables are the result of domestic policies, such as commercial policies that affect an economy's degree of openness, as well as the management of fiscal and

monetary policies that determine its degree of nominal stability and thus the presence or absence of incentives for liability dollarization. The evidence suggests, therefore, that the extreme macroeconomic instability associated with sudden stops is but another price that developing countries pay for inappropriate domestic macroeconomic institutions.

Conclusions

The macroeconomic experiences of today's emerging market economies provide important lessons for low-income developing countries that aspire to become attractive destinations for private external funding. This is an important development issue, because the 'emerging market' route is not the only path to development, as the example of China makes clear. Low-income developing countries that currently retain restrictions on capital movements will have to decide whether to remove such restrictions early in the development process, thus opting for an emerging-market route to development, or to retain them until the development process is far advanced, following the Chinese model.

An important consideration in making this choice is that short-run macroeconomic instability can have important impacts on long-run economic growth. Macroeconomic instability can reduce the slope of the growth path, and thus is of first-order importance for long-run economic development. Thus the choice of development model depends in part on which of the options offers the most favorable prospect for producing a stable macroeconomic environment within which growth can take place.

To secure macroeconomic stability under the 'emerging economy' model requires not just good macroeconomic performance, but the development of appropriate macroeconomic institutions. Such institutions are critical to the achievement of an optimal trade-off between credibility and flexibility in macroeconomic policy. In the fiscal area, this involves the design of budgetary procedures and institutions that promote a credible perception of sustainable fiscal solvency without excessive reliance on seignorage revenues, thus gaining the option of deploying countercyclical fiscal policy without impairing creditor confidence. In the monetary area, it involves an independent central bank with a sufficiently credible commitment to price stability – perhaps through a monetary policy regime such as inflation targeting – as to allow monetary policy to be deployed flexibly to stabilize the economy in response to shocks. In the area of exchange rate policy, it involves an exchange rate regime that is appropriate to the economy's circumstances, with the recognition that 'one size fits all' does not apply in this important area of economic policy. With respect to the financial sector, it involves putting in place an institutional framework that minimizes the risk of moral hazard-based lending. With such an institutional framework in

place, the ‘emerging market’ route to development has the potential of yielding the substantial economic benefits that are promised by an open capital account without the growth-destroying disruptions that currency crises and ‘sudden stops’ can cause.

Notes

1. Reinhart et al. (2003) refer to this situation as one of ‘debt intolerance’.
2. This generates the ‘voracity effects’ mentioned above.
3. The list of inflation targeters is from Fraga et al. (2003).
4. See, for example, the comprehensive survey by Edison et al. (2004).

References

- Broda, Christian (2004), ‘Terms of Trade and Exchange Rate Regimes in Developing Countries,’ *Journal of International Economics*, 63: 31–58.
- Calvo, Guillermo (1998), ‘Capital Flows and Capital-Market Crises: The Simple Economics of Sudden Stops,’ *Journal of Applied Economics*, 1 (1): 35–54.
- Calvo, Guillermo and Carmen Reinhart (2000), ‘When Capital Inflows Come to a Sudden Stop: Consequences and Policy Options’, in P. Kenen and A. Swoboda (eds), *Reforming the International Monetary and Financial System*, Washington, DC: International Monetary Fund.
- Calvo, Guillermo and Carmen Reinhart (2002), ‘A fear of floating’, *Quarterly Journal of Economics*, 117 (2): 379–408.
- Calvo, Guillermo A. and Frederic S. Mishkin (2003), ‘The Mirage of Exchange Rate Regimes in Emerging Market Countries’, *Journal of Economic Perspectives*, 17: 99–118.
- Calvo, Guillermo A., Alejandro Izquierdo and Luis-Fernando Mejia (2004), ‘On the Empirics of Sudden Stops: The Relevance of Balance-Sheet Effects’, mimeo, IDB.
- Caprio, Gerard and Daniela Klingebiel (1997), ‘Bank Insolvency: Bad Luck, Bad Policy, or Bad Banking?’ World Bank, *Annual Bank Conference on Development Economics 1996*, (April), pp. 1–26.
- Cerra, Valerie and Sweta Chaman Saxena (2005), ‘Growth Dynamics: the Myth of Economic Recovery’, IMF, mimeo, February.
- Cukierman, Alex, Steven B. Webb and Bilin Neyapti (1994), ‘Measuring Central Bank Independence and Its Effect on Policy Outcomes’ International Center for Economic Growth Occasional Paper Number 58.
- Edison, Hali, Michael Klein, Luca Ricci and Torsten Slok (2004), ‘Capital Account Liberalization and Economic Performance: Survey and Synthesis’, IMF Staff Papers, 51 (2): 220–56.
- Eichengreen, Barry (1994), *International Monetary Arrangements for the 21st Century*, Washington, DC: Brookings Institution.
- Fraga, Arminio, Ilan Goldfajn and Andre Minella (2003), ‘Inflation Targeting in Emerging Market Economies’, NBER Working Paper 10019.
- Frankel, Jeffrey and Shang-Jin Wei (2004), ‘Managing Macroeconomic Crises: Policy Lessons’, in J. Aizenman and B. Pinto (eds), *Managing Volatility and Crises: A Practitioner’s Guide*, Cambridge: Cambridge University Press, pp. 315–405.
- Gavin, Michael and Ricardo Hausmann (1998), ‘The Roots of Banking Crises: The Macroeconomic Context’, in R. Hausmann and L. Rojas-Suarez (eds), *Banking Crises in Latin America*, Baltimore, MD: Johns Hopkins University Press, pp. 27–63.
- Hausmann, Ricardo (2004), ‘Good Credit Ratios, Bad Credit Ratings: The Role of Debt Structure’, in G. Kopits (ed.), *Rules-Based Fiscal Policy in Emerging Markets*, Washington, DC: International Monetary Fund, pp. 30–52.
- Hausmann, Ricardo, Michael Gavin, Carmen Pages-Serras and Ernesto Stein (1999), ‘Financial Turmoil and the Choice of Exchange Rate Regime,’ IDB Working Paper 400.

- Hnatkovska, Viktoria and Norman Loayza (2004), 'Volatility and Growth,' in J. Aizenman and B. Pinto (eds), *Managing Volatility and Crises: A Practitioner's Guide*, Cambridge: Cambridge University Press, pp. 65–100.
- Ho, Corinne and Robert McCauley (2003), 'Living with Flexible Exchange Rates: Issues and Recent Experience in Inflation Targeting Emerging Market Economies', BIS Working Paper No. 130.
- Kaminsky, Graciela, Carmen Reinhart and Carlos Vegh (2004), 'When It Rains, It Pours: Procyclical Capital Flows and Macroeconomic Policies', NBER Working Paper 10780, September.
- Khan, Mohsin S. and Abdelhak S. Senchadji (2001), 'Threshold Effects in the Relationship between Inflation and Growth', *IMF Staff Papers*, 48 (1): 1–21.
- Kose, Ayhan (2002), 'Explaining Business Cycles In Small Open Economies: How Much Do World Prices Matter?' *Journal of International Economics*, 56 (2): 299–327.
- Levine, Ross (2004), 'Finance and Growth: Theory, Evidence, and Mechanisms', in P. Aghion and S. Durlauf (eds), *Handbook of Economic Growth*, Amsterdam: North-Holland, pp. 865–934.
- McKinnon, Ronald I. (1973), *Money and Capital in Economic Development*, Washington, DC: Brookings Institution.
- Mishkin, Frederic S. (2004), 'Can Inflation Targeting Work in Emerging Market Countries?' NBER Working Paper 10646, July.
- Mohanty, M.S. and Marc Klau (2004), 'Monetary Policy Rules in Emerging Market Economies: Issues and Evidence', BIS Working Paper No. 149, March.
- Reinhart, Carmen, Kenneth Rogoff and Miguel Savastano (2003), 'Debt Intolerance', NBER Working Paper No. 9908, August.
- Shambaugh, Jay (2004), 'The Effect of Fixed Exchange Rates on Monetary Policy', *Quarterly Journal of Economics*, 119 (1): 301–52.
- Shaw, Edward S. (1973), *Financial Deepening in Economic Development*, New York: Oxford University Press.
- Stein, Ernesto, Ernesto Talvi and Alejandro Grisanti (1998), 'Institutional Arrangements and Fiscal Performance: The Latin American Experience', NBER Working Paper 6358, January.
- Tornell, Aaron and Philip Lane (1998), 'Are Windfalls a Curse?' *Journal of International Economics*, 44: 83–112.
- Wachtel, Paul (2003), 'How Much Do We Really Know about Growth and Finance?' *Economic Review*, Federal Reserve Bank of Atlanta (First Quarter): 33–47.
- Williamson, John and Molly Mahar (1998), *A Survey of Financial Liberalization*, Essays in International Finance No. 211, International Finance Section, Department of Economics, Princeton University.

17 Sectoral interactions in development

*Jørn Rattsø*¹

Introduction

Economic growth out of backwardness means structural transformation of the economy, typically industrialization from an agricultural base. The analysis of sectoral interaction captures this transformation. Low-income countries are often described as dual economies and the duality refers to the different economic conditions in traditional and modern economic activities. Sectoral balance in economic development can represent various sectoral disaggregations, but is primarily about agriculture versus industry. Although land can be expanded and the returns to land can be increased by irrigation and fertilizer, agriculture in the end faces decreasing returns. In modern growth terms the transformation is described as a movement of resources from decreasing-return to increasing-return activities. Increasing returns in industry can take various forms, and in economic analysis there is a shift from emphasis on traditional scale effects to externalities and productivity growth. The sources of productivity growth in agriculture and industry are important determinants of the sectoral interaction and transformation. Compared to earlier overviews of sectoral interaction in development I will concentrate on the productivity dynamics here.

The background understanding of my emphasis on productivity is that income differences across countries are not primarily the result of different availability of production factors. The empirical evidence that capital stocks per worker explain a limited part of the income differences among nations now is widely accepted. The attention therefore is turned to productivity, and productivity differences between countries are substantial, as documented by Hall and Jones (1999). The observation is consistent with a stable world income distribution over time. Differences in income levels are permanent, while differences in growth rates are mostly transitory. Movements in the world income distribution are associated with structural transformation and productivity growth.

In a closed economy setting, the conflict between agriculture and industry is obvious. The two sectors compete for resources, both labor and capital. The allocation of resources between agriculture and industry was addressed in stark form under early Soviet Union planning. When planning for growth and with central control of the investment allocation, the

government had to evaluate the advantages of the two sectors. Two conflicting views led to dramatic political battles about development strategy, associated with Preobrazhensky (1926 [1965]) and Bukharin (1971). Preobrazhensky argued that agricultural prices should be squeezed to give room for industry. When agriculture and industry compete for resources, agriculture must be held back to give room for industry. Bukharin argued for complementarity between agriculture and industry: agricultural income expansion stimulates industry via demand.

Preobrazhensky's argument about the terms of exchange between 'town and country' has been an influential basis for a widespread policy of discriminating agriculture. He wanted the peasants to carry more of the burden of industrialization to the advantage of the industrial proletariat. This cheap food policy was assumed to transfer agricultural surplus to finance industrial investment. In the Soviet Union, the policy included collectivization of agriculture, which should increase the availability of labor to industry without hurting agricultural output. The relocation of the rural population to villages in Tanzania can be seen as a sub-Saharan Africa counterpart to the Soviet collectivization. More usually the price discrimination of agriculture is combined with various measures to stimulate food supply. Sub-Saharan Africa shows how economies broadly failed to industrialize when agriculture was discriminated. It is hard to find successful growth experiences with such discrimination, except for the cases where export-oriented industrialization have financed food imports.

The view that agriculture and industry are in conflict for resources has been given a nice analytical presentation by Matsuyama (1992). In his model of an open economy 'the productive agricultural sector squeezes out the manufacturing sector and the economy will deindustrialize over time' (p. 326). The obvious policy conclusion is to smash machinery in agriculture to give room for expansion of industry with productivity growth. In this chapter I will discuss competition and complementarity between agriculture and industry. Starting out from the core interactions in the closed static economy I develop the understanding into open economies and dynamics. Old literature on dual models is linked to recent contributions about productivity dynamics.

This chapter discusses analytical approaches and results concerning sectoral balance.² I start out with a brief overview of sectoral balance issues in the next section. The following section discusses static interaction in models with a focus on competition and complementarity between agriculture and industry. The subsequent section extends to more recent analyses of the dynamics of productivity growth. The final section discusses future research.

Sectoral balance: a brief overview

Sectoral interaction appears as an issue when one goes behind the aggregate growth models to capture the underlying growth process of the economy. The disaggregation in question will differ dependent on the issue addressed. An overview of sectoral balance approaches is given by Taylor (1989). Separation between consumption goods and capital goods only makes the sources of consumption and investment explicit. In the development literature the dominating interest has been the transformation from an agriculture-based economy to industrialization, and I concentrate on this separation between agriculture and industry (the concept typically used to describe the modern economy). In the analysis below I emphasize the interaction between decreasing returns to scale (due to land) and increasing returns. Authors describe this duality with different concepts including capitalist and non-capitalist sectors and the rural–urban divide. Other disaggregations are investigated in the literature, and I will mention the separation between wage goods and luxury goods in the context of income distribution.

Growth theory is developed for the closed economy, but growth experiences for the many small countries are strongly linked to the world market. New disaggregations appear in the open economy. Trade analysis suggests a separation between labor-intensive and capital-intensive goods, while more macro-oriented approaches study traded and non-traded goods, often with a disaggregation of traded goods into exportables and importables.

Sectoral interactions across different disaggregations are based on product markets and factor markets. Sectors strengthen each other through income generation and demand, while they compete for resources in factor markets. Given balanced consumption demand between sectors it is hard to avoid balanced growth in the closed economy. The open economy allows specialization and unbalanced growth is more likely. This is the main thesis investigated below in the context of agriculture and industry.

Starting out from a low income level, development in most cases will involve a transformation of production from agriculture to industry. But there are many ways of arranging this transformation. As mentioned in the introduction, conflicting strategies regarding the priority of agriculture and industry came out starkly in the early Soviet Union planning. Preobrazhensky emphasized competition while Bukharin believed in complementarity. Both strategies imply that agriculture contributes to industrialization. In short agricultural contributions can be summed up as food, inputs, labor, savings, demand and foreign exchange for industry. Cheap food policy is argued to stimulate industry by holding down the real wage. Early industrialization is often based on processing of agricultural goods.

Labor is released from agriculture to industry. Agricultural profits and savings must be channeled to industrial investment. Agricultural income forms the basis of domestic demand for industrial goods. Agricultural exports may finance imported intermediate inputs to industry. The relationship goes the other way also: in particular, industrial inputs improve productivity in agriculture.

Arthur Lewis's 1954 article in the *Manchester School of Economic and Social Studies* on development with unlimited supplies of labor started up the modern literature analyzing sectoral balance and economic growth. His understanding of the roles of non-capitalist and capitalist sectors in the dual economy has generated an enormous literature on the transformation from economic backwardness as sectoral interaction. After endogenous growth theory this is now often understood as an interplay between decreasing return and increasing return production sectors. Rosenstein-Rodan (1943) and Nurkse (1953) had already analyzed the importance of increasing returns to scale, also involving multiple equilibria and poverty trap. Indeed most of the insights of endogenous growth theory can be found here. Ros (2000) shows the mechanisms at work in a comprehensive analytical discussion relating to models of endogenous growth. Temple (2005) gives a recent survey of dual economy models.

The first generation of contributions addressed controversial aspects and shortcomings of the original Lewis model, notably the working of agriculture, the labor market and surplus labor, and underlying migration from rural to urban areas. Dutt (1990) offers a nice overview of the many extensions of the Lewis model. A special issue of the *Manchester School* in December 2004 also reviews the literature. Gradually the literature specialized in different topics where agriculture–industry interactions are of relevance. This broad literature takes a closer look at particular aspects of the dual model such as the role of informal urban and rural markets, the importance of income distribution and household heterogeneity, and how macroeconomic closure affects the sectoral interactions. This chapter can be seen as an extension into one such particular aspect, the determination of productivity growth and thereby overall growth.

Complementarity and competition for resources

To clarify how sectoral interactions affect growth I will make use of a simple analytical framework. The aim here is not to present a history of thought, but to establish key insights and issues that need further attention. Since the focus is on productivity growth, it seems natural to apply the Matsuyama (1992) model as a benchmark and I apply the simplifying explicit functional forms of Rodriguez and Rodrik (2000). Full employment of labor is assumed in this analysis worked out to understand long-run growth. The

modifications compared to surplus labor are discussed. In this section I only deal with static market interactions.

Production in agriculture (X^A) and industry (X^M) are written in Cobb–Douglas form dependent of productivity level (A and M respectively) and the labor share. The labor share n is defined as the share employed in the manufacturing sector and consequently measures the size of industry. For simplicity of exposition I assume constant labor force (set at unity), and elasticity α is common to the two sectors. The production structure is:

$$X_t^A = A_t(1 - n_t)^\alpha \quad (17.1)$$

$$X_t^M = M_t n_t^\alpha \quad (17.2)$$

The above assumptions imply stark competition for labor (share) between agriculture and industry and that the development of productivity levels will determine growth. Productivity substitutes for investment, and the argument for this simplification, is that competition for capital will work very similarly to competition for labor. Allocation of labor is assumed to be competitive, and the marginal productivities of the two sectors are equal (with p measuring the relative price of manufacturing goods). The labor market equilibrium obeys:

$$A_t(1 - n_t)^{\alpha-1} = p_t M_t n_t^{\alpha-1} \quad (17.3)$$

This is the first key element of the model: competition for labor contributing to unbalanced growth. If one sector takes off, the other must suffer, since the winner absorbs labor (resources) from the loser. At this stage, the productivity levels are exogenous. If the productivity level in agriculture goes up, labor will shift from industry to agriculture (given the relative price p). The productivity level in agriculture is a threat to industry at the supply side of the economy. This is no longer true when I complete the demand side of a full equilibrium, to be shown below.

On the demand side I assume identical household preferences for agricultural and manufacturing goods. A parameter γ is introduced to represent the subsistence level of agricultural goods (food) consumption, and β measures the relative share of agricultural goods. Aggregating over all households in the closed economy (where consumption is equal to production), I get:

$$X_t^A = \gamma L_t + \beta p_t X_t^M \quad (17.4)$$

This second key element of the model reflects complementarity. At the demand side there is positive relationship between production in the two sectors, since they represent income and thereby a market for each other. Expansion of agricultural production allows a larger market for industrial goods.

Combining the labor market equilibrium and the demand condition, I reach one equation determining the labor share in industry (given the labor supply and the productivity level in agriculture):

$$\frac{\gamma L_t}{A_t} = (1 - n_t)^\alpha \left(1 - \beta \frac{n_t}{(1 - n_t)} \right) \quad (17.5)$$

I reach the core insight that a closed economy must have balanced growth (under these assumptions). When households demand both agricultural and industrial goods, even when an Engel effect is allowed, the closed economy must produce both goods. And when one sector expands, the income effect forces the economy to come up with more production from the other sector. The relative price must adjust to see that this happens. It follows that there is a positive relationship between the industrial labor share and the agricultural productivity level in the closed economy. The intuition is that agricultural productivity helps create domestic market expansion for industrial goods. This is consistent with the labor market equilibrium constraint since the relative price of industrial goods increases to stimulate labor flow into industry.

I have established complementarity between agriculture and industry even in a model with stark competition for labor between the two sectors. If one relaxes the assumption of full employment, one expects the complementarity between the sectors to be even stronger, since the employment level can increase in both sectors. The literature shows that the interaction depends on macroeconomic closure of the model in the case of labor surplus. The sectoral balance under various assumptions has been analyzed related to the Soviet Union 'price-scissors' debate, where the gap of the scissors indicates the relative price of agricultural to non-agricultural goods. In a formalization of the debate, Sah and Stiglitz (1984) confirm the Preobrazhensky position that agricultural terms of trade should be turned against agriculture to stimulate industry. The logic of the model explains the result: lower agricultural prices allow lower wage levels and thereby more savings to finance investment in industry. Rattsø (1988) shows that such competition between sectors is not a necessary outcome under other closures. Lower agricultural prices are likely to draw down food stocks and/or reduce agricultural production and the repercussions to industry are not necessarily positive. The working of the labor market under unemployment will be an important determinant.

Open economy extensions can be made in various ways, but I keep my focus on stark differences here. The small open economy has given world market prices for agricultural and industrial goods and set the price rate $p = 1$. The arguments below assume that the country has comparative advantage in agriculture. To facilitate discussion of the openness, I introduce the (ad valorem) tariff rate on manufacturing goods τ . The labor market equilibrium (17.3) in this case can be rewritten:

$$A_t(1 - n_t)^{\alpha-1} = (1 + \tau_t)M_t n_t^{\alpha-1} \quad (17.6)$$

Since production and demand are completely separated in the small open economy, given the budget constraints of the economy, no demand side is needed to derive the open economy relationship between agricultural productivity and industrial labor share. According to (17.6) it is all about competition for labor, and higher agricultural productivity draws labor to agriculture. The conflict between the sectors is sharpened in the open economy since the world market takes care of demand.

The broad conclusion is that comparative advantage in the open economy dramatically changes the intersectoral allocation. While agriculture and industry are complementary in the closed economy, they easily turn competitive in the open economy. When the economy specializes according to static comparative advantage, one sector can expand at the cost of the other. The demand side is satisfied by foreign trade. The static allocation gain of this basic model is well known. But the dynamic gain is open for discussion. The worry in the literature is that specialization according to comparative advantage moves low-income countries to low-income equilibria. The rest of this chapter investigates the growth aspects of sectoral interaction.

Dynamic interactions with endogenous productivity growth

Sources of productivity growth and productivity linkages between sectors and to the world market represent an area of active research. Recent contributions are based on two innovations in the understanding of productivity growth: Arrow-type (1962) learning by doing (LBD) that is external to the firms, and the Nelson and Phelps (1966) catching-up mechanism to the technology frontier. Benhabib and Spiegel (2005) and Klenow and Rodriguez-Clare (2005) present nice overviews. Aghion and Howitt (2005) show how the catching-up mechanism can be integrated into a Schumpeterian growth model. The origins of this thinking in the development literature are old and a key element is the advantage of backwardness, called the Veblen–Gerschenkron effect. Productivity growth has not been much studied in the dual-economy framework.

I concentrate here on the LBD mechanism, and influential early contributions are Krugman (1981) and van Wijnbergen (1984). These models postulate that only employment in the industrial sector contributes to LBD. It follows that the more one succeeds in squeezing employment outside industry, the higher is the growth rate. Empirical evidence offers support to the notion that industrial employment is the principal source of LBD, and I start out with this assumption. But the empirical literature also shows that there are spillovers between sectors, in particular that learning experiences in manufacturing are useful in agriculture. Pieper (1998, p. 38) offers some empirical support and finds ‘a leading role for industry in determining the level and trend of aggregate productivity growth’. Her results are consistent with the concluding remarks of Matsuyama (1992, p. 330) where he acknowledges that ‘learning experiences in manufacturing should be useful in agriculture’. An alternative approach to productivity growth is the Kaldorian tradition based on returns to scale (see overview by Skott, 1999). Agriculture is a growth constraint because of decreasing returns to scale. In the models of Canning (1988) and Skott (1999), increasing returns to scale in industry relax the long-run constraint agriculture places on growth. Assuming constant returns to scale in industry, Thirlwall (1986) finds that only productivity growth in agriculture matters for economic growth.

Below I will formalize the case of LBD in industry and spillover from industry to agriculture. This should be seen as only a starting point for future research in different intersectoral and international spillovers. Sachs and Warner (1995) assume industry as the engine of growth, and allow perfect spillover from industry to agriculture. Here I assume a technology gap whereby agriculture is catching up on industrial productivity along the lines of Rattsø and Torvik (2003). Productivity growth in agriculture depends on the technology gap between industrial and agricultural productivity. This endogenization of agricultural productivity growth takes advantage of the technology gap concept as introduced by Nelson and Phelps for the world technology frontier and applied by Krugman (1979) in a North–South analysis. I model the technology gap between sectors within a country, rather than between countries. Technically I will use the gap specification of Benhabib and Spiegel (2005).

Compared to the static analysis in the previous section, the dynamics of sectoral balance typically involves more competition. The standard extension into dynamics is to add investment allocation and capacity-building in the sectors. The sectors now compete for scarce investment funds in much the same way as they compete for labor. Since factor market competition is already included in the labor market, we concentrate on productivity growth which has richer dynamics.

The set of equations describing the productivity growth in industry and agriculture are given by (17.7) and (17.8). Productivity growth in industry is related to the industrial labor share n , where α and δ are parameters. α is the elasticity of productivity growth with respect to the labor share. δ is a learning parameter determining the level of productivity growth:

$$\frac{\dot{M}_t}{M_t} = \delta n_t^\alpha \quad (17.7)$$

The productivity growth in agriculture increases by u units per unit rise in the technology gap $G = 1 - A/M$. The gap measures the relative distance of productivity levels between agriculture and industry:

$$\frac{\dot{A}_t}{A_t} = uG_t \quad (17.8)$$

I start out by reproducing the main results of Matsuyama (1992) and then move on to discuss balanced growth in the more general model with endogenous productivity growth in agriculture. Matsuyama applies a special case of my formulation by assuming $u = 0$.

When the agricultural productivity growth is given, equation (17.7) determines the growth rate of the economy. The equilibrium employment share n^* determined by (17.5) feeds into the dynamics of the economy. Since the labor share in industry is rising with the productivity level of agriculture, the growth rate of the economy is increasing in the agricultural productivity level. The understanding is that learning by doing in industry implies complementarity between agriculture and industry. Improved agricultural productivity contributes to industrial productivity growth. Compared to the post-revolutionary controversy in the Soviet Union, Bukharin seems to hold the upper hand. Agriculture and industry feed each other during the growth process. The closed economy model is realistic for large countries like the old Soviet Union. But most countries are small and open, and I move on to discuss them.

Given a constant tariff rate, the growth rate of the industrial labor share is:

$$\frac{\dot{n}}{n} = \left(\frac{\delta}{1 - \alpha} \right) (1 - n_t) n_t^\alpha \quad (17.9)$$

The growth rate of the industrial labor share is determined by the learning process in industry (δ) and the level of the labor share. Introducing γ as the manufacturing output share of total output, the growth rate of the economy output Y_t then follows:

$$\frac{\dot{Y}_t}{Y_t} = \delta \left[\gamma_t + \left(\frac{\alpha}{1-\alpha} \right) (\gamma_t - n_t) \right] n_t^\alpha \quad (17.10)$$

The economy-wide growth rate is basically determined by the learning coefficient δ and the manufacturing labor share n .

Openness represented by tariff policy τ has two conflicting effects on economic growth. In standard fashion, tariffs imply a static allocation loss. The second part of the parenthesis represents this static effect reducing the growth rate in the short run. Positive tariffs are associated with a manufacturing output share lower than the labor share, $\gamma < n$, and consequently the manufacturing productivity growth has less impact economy-wide. On the other hand, tariffs generate a dynamic gain by increasing the labor share in industry. The consequences of openness depend on the empirical parameterization. A possible relationship between the tariff rate and the growth rate is increasing growth with more protection for low values of the tariff, then moving to an area of diminishing growth with more protection for high values of the tariff.

Let us turn to the more general model with endogenous agricultural productivity growth. The dynamics of the model are now best investigated by analyzing the development of the productivity gap G :

$$\frac{\dot{G}}{G} = \frac{\dot{M}}{M} - \frac{\dot{A}}{A} = \delta n^\alpha - uG \quad (17.11)$$

The stability condition for the dynamics of G follows as:

$$\frac{d\left(\frac{\dot{G}}{G}\right)}{dG} = \alpha \delta n^{\alpha-1} \frac{dn}{dG} u < 0 \quad (17.12)$$

Balanced growth follows when equation (17.11) is set to zero. Equation (17.12) is the condition for a stable balanced growth. The framework offers a basis for a discussion of balanced growth. Balanced growth implies that more growth in one sector brings the other sector with it. The productivity gap formulation of the agricultural productivity growth is a mechanism contributing to balanced growth. Technically this is captured by the negative effect of the gap parameter u in (17.12). Given this effect, the stability depends on the relationship between the productivity gap and the labor allocation, as discussed in the previous section. In the closed economy I have shown that higher relative productivity in industry is expected to reduce relative industrial employment. Agriculture must be allowed to increase its production to satisfy the demand effect of more industrial productivity and income. In this case also the first term in (17.12) is negative

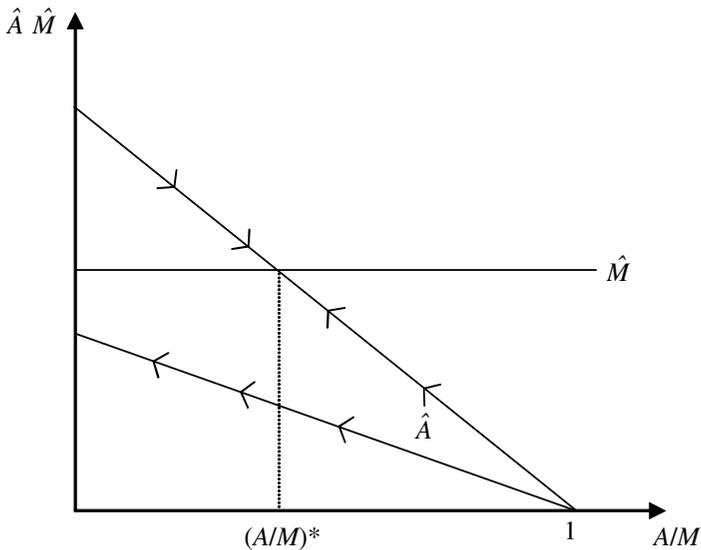


Figure 17.1 Sectoral productivity dynamics in agriculture (A) and industry (M)

and the interpretation is that both sectors contribute to stable steady-state balanced growth in the closed economy. On the one hand agriculture catches up with industry, and on the other hand the LBD in industry responds to the labor allocation. The dynamics here reinforces the conclusion of balanced growth in a closed economy. In the long term the economy grows with the same rate in both sectors determined by LBD in industry.

The economic adjustments involved are described in Figure 17.1. The horizontal axis shows the gap measured as relative productivity in agriculture versus industry. When $A/M = 1$, there is no gap. The further to the left the economy is positioned, the larger is the technology gap. The sectoral productivity growth rates are given on the vertical axis. Productivity growth in manufacturing is determined by LBD and is independent of productivity in agriculture, here written as a constant. When agricultural productivity growth exceeds the growth rate in industry, agriculture is catching up and the gap to industry decreases. The dynamic adjustments are described by the arrows. Balanced growth implies a catching-up process where agricultural productivity growth declines towards the industrial productivity growth and the equilibrium relative productivity $(A/M)^*$ is reached. Unbalanced growth occurs when the agricultural productivity growth line is below the industrial productivity growth for any value of the relative productivity. In this case

relative agricultural to industry productivity is gradually reduced over time and growth of industry overtakes the overall growth. Agriculture is not in a position to catch up the productivity gap to industry.

Economic transformation and growth work differently in the open economy. The dynamics of the open dual economy is much less investigated, in particular taking into account the recent advances in productivity growth modeling. We will emphasize the core differences between closed and open economies here and keep the productivity growth relationships (17.7) and (17.8) constant. The dominating competition for resources in the open economy carries over to the dynamics. As clarified in the second section, the industrial labor share is increasing with industrial productivity according to competitive advantage. When this effect is strong enough the balanced growth condition (17.12) breaks down. The industry sector may take off while agriculture stagnates. This is the type of transformation process discussed by Rauch (1997). Trade liberalization changes the dynamics from balanced to unbalanced growth since openness allows specialization.

It should be noticed that this endogenously determined balanced–unbalanced productivity growth mechanism differs from earlier two-sector models. Van Wijnbergen (1984), Krugman (1987) and Matsuyama (1992) have unbalanced productivity growth by definition, since they have positive productivity growth in one sector and exogenous productivity in the other. Sachs and Warner (1995) have balanced productivity growth by assumption. Rauch (1997) and Torvik (2001) have balanced productivity growth when the elasticity of substitution in consumption is less than one. Higher productivity growth in one sector relative to another increases labor use and learning by doing in the sector with the lowest productivity growth, and in this way contributes to balanced growth. In our setting this mechanism is not present since the elasticity of substitution is set to one.

The sectoral competition for a given labor pool follows from the long-run full-employment assumption. In the medium run surplus labor probably represents a more realistic description of labor markets in developing countries. Under unemployment or surplus labor, policy choices may have more dramatic and different growth effects than under full employment. In my setting surplus labor may hold back the process of reallocating labor from agriculture to industry and consequently higher agricultural productivity is expected to give less effect on the industrial labor share and growth.

Concluding remarks

I have shown how the relationship between agriculture and industry involves both competition and complementarity. The structural transformation from agriculture to industry observed during growth reflects competition for resources in factor markets. Labor and investment funds must

be transferred to expanding industry. But this competition does not fully describe the relationship between agriculture and industry. Agricultural production and growth also contributes to expansion of industry. The key complementary aspects include income generation and demand, cheap food to hold down wage costs, input deliveries for processing, savings to finance investment, and foreign exchange to finance imported inputs.

I have concentrated on the dynamics of agriculture – industry interaction essentially related to intersectoral productivity spillovers. I have challenged the stark conclusion of Matsuyama (1992) that economic growth can be stimulated if production capacity in agriculture is smashed. This can be called the ‘extreme competition between sectors’ result. The argument is easy to understand in an open economy setting. High agricultural productivity attracts resources away from industry with productivity growth based on comparative advantage. The balance between sectors during productivity growth depends on how productivity growth is generated and how it spills over between sectors. I have discussed conditions for balanced and unbalanced growth under the assumptions of LBD in industry and productivity catch-up in agriculture, and acknowledge that the conditions will look different depending on how agricultural productivity is formed.

Recent advances in the understanding of productivity growth can throw more light on the sectoral transformation in a growth process. Future research can do more to marry dualism and productivity growth. Also the basic understanding of both LBD and catching-up are conceptually successful, but with weak (or too many) underpinnings. Productivity growth has been mostly addressed in the developed-economy setting, and the determinants of productivity under backwardness are still an area where research progress can be helpful.

Notes

1. I appreciate collaboration and discussion with Xinshen Diao, Hildegunn Stokke and Ragnar Torvik, and comments and suggestions from the editors.
2. The stylized facts of sectoral issues related to structural change are treated by Moshe Syrquin in Chapter 4 on ‘Structural change’.

References

- Aghion, P. and P. Howitt (2005), ‘Appropriate Growth Policy: A Unifying Framework’, mimeo, Harvard University and Brown University.
- Arrow, Kenneth J. (1962), ‘The Economic Implications of Learning By Doing’, *Review of Economic Studies*, 29: 155–73.
- Benhabib, J. and M. Spiegel (2005), ‘Human Capital and Technology Diffusion’, in P. Aghion and S. Durlauf (eds), *Handbook of Economic Growth*, Vol. 1A, Amsterdam: North-Holland.
- Bukharin, N.I. (1971), *Economics of the Transformation Period*, New York: Bergman Publishers.
- Canning, David (1988), ‘Increasing Returns in Industry and The Role of Agriculture in Growth’, *Oxford Economic Papers*, 40: 463–76.

- Dutt, Amitava (1990), 'Sectoral Balance in Development: A Survey', *World Development*, 18: 915–30.
- Hall, R. and C. Jones (1999), 'Why Do some Countries Produce so much more Output per Worker than Others?', *Quarterly Journal of Economics*, 114 (1): 83–116.
- Klenow, P. and A. Rodriguez-Clare (2005), 'Externalities and Growth', in P. Aghion and S. Durlauf (eds), *Handbook of Economic Growth*, Vol. 1A, Amsterdam: North-Holland.
- Krugman, Paul R. (1979), 'A Model of Innovation, Technology Transfer, and The World Distribution of Income', *Journal of Political Economy*, 87: 253–66.
- Krugman, Paul R. (1981), 'Trade, Accumulation, and Uneven Development', *Journal of Development Economics*, 8: 149–61.
- Krugman, Paul R. (1987), 'The Narrow Moving Band, The Dutch Disease, and The Competitive Consequences of Mrs Thatcher: Notes on Trade in The Presence of Dynamic Scale Economics', *Journal of Development Economics*, 27: 41–55.
- Lewis, W. Arthur (1954), 'Economic Development with Unlimited Supplies of Labour', *Manchester School of Economic and Social Studies*, 22: 139–91.
- Matsuyama, Kiminori (1992), 'Agricultural Productivity, Comparative Advantage and Economic Growth', *Journal of Economic Theory*, 58: 317–34.
- Nelson, R. and E. Phelps (1966), 'Investment in Humans, Technology Diffusion and Economic Growth', *American Economic Review*, 56 (1/2): 69–75.
- Nurkse, R. (1953), *Problems of Capital Formation in Underdeveloped Countries*, Oxford: Oxford University Press.
- Pieper, Ute (1998), 'Openness and Structural Dynamics of Productivity and Employment in Developing Countries: A Case of De-Industrialization?' Employment and Training Papers 14, ILO.
- Preobrazhensky, E. (1926), *The New Economics*, Oxford: Clarendon Press; reprinted 1965.
- Rattsø, Jørn (1988), 'Macroeconomic Adjustments in A Dual Economy Under Policy Controlled Domestic Terms of Trade', *Indian Economic Review*, 23: 45–59.
- Rattsø, J. and R. Torvik (2003), 'Interactions Between Agriculture and Industry: Theoretical Analysis of the Consequences of Discriminating Agriculture in Sub-Saharan Africa', *Review of Development Economics*, 7 (1): 138–51.
- Rauch, James E. (1997), 'Balanced and Unbalanced Growth', *Journal of Development Economics*, 53: 41–66.
- Rodriguez, F. and D. Rodrik (2000), 'Trade Policy and Economic Growth: A Skeptic's Guide to the Cross-National Evidence', *NBER Macroeconomics Annual*, 15: 261–325.
- Ros, J. (2000), *Development Theory and the Economics of Growth*, Ann Arbor, MI: University of Michigan Press.
- Rosenstein-Rodan, P. (1943), 'Problems of Industrialization of Eastern and South-Eastern Europe', *Economic Journal*, 53: 202–11.
- Sachs, Jeffrey and Andrew Warner (1995), 'Natural Resource Abundance and Economic Growth', NBER Working Paper 5398.
- Sah, Raaj and Joseph Stiglitz (1984), 'The Economics of Price Scissors', *American Economic Review*, 74: 125–38.
- Skott, Peter (1999), 'Growth and Stagnation in a Two-Sector Model: Kaldor's Mattioli Lectures', *Cambridge Journal of Economics*, 21: 353–70.
- Taylor, Lance (1989), 'Theories of Sectoral Balance', in J. Williamson and V. Panchamukhi (eds), *The Balance Between Agriculture and Industry in Economic Development*, New York: St Martin's Press/IEA, Vol. 2, pp. 3–31.
- Temple, J. (2005), 'Dual Economy Models: A Primer for Growth Economists', *Manchester School*, 73 (4): 435–78.
- Thirlwall, A.P. (1986), 'A General Model of Growth and Development on Kaldorian Lines', *Oxford Economic Papers*, 38: 199–219.
- Torvik, R. (2001), 'Learning By Doing and the Dutch Disease', *European Economic Review*, 45: 285–306.
- van Wijnbergen, Sweder (1984), 'The "Dutch Disease": A Disease After All?' *Economic Journal*, 94: 41–55.

18 Open-economy issues in development

José Antonio Ocampo

Open-economy concerns have been at the centre of development economics since its origins in the 1940s. They involve two different but connected issues. The first is of a long-term character: the interrelations between economic growth and the patterns of international specialization. The second is of shorter-term nature and involves the source of the business cycle: the central role that trade and capital account shocks play in business cycle dynamics. The debates on these issues over more than half a century have featured different conceptions of the world economy advanced by different schools of economic thought. We thus begin this brief survey of open-economy issues by examining the competing ‘orthodox’ and ‘structuralist’ views, bearing in mind that each of these views is heterogeneous and that some (otherwise) orthodox thinkers have at times expressed ‘structuralist’ views. We then analyze foreign exchange gaps and industrialization strategies, the role of exchange rate in open economies, and capital account issues and their relation to the choice of the exchange rate regime. As a way of conclusion, we draw some implications of the analysis for reform of the international economic system.

A symmetric or an asymmetric world economy?

Traditional trade theory since Ricardo considers a world in which countries specialize according to their comparative advantages. The relative size of the economies involved may matter, and they have to differ in terms of factor endowments and/or technological capacity for specialization to take place. Nonetheless, the economies are viewed as essentially equal partners in their trade relations. The world economy is thus essentially a ‘level playing field’ – although altered by normative rules, which must then be harmonized internationally. A fundamental implication of this analysis is that success in development is essentially determined by each country’s domestic economic management. This is also the dominant view in most orthodox analysis of macroeconomic performance or of the determinants of different rates of growth among developing countries.

An entirely different conception of the world economy is advanced by ‘structuralist’ schools of economic thought, which view the world economy as a ‘center–periphery’ system, to use the terminology of Raúl Prebisch (1950 [1962]). This implies that, although national economic, social and

institutional factors obviously do matter, economic opportunities are largely determined by the position that countries occupy within that world hierarchy. Thus, according to this view, the global economy is not a 'level playing field', and, unless such asymmetries are systemically addressed, world inequalities will persist or may deepen over time.

In the early days of development economics, structuralists tended to distinguish between a centre of the world economy that dominated the production of manufactures and a periphery specialized in the production of primary goods. Because the share of primary goods in gross domestic product (GDP) tends to fall as income grows, due to the low income-elasticity of demand for raw materials (particularly agricultural goods), this pattern of specialization was viewed as leading to lower economic growth in developing countries unless these countries industrialized. Industrialization also offered better opportunities for technological externalities and for technological change in general, and thus for income growth. Both views were expressed early on by Prebisch (1950 [1962]) and Singer (1950). A related issue was that specialization in primary goods subjected developing countries to the sharp cyclical swings of raw material prices and possibly to a long-term downward trend of these prices *vis-à-vis* manufactures (a heated debate, reviewed elsewhere in this *Handbook*).

Over time, the nature of the asymmetries changed. Today, in the early twenty-first century, they may be thought as involving three different issues: (1) technical change and its diffusion throughout the world economy; (2) the way financial markets treat different economies and the degree of macroeconomic policy autonomy that this provides; and (3) the greater international mobility of some factors of production, such as capital and skilled labor, relative to others, particularly to unskilled labor (Ocampo and Martin, 2003). The first two issues will be considered below; the third falls outside the scope of this chapter.

The first asymmetry reflects the high concentration of technological progress in the developed countries. This means that the major engine of world economic growth, technological change, is characterized by its transfer from the centre to the periphery (or peripheries), a process that may be 'slow and uneven' according to Prebisch's predicament. The 'product cycle' literature of the 1960s analyzed some of the features of this transfer (for example, Vernon, 1966). A related literature showed that technology gaps generate income differentials among countries (for example, Krugman, 1990, Chapter 9). As control of technology is one of the most distinctive advantages of firms, the concentration of technological change in the developed countries will also be reflected in the clustering of transnational enterprises in the industrial centers (financial channels may also be relevant in this regard).

The major implication of this is that growth in developing countries is primarily associated with the spread of new sectors and products, technologies and organizational or commercial strategies previously developed in the industrial centers (Ocampo, 2005). In the developing countries this has involved strategies of import substitution, export promotion or a mix of them. These strategies may actually be complementary. Because adapting and mastering new technologies involves a learning process (or, more generally, dynamic economies of scale), import substitution may be necessary for successful export performance. Indeed, Chenery et al. (1986) found that all successful manufacturing exporters had experienced a previous phase of import substitution, and Krugman (1990, Chapter 12) formalized this argument as the case of 'import substitution as export promotion'. The process involves not only mastering and adapting technology, but also generating market information and building a reputation that allows developing-country firms successfully to break into established production and marketing channels. Entry costs may turn out to be prohibitive for new firms; in this case, the possibilities open to developing countries will be limited to attracting established multinationals that are searching for new places to locate their production activities.

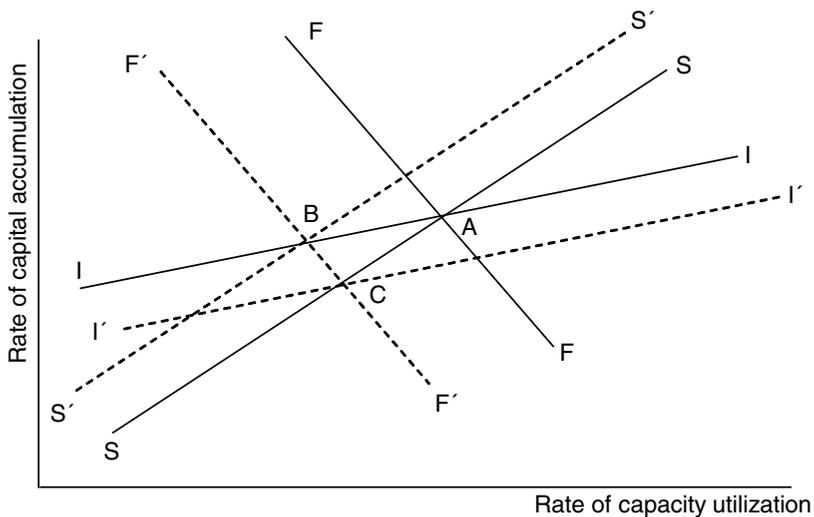
The second asymmetry is associated with the greater financial and macroeconomic vulnerability of developing countries to external shocks. This is an essential issue to the extent that external shocks play a prominent role in business cycle dynamics in developing countries. The nature of this vulnerability has also been changing over time. While the transmission of external shocks through trade (particularly through the commodity terms of trade) remains important, procyclical private capital flows have come to play a more prominent role in recent decades. In the financial area, such asymmetries reflect three basic features: (1) the incapacity of most developing countries to issue liabilities in their own currencies, a phenomenon that has come to be referred to as the 'original sin' (Eichengreen et al., 2003); (2) differences in the degree of domestic financial and capital market development, which leads to a domestic undersupply of long-term financial instruments in developing countries; and (3) the small size of developing countries' domestic financial markets *vis-à-vis* the magnitude of the speculative pressures they may face. This implies that domestic financial markets in the developing world are significantly more 'incomplete' than those in the industrial world and thus that some financial intermediation must necessarily be conducted through international markets. As a result, developing countries are plagued by variable mixes of currency and maturity mismatches in the balance sheets of their economic agents. Furthermore, assets that are classified by markets as risky are subject to cyclical swings in the 'appetite for risk' and thus tend to behave in a procyclical pattern.

Financial asymmetries generate, in turn, macroeconomic asymmetries, particularly in the capacity of developing countries to undertake countercyclical macroeconomic policies. Industrialized countries, whose currencies are the international currencies, have larger degrees of freedom to undertake countercyclical macroeconomic policies and induce a stabilizing response from markets. In contrast, developing countries have more limited degrees of freedom to do so, and face procyclical pressures from financial markets, which generally lead them to adopt procyclical macroeconomic policies (Kaminsky et al., 2004). In particular, they have limited access to private external financing during crises, thus limiting their capacity to undertake active demand management.

Foreign exchange gaps and the industrialization strategy

The existence of asymmetries in the functioning of the world economy was at the center of the early post-World War II literature on development, which emphasized foreign exchange constraints, or ‘external strangulation’, the term coined by Latin American structuralists. This literature had similarities with the contemporary emphasis in Europe on the ‘dollar shortage’, but the emphasis was placed in the developing countries on dependence on primary goods. In the simple Keynesian formalization presented by Johnson (1967), the essential issue was that the lower income-elasticity of the demand for raw materials meant that commodity-dependent countries would either face slower economic growth or a tendency for raw material prices to decline, with the price-elasticity of demand for raw materials determining how large this effect was. Cyclical swings in commodity prices represented an additional challenge, as it became clear with the downswing of commodity prices following the Korean War.

The literature on the ‘foreign exchange gap’ formalized the nature of constraints and policy choices involved (Chenery and Bruno, 1962). The corresponding macroeconomic dynamics can be shown in a simple diagram linking the rate of capital accumulation (or the rate of potential output growth, if capital–output ratios are constant) and the rate of capacity utilization (see Figure 18.1). The savings function *SS* slopes upwards due to the effects of higher private income on private savings, higher tax revenues on public savings, and higher capacity utilization on the current account deficit and thus foreign savings. The investment function *II* also slopes upwards due to the accelerator effect of capacity utilization on investment. Stability requires a steeper *SS* curve. On the contrary, the balance-of-payments equilibrium curve, *FF*, slopes downwards, as the additional demand for consumer and intermediate goods generated by additional income reduces the foreign exchange available to finance the demand for capital goods; it is also possible that additional domestic spending may crowd out the available



Source: Based on Taylor (1994).

Figure 18.1

supply of exportable goods and services as income grows. External financing is assumed to be exogenous. Among the factors held constant in drawing the curves are the exchange rate and income distribution.

Full equilibrium requires that all curves intersect, at point A . Suppose, however, that export prices fall or external financing declines. FF shifts towards $F'F'$, generating a balance of payments deficit at the initial equilibrium. Investment will fall, due to lower profitability of export activities or reduced financing, but equilibrium may not be restored. (The savings function could also be affected, but for simplicity we will assume that it is not.) If the deficit is not stopped, the drain on foreign exchange reserves will be reflected in a reduction in the money supply and in domestic credit, further reducing investment. If these automatic adjustment mechanisms are allowed to operate without any policy intervention, equilibrium will be eventually restored at point C , with lower capacity utilization and capital accumulation.

Alternatively, the restoration of equilibrium may involve an upward shift of the SS function, towards $S'S'$. The mechanism is likely to be, in this case, the inflationary effects of the nominal devaluation induced by the excess demand for foreign exchange, which would generate both an inflation tax on monetary balances and 'forced savings' (reduced real wages, which imply that income is shifted away from workers, who have a higher

propensity to consume). In this case, equilibrium would be restored at point B: the rate of capital accumulation is kept at a higher level, but only at the cost of a stronger domestic recession.

This implies that in economies subject to external shocks and limited availability of external financing, Keynesian stabilization instruments are ineffective. Reduced foreign exchange availability will lead to both domestic recession and investment (potential output growth). The room for maneuver of the authorities would be limited to choosing the mix between the current level of economic activity and the potential growth rate. This trade-off would in practice be especially stringent given the large import content of intermediate and capital goods (particularly machinery and equipment) that characterizes developing countries. To increase the room for maneuver, demand management must be substituted by policies aimed at the direct determinants of the foreign exchange constraint.

Through the first decades of the post-World War II period, the typical response of developing countries was import substitution; as the growth of international trade began to open new opportunities in the 1960s, several countries added export promotion to the strategy, generally mixed still with some degree of import substitution. The basic rationality of both strategies is that they allowed the FF function to shift outwards. Also, substituting imports of consumer and intermediate goods allowed greater room for imports of capital goods, thereby reducing the stringency of the trade-off between maintaining the current level of economic activity and the potential growth rate. Industrialization also offered the opportunity to accelerate economic growth through the greater opportunities for technical change and externalities, that is, through speeding up the transfer of technical change from the industrial center.

This 'structuralist' industrialization strategy was criticized on several grounds. In macroeconomic terms, the first was that many of the new import-substitution activities had high import requirements, so that the strategy showed diminishing returns in terms of shifting the FF curve. This criticism was basically accepted by defenders of the import-substitution strategy, particularly in small economies, and it led to greater emphasis on export diversification and economic integration among developing countries. A second argument was that protection led to a redistribution of factors of production from export into import-substitution activities, thus generating an 'anti-export bias', so that the effect on FF was uncertain. This criticism had three major flaws: (1) the strategy was aimed at increasing available resources (this was, indeed, its major justification), which means that assuming they were constant was inappropriate; (2) as we have seen, through learning, the strategy could in fact generate new export sectors in the long run; and (3) import substitution could be undertaken at

the cost of reducing the production of non-tradable goods and services. As the first two elements imply, the 'structuralist' defense of industrialization was a dynamic argument, and not a static one.

Regarding the longer-term effect on overall technological change, the critics basically argued that the cumbersome structure of incentives to enhance import substitution, as well as export promotion, generated inefficiencies, which could well swamp the assumed positive effects of the strategy in terms of technological change. This would be even more so if quantitative restrictions (import quotas or export requirements) rather than price instruments were used, and if state firms were involved in developing the new sectors. Furthermore, the search for protection led resources to be allocated to 'rent-seeking' rather than to production activities. In short, as in the previous case, static efficiency arguments were used against an essentially dynamic case for industrialization.

Despite all the criticisms, the historical record indicates that the 'structuralist' strategy left a legacy of strong industrialization and rapid growth in most of the developing world up until the oil shock of 1973 and, in several regions, until the end of that decade. Moreover, this development was not inconsistent with the fact that a growing number of countries had been benefiting since the mid-1960s from expanding world markets for manufactures. Thus, according to the structuralist interpretation, import substitution created the industrialization base that allowed many (particularly middle-income) developing countries to benefit from the new opportunities provided by the expansion of world trade for manufacturing goods.

The exchange rate and relative prices

The call for a greater focus on static efficiency was a call to 'get the prices right'. This basically implied reducing the role of non-market mechanisms (quantitative restrictions and state firms), using the exchange rate rather than trade policy as the essential mechanism to manage the balance of payments, and generally letting resource allocation work to guarantee the specialization of the developing countries according to their comparative advantage. This call contained an important grain of truth, as the apparatus of trade interventions had many times led to a reduced focus on exchange rate management, or even transformed exchange rate policy into a mere appendix of trade policy through the common use of multiple exchange rates. It was recognized that the transition to a more 'neutral' policy stance could involve costs, although this warning was frequently ignored in specific liberalization packages.

In macroeconomic terms, what this criticism of structuralist policies implied is that there was not, in fact, a 'foreign exchange gap'. Rather, any disequilibrium in the balance of payments was the result either of

excessively expansionary demand policies or of overvaluation. This coincided with the views that have been expressed by the International Monetary Fund (IMF) in its financial programming exercises introduced in the 1950s. If the problem was excess demand, the answer was austerity; if it was overvaluation, the answer was exchange rate devaluation. According to this view, aggregate demand policies could be used to shift the SS and II functions so that they would cross at full employment; the exchange rate could then be fixed at a level which would guarantee that the FF function would cross the other curves at the full employment equilibrium.

There are, however, several complications to this argument. The first and most obvious is that there may not, in fact, be a feasible FF that can achieve this result, due to certain constraints: existing capacity in exportable sectors (and existing marketing networks and external demand under conditions of imperfect competition for exportable goods) may be limited; given specialization patterns, the demand for imports, particularly for intermediate and capital goods, may be too high; and available external financing may be insufficient. This complication thus manifests the strong meaning of a 'structural' foreign exchange constraint.

A second complication is that the authorities control the nominal, not the real exchange rate (or the relative price of tradable to non-tradable goods and services, which is the favorite measure of the exchange rate in the theoretical literature). If the exchange rate has supply and not only demand effects, domestic price dynamics will be closely linked to the nominal exchange rate, particularly through the effects of imported intermediate and capital goods on costs of production; this effect will be amplified if other domestic prices respond to the initial surge in inflation and, particularly, if the exchange rates and the accompanying inflation lead to a rise in nominal wages or in inflationary expectations. Real devaluation is thus limited by the 'pass-through' of devaluation to inflation, which tends to be higher in economies with strong inflationary traditions, as well as in small economies, which have a smaller array of non-tradable goods and services. As inflation has come down in the developing world in the 1990s, the pass-through coefficients have tended to fall, rendering this argument somewhat less relevant.

A third set of complications were identified by the literature on the contractionary effects of devaluation. The basic rationale in this regard is that devaluation also affects aggregate demand: it can shift the SS and II schedules in an adverse manner following devaluation, reducing the current level of economic activity. Two basic arguments were laid down in the early literature (Krugman and Taylor, 1978). First, if an initial trade deficit exists (as is typical when countries devalue), the deficit will initially increase when measured in the domestic currency, generating a larger demand leakage.

Second, the inflationary effect of devaluation reduces the real income of workers and can thus result in 'forced savings'. Both effects shift the SS curve upwards. A third argument highlights a wealth effect and has been the focus of the recent literature. Because the liabilities of developing countries tend to be denominated in foreign currencies, real devaluation generates a negative wealth effect that also shifts the SS curve upwards and forces firms and the government to clean their balance sheets at the cost of reduced investment (II shifts downwards). An interesting corollary of these arguments is that devaluation may contribute to adjustment less through its substitution effects (that shift the FF curve to the right) than through its contractionary effects (that shift the point at which SS and II cross to the left).

These complications may not, in any case, eliminate the rationale for a more active exchange rate management. Consequently, exchange rate overvaluation has come to be seen as a major source of macroeconomic crises in developing countries (see, for example, Edwards, 1989). As shown below, the real issue here is the effective autonomy that developing countries have in their macroeconomic policy to manage externally generated cyclical swings. Exchange rate equilibrium (a concept that may be difficult to operationalize) or even some degree of exchange rate undervaluation is seen as having long-term development benefits. Certain successful development experiences, particularly in Asia, have been read by some in this light.

This issue came across most clearly, although indirectly, in the literature on the 'Dutch disease'. The essential issue stressed in this regard is that a discovery of new natural resources or a boom in commodity prices tends to generate an appreciation of the real exchange rate through either nominal appreciation or domestic inflation. In the latter case, the essential mechanism is the increased demand for non-tradable goods and services (for example, construction and services), which also raises their prices relative to those tradable goods that are not subject to the price boom, particularly manufacturing exports and importables. Relative price shifts and their induced reallocation of resources towards exportable natural resources and non-tradable sectors have adverse effects on the production of importable goods, particularly manufacturing, as well as on manufacturing exports. Such deindustrialization could have adverse long-term effects if industrial activities are subject to dynamic economies of scale (Krugman, 1990, Chapter 7; van Wijnbergen, 1984). This analysis could be extended to any form of foreign exchange abundance, including excessive private capital flows, external aid or workers' remittances.

Although generally correct, the emphasis on an adequate management of the real exchange rate versus the reliance on trade policies for development purposes tended to overlook one major point: the two are quite

different instruments. The exchange rate is a macroeconomic instrument that affects not only resource allocation but also the price level, capital flows and balance sheet valuations. Trade policies lack these macroeconomic effects (they have at most weak effects on the overall price level); most importantly, they can be selective in their effects. If dynamic economies of scale (learning and market penetration) are a feature of only certain sectors (for example, infant industries or new economic activities more generally), exchange rate policies would be a rather blunt instrument to promote these sectors. And if only limited autonomy exists to manage the exchange rate, substituting more neutral exchange rate policy for trade instruments would imply a net loss of available policy instruments to manage external shocks and external sources of business cycles.

Capital account issues and the choice of the exchange rate regime

Previous debates focused mainly on the patterns of specialization and the current account of the balance of payments. Since the 1970s, with the return of private capital flows to developing countries, the focus shifted to how to manage the boom–bust pattern that characterizes these flows. This debate was closely linked to that associated with the degree of macroeconomic policy autonomy when capital is highly mobile and to the debate on what this implies for the choice of the exchange rate regime.

A common way to express the challenges involved is the ‘impossible trinity’ depicted in Figure 18.2. According to this framework, in the presence of capital mobility, countries can manage the exchange rate while giving up any attempt to manage interest rates (side A of the triangle), or they can manage the interest rate while giving up any control of the exchange rate (side B). This analysis easily leads to the view that only two

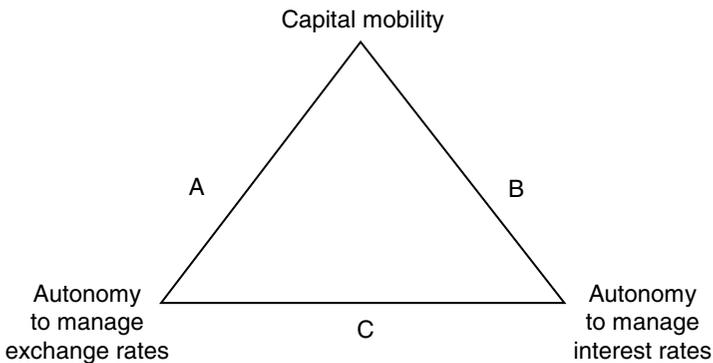


Figure 18.2 The ‘impossible trinity’

'polar' exchange rate regimes are sustainable in a world of capital mobility: either a hard peg or fully flexible exchange rates. Of course, countries could maintain capital controls and then attempt to strike a balance between some degree of exchange rate and monetary autonomy (side C of the triangle). This is the policy choice that many heterodox economists would prefer (see, for example, Stiglitz et al., 2006). But orthodox economies have significant concerns about the capacity of authorities to manage capital controls, as well as their convenience (see the review of this issue elsewhere in this *Handbook*).

The limitations on policy choices are severe in the face of the procyclical character of capital flows to developing countries. In the case of a hard peg, authorities would have to give up countercyclical monetary policy altogether. Fiscal policy could still be used as a countercyclical tool, but experience indicates that this possibility is also limited. Particularly during crises, authorities in developing countries tend to adopt procyclical fiscal policies, due to reduced revenues and limited (non-inflationary) financing. This implies that, in the face of capital mobility, authorities would have to forego any hope of counteracting the procyclical effects of capital flows and the associated real macroeconomic volatility.

There is widespread evidence that capital account liberalization brings additional macroeconomic volatility without any clear benefits in terms of additional economic growth (Prasad et al., 2003). Such volatility is important in its own right, as it gives rise to real income and employment volatility. But it may also affect long-term growth because it increases the risk and uncertainty surrounding investment decisions, reduces the average utilization of existing production capacity and thus also reduces the incentives to invest. Moreover, in the presence of dynamic economies of scale, adverse macroeconomic shocks may change growth trajectories.

If, under the alternative choice, exchange rates are allowed to swing at the rhythm of procyclical capital flows, different types of costs would be incurred. According to the traditional argument against fixed exchange rates, exchange rate volatility implies that investors in tradable sectors face unstable incentives, which reduces the benefits of international specialization. In macroeconomic terms, capital account booms could have negative 'Dutch disease' effects that may affect macroeconomic dynamics in the subsequent bust, as well as long-term growth. In turn, the inflationary effects of devaluation in the face of capital outflows may force authorities to adopt, in any case, contractionary monetary policies, thus implying that monetary autonomy is also limited.

Some of the dilemmas posed by the polar regimes can also be seen in the light of the controversy over monetary rules. The crucial issue in this regard is the link between the exchange rate and the domestic price level in open

economies – that is, the supply effects rather than the demand effects of exchange rates. Under these conditions, two commonly used rules are procyclical: anchoring the price level to a fixed exchange rate during periods of foreign exchange abundance, and counterbalancing the inflationary effects of devaluation through contractionary monetary policies during periods of foreign exchange scarcity. Expressed in terms of the literature on open-economy inflation targeting, strict inflation targeting will generate more output volatility than flexible inflation targeting, which takes into account other objectives of monetary policy, particularly reducing the output gap (Svensson, 2000).

These considerations imply that in today's open developing economies, the exchange rate regime is subject to conflicting and not easily reconcilable demands (Ocampo, 2003). The first is a demand for stability that comes from trade, but also from concern to maintain domestic price stability and to avoid the procyclical wealth effects of exchange rate fluctuations. The demand for flexibility comes from the need to have some degree of freedom to manage trade and capital account shocks. Authorities will thus tend to choose the exchange rate regime based on their preferences, but also on the relative benefits of flexibility versus stability, which are determined by both the stability and instability of the external environment and by domestic factors (particularly the size of the economy involved).

Another way to view these conflicting demands is to recognize that a broad framework for stability implies that macroeconomic authorities have, in fact, multiple objectives: low inflation but also smoother business cycles; competitive real exchange rates; stable long-term interest rates; and sound balance sheets. The frequency of and the case for 'intermediate' exchange rate regimes (Williamson, 2000) may be read as a reflection of the revealed preference of authorities in the developing world to strike a balance among conflicting objectives. Such regimes can take several forms: (1) quasi-fixed exchange rate regimes with large central bank interventions in foreign exchange markets; (2) managed exchange rates, such as crawling pegs and bands; and (3) dirty floats. All these regimes could be understood as integrating an element of 'real exchange rate targeting' into the design of macroeconomic policy, and many or most of them are also mixed with different capital account regulations. To the extent that the contractionary effects of devaluation are effective, smoothing out real exchange rate fluctuations has a countercyclical effect. Therefore, under these conditions, 'real exchange rate targeting' turns out to be complementary with the objective of smoothing output volatility.

Intermediate regimes may thus provide a better framework for effective macroeconomic (and, particularly, monetary) policy autonomy than do floating exchange rates. This approach implies, of course, that monetary

authorities will not have a single objective and that they will coordinate their actions with the fiscal authorities. In any case, however, the scope for policy autonomy is limited. First of all, such autonomy will depend on the effectiveness of capital account regulations. Secondly, intermediate regimes will generally require sterilized intervention in foreign exchange markets to avoid swings in international reserves from being reflected in domestic monetary aggregates. Such intervention, however, may be partly ineffective (as it increases domestic interest rates when reserve accumulation is sterilized, thus inducing additional capital flows, while having the opposite effect during crises). It also generates quasi-fiscal losses. Thirdly, all intermediate options are subject to speculative pressures if they do not generate credibility in markets, and the costs of defending the exchange rate may be high in this context. This is particularly true of any preannouncement (of the rate of the crawl, of a band, or of a specific exchange rate target). Nonetheless, one of the advantages of intermediate regimes is that flexibility can be graduated, depending on the relative benefits of stability versus flexibility, which change throughout the business cycle.

Reform of the international economic system

While an examination of international economic reform *vis-à-vis* developing countries lies outside the scope of this chapter, our survey of open-economy issues gives rise to a few closing observations on the features of the international economic system. An important implication of the view that the international system has asymmetric features is that successful development outcomes will depend on global economic reforms, not only on country-specific strategies and policies. In particular, the first asymmetry implies that the multilateral trade system must facilitate the smooth transfer to developing countries of the production of primary commodities, technologically mature manufacturing activities and standardized services. The system should, therefore, avoid erecting obstacles to such transfers through protection or subsidies. Moreover, it must also accelerate developing countries' access to technology and ensure their increasing participation in the generation of technology and in the production of goods and services with high technological content.

The second asymmetry in turn implies that, from the perspective of the developing countries, the essential function of the international financial institutions is to compensate for the procyclical impact of financial markets, by smoothing financial boom and bust at its source, through adequate regulation, and by providing a larger degree of freedom for countries to adopt countercyclical macroeconomic policies. An additional and equally essential function is to act as a countervailing force to the concentration of credit in private capital markets, making resources available to

countries and economic agents with limited access to credit in international capital markets.

References

- Chenery, Hollis and Michael Bruno (1962), 'Development Alternatives in an Open Economy: The Case of Israel', *Economic Journal*, 72.
- Chenery, Hollis, Sherman Robinson and Moshe Syrquin (1986), *Industrialization and Growth: A Comparative Study*, New York: Oxford University Press for the World Bank.
- Edwards, Sebastian (1989), *Real Exchange Rates, Devaluation, and Adjustment: Exchange Rate Policy in Developing Countries*, Cambridge, MA: MIT Press.
- Eichengreen, Barry, Ricardo Hausman and Ugo Panizza (2003), 'Currency Mismatches, Debt Intolerance and Original Sin: Why they are not the same and why they matter', NBER Working Paper 10036, October.
- Johnson, Harry G. (1967), 'Increasing Productivity, Income-Price Trends and the Trade Balance', *International Trade and Economic Growth*, Cambridge, MA: Harvard University Press.
- Kaminsky, Graciela L., Carmen M. Reinhart and Carlos A. Végh (2004), 'When It Rains, It Pours: Procyclical Capital Flows and Macroeconomic Policies', NBER Working Paper Series, No. 10780, September.
- Krugman, Paul R. (1990), *Rethinking International Trade*, Cambridge, MA: MIT Press.
- Krugman, Paul R. and Lance Taylor (1978), 'Contractionary effects of devaluations', *Journal of International Economics*, 8.
- Ocampo, José Antonio (2003), 'Developing Countries, Anti-Cyclical Policies in a Globalized World', in Amitava K. Dutt and Jaime Ros (eds), *Development Economics and Structuralist Macroeconomics: Essays in Honor of Lance Taylor*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.
- Ocampo, José Antonio (2005), 'The Quest for Dynamic Efficiency: Structural Dynamics and Economic Growth in Developing Countries', in José Antonio Ocampo (ed.), *Beyond Reforms: Structural Dynamics and Macroeconomic Vulnerability*, Palo Alto: Stanford University Press and Economic Commission for Latin America and the Caribbean, Chapter 1.
- Ocampo, José Antonio and Juan Martín (2003), *Globalization and Development*, Palo Alto: Stanford University Press and Economic Commission for Latin America and the Caribbean.
- Prasad, Eswar S., Kenneth Rogoff, Shang-Jin Wei and M. Ayhan Kose (2003), 'Effects of Financial Globalization on Developing Countries: Some Empirical Evidence', Washington, IMF, Occasional Paper 220.
- Prebisch, Raúl (1950), *The Economic Development of Latin America and its Principal Problems*, New York: United Nations; reprinted (1962), in *Economic Bulletin for Latin America*, 7.
- Singer, Hans W. (1950), 'The Distribution of Gains between Investing and Borrowing Countries', *American Economic Review*, 40.
- Stiglitz, Joseph, José Antonio Ocampo, Shari Spiegel, Ricardo Ffrench-Davis and Deepak Nayyar (2006), *Stability with Growth: Macroeconomics for Development*, New York: Oxford University Press.
- Svensson, Lars E.O. (2000), 'Open-Economy Inflation Targeting', *Journal of International Economics*, 50.
- Taylor, Lance (1994), 'Gap Models', *Journal of Development Economics*, 45.
- van Wijnbergen, Sweder (1984), 'The Dutch Disease: A Disease After All?' *Economic Journal*, 94: 41–55.
- Vernon, Raymond (1966), 'International Investment and International Trade in the Product Cycle', *Quarterly Journal of Economics*, 80.
- Williamson, John (2000), 'Exchange Rate Regimes for Emerging Markets: Reviving the Intermediate Option', *Policy Analyses in International Economics*, 60, Washington, DC: Institute for International Economics, September.

PART IV

FACTORS IN DEVELOPMENT

19 Savings, investment and capital accumulation

*Andrés Solimano and Mario Gutierrez*¹

One of the most complex and empirically unsettled subjects in economics is the explanation of the process of economic growth. As the creation of wealth is of critical importance for the welfare of most people around the world the current disarray in growth economics is not only a topic of analytical interest but also of practical importance. One of the controversies in growth analysis is the relative role of capital accumulation and productivity growth in driving output growth. As we interpret the evidence, discussed in this chapter, part of the controversy on the role of capital accumulation in the growth process is due to the time span of the analysis (growth transitions versus steady states and long-run growth). In fact, the empirical importance of various growth determinants will depend on what we want to explain: long-run growth, say growth over half a century or a century as compared to growth dynamics over one or two decades. New evidence is showing that growth fluctuations at frequencies of a decade or so are a very important part of the growth story for most countries, except probably high per capita income economies. Growth is an irregular and volatile process in which the same country over a period of several decades may experience various shifts in growth regimes that can entail growth take-off and booms, stagnation and/or growth collapses. The description of steady growth around a well-defined and stable trend is clearly not a good description of the actual growth experience for most economies in the world, certainly not for developing countries. In this chapter we review and examine the main determinants of savings and investment and their links to economic growth. A growth boom can be driven by a positive terms-of-trade shock, the discovery of natural resources or the adoption of pro-growth economic policies. To support and consolidate growth beyond a boom phase, investment is a critical vehicle to create productive capacities and probably generate knowledge spillovers and new technologies. At the same time, ensuring an adequate level of national savings is important as foreign savings can be volatile and lead to ‘sudden stops’ that force costly macro-economic adjustment and eventually growth crises.

The relation between savings and investment involves analytically important and critical policy issues of great relevance. First, the discrepancies

between intended savings and desired investment creates macroeconomic fluctuations and growth cycles in a world of less than perfect price and wage flexibility. Second, the causality between savings, investment and growth can run in various directions, depending on how the economist theorist views the working of the economic system at macro level. Third, in a world of capital mobility we want to know how close is the relationship between domestic savings and domestic investment.

This chapter examines various topics around savings, investment, their determinants and the relationship between them (particularly in a world of increased international capital mobility) and to economic growth. The chapter first discusses, briefly, alternative causality lines in the relationship between these three variables, putting them in the perspective of macroeconomic theory and growth economics. In addition, we show how different schools of economic thought 'close' the relevant economic model. Second, the chapter looks at the main determinants of savings and investment from a national point of view, highlighting transmission channels and empirical evidence that are more relevant for developing countries. Third, the chapter reviews recent empirical evidence on the role of capital accumulation in accounting for growth both during shifts between different growth regimes and in the medium and long run. Fourth, the chapter discusses the relationship between domestic savings and domestic investment in a world of capital mobility, the so-called Feldstein–Horioka 'puzzle'. The chapter also discusses the evolution of global savings–investment balances in a historical perspective starting from the period of the gold standard and the first wave of globalization of the second half of the nineteenth century until World War I, the inter-war period and the post-1970s to late twentieth-century wave of financial globalization that dominates the international economy today. It shows the changing pattern of savings and investment in main economies and the role of savings flows to and from developing countries. The chapter closes with some final remarks on the analytical and empirical results examined in the chapter as well as policy implications for the savings and investment process from a pro-growth perspective.

National growth, savings and investment: causality issues

In the Keynesian and post-Keynesian traditions investment plays a critical role both as a component of aggregate demand (probably the most volatile) as well as a vehicle of creation of productive capacity on the supply side. In post-Keynesian demand-driven models investment still plays a crucial role in determining medium-run growth rates. Most of these models assume unemployment and idle productive capacities. A variant, but assuming full employment of labour, is provided by Nicholas Kaldor who postulated

growth models with changes in functional income distribution as a mechanism of macroeconomic adjustment acting through national savings in which capitalists have a greater marginal propensity to save than workers.

In a different vein we have the Austrian School of Von Mises, Hayek and others. In this school, the real interest rate (relative to the prospective return on physical assets) is the equilibrating variable between the supply of loans (savings) and the demand of loans for productive purposes (investment). An investment boom is created when banks or monetary policy keep the interest rate below the 'natural rate' (a concept developed by the Swedish economist Knut Wicksell), that is to say the interest rate which equilibrates the demand for loans (investment) with the supply of funds (savings).

In the 1950s neoclassical economics gave rise to celebrated long-run, supply-driven, growth models such as Solow (1956). In this model, the rate of technical change, the savings ratio and the rate of population growth are the three parameters that determine the rate of growth of the economy in a steady state. In this model, the investment ratio plays a role only in the transition between steady states (in practice that transition may take a few decades), but not in the configuration of long-run growth equilibrium of the economy. We will see that these transitions are empirically very relevant; in fact, new papers in growth economics are starting to focus more on this rather than on long-run growth. In the Solow model, as said before, there is no independent investment function (a concept central to the Keynes (1997 [1936]) of the *General Theory*). Full wage–price flexibility solves any *ex ante* discrepancy between intended savings and desired investment, avoiding the sort of macroeconomic fluctuations that were the concern of Keynes and Austrian economists alike. In the 'endogenous' growth theory developed since the mid-1980s a new role was recreated for investment to affect long-run growth by making the rate of technical change and productivity growth linked either to the accumulation of physical capital or the accumulation of human capital.

The issue of causality between savings, investment and growth has plagued growth economics since the start. The controversy can be cast in terms of two leading theoretical perspectives: the 'Marx–Schumpeter–Keynes view' versus the 'Mill–Marshall–Solow view' (see Chakravarty, 1993; Solimano, 1997). The first view posits that investment (Keynes, and to some extent, Marx) and innovation (Schumpeter, Marx) are the two variables that drive output growth. In this context, savings adjusts passively to meet the level of investment required to hold macroeconomic equilibrium and deliver a certain growth rate of output. In this view growth leads savings. In contrast, in the Mill–Marshall–Solow approach that channel of causality is reversed as it assumes that all savings are automatically invested and translated into output growth under wage–price flexibility and full employment. As a result,

in the Mill–Marshall–Solow approach savings leads economic growth. The two schools deliver alternative lines of causality between savings, investment, innovation and growth. These causality issues are still relevant in an open economy with capital mobility, as we shall see in a later section.

Determinants of savings

The research on savings has identified some key factors explaining savings rates such as income (both level and growth rate), the degree of macroeconomic stability, foreign borrowing constraints, financial and demographic variables, and income distribution. We also discuss the evidence on the relation between government savings and private savings, and government savings and national savings.

Savings and income

A positive association between national savings and current income levels is observed in both time series and cross-section data (micro and aggregate) as savings (as a proportion of gross domestic product) rises with the level of income per capita. The evidence has found a type of inverted ‘U’ relation between savings and the level of income per capita (Masson et al., 1998). It has become an accepted stylized fact that savings rates rise at the initial stages of development (although not at very low per capita income levels) and decline as the countries reach higher per capita income and more mature development levels (see also Ogaki et al., 1995). In low-income countries that are closer to subsistence levels we may expect that most income be consumed (with little left for savings). Savings rates rise with the level of income although at decreasing rates in line with a decline of investment and growth opportunities at home. Also factors that tend to be associated with lower savings are the ageing of the population, and lower fertility rates that tend to be observed in countries with higher per capital income levels.

Evidence is also extensive on the positive association between savings and growth (see Carrrol and Weil, 1994; Edwards, 1996; Loayza et al., 2000). The permanent income theory implies that consumption is determined by permanent (long-run) income, implying that savings follows current (transitory) growth.² The life-cycle model, first developed by Franco Modigliani, argues that productivity growth makes the working young richer than the old, and that the young will be saving more than the old are dis-saving. Aggregate income growth would follow from increasing the lifetime profiles for succeeding generations.³ In turn, habit formation in consumption is a factor that helps to rationalize the positive correlation between savings and growth. Carroll and Weil (1994) argued that people adjust consumption habits slowly, which makes savings positively related with current growth of income.

Foreign credit constraints

Theory says that one of the purposes of borrowing is to allow people to smooth consumption in the face of shocks. However, consumption will follow current income more closely at low income levels because credit constraints are more binding at those income levels. In contrast, consumption is expected to follow permanent (or expected income) more closely at higher income levels. Foreign credit restrictions are more relevant for low-income and financially distressed middle-income countries; in those cases we should expect that consumption would adjust more to shocks, as smoothing is more difficult. In the context of foreign borrowing constraints, additional foreign savings is likely to lead to higher consumption and, *ceteris paribus*, lower national savings. There is evidence about a negative relationship between national and foreign savings, with the offsetting effect ranging between 50 per cent and 70 per cent (see Schmidt-Hebbel and Serven, 1999).

Financial development, domestic credit constraints, and interest rates

The research on financial development has found an ambiguous effect of financial variables on national savings. Deeper financial markets and strengthened prudential regulation of financial institutions help to enhance saving (and investment) opportunities by offering a wider variety of financial instruments to channel savings and also by providing more security (in the case of effective regulation) to investors. However, financial development is also often associated with an increased availability of credit for consumption, relaxing domestic liquidity constraints. Savings can be discouraged as more credit becomes available, particularly credit for consumption.⁴

The association between interest rates and savings is also ambiguous theoretically (income and substitution effects may work in opposite directions). The income effect produced by higher interest rates may be positive or negative depending whether the saver is a net wealth holder or a net debtor. The (positive) income effect of an increase in interest rates for a net wealth-holder may run in the opposite direction to the substitution effect that induces a cut in current consumption (substituting for future consumption). The empirical evidence on the effects of interest rates on savings has proven to be inconclusive (see Schmidt-Hebbel and Serven, 1999). Some have explored the sensitivity of savings to the rate of interest as a function of income levels. Ogaki et al., (1995) provided evidence showing that savings are more responsive to rates of return at higher income levels. At lower income levels people cannot smooth consumption over time. At higher income levels it is possible to save and dis-save. Thus, according to this evidence the inter-temporal elasticity of substitution between present and future consumption varies with the level of wealth.

Macroeconomic uncertainty

In the literature an important reason to save is the precautionary motive, as people save more at times of uncertainty to anticipate the possibility of difficult times. One such source of uncertainty is of a macroeconomic nature. This can be reflected in high and erratic inflation, exchange rate volatility, cycles of boom and contraction, or instability in the financial system. Capital flight is an economic response to these uncertainties, as people seek to place their assets abroad to escape internal uncertainty and volatility (Edwards, 1996; Taylor, 1996).⁵

Inflation has been a factor strongly associated with macroeconomic instability; however, the effects of low to moderate inflation on savings is bound to be very different from the impact of high or even explosive inflation of the type that destroys the payments and banking systems and financial savings along the way. The classic example is the hyperinflation of Germany in 1923, although there are more recent cases such as Argentina during the hyperinflation of the late 1980s and Brazil in the early 1990s. In 2001–02 Argentina suffered a banking crisis following the abandonment of the currency board adopted in 1991. In this later banking crisis, people (mainly from the middle class) that believed in the system and had deposits in the banks experienced the loss of part of their financial savings.

Fiscal policy

The stance of fiscal policy is expected to affect savings. One channel is the size of the fiscal deficit or surpluses that has been found to affect the level of national saving rates. Low fiscal deficits or surpluses contribute to national savings, as complete Ricardian equivalence has been refuted empirically (that is, an increase in public savings is not fully offset by a decline in private savings). This effect is stronger in developing countries subjected to subsistence consumption and liquidity constraints (see Corbo and Schmidt-Hebbel, 1991). The evidence confirms the partial offset between government and private savings (with the offset coefficient in the range of 40 per cent to 70 per cent. This means that 1 per cent of additional government savings (in terms of gross domestic product) adds about 0.5 per cent of gross domestic product (GDP) to national savings. Another fiscal policy channel is to enact taxation on factors that affect savings such as interest rates, dividends of firms and other variables.

Demographics

The age structure of population is another determinant of national savings. According to the life-cycle hypothesis a larger working population relative to the older population (or young family dependents) contributes to raise national savings. The working young are net savers and the retired elderly

often have negative savings. In economies with higher proportions of working populations the national savings rates would be higher than in ageing economies with higher shares of old people in their populations. Studies using cross-country data have been more successful in confirming the negative effect of dependency ratios (say the share of population below age 15 and above 65) on saving, probably because demographic variables change slowly over time (Masson et al., 1998). Some microeconomic studies conflict with the findings at country levels, which may be partly due to the aggregation of cohorts of different ages in macro studies. Bequests may also contribute to reduce aggregate savings even if the old do not dissave (Carroll and Weil, 1994; Deaton and Paxson, 2000). The literature mostly agrees on a negative correlation between age dependency ratios and national savings, confirming the theory and empirical evidence.

Income distribution and savings

Richer people are expected to save more as a proportion of their income than poor people (savings are, in a way, a superior good). Some formulations make savings dependent upon functional income distribution (for example, Nicholas Kaldor assumed that capitalists have a higher propensity to save than workers) whereas others make a link between personal income distribution and saving. While for the most part, the empirical literature based on cross-section micro-data suggests a positive relation between personal income inequality and overall personal savings, the evidence on this issue is more mixed at aggregate, country level. Empirical studies as Schmidt-Hebbel and Servén (1998) indicate that cross-country data do not reveal a strong association between personal income distribution and aggregate saving. The authors show that this relation holds for samples of developing and developed countries, and is robust to alternative saving measures, income distribution indicators and functional forms.

New political economy literature emphasizes that regressive income distributions are a factor contributing to political instability, and through this channel they may depress both growth and savings. Lower growth contributes to reduce savings through the growth–savings link, but political instability may also discourage savings because of the uncertainties about saving prospects.

National savings and growth

Empirically, national savings and growth are positively associated, especially in the case of developing countries. The evidence also shows that investment and national savings are positively related, reflecting the existence of foreign credit constraints, an issue we take up later. In terms of causality, the research on the determinants of savings has generally

considered growth as a determinant of national savings, suggesting that the causality runs from growth to national savings (the typical regression is one in which national savings is the dependent variable of the regression and GDP growth is a right-hand-side explanatory variable).⁶ The evidence on the association of GDP growth and foreign savings is mixed: there are episodes of high growth with relatively low levels of foreign saving rates (that is, some East Asian economies) and episodes of low growth episodes and high foreign savings (that is, low-income countries in Africa and Latin America that receive sizeable levels of foreign aid).⁷

The issue of causality between savings and growth is more controversial, as discussed before. In the neoclassical growth model à la Solow, saving is exogenously given. In contrast, in the Keynesian school saving is endogenously determined as a result of the interactions between income and consumption. Higher growth generates higher incomes that lead, in turn, to higher savings (as the propensity to consume out of income is less than one). Carroll and Weil (1994) provided strong evidence that growth causes saving (Granger causality), but Attanasio et al. (2000) questioned Carroll and Weil's results, showing that the causality may go both ways depending on the data (sample and frequency of the data) and the econometric technique used to estimate the relationship between both variables.

The determinants of investment

Wicksell, Bohm-Bawerk, Fisher and others (see Patinkin, 1982) developed a capital theory in the late nineteenth century and early twentieth century. However, John Maynard Keynes in *The General Theory* (1997 [1936]) was among the first that postulated an independent investment equation in a demand-driven macro model.⁸ Keynes emphasized that the determinants of savings were of a different nature than the determinants of investment, challenging the classic view, prevailing at the time, that assumed the real interest rate was the key variable that equilibrated savings and investment.

In Keynes it was disposable income and possibly wealth that were the main determinants of savings, whereas investment depended upon the difference between the real cost of capital relevant for firms and the marginal efficiency of capital (or productivity of capital). Expectations were critical in the determination of investment as it was the prospective estimate of the future profitability of capital that mattered for investment decisions. The investment function experienced several refinements and reformulations after Keynes's original formulation. Nowadays, if we want to understand the effect of policies of macroeconomic adjustment and structural reform on investment and growth we need some reformulation of the investment function to incorporate new relevant variables.

Profitability and appropriability

The return of investment is obviously important in investment decisions, but the capacity of investors to appropriate those returns is also ultimately very relevant.⁹ If property rights are weakly enforced, potentially high returns may not induce higher actual investment as the appropriability or internalization of those returns may not take place. The respect of property rights and the capacity to internalize returns from investment require a certain level of trust that guides economic transactions and also a judiciary system that allows contracts to be drawn up and respected at a reasonable cost. Recently, new attention has been devoted to analyze issues of profitability of investment in terms of the 'cost of doing business', a concept that involves the cost of obtaining permits, licences and other requisites to set up a plant and start business. The profitability variable, in turn, is affected by factors such as cash flows, corporate income taxes, depreciation rules and other fiscal policy variables (see Schmidt-Hebbel et al., 1996; Alesina et al., 2002). Political economy variables are also important in the determination of investment. In fact, political stability and social peace are also factors that private investors – national or foreign – attach great importance to. For the returns of capital to accrue to capitalists (that is, to make them appropriate), the risks of destabilizing policies and/or confiscator actions by governments – that have the monopoly of force and law enforcement – have to be low. Also, labour–capital relations must be reasonably harmonious, or at least not conflictive, to ensure social peace. In this vein, a social equilibrium characterized by high inequality, political polarization and/or conflictive labour–capital relations tends to lead to policies that are ultimately against capital and therefore penalize investment. Counterproductive policies are those that try to buy social peace in the short run by artificially raising real wages (that is, through fixing an overvalued exchange rate), or through higher taxation. Also, downright hostility to private property in highly unequal societies may develop with negative consequences for private investment.

Growth cycles and capacity utilization

As indicated before, a stylized fact of the process of economic growth, particularly for developing countries, is the high frequency of cycles of growth take-offs, growth collapses and stagnation. In other words, past growth is often a poor predictor of future growth for a given country.¹⁰ The behaviour of investment in growth cycles (see later) is important. Investment is affected by, and also affects, the type and duration of those cycles in a double-causation fashion. The literature on investment has emphasized the effects of capacity utilization on investment, stressing the fact that investment is deterred in an economy with unused productive capacity and

possibly uncertain expectations by the private sector on the duration and intensity of various stagnationary and recessive cycles (Severn and Solimano, 1993). Empirically, the effects of capacity utilization variables in empirical investment equations are often very strong and statistically significant, but caution is required when establishing causality between investment and the degree of unused capacity.

Fiscal policy and investment

The effect of fiscal policy on private investment acts through at least three channels. Firstly, the fiscal deficit in general tends to reduce private investment through its effects on real interest rates and the absorption of private savings to finance the deficits. It also signals a lack of sustainability of fiscal policy that deters private investment. Secondly, another channel is through the level, composition and quality of public investment that determines the extent to which public investment complements or substitutes private investment (see Schmidt-Hebbel et al., 1996). Thirdly, the level of taxation on corporate earnings and depreciations rules.

The role of uncertainty and irreversibility

Another important topic is the effect of uncertainty on private investment. To explain the channels through which uncertainty affects investment, research in the 1980s and 1990s has developed and tested new theories that highlight the role of irreversibility on investment. As physical capital once installed in a particular firm or sector cannot be changed or disinvested, except at a large cost, in a sense capital once installed becomes 'irreversible' (see Dixit and Pindyck, 1994). This feature of investment makes it very sensitive to risk and uncertainty. In general there is a high 'value of waiting' in an uncertain environment, as firms do not wish to get stuck with an excessive stock of capital in the event that conditions affecting profitability change. In the context of developing countries this is very relevant as economic structures and policies are often more volatile and unstable than in advanced countries. Pindyck and Solimano (1993) investigated the effect of macroeconomic volatility – as measured by level and variance of inflation rates – on the marginal profitability of investment, using a formulation of irreversibility investment constraints (political instability variables were also tried, with no significant statistical results). The paper also studied the slow response of investment after stabilization in several countries suffering from high inflation in the 1970s, 1980s and early 1990s. The long investment pause that led to a slow recovery of economic growth, has been considered in the aftermath of stabilization as a case of increased value of waiting after large macro shocks take place.¹¹

A considerable literature has studied the effect of macroeconomic uncertainty on investment and growth in developing countries (see Serven and Solimano, 1993; Schmidt-Hebbel et al., 1994, among others). In general, topics of interest have been the effects of unanticipated currency devaluations, external shocks, debt problems, financial crises, and other shocks on investment.

Finance and investment

The effect of financial constraints and the structure of finance has been another topic of research on investment (see Summers, 1981; Fazzari et al., 1988). In general, firms have two sources of finance: external (equity, bank loans, bonds) and internal (retained profits, accelerated depreciation). At the margin, the optimal capital structure among different sources of finance is the one in which the marginal cost of different types of finance (adjusted by risk, taxes and currency denomination) is equal among different sources. The problem is that the supply of external financing in developing countries is restricted, particularly for small and medium-sized enterprises. External borrowing may relax internal credit constraints, but again mainly for large and well-connected firms that have access to foreign borrowing. Given the constraints on borrowing and the imperfections of capital markets, retained profits are a main source of investment financing by firms in developing countries.

The composition of investment

The structure of investment by type of assets matters for economic growth because the different types of investment goods have different effects on productivity and growth. Some quantitative studies have emphasized the role of machinery and equipment investment in augmenting the role of physical capital (and labour) in the growth processes. Since the Industrial Revolution machinery investment has played a key role – directly as a production factor, and also as a mean of acquisition and transmission of technological improvements across countries and within countries. De Long and Summers (1991, 1993) found evidence of high social returns from investments in machinery, assigning to machinery investment a primary role in boosting productivity growth (proxied by per capita GDP). They showed that high rates of machinery investment accounted for most of Japan's successful growth experience after World War II. They also concluded that fast-growing countries were those with favourable supply conditions for machinery investments, and that developing countries benefited as much as richer economies from the technologies embodied in machinery. Building projects are usually less effective in promoting growth because the technologies embodied in construction have a lower potential of being

transmitted across the production process. In addition, the output of the construction sector is mostly non-tradeable and technologically less dynamic.

There are also potential complementarities between private investment and public investment (Khan and Kumar, 1997; Khan and Reinhart, 1990; Serven and Solimano, 1992; Greene and Villanueva, 1991), mainly public investment in infrastructure and education.¹² The roles played by foreign direct investment (FDI) have been addressed by Borensztein et al. (1998), Olofsdotter (1998) and Lim (2001).

The empirical role of investment in long-run growth and in growth transitions

In evaluating the impact of capital accumulation and investment on output growth it is useful to draw a distinction between medium to long-run growth processes and growth transitions. In addressing the first issue a strand of the literature tends to attach a greater role to total factor productivity (TFP) growth than to capital accumulation in accounting for output growth. In the words of Easterly and Levine (2001): 'although physical and human capital accumulation may play key roles in igniting and accounting for economic progress in some countries, something else – TFP – accounts for the bulk of cross country differences in the level and growth of GDP per capita in a broad cross section of countries'. The authors find that the contribution of capital growth typically explains less than half of output growth and that the share of TFP is usually larger for fast-growing economies. The issue of causality is important here and growth accounting does not imply causality. Disagreement persists about the role of investment in the growth process. Some authors have concluded that investment has been the main factor explaining economic growth. In a study for East Asia, Young (1994) concluded that investment was the main source of growth in the experience of the East Asian economies, downplaying the importance of TFP growth in the Asian case. Other economists have acknowledged the important role played by fixed investment but argued that productivity has been the engine that has marked the difference between fast and slow growth experiences (Blomstrom et al., 1996; Harberger, 1996, 1998; Klenow and Rodriguez-Clarke, 1997). Elias (1992) produced evidence showing that total factor productivity explained about one-third of GDP growth in Latin America during 1940–85.

Some studies find that output growth causes, in the Granger sense, investment rather than the other way around (Blomstrom et al., 1996). Also as mentioned before in this chapter, Carroll and Weil (1993) show that causality runs from output growth to savings rather than the other way around. Departing from the standard, Barro type of cross-country growth

regressions methodology, new studies have investigated ‘growth transitions’, that is to say processes in which the growth rate of output changes upwards or downwards, that is, growth accelerations and/or growth collapses or growth crises. These studies are Hausmann et al. (2004), Jones and Olken (2005) and Solimano and Soto (2006). They investigate the role of investment and capital accumulation in the transition from one growth regime to another. In Hausmann et al. (2004), growth accelerations (say a significant increase in growth rates relative to a decade or so before a certain turning-year – say the year in which the rate of growth accelerates according to a certain threshold or benchmark criteria) that often last near a decade have been accompanied by an increase in investment and trade and also come with real exchange rate depreciations. In general the pattern seems to be that growth accelerations are correlated with increases in exports, imports and investment ratios but do not seem to be driven by pure accelerations in total factor productivity. The study by Jones and Olken concludes that changes in the rate of factor accumulation (including, of course, capital) explain relatively little about growth reversals, especially growth accelerations; in contrast, according to these authors reversals are ‘largely due to shifts in the growth rate of productivity’. For these authors the weak role of capital accumulation in growth transitions suggests an efficiency story. In fact, Jones and Olken find that growth accelerations coincide with major expansions in international trade (exports and imports). The authors attribute the accelerations in output growth rates to sector reallocations of labour (and other factors of production) towards higher productivity sectors. However the authors detect an asymmetry in accelerations and decelerations with a much larger change in investment in growth decelerations than in accelerations. Solimano and Soto (2006) focused on Latin American growth experiences and cast the analysis in terms of growth cycles and sustained growth episodes. The authors find a higher incidence of growth crises (negative growth) in the 1981–2003 period than in the 1960–80 period; in addition, they show that the countries that were rapid growers before 1980 (that is, Brazil and Mexico) are not the same as those that grew faster after 1980 (that is, Chile and Dominican Republic). The study shows a relatively even importance between capital accumulation and TFP growth in changes in growth regimes, and emphasizes that the TFP story tends to be more of a long-run nature.

Summing up, the empirical evidence on the role of investment in explaining output growth is far from conclusive. Investment plays a greater role in explaining growth transitions (that last around a decade or so) than in accounting for medium-term and long-run growth paths (that last several decades). In turn, the determinants of long-run growth seem to be more in line with the Solow model (and to some extent the endogenous growth

theories) that stress the role of TFP growth in driving long-term GDP growth and highlight that investment is important in the transitions between steady states.

National savings and investment under international capital mobility

In an era of globalization, another important theme is the correlation between domestic savings and domestic investment under international capital mobility. In an influential paper Feldstein and Horioka (1980) (hereafter FH) argued that in a world with perfect capital mobility domestic savers would seek the higher rate of return irrespective of the home or foreign origin of the assets to be invested. In turn, attractive investment projects would find adequate financing irrespective of whether the funds came from the pool of national savings or from foreign savings. The authors pose that under perfect capital mobility, national savings and domestic investment would be largely uncorrelated. However, FH found empirically that, contrary to the predictions of perfect capital mobility theory, there was a strong correlation (which was statistically significant) between domestic savings and domestic investment (a high 'savings retention coefficient') when the relation was tested for cross-section data of industrial economies with (five years' average) data of the 1960s and 1970s. Other authors that tested the relation between national savings and domestic investment using a larger sample of countries and longer time periods further investigated the results of Feldstein and Horioka. Taylor (1996) reports those results of various studies including his own that basically find a close correlation between national savings and national investment, a finding that is relatively robust across space and time although it varies in periods of higher capital mobility (that is, during the gold standard and since the 1970s, a second period of financial globalization). The high correlation of national investments and domestic savings demonstrates that the financial markets are not more integrated today than at the beginning of the twentieth century, although a change occurred between the two periods in the composition of capital flows – especially an increase of the short-term capital flows relative to long-term capital flows (Taylor, 1996; Baldwin and Martin, 1999). In any case, the results of the FH tests reported by Taylor (1996) suggest that the existence of 'home bias' in terms of the allocation of savings towards national assets and towards national investment projects seems to hold.

Let us now briefly review some historical evidence pertaining to this topic. One feature is that countries change their position of net exporter (or net importer) of capital over time.¹³ From the nineteenth century until the 1980s the United States was, on average, a net exporter of capital. After World War I British financial hegemony was replaced by that of the United

States as the main capital exporter of the world economy. The US role as a net capital exporter lasted until the early 1980s when it started to run current account deficits, importing savings from the rest of the world to finance a level of expenditure above its real output,¹⁴ financing the gap with savings from the rest of the world, mainly from positive net savings economies in Asia and also from international reserves held by central banks in developing countries, held mostly in US securities. In addition, the US became a net debtor as its foreign liabilities exceed its net foreign assets. Interestingly, under current conditions, there is a transfer of savings from developing countries (and from 'emerging economies') to the richest economy in the world, that spends more than its income generated by nationally owned factors of production. Thus, national savings are diverted from the financing of growth at home to finance consumption and investment in the richest world economy. In the nineteenth century and up to World War I, a period known as the first wave of globalization, the most important flow of capital occurred from Great Britain to a group of countries known as the 'new World countries' (Argentina, Australia, Canada, New Zealand, the United States). London constituted the financial centre of the global capital market and was called the 'banker of the world'. It is estimated that the surplus of domestic savings over investment in the UK was around 50 per cent in the first decade of the twentieth century (Obstfeld and Taylor, 2004). The British pound was the dominant currency in the context of the international gold standard. The United Kingdom contributed to a peak average of 80 per cent of total global foreign investment.¹⁵

In the early twentieth century capital flows were characterized by the accumulation of enormous one-way positions and a great portfolio diversification by the principal creditor countries, in particular Great Britain, and inversely little diversification and high foreign capital 'dependence' by the debtor New World countries.¹⁶ It is interesting to note that capital flew to rich and labour-scarce New World countries instead of going to poor and labour-abundant Asian and African countries, where it could, in principle, have been more profitable given the abundance of cheap labour. This is the so-called 'Lucas Paradox'.¹⁷ In today's global capital markets, capital flows and foreign investment aim for risk-sharing and diversification instead of long-term financing to build infrastructure and housing, as was the case in the pre-1914 world. Regarding the direction of international capital flows we also face the 'Lucas Paradox', in which too little capital flows to capital-scarce, poor countries. We may think of various reasons why capital does not go to low-income countries: the lack of educated and properly trained workforces in poor countries, the lack of enforceable property rights, bureaucracy, political instability, weak institutions, small domestic

markets and other factors. The literature of growth under increasing returns suggests that capital, skilled labour and superior institutions tend to go together, and concentrate in a certain group of countries (Easterly, 2001) in which they find favourable conditions for international investment. Another difference between the first wave of globalization and contemporaneous financial globalization is the importance of capital flows as a proportion of savings and investment in both source and receiving countries. Although financial globalization since the 1970s and 1980s has expanded very rapidly in relative terms, it is lower than in the pre-1914 world. In fact, Obstfeld and Taylor (2004) report that in 1900–1913 overseas investment represented about one-half of domestic savings of the UK (and one-third, on average, between 1870 and 1914). In other capital-exporting countries such as Germany, overseas investment represented about 10 per cent of national savings in 1910–13. In turn, as said before, around 50 per cent of the capital stock of Argentina in 1914 was in the hands of foreigners (in Canada and Australia that percentage was in the range of 20–30 per cent). These numbers are lower in the new wave of globalization. After 1970 the ratio of net capital outflows over savings in the capital-exporting countries never exceeded 5 per cent (this is influenced by the large current account deficits of the United States). In turn, capital inflows, on average, in the same period never exceeded 15 per cent of investment in capital-importing countries (Obstfeld and Taylor, 2004).

In 2005 the current account deficit of the US was about 6 per cent of its gross domestic product, or near US\$600 billion. In contrast, countries such as Japan, China, Korea are running large current account surpluses, contributing to finance the savings shortfall of the United States.

Final remarks and policy implications

Recent literature on economic growth emphasizes the role of productivity growth in determining output growth, thereby downplaying the contribution of factor accumulation in this process. In this chapter we argue that the role of investment (and factor accumulation in general) is different if the focus is on growth transitions rather than long-run growth. Now, the empirical relevance of growth transitions is highlighted by the fact that the growth process is more a shift between different growth regimes over time rather than steady growth around a stable trend. Growth is characterized by volatility and low correlation between current and past growth rates (low time persistence), particularly in developing countries and transition economies. In this context, the role of investment in these growth transitions is bound to be important. The efficiency story of productivity growth is more appropriate to explain long-run growth within countries, and in explaining cross-country differences in growth performance. Still, a reform

process can trigger short-term productivity gains leading to faster growth initially if economies start from very distorted levels.

This chapter reviews the various determinants of savings such as income, wealth, age structure of the population, credit constraints, macroeconomic volatility and inequality of income and wealth. On the investment side, we underline the role of profitability and appropriability of investment returns and stress the influence of property rights, the cost of doing business, political stability, inequality and quality of capital–labour relations as background factors that affect the appropriability of the returns on capital investment. Other factors that affect investment are macroeconomic uncertainty and volatility, fiscal policy, anticipated and unanticipated changes in policy regimes, and credit constraints. Another topic analysed in the chapter is the extent to which increased capital mobility affects the correlation between domestic savings and domestic investment. Empirical studies show that in spite of growing financial integration there is still a high and significant correlation between national savings and domestic investment, both in time series national data as well as in cross-country data, contrary to the predictions of perfect capital mobility theory. The evidence confirms the existence of home biases in the savings–investment process.

In general, it is apparent that the benefits of international financial intermediation go more to advanced, financially mature economies rather than countries with limited access to private capital financing. Financial integration in a context of speculative and procyclical capital flows can induce macroeconomic volatility and financial crises, disrupting orderly investment processes. Finally, current global economic imbalances in which rich economies have become net capital importers affect the global allocation of savings, and therefore the financing of investment needed for growth in developing countries. Several emerging economies and developing countries have become exporters of capital to developed countries (particularly to the USA). The consequences for global growth of these new patterns of allocation of savings across countries remain to be seen.

From a policy perspective it is important to identify the factors that: (1) accelerate economic growth; (2) maintain a growth momentum once it is reached; and (3) help to avoid traumatic stops of growth (such as growth collapses or growth crises). A main mechanism for igniting growth and possibly generating new knowledge and productivity growth is investment. But investment is still an ‘intermediate variable’ that will be activated if new opportunities are opened by policy reforms, by growing international integration or by the discovery of valuable natural resources. In turn, factors that can boost investment refer to profitability, adequate property rights, reasonable cost of capital, predictable policy environments, absence of

acute social conflict, and others. In the sequence after growth momentum is set in motion, it is important to ensure macro stability and the absence of macro imbalances whose sharp corrections often derail growth. In addition, investment has to be financed some way or another and therefore savings enter into the picture. Assuring an adequate level of national savings is critical, as an excessive reliance on foreign capital can be a risky course of action in a world of imperfect international capital markets and often volatile capital flows. Public savings can be a mechanism to spur national savings, given the empirical evidence showing that an increase in public savings is less than fully offset by a decline in private savings. This analysis illustrates that two critical variables through which public policies can affect growth are savings and investment. The trick is to mobilize the adequate policy instruments that will affect these variables in a desired direction during the different phases of the growth process.

Notes

1. Comments by A. Dutt and J. Ros to a first version of this chapter are greatly appreciated.
2. The terms-of-trade effect is viewed as a transitory deviation of national income from its trend. The Milton Friedman's consumption hypothesis would argue that the additional income resulting from transitory improvements in the countries' terms of trade would be mostly saved.
3. In more extensive models of consumer behaviour the relationship is theoretically ambiguous (Carroll and Weil, 1994).
4. Most likely both effects interact, affecting the results of the effects of financial development on savings (Piles and Reinhart, 1999).
5. Precautionary motives may help to explain the positive association between saving and consumption of young consumers (who expect positive but uncertain future income growth) and the positive saving of retired people (Loayza et al., 2000).
6. In the case of the link between investment and growth, growth in most cases is regressed against investment, therefore implicitly assuming that investment causes growth. A mutual reinforcing process between national savings and growth, and investment and growth, is assumed in the literature (see for example Attanasio et al., 2000; Hausmann et al., 2004; Gutiérrez, 2006; Solimano, 2006).
7. See Gutiérrez (2007) for Latin America 1990–2003.
8. In the early 1930s the Polish economist Michael Kalecki put forward, independently, a somewhat similar formulation to Keynes. See Don Patinkin (1982) for a view in which Keynes and Kalecki formulations of investment are interpreted to be quite different in scope.
9. Rodrik (2006) discusses the role of appropriability in formulating successful growth strategies.
10. This fact was brought to attention by Easterly et al. (1993) and confirmed by subsequent empirical work on growth.
11. In the context of market-based economic reforms in Latin America, Eastern Europe and the former Soviet Union a relatively slow initial reaction of private investment was also observed. This may reflect the lack of a private sector in the former socialist countries but also reflects the effects of uncertainty on the consolidation of largely untried reform packages, again an increased value of waiting at work.
12. The roles of infrastructure have been addressed by Easterly and Pack (2001) for Africa and Moguillansky and Bielchowsky (2000) for Latin America.
13. This analysis draws from Solimano and Watts (2005).

14. In the 1980s and up to 1993 and after 2000, US public sector deficits contributed significantly to create the current account deficits.
15. Between 1907 and 1913, Britain's foreign assets were estimated at £1127 million, from which 61 per cent or £689 million went to Canada, Australasia, Argentina and the United States. This percentage rises to 76 per cent or £857 million if we add the other countries of Latin America (see Taylor and Williamson, 1994; Taylor, 1999).
16. For example, foreigners held one-fifth of the capital stock of Australia and owned almost half of the capital stock of Argentina. Even the United States presented high levels of foreign capital dependence at the end of the nineteenth century, in spite of its increasing domestic savings and investments since the 1830s (O'Rourke and Williamson, 2000, p. 209). Thus, gross assets during this period were almost equal to net assets. Also, investments took the form of long-term finance to less-developed countries, what Obstfeld and Taylor (2004) called 'development finance'. For example, in 1900, one-third of global assets went to countries in Latin America and, to a lesser extent, Asia and Africa.
17. Indeed, the labour-scarce New World countries, where only a tenth of the world's population lived, received two-thirds of the British capital in 1913–14, while labour-abundant Asia and Africa, accounting for two-thirds of the world's population, only received a quarter of European foreign investment (Clemens and Williamson, 2000).

References

- Alesina, Alberto F., Silvia Ardagna and Roberto Perotti (2002), 'Fiscal Policy, Profits and Investment', *American Economic Review*, 92 (3): 571–89.
- Attanasio, O., L. Picci and A. Scorcu (2000), 'Saving, Growth, and Investment', *Review of Economics and Statistics*, 82 (2): 182–211.
- Baldwin, Richard E. and Philippe Martin (1999), 'Two Waves of Globalization: Superficial Similarities, Fundamental Differences', NBER Working Paper Series, no. 6904
- Blomstrom, M., R. Lipsey and M. Zejan (1996), 'Is Fixed investment the Key to Economic Growth?', *Quarterly Journal of Economics*, 111 (1): 296–76.
- Borensztein, E., J. De Gregorio and J. Lee (1998), 'How Does Foreign Direct Investment Affect Economic Growth?', *Journal of International Economics*, 45: 115–35.
- Carroll, C. and D. Weil (1993), 'Saving and Growth: A Reinterpretation', NBER Working Paper No. W4470.
- Carroll, C. and D. Weil (1994), 'Saving and Growth: A Reinterpretation', *Carnegie-Rochester Conference Series on Public Policy*, 40: 133–92.
- Chakravarty, Sukhamoy (1993), 'Alternative Approaches to a Theory of Economic Growth: Marx, Marshall and Schumpeter', in Sukhamoy Chakravarty, *Selected Economic Writings*, New Delhi: Oxford University Press, pp. 51–79.
- Clemens, Michael A. and Jeffrey G. Williamson (2000), 'Where Did British Foreign Capital Go? Fundamentals, Failures and the Lucas Paradox, 1870–1913', NBER Working Paper Series, no. 8028, December.
- Corbo, V. and K. Schmidt-Hebbel (1991), 'Public Policies and Saving in Developing Countries', *Journal of Development Economics*, 36 (1): 89–115.
- Deaton, A. and C. Paxson (2000), 'Growth and Saving Among Individuals and Households', *Review of Economics and Statistics*, 82 (2): 212–25.
- De Long, B. and L. Summers (1991), 'Equipment Investment and Economic Growth', *Quarterly Journal of Economics*, 106 (2): 445–502.
- De Long, B. and L. Summers (1993), 'How Strongly do Developing Economies Benefit from Equipment Investment?', *Journal of Monetary Economics*, 32: 395–415.
- Dixit, A. and R. Pindyk (1994), *Investment under Uncertainty*, Princeton, NJ: Princeton University Press.
- Easterly, William (2001), *The Elusive Quest for Growth*, Cambridge, MA: MIT Press.
- Easterly, W., M. Kremer, L. Pritchett and L.H. Summers (1993), 'Good Policy or Good Luck? Country growth performance and temporary shocks', *Journal of Monetary Economics*, 32: 459–83.

- Easterly, W. and R. Levine (2001), 'It's not Factor Accumulation: Stylized Facts and Growth Models', *World Bank Economic Review*, 15 (2): 177–220.
- Easterly, W. and H. Pack (2001), 'Is Investment in Africa too Low or too High? Macro and Micro Evidence', Policy Research Working Paper, 2519, World Bank.
- Edwards, S. (1996), 'Why are Latin America's Savings Rates so Low? An International Comparative Analysis', *Journal of Development Economics*, 51 (1): 5–44.
- Elias, V. (1992), *Sources of Growth: A Study of Seven Latin American Economies*, San Francisco, CA: ICS Press.
- Fazzari, S., Hubbard R. and B. Peterson (1988), 'Financing Constraints and Corporate Investment', *Brookings Papers on Economic Activity*, 1: 147–95.
- Feldstein, Martin and Charles Horioka (1980), 'Domestic Saving and International Capital Flows', *Economic Journal*, 90: 314–29.
- Greene, J. and D. Villanueva (1991), 'Private investment in developing countries: an empirical analysis', *IMF Staff Papers*, 38 (1): 33–58.
- Gutiérrez. M. (2006), 'Economic Growth in Latin America: The Role of Investment and other Growth Sources', Eclac/Cepal-United Nations, Macroeconomics and Development Series 36; also in A. Solimano (ed.) *Vanishing Growth in Latin America: The Experience of the late XX Century*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 161–236.
- Gutiérrez. M. (2007), 'Savings in Latin America after the Mid-1990s: Constraints, Determinants, and Policies', Eclac/Cepal-United Nations, Macroeconomic and Development Series 57.
- Harberger, A. (1996), 'Reflections on Economic Growth in Asia and the Pacific', *Journal of Asian Economics*, 7 (3): 365–92.
- Harberger, A.C. (1998), 'A Vision of the Growth Process', *American Economic Review*, 88: 1–32.
- Hausmann, R., L. Pritchett and D. Rodrik (2004), 'Growth Accelerations', NBER Working Paper 10566.
- Jones, B.J. and B.A. Olken (2005), 'The Anatomy of Start-Stop Growth', NBER Working Paper Series 11528.
- Keynes, John M. (1997 [1936]), *The General Theory of Employment, Interest, and Money*, Amherst, New York: Prometheus Books.
- Khan, M. and M. Kumar (1997), 'Public and Private Investment and the Growth Process in Developing Countries', *Oxford Bulletin of Economics and Statistics*, 59 (1): 69–88.
- Khan, M. and C. Reinhart (1990), 'Private Investment and Economic Growth in Developing Countries', *World Development*, 18 (January): 19–27.
- Klenow, P.J. and A. Rodríguez-Clare (1997), 'The Neoclassical Revival in Growth Economics: Has It Gone too Far?', *NBER Macroeconomics Annual 1997*, Cambridge, MA: MIT Press, pp. 73–103.
- Lim, E. (2001), 'Determinants Of, and the Relation Between, Foreign Direct Investment and Growth: A Summary of the Recent Literature', Working Paper, No. 01/175 (November), International Monetary Fund, Washington, DC.
- Loayza, N., K. Schmidt-Hebbel and L. Servén (2000), 'What Drives Private Saving Across the World?', *Review of Economics and Statistics*, 82 (2): 165–81.
- Masson, P., T. Bayoumi and H. Samiei (1998), 'International Evidence on the Determinants of Private Saving', *World Bank Economic Review*, 12 (3): 483–501.
- Moguillansky, G. and R. Bielschowsky (2000), *Inversión y reformas económicas en América Latina*, Santiago: Fondo de Cultura Económica-Cepal.
- Obstfeld, Maurice and Alan M. Taylor (2004), *Global Capital Markets, Integration, Crisis and Growth*, Cambridge and New York: Cambridge University Press.
- Ogaki, M., J. Ostry and C.M. Reinhart (1995), 'Saving Behavior in Low-and Middle Income Developing Countries: A Comparison', Working Paper WP/95/3, IMF.
- Olofsdotter, K. (1998), 'Foreign Direct Investment, Country Capabilities and Economic Growth', *Weltwirtschaftliches*, 134 (3): 534–47.
- O'Rourke, Kevin H. and Jeffrey G. Williamson (2000), *Globalization and History: The Evolution of a Nineteenth Century Atlantic Economy*, Cambridge, MA: MIT Press.
- Patinkin, Don (1982), *Anticipations of the General Theory? And Other Essays on Keynes*, Chicago, IL: University of Chicago Press.

- Piles, W. and C. Reinhart (1999), 'Savings in Latin America and Lessons from Europe: An Overview', *Accounting for Saving: Financial Liberalization, Capital Flows and Growth in Latin America and Europe*, Washington, DC: Johns Hopkins University Press and Inter-American Development Bank.
- Pindyck, R. and A. Solimano (1993), 'Economic Instability and Aggregate Investment', in O. J. Blanchard and S. Fischer (eds), *NBER Macroeconomic Annual*, Cambridge, MA: MIT Press, pp. 259–303.
- Rodrik, D. (2006), 'Goodbye Washington Consensus, Hello Washington Confusion? A Review of the World Bank's Economic Growth in the 1990s: Learning from a Decade of Reform', *Journal of Economic Literature*, 44 (4): 973–87.
- Schmidt-Hebbel, K. and L. Servén (1998), 'Income Inequality and Aggregate Saving: The Cross Country Evidence', in Andrés Solimano (ed.), *Social Inequality, Values, Growth and the State*, Ann Arbor, MI: University of Michigan Press, pp. 82–119.
- Schmidt-Hebbel, K. and L. Servén (1999), *The Economics of Saving and Growth: Theory, Evidence, and Implications for Policy*, Cambridge: Cambridge University Press.
- Schmidt-Hebbel, K. and L. Servén (2000), 'Does Income Inequality raise Aggregate Saving?', *Journal of Development Economics*, 61: 417–46.
- Schmidt-Hebbel et al. (1994), 'Saving, Investment and Growth in Developing Countries: An Overview', Policy Research Working Paper 1382, Washington, DC: The World Bank.
- Schmidt-Hebbel, K., L. Servén and A. Solimano (1996), 'Saving and Investment: Paradigms, Puzzles, Policies', *World Bank Research Observer*, 11 (1): 87–117.
- Schmidt-Hebbel, K., S. Webb and G. Corsetti (1992), 'Household Saving in Developing Countries', *World Bank Economic Review*, 6 (3): 529–47.
- Severn, Luis and Andrés Solimano (1992), 'Private Investment and Macroeconomic Adjustment: A Survey', *World Bank Research Observer*, Oxford: Oxford University Press, 7 (1), 95–114.
- Servén, L. and A. Solimano (1993), *Striving for Growth After Adjustment. The Role of Capital Formation*, World Bank Regional and Sectoral Studies, Washington, DC: World Bank.
- Solimano, Andrés (ed.) (1996), *Roadmaps to Prosperity: Essays on Growth and Development*, Ann Arbor, MI: University of Michigan Press.
- Solimano, Andrés (1997), Commentary to Chapter 5: 'Why are Latin America's Saving Rates so Low?' (pp. 131–58) by Sebastian Edwards in Nancy Birdsall and Frederick Jaspersen (eds), *Pathways to Growth: Comparing East Asia and Latin America*, Inter-American Development Bank, pp. 159–60.
- Solimano, Andrés (ed.) (2006), *Vanishing Growth in Latin America. The Experience of the Late XX Century*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.
- Solimano, A. and R. Soto (2006), 'Economic Growth in Latin America in the Late twentieth Century: Evidence and Interpretation', in A. Solimano (ed.), *Vanishing Growth in Latin America: The Experience of the Late XX Century*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 11–45.
- Solimano, Andrés and Nathalie Watts (2005), 'International Migration, Capital Flows and the Global Economy: a Long Run View', *Macroeconomics of Development Series*, UN-ECLAC.
- Solow, Robert (1956), 'A Contribution to the Theory of Economic Growth', *Quarterly Journal of Economics*, 70: 65–94.
- Summers, L. (1981), 'Taxation and Corporate Investment: A q-Theory Approach', *Brookings Papers on Economic Activity*, 1: 67–140.
- Taylor, A.M. (1996), 'International Capital Mobility in History: The Saving–Investment Relationship', NBER Working Paper Series, no. 5743.
- Taylor, A.M. (1999), 'Latin America and Foreign Capital in the Twentieth Century: Economics, Politics, and Institutional Change', NBER Working Paper Series, no. 7394, October.
- Taylor, Alan M. and Jeffrey G. Williamson (1994), 'Capital Flows to the New World as an Intergenerational Transfer', *Journal of Political Economy*, 102 (2): 348–71.
- Young, A. (1994), 'Lessons from the East Asian NICs: A Contrarian View', *European Economic Review*, 38 (3–4): 964–73.

20 Role of finance and credit in economic development

Philip Arestis and Santonu Basu

Introduction

Finance and credit play an important role in the context of economic development. However, the role that financial institutions play in developed countries is very different from the one they play in developing countries. In developed countries financial institutions largely emerged within the process of industrialization. The process of industrialization increased the demand for finance, and many entrepreneurs recognized that there was an opportunity to make a profit from the intermediation between savers and investors or between lenders and borrowers. That, then, led to the growth of varieties of financial institutions. There was, thus, a mutual feedback between the two, arising from mutual benefit. It was observed by Goldsmith (1969) that there was an upward drift in the ratio of financial institutions' assets to gross national product (GNP) for 35 countries, including both developed and developing countries, between 1860 and 1963. He also observed that the banks played an important role in financing the early stage of development, but at the later stage of development the importance of banks was reduced to some extent due to the growth of non-bank financial intermediaries (NBFIs) and the stock market. This might have given the impression to a number of authors, such as Greenwood and Jovanovic (1990), King and Levine (1993), Bencivenga et al. (1996), Levine and Zervos (1996) and Fry (1997), but not all (see, for example, Arestis and Demetriades, 1997), that the financial markets, especially the stock markets, play a crucial role in promoting growth via bringing discipline to firms (see, also, Arestis et al., 2001; Arestis and Sawyer, 2005; Arestis, 2006) Accordingly they argue that financial repression not only lowers the savings rate but also discourages investors from undertaking high-return investment, thereby producing a lower growth rate than could be achieved under the liberalized system. Following these arguments, especially in the presence of the unsatisfactory state of the financial institutions and their performance under a prolonged regulatory regime, many developing countries, including South Korea, embraced the process of financial liberalization with the anticipation that this might bring efficiency in the financial sector and a higher growth rate. But soon it

brought a financial crisis in South Korea and many other countries, the severity and frequency of which was unparalleled in monetary history. The problem is that the financial market does not operate the way these authors have assumed (see Basu, 2002, for more details on this issue). Beside this, there exists an important difference between the developed and the developing countries' process of industrialization. In developing countries the process of industrialization was not a natural process of transition from a backward state to an advanced industrial state; rather it was the respective governments' deliberate attempts to reach such a state. Thus, there was very little scope, if any, to make profit from engaging in financial intermediation. Yet the financial institutions had a very important role to play in fostering the process of industrialization via the coordination between savers and investors.

The rest of the chapter is organized as follows. In the first section, drawing on the experience of India and South Korea, we investigate and demonstrate that it is not possible for private banks, or what is commonly referred to as the market, to take the initiative of their own accord to finance the process of development with some assistance from the government; direct intervention by the government is absolutely necessary. In the second section we investigate what potential problems intervention may produce. Finally, we summarize the argument and conclude.

Is the market effective or is intervention required?

The South Korean government recognized in 1961 that private banks, whose principal objective was to make profit, were unlikely to participate in the development programme, and nationalized the major banks. By contrast, in India the government initially chose to follow the path of persuasion with the market being allowed to perform its usual function. South Korea used the interest rate as a vehicle for the mobilization of savings. The interest rate on time deposits was raised from 15 to 30 per cent in 1965, and although the rate gradually declined it remained above 20 per cent up to 1971 (Amsden, 1989 and Cho, 1989). As a result of this high rate on deposits, the M2–GNP ratio (where M2 is money stock defined as currency held by the public plus demand deposits) rose from less than 9 per cent to a little over 33 per cent between 1964 and 1971 (see, for example, Cole and Park, 1983; Cho, 1989).¹ That higher rate also helped to increase the capital inflow, especially from Japan. Needless to say, that higher rate of mobilization was made at the cost of banks, as the rates on time deposits remained higher than those on non-preferential loans between 1965 and 1968. The implication was that the banks were running at a loss, independent of the performance of the project for which the banks advanced the loans (Cole and Park, 1983).

By contrast, the Indian government mainly used the extension of branch facilities as a vehicle to mobilize savings. However, it recognized that given the smaller size of operation of these banks, individually they would not have sufficient resources to open branch facilities outside the metropolitan area. Thus, banks were encouraged by the Reserve Bank of India to merge with other banks, and as a result the total number of banks fell from 605 to 85 between 1950 and 1969. The number of bank branches increased from 4151 to 8262, and gross domestic savings (GDS) as a percentage of gross domestic product (GDP) rose from 10.2 to 15.7 per cent during the same period. But the banks that opened branch facilities in the rural areas were mostly government-sponsored banks such as cooperatives, land mortgage banks and the State Bank of India (SBI); private banks remained reluctant to participate in the government venture.² In fact, in 1969 only 22.2 per cent of the banks' total branch facilities were located in the rural areas. The banks' performance, however improved following the nationalization of banks in 1969, both in terms of the extension of branch facilities outside the metropolitan areas and in terms of the mobilization of savings. But as most of these deposits were small in size, the result was a rise in the administrative cost per unit of deposit, thereby increasing the operating costs of banks. Thus, although both countries were successful in mobilizing savings, banks' profitability was adversely affected in the process (see Krishnaswamy et al., 1987 for more details).

In South Korea, as the government nationalized private banks, no problem was observed in the allocation of loans. It is interesting to note that despite having control over the allocation of loans, the government neither implemented any quota nor imposed any restrictions on the allocation of loans; instead it introduced an incentive mechanism to attract market participants to join in the government development programme. For example, export and infant industries received a preferential loan rate of 6 per cent and this rate remained below 10 per cent up to 1980. The rest of the economy received loans at a non-preferential rate. This rate rose from 16.9 per cent to 26 per cent between 1964 and 1965 and remained roughly around 24.4 per cent between 1966 and 1970. This rate came down to 17 per cent by the end of 1971, and remained on average between 17 and 18 per cent up to 1980 (Amsden, 1989; Cho, 1989; Amsden and Euh, 1993). The main aim of the differential interest rate policy was to reduce the net return of those projects that are not warranted by the government. As there was hardly any difference in the deposit rate and interest rate on non-preferential loans, this suggests that profit from loan portfolios was not the concern of the government. It appears that the government used differential interest rates to induce market participants to join the government-directed development programme. The policy could be considered as important, earmarked to

change the state of the economy from a backward to an advanced industrial state. As the number of firms started to join the government-directed development programme, so did South Korea's growth rate, which started to increase at a rapid rate. In fact, South Korea enjoyed an unprecedented growth rate between the mid-1960s and late 1990s. It reached 8 per cent in the 1980s and continued to grow at the same rate until the early 1990s. In the process, however, there was the financial crisis in 1997, which put a break to that continuous growth process (see Arestis and Glickman, 2002, for more details on the crisis).

India, on the other hand, from the very beginning faced problems in allocating loans the way the government wanted. It was confronted with problems both in allocating loans for socially more productive areas of the economy, and in improving smaller and marginal borrowers' access to the loan market. Initially, it attempted to channel banks' credit into socially more productive areas of the economy via the Reserve Bank of India regulations and incentives. To provide long-term credit, financial corporations and government sponsored banks were established, such as the Industrial Finance Corporation (IFC), Industrial Development Bank of India (IDBI), cooperative banks and land mortgage banks. In order to provide assistance to small-scale enterprises, similar institutions were developed. The IFC and IDBI mainly provided credit for the industrial sector's development; their clients in general were large borrowers, and no appreciable problems were observed in relation to these borrowers' ability either to raise large long-term loans or to meet large working capital requirements from the banks. But the problem arose in the government's effort to divert credit to the rural areas and small-scale enterprises. The Reserve Bank of India regulations and incentives largely remained ineffective.³ It was observed that much of the banks' credit was still being received by private traders, especially wholesale traders and large entrepreneurs. Wholesale traders used this credit for the purchase of food grains, edible oils, oil seeds, raw cotton, sugar, and so on, with the expectation that they would make a wind-fall gain from future changes in the prices of these items. In the case of industry, the banks' finance principally went to maintain inventories but they remained reluctant to offer credit for fixed investments.

The Indian experience sheds some light on why it is not possible for private banks to assist in the development programme. Private banks operate on the condition of profit; participating in the development programme meant banks not only have to make sacrifices but also have to carry high credit risk. In the early days of independence, the financial sector was not developed; there were neither NBFIs nor a well-developed stock market; banks were the only player in such markets. Consequently, they used to enjoy monopoly power over all borrowers irrespective of

their size of operation. As a result, they were not used to taking any known credit risk that developed countries' counterparts quite often had to take, especially when operating in the large borrowers' market (Basu, 2002). Bankers were neither trained on how to offer a credit-risk-adjusted interest rate nor did they know how to manage it. But to cooperate with the government meant they would have to undertake a substantial amount of credit risk for the reasons already explained. Furthermore, in most cases banks knew that many entrepreneurs would be reluctant to borrow if the credit risk-adjusted interest rates were implemented. This is mainly because entrepreneurs themselves do not know the expected return from these projects. In the case of agriculture, as the loans are supposed to be allocated for the purpose of reducing poverty, the issue of credit-risk-adjusted interest rate does not arise. This means profit from these loans will be low; therefore, banks will have little incentive to participate in such projects.

Furthermore, banks did not have much information about the past performances of these projects, whether we speak of agriculture or industry. This meant bankers themselves did not know the approximate level of financial return at regular intervals. The involvement in such projects meant that they had to carry a high liquidity risk, arising from the short-term nature of deposits (with a smaller deposit base) and a long-term commitment to investment. In this situation, if a bank had to take a higher credit risk then it could not avoid liquidity risk, arising from the possibility of an irregular financial flow. Under such circumstances, the return from the entrepreneurs' equity may not be sufficient to meet the shortfall that may arise in meeting the regular debt repayment. This problem particularly arises from the fact that banks are required to relax their credit standard requirements. Therefore, from the point of view of the banks, cooperating with the government means they cannot avoid liquidity risk, while their survival depends upon liquidity. Consequently, we observe that despite various attempts by the government to encourage banks to offer long-term loans, banks' advancement policy essentially remained short term (Basu, 2006a, 2006b).

In 1969 India too nationalized its major commercial banks. The experience with that experiment produced an enlarged size of the banking market, reaching almost one-third of the population. But in the process the banking sector became fragile, carrying the cost of a number of non-performing loans, and India too recognized that it could not continue with that method of financing relevant projects. This raises an interesting question as to why, in both countries, the process of development adversely affected the banking sector, which plays such an important role in economic development.

Does intervention per se produce vulnerability?

It is important to note that at the time when the two countries undertook their development programmes the subject of finance as an academic discipline was in its infancy.⁴ The understanding was that financial markets operated within the framework of a perfectly competitive market. Essentially this means that market forces determine the rate of interest and it is this variable that governs the allocation of loans. The response to the suggestion that the rate of interest might be unable to govern the allocation of loans was that there must remain some imperfection in the functioning of the loan market. This imperfection arises either from the existence of monopolistic elements or from the market being underdeveloped. No investigation was undertaken in relation to how the loan market actually operated. The concepts of uncertainty, credit standard and credit risk were largely unknown to the economists. Even today not only do many economists still continue to ignore the importance of these concepts, but more importantly the theory of financial liberalization does not fundamentally differ from the assumption of perfectly competitive markets. Yet, the fundamental problem for both countries principally emerged from the fact that South Korea overlooked the importance of the credit standard, while India underestimated the importance of the credit-risk-adjusted interest rate and furthermore refused to close down many non-profit-making firms.

The evidence of the last section suggests that neither government was much worried about the profitability of the banks. While South Korea was mainly concerned with growth, India was more focused on the enlargement of the market. Consequently, neither paid adequate attention to the question of whether the process of financing development projects would expose banks' loan capital to a very high level of credit risk. Yet both countries' problems principally emerged from ignorance of this factor. The issue of credit risk principally arises from the fact that in the early days of independence most of the entrepreneurs from developing countries had neither adequate capital nor sufficient assets to meet the banks' credit standard. Both were necessary for obtaining the size of loan required to undertake the projects that their respective governments had planned. Under such circumstances, banks had to take a very high level of credit risk, which effectively meant that the banks became the principal investors. In short, banks' share of loan capital in the total investment often exceeded the entrepreneurs' share of investment. That introduced the possibility that in the event of a project failure, or if the return from the project was not sufficient to maintain the debt obligation, the return from the entrepreneurs' equity might not be sufficient to meet this shortfall. The implication being that from the very outset of the implementation of development programmes, banks had to carry a relatively high degree of fragility compared

to banks from other developed countries. The interesting question, therefore, is how in the process of development would one try to reduce this fragility over time.

In order to answer this question it is necessary to examine the composition of firms' investment funds, that is, the combination of internal (that is, entrepreneurial equity) and external funds (that is, loanable funds). This should give us some indication of whether, if the project performs adversely, the return from the internal fund will be sufficient to meet the shortfall in the debt repayment rates. It will also allow us to determine the extent of the fragility of the banking sector in terms of whether it is high or low. For example, in the case of South Korea, firms' internal funds initially constituted 47.7 per cent, while external funds constituted 52.3 per cent in 1963–64 (Amsden and Euh, 1993). In the case of India this figure was 60.1 per cent and 39.9 per cent respectively (Singh, 1995). Although both of these figures are quite high compared to, say, the average UK corporations' long-term debt of 26.6 per cent of the total investment (Cobham and Subramaniam, 1998), it is reasonable to assume that if the project performed adversely, any shortfall in the repayment could be recouped from the return on the equity finance. Furthermore, if the project failed, then the bank could sell all the company's assets including the entrepreneur's share, to recoup the principal. Thus, the question turns out to be how the problem emerged.

South Korea's problem emerged from its very success, and as the economy grew, so did its ability to service the contractual debt commitment, all things being equal. This steady flow of return on loans in turn not only increased firms' credit ratings, but also enhanced the confidence of the policy-makers. Consequently, when the lenders' willingness to offer larger loans to these firms increased, policy-makers overlooked the fact that the process could overexpose the banks' capital. With the increasing access to credit, investors (that is, borrowers) also did not feel the necessity to rely on internal funds to any great extent for growth, and as a result their share of internal funds in relation to total funds shrank. This in turn caused the debt–equity ratio to rise with the growth of the firm, thereby causing the debt service ratio to rise, which in turn, in the absence of an appropriate credit standard, exposed banks' capital to very high levels of credit risk.

As soon as the growth in export earnings slowed down, South Korean firms had a problem in maintaining their debt commitment. Although the government rescued the troubled firms (Cho, 1989), it did not give attention to the source of the problem. By the early 1980s, the economy faced the same problem again as it slowed down, mainly as a result of the collapse of foreign markets in construction, shipping and shipbuilding. The GDP growth rate turned negative for the first time since the Korean War. Firms started to face difficulties in maintaining the debt commitment from

their own returns, and started to borrow in order to meet the shortfall in debt commitment. As a result, their share of internal funds shrank from an average of 21.1 per cent during 1975–79 to 16.4 per cent by 1980, while their share of external funds rose from 78.9 per cent to 83.6 per cent (Choong-Hwan, 1990). By 1982, a growing number of highly indebted firms found it difficult to service their debt. The South Korean government once again had to organize assistance in restructuring industrial firms that faced financial difficulties. Thus, the problem in 1972 repeated itself: that is, the problem of overinvestment.

Following the crisis, the government decided to abstain from further credit-directed programmes and abolished the preferential lending rates. In addition, the NBFIs were deregulated and corporations were allowed to issue bonds with a guarantee by commercial banks. The government privatized the commercial banks, but in the presence of large non-performing loans it could not withdraw from maintaining control over the banking sector. That was so since the restructuring of the industrial sector required government supervision of credit allocation. The government maintained its control over the interest rate and the credit allocation of the banking sector, which was revoked in 1991, along with its control over the foreign capital inflow. Although the external share as opposed to the internal share of the total investment improved from its early 1980s position, it still remained very high. For example, between 1987 and 1991, the share of external capital funds constituted 73.6 per cent, while the internal share constituted merely 26.4 per cent of the total investment (Amsden and Euh, 1993). Thus, the banking sector remained overexposed. By 1994, banks had to increase their allocated funds in order to make provision for non-performing loans, but it appears that inadequate attempts were made to reduce the debt–equity ratio especially for *chaebols*. For example, even in 1996 the average debt to equity ratio for the top 30 *chaebols* was 898.49 per cent. Fourteen of these 30 *chaebols* were making negative profits in 1996, while for those who were making a positive profit, this remained marginal compared with the total assets, including loans that were invested (Lee, 1997). Thus, it was no wonder that as soon as the economy slowed down it led to a banking crisis.⁵

Although India at the initial stage had taken a more conservative approach, compared to South Korea, in financing its large economic units, it could not maintain that position because the size of the market remained very modest for some time. This was largely because of the fact that the government was unable to make any appreciable inroad in improving a sizeable portion of the population's livelihood. As a result, not only did effective demand remain quite low even in the mid-1970s, but more importantly future growth in the pattern of domestic demand remained unnoticed for some time. Consequently, many firms, especially the large ones, were often

either unable to capture a sufficient share of the market that was necessary to achieve economies of scale, or were running at a loss. The government neither allowed these loss-making firms to close down nor encouraged them to search for an export market. The government did not encourage these firms to search for an export market because it was of the opinion that those firms may not be able to capture a sufficient share of the market (Chakravarty, 1987).

Summary and conclusions

It is argued in this chapter that banks play a very important role in financing the process of development. But this role cannot be initiated by private sector banks; government intervention will be required in the operation of this market. In fact, the Indian experience highlights a very important lesson. This is that at the early stage of development the risk in financing development is so high that it is not possible for the private sector to bear it. India, which is known for its over-regulation, did not begin as an over-regulated economy; on the contrary, and as shown in the second section, it started with a market-friendly approach. It encouraged market participants that is, banks, through persuasion and incentive mechanisms to participate in the government development programmes. But the market did not cooperate with the government in those programmes which reduced their expected profitability and increased their risk of survival. It was the refusal of the market to cooperate with the government that brought one regulation after another over the financial system without investigating the consequences of those regulations on the system, and made India one of the most over- or ill-regulated economies. But the government did not realize that the refusal to cooperate principally arose from the fact that such cooperation would only increase banks' vulnerability.

Following nationalization, governments in both countries used their banks to finance their development agendas. While South Korea became a success story, India's performance remained very modest till the early to mid-1970s, and only started to improve following the establishment of a firm control over the allocation of credit. But both countries overlooked the fact that having control over the allocation of loans does not give any control over the repayment of loans. Without the latter, banks cannot survive; hence the importance of the credit standard emerges in order to ensure that the fate of the banks' loanable funds should not be tied to the borrowers' projects. As both governments did not realize the importance of this issue, both continued their respective agendas by exposing banks' loan capital. In the case of South Korea the problem only emerged when the economy slowed down, so concentration was given to the growth rate, which further exposed banks' loan capital. India's problem principally

came from the fact that it had not offered a credit-risk-adjusted interest rate, and consequently there remained a very small margin between the deposit rate and the interest rate on loans. That meant that if a small fraction of the loans functioned adversely, banks would be in trouble, and this problem was further magnified by the presence of weak bankruptcy laws. Consequently, both countries had to undertake capital market reform. Needless to say, in both countries, banks' finance played a very important role in the context of development.

Notes

1. Amsden (1989) pointed out that household savings as a percentage of gross domestic product (GDP) increased from 0.18 per cent in 1965 to 4.15 per cent in 1966, but declined in the following year. From there onwards no systematic relationship can be observed between interest rate and saving behaviour, suggesting that higher rates perhaps mainly helped to transfer savings that were previously held in an unproductive form.
2. For example, in 1956 the SBI and its associates were directed to open 400 branch facilities in rural and semi-urban areas in the next five years (Reserve Bank of India, 1969).
3. It is important to note that India is not the only country which was confronted with this problem; in fact all countries across the globe faced similar problems, irrespective of whether they were developed or developing countries. For more on this issue see Basu (1986, 1989).
4. It is important to note that in economics, the role of finance in general is mixed. Robinson (1979) suggests that: 'by and large, it seems to be the case that where enterprise leads finance follows' (p. 20). The role of finance in the context of economic development is also a much-neglected area in the literature on economic development (Arndt, 1987). Lucas (1988) considered the relationship between financial and economic development to be 'overstressed'. Chandravarkar (1992) provides an impressive list of authors who are pioneers in economic development, including three Nobel laureates, who did not even mention finance as a factor in economic development. However, following McKinnon's (1973) contribution, finance has been given a more active role (see, for example, Levine, 2004).
5. In fact, the South Korean growth rate came down from 8 per cent to 4 per cent just prior to the crisis (Arestis and Glickman, 2002).

References

- Amsden, A.H. (1989), *Asia's Next Giant: South Korea and Late Industrialisation*, New York: Oxford University Press.
- Amsden, A.H. and Yoon-Dae Euh (1993), 'South Korea's Financial Reforms: Good-Bye Financial Repression (Maybe), Hello New Institutional Restraints', *World Development*, 21: 379–90.
- Arestis, P. (2006), 'Financial Liberalization and the Relationship between Finance and Growth', in P. Arestis and M.C. Sawyer (eds), *A Handbook of Alternative Monetary Economics*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 346–64.
- Arestis, P. and P. Demetriades (1997), 'Financial Development and Economic Growth: Assessing the Evidence', *Economic Journal*, 107 (442): 783–99.
- Arestis, P. and M. Glickman (2002), 'Financial Crisis in Southeast Asia: Dispelling Illusion the Minskyan Way', *Cambridge Journal of Economics*, 26 (2): 237–60.
- Arestis, P. and M.C. Sawyer (2005), 'Financial Liberalization and the Finance–Growth Nexus: What Have We Learned', in P. Arestis and M. Sawyers (eds), *Financial Liberalization: Beyond Orthodox Concerns*, New Series and International Papers in Political Economy, Basingstoke: Palgrave Macmillan, pp. 1–42.
- Arestis, P., P. Demetriades and K. Luintel (2001), 'Financial Development and Economic Growth: The Role of Stock Markets', *Journal of Money, Credit, and Banking*, 33 (1): 16–41.

- Arndt, H.W. (1987), *Economic Development: The History of an Idea*, Chicago, IL and London: Chicago University Press.
- Basu, S. (1986), 'Problems of Small Business', *Economic Papers*, 5: 92–110.
- Basu, S. (1989), 'Deregulation: Small Business Access to the Capital Market – Theoretical Issues with Special Reference to Australian Bank Finance', *Australian Economic Papers*, 28: 141–59.
- Basu, S. (2002), *Financial Liberalisation and Intervention: A New Analysis of Credit Rationing*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.
- Basu, S. (2006a), 'Structural Problems in Financing Development: Issues Related to India', *International Review of Applied Economics*, 20: 85–101.
- Basu, S. (2006b), 'The Role of Banks in the Context of Economic Development with Reference to South Korea and India', in P. Arestis and M.C. Sawyer (eds), *A Handbook of Alternative Monetary Economics*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 291–306.
- Bencivenga, V.R., B.D. Smith and R.M. Starr (1996), 'Equity Markets, Transactions Cost and Capital: An Illustration', *World Bank Economic Review*, 10: 241–65.
- Chakravarty, S. (1987), *Development Planning: The Indian Experience*, Oxford: Oxford University Press.
- Chandravarkar, A.G. (1992), 'Of Finance and Development: Neglected and Unsettled Questions', *World Development*, 20: 133–42.
- Cho, Y. Je (1989), 'Finance and Development: The Korean Approach', *Oxford Review of Economic Policy*, 5: 88–102.
- Choong-Hwan, R. (1990), 'Korean Corporate Financing', *Monthly Review, Korean Exchange Bank*, 24 April: 3–13.
- Cobham, D. and R. Subramaniam (1998), 'Corporate Finance in Developing Countries: New Evidence for India', *World Development*, 26: 1033–47.
- Cole, D. and Y.C. Park (1983), *Financial Development in Korea, 1945–1978*, Cambridge, MA: Harvard University Press.
- Fry, M.J. (1997), 'In Favour of Financial Liberalisation', *Economic Journal*, 107: 754–70.
- Goldsmith, R.W. (1969), *Financial Structure and Development*, New Haven, CT: Yale University Press.
- Greenwood, J. and B. Jovanovic (1990), 'Financial Development, Growth, and the Distribution of Income', *Journal of Political Economy*, 98: 1076–1108.
- King, R.G. and R. Levine (1993), 'Finance, Entrepreneurship, and Growth: Theory and Evidence', *Journal of Monetary Economics*, 32: 513–42.
- Krishnaswamy, K.S., K. Krisnamurthy and P.D. Sharma (1987), *Improving Domestic Resource Mobilisation through Financial Development: India*, Manila: Asian Development Bank.
- Lee, S.J. (1997), 'Financial Crisis in Korea', mimeograph, New Haven, CT: Yale University.
- Levine, R. (2004), 'Finance and Growth: Theory and Evidence', NBER Working Paper Series, No. 10766.
- Levine, R. and S. Zervos (1996), 'Stock Market Development and Long-run Growth', *World Bank Economic Review*, 10: 323–39.
- Lucas, R.E., Jr (1988), 'On Mechanics of Economic Development', *Journal of Monetary Economics*, 22: 3–42.
- McKinnon, R.I. (1973), *Money and Capital in Economic Development*, Washington, DC: Brookings Institutions.
- Reserve Bank of India (1969), *Review Committee*, Reports of the All-India Rural Credit, Bombay: Reserve Bank of India.
- Robinson, J. (1979), 'The Generalisation of the General Theory', in J. Robinson (ed.), *The Generalisation of the General Theory and Other Essays*, 2nd edn, London: Macmillan, pp. 1–52.
- Singh, A. (1995), *Corporate Financial Patterns in Industrialising Economies: A Comparative International Study*, IFC Technical Paper No. 2, Washington, DC: World Bank.

21 Physical infrastructure

Gregory K. Ingram and Marianne Fay¹

Perceptions of infrastructure's role in economic development and of desired modes for infrastructure provision have evolved over recent decades. In the mid-twentieth century, infrastructure was seen as a key determinant of economic development – the 'commanding heights' of the economy – and market failures in infrastructure provision were thought to be endemic. Accordingly, public sector involvement in infrastructure became pervasive and was advocated by many development agencies. In the 1980s and 1990s, poor performance of public infrastructure agencies, concerns about government failure, and large investment needs stimulated interest in private participation in infrastructure. Concurrently, many development agencies turned their attention much more to investments in education, the environment and health. The results from private participation in infrastructure have been mixed, and a more balanced and country-specific approach is emerging with renewed interest in infrastructure from development agencies. Meanwhile, research on the links between infrastructure and development has benefited from improvements in both data and technique. Within this evolving context, this chapter addresses three questions: How does infrastructure relate to development? How can infrastructure services best be delivered? What are the key barriers to progress?

Infrastructure is imprecisely defined by development economists and originally encompassed elements of social overhead capital. In this chapter, physical or economic infrastructure includes: public utilities (power, telecommunications, piped water, sanitation and sewerage, solid waste collection and disposal, and piped gas); public works (roads, major dams, and canals for irrigation and drainage); and other transport (urban and inter-urban railways, urban transport, ports and waterways, and airports). Many of these activities have in common technical features such as economies of scale and economic features such as externalities and aspects of public goods. Developing economies include low-income economies (per capita gross national product below \$765 in 2003) and middle-income economies (per capita gross national product equal or more than \$765 but less than \$9386 in 2004) (World Bank, 2005a).

Special characteristics of infrastructure

Infrastructure services range across the spectrum of public and private goods, but many have two key aspects of private goods: they are rival in consumption (services consumed by a user reduce the supply available to others) and excludable (usage can be prevented). This is the case for most public utilities, ports, airports and railways wherein use depends on gaining access to a facility or network, and consumption can be metered and/or charged for. Users of these services can and do impose congestion costs on other users, and negative externalities (pollution, noise) on non-users. Roads are more likely to be public goods, especially uncongested rural roads. Congestion makes urban roads rival, and – if users cannot be excluded – urban roads are common property goods (rival but not excludable), as is also generally the case for groundwater extraction. Electronic tolling systems can make road use excludable. Toll roads exclude users and charge for usage, making them club goods (excludable but not rival) as long as they are uncongested.

Some elements of infrastructure – particularly public utilities and transport systems – have economies of scale related to the networks used to supply services. Since a network is usually dedicated to carry a single good (for example, water), and it is uneconomic to convert it to another use or to move it once it is constructed, network investments are largely sunk costs. Flows on the network must be coordinated for efficient operation. Networks often exhibit economies of scale because of technical reasons (for example, flow resistance decreases as pipe diameter increases) or network effects (for example, adding users to a phone system benefits all users). Scale economies and the high cost of constructing the networks make them natural monopolies. The spatial reach of network natural monopolies varies by sector and is typically exhausted at the municipal level for water supply, but can extend to the national or regional level for a high-voltage electrical grid. When network costs are high and unrecoverable, it is difficult for other suppliers to contest the market for network services.

In many countries public utility and transport services became the domain of vertically integrated monopolies whose existence was justified by economies of scale that only occur in the underlying delivery networks. However, network-based economies of scale do not imply the absence of all competition or the need for vertically integrated monopolies. Many activities associated with network operation can be unbundled and provided competitively. For example, having a single national high-voltage grid does not require common ownership of electricity generating plants, which can compete to provide power over the grid. Similarly, truckers can compete to provide services over a common road network. In addition,

some infrastructure sectors compete with each other, as when road-based transport competes with rail-based transport. Many infrastructure reforms implemented in developing countries since the mid-1990s have unbundled large monopolies, for example by separating long-distance from local phone service provision, power generation from distribution, port operations from port ownership, or natural gas production from distribution.

Appropriate technology may differ across users or change over time. Service technology may vary with household ability to pay (for example, standpipes versus house connections for water supply) or circumstances (for example, network-based power may be uneconomic in remote rural villages). In addition, technical characteristics of infrastructure sectors often change over time. In telecommunications, the advent of microwave and satellite transmission has removed the need for long-distance cable networks, and cellular service has reduced dependence on local networks. In power generation, combined cycle gas turbines have greatly reduced the minimum efficient size of power generators. In transport, electronic sensors have reduced the cost of collecting vehicle tolls. Accordingly, while infrastructure sectors have some common features, technology within a sector may differ by user group, and sectors differ dramatically in their technical and related economic characteristics – both of which change over time.

Regulation is needed when competition is absent. When competition is not possible in providing network services, regulatory oversight is necessary. Experience has shown that regulation is a costly and imperfect business requiring good information and trained staff – inputs that are scarce in many developing countries. Yet many developing countries have put regulatory capacity in place since the mid-1990s. For example, three out of four African countries have an independent regulatory agency for the telecom sector and one out of three for the electricity sector (World Bank, 2005c). Countries have experimented with a variety of regulatory approaches, including price caps, benchmark comparisons and rate-of-return limits. Experience has been mixed, and earlier optimism about implementing regulation has proven to be naive (Gomez-Ibanez, 2003; Kessides, 2004; Guasch, 2005). Regulation continues to be a special challenge in developing countries.

Why does infrastructure matter?

Good data are now available on physical stocks of infrastructure by sector across developing countries and over time (Canning, 1998; Calderón and Servén, 2004b), and per capita infrastructure stocks are strongly related to per capita gross domestic product (GDP). Analyses of expansion paths yield simple elasticities of per capita sectoral stocks with per capita GDP of 0.6 for roads, 1.2 for paved roads, 1.2 for power generation and 1.4 for

phone service (Canning, 1998). These differing sectoral elasticities reflect the systematic change in the composition of infrastructure between low- and middle-income countries. In low-income countries basic services such as water, irrigation and transport comprise most infrastructure, while in middle-income countries telecommunications and especially electric power become more important. Aggregate infrastructure stocks are roughly unit-elastic with GDP (World Bank, 1994). Within countries, there is some evidence that infrastructure investment is higher in richer than poorer regions (Banerjee, 2004).

While these expansion paths are averages, the relation between infrastructure stocks and GDP per capita has been used to identify countries that are outliers. For example, Latin American countries made modest investments in infrastructure starting in the mid-1990s and now have less than the average aggregate stocks for their income levels (World Bank, 2005b). Less than average stocks are also reported for sub-Saharan African countries for all sectors except telephone service where widespread availability of cellular service has moved African countries above the sectoral average for their income level (World Bank, 2005c). Country-specific deviations from cross-country averages do not have unambiguous normative content, but they are often used as a diagnostic tool in the exploration of cross-sectoral investment priorities.

Income growth is related to infrastructure level and investment. Aschauer's (1989) seminal analysis of the relation between overall infrastructure, investment and economic growth found that infrastructure capital increased investment and made a large contribution to total factor productivity in the USA. Subsequent work found his estimates overstated infrastructure's economic impact and were sensitive to econometric technique and the level of aggregation (Holtz-Eakin, 1994; Baltagi and Pinnoi, 1995; Cashin, 1995). Gramlich (1994) provides an overview. More recent work using cross-country panel data from developing countries addresses simultaneity issues and finds significant contributions of infrastructure to economic growth. Returns to infrastructure are likely the highest at early development stages when basic networks are still incomplete. Returns to infrastructure tend to fall as economies reach maturity. Canning (1999) finds above-average returns to telephone networks and average returns for power generation and transport networks. Canning and Bennathan (2000) find complementarity across infrastructure sectors with diminishing returns when one sector's capacity is increased in isolation, implying an optimal mix of capital stocks across sectors. Controlling for endogeneity of infrastructure accumulation, Calderón and Servén (2003) find positive and significant output contributions from telecommunications, transport and power.

Growth is also related to infrastructure efficiency and service quality, so analysis of infrastructure stocks alone is incomplete. The efficiency and quality of infrastructure services vary across countries. For example, distribution losses in power delivery vary from 5 percent to 30 percent, and unaccounted-for-water ranges from 12 to 45 percent across developing countries. Hulten (1996) finds that differences in the efficiency of use of infrastructure stocks explain a quarter of the growth difference between Africa and East Asia and 40 percent of the growth difference between low- and high-growth developing countries. Using panel data from 1960 to 2000 for 100 countries, Calderón and Servén (2004a) find that both infrastructure quantity and quality affect economic growth – where quality is measured in terms of frequency of service interruptions, faults per phone line, and so on. Moreover, the efficiency and quality of service provision is not closely related to a country's GDP per capita. Organizational effectiveness, adequate maintenance and sufficient operating revenues are important determinants of the efficiency of service provision that countries can control, and better management of infrastructure stocks can substitute for increased investment when efficiency is low.

Infrastructure helps reduce poverty. Households in the lowest income quintile are much less likely to be served by infrastructure than those in the highest quintile. Rural households are much less likely to receive services than urban households. In sub-Saharan Africa the urban–rural service gap is 46 percentage points for grid electricity, 29 for improved water, 26 for improved sanitation and 9 for telephone service. Extending infrastructure to the poor gives them access to productive opportunities, services, education and health care. Improving the communication and transport access of farmers to markets improves their economic condition and enhances diversification outside of agriculture (Jacoby, 2002). Providing the level of infrastructure enjoyed by the highest income quintile to those in the lowest quintile would reduce child mortality by 8 percent and stunting by 14 percent (Fay et al., 2005). Improved transport obviously facilitates access to health care, and the access of health care workers to rural clinics. Some impacts are less obvious. For example, access to piped water promotes attendance of girls at school because they need not fetch water; electric lights allow more time for study; clean water improves health and reduces school absences (Brenneman, 2002).

Infrastructure is linked to income inequality. Studies at the macro level have broadened their focus beyond growth to include the relation between infrastructure and the distribution of income. Their results reinforce the micro-level linkages between infrastructure and poverty. Calderón and Servén (2004a) find that infrastructure access and quality have a significant

impact on overall inequality. For example, they find that an improvement of one standard deviation in the infrastructure stock reduces the Gini coefficient for income by 0.06 points (the average Gini for developing countries with available data is 0.41).

In developing countries, firms are large users of infrastructure services, using half as intermediate inputs while households consume the balance (Prud'homme, 2004). Good infrastructure makes firms more productive and more competitive internationally. This is demonstrated at the macro level where infrastructure performance influences total factor productivity (Krugman, 1994) and infrastructure stimulates private investment (Taylor, 2001), and by long-standing micro-level studies of infrastructure use by enterprises (Lee and Anas, 1992). It is also supported by interviews with managers and investors used to develop indices of competitiveness and investment climate (World Economic Forum, 2004; International Institute for Management Development, 2005; World Bank, 2005d). In Latin American and Middle Eastern countries, 55 percent of firm managers consider infrastructure to be a major or severe obstacle to the operation and growth of their business (World Bank, 2004). High logistics costs impede exports and raise costs in many countries. Indirect costs are 10 to 12 percent of total production costs in strong export performers. In sub-Saharan Africa they run as high as 20 to 30 percent of production costs, and half are infrastructure-related, with transport costs in Africa being twice as high as those in South and East Asia (Ndulu, 2004).

What is the recent progress on infrastructure service delivery?

Infrastructure provision has increased significantly in developing countries. The leader by far is telecommunication where teledensity increased sixteenfold between 1980 and 2003 (Table 21.1). This reflects the explosion of mobile coverage (from zero in 1980 to 136 lines per 1000 persons in 2003 in developing countries), but fixed-line coverage also increased substantially (from 16 to 112 lines per 1000 people). Telecommunication is now the infrastructure sector with the narrowest divide between rich and poor countries, although teledensity remains much lower in the developing world. Road density in low- and middle-income countries increased by about 180 percent from 0.027 km/km² in 1980 to 0.049 km/km² in 2001, mostly due to a doubling in coverage in East Asia, South Asia, the Middle East and North Africa (MENA) (Ingram and Liu, 1999). Electricity-generating capacity increased by 56 percent in developing countries (from 0.18 to 0.28 kW per person). Again, this was driven by increases in East Asia (where capacity per person almost tripled) and South Asia and MENA (where it doubled), while Latin America's capacity per capita actually declined and sub-Saharan Africa's stagnated.

Table 21.1 Access to utilities and roads

	Telephone mainlines (per 1000 people)		Mobile phones (per 1000 people)		Paved road density (km per km ²)		Electricity generating capacity (KW per person)	
	1980	2003	1980	2003	1980	2001	1980	2001
	All developing countries	15.55	111.74	0.00	135.97	0.027	0.049	0.18
High income	317.11	549.90	0.03	698.29	0.225	0.285	–	–
East Asia & Pacific	3.02	161.36	0.00	195.38	0.024	0.052	0.07	0.25
Europe & Central Asia	64.42	228.06	0.00	300.75	–	0.052	–	0.95
Latin America & Caribbean	39.86	169.76	0.00	246.35	0.015	0.021	0.26	0.25
Middle East & North Africa	17.79	133.34	0.00	84.77	0.017	0.037	0.16	0.34
South Asia	2.91	38.88	0.00	22.53	0.157	0.380	0.04	0.10
Sub-Saharan Africa	8.13	10.67	0.00	51.27	0.008	0.013	0.08	0.10

Source: World Bank (2005a) for phone data; Calderón and Servén (2004a) for road and electricity data.

Households' access to services has also improved, but more modestly. Although data on access to clean water and sanitation is poor and uneven in coverage, overall access to clean water improved modestly from 71 percent to 79 percent between 1990 and 2002, with large variations in coverage (Table 21.2). Sanitation coverage remains below 50 percent, varying from a mere 35 or 36 percent in sub-Saharan Africa and South Asia to a high of 82 percent in Eastern Europe and Central Asia. Historical data on access to electricity is not available, but coverage is estimated at close to 55 percent in 2000. The substantial regional variations are not well explained by differences in urbanization (urban electricity coverage is quasi universal): the MENA region has the highest electricity access rate despite being substantially less urbanized than Latin America.

Evidence on improvements in the quality of services is more limited. Data on performance are only widely available for telecommunications and electricity. They show a substantial improvement for telephones but no clear changes for electricity (Table 21.3). Telephone faults declined from an average of 97 per 100 mainlines in 1990–95 to 65 in 1996–2003. This average masks large variations across regions from 35 in Latin America and MENA

Table 21.2 *Access to water, sanitation and electricity*

	Improved sanitation facilities (% of population with access)		Improved water source		Electricity 2000
	1990	2002	1990	2002	
All developing countries	36.75	49.62	71.49	79.40	54.69
High income	99*	99*	99*	99.42	98*
East Asia & Pacific	29.74	48.70	70.98	77.60	62.50
Europe & Central Asia	86.36	81.95	–	91.32	–
Latin America & Caribbean	68.09	74.48	82.22	88.87	78.49
Middle East & North Africa	69.36	74.80	87.11	87.79	89.18
South Asia	16.50	34.59	69.79	83.73	32.62
Sub-Saharan Africa	32.31	36.03	48.81	58.16	24.02

Note: *: estimated.

Source: World Bank (2005a).

to 130 for East and South Asia. The greatest improvements occurred in MENA while performance declined in East Asia and the Pacific. Transmission and distribution losses worsened somewhat for developing countries to nearly 18 percent of electricity output. East Asia is the only region with a noticeable improvement.

Overall, developing countries still have far to go in stocks, service and quality. Despite major improvements, stocks and coverage of infrastructure remain a fraction of what they are in developed countries: one-fifth for teledensity and one-sixth for road density. Overall, reliability remains problematic: one-fifth of managers responding to an investment climate survey considered electricity as a major constraint to doing business in developing countries.

What are the key barriers to progress?

Pricing of services rarely covers costs, except for telecommunications, and performance varies systematically by region and subsector, with water having the least cost recovery (World Bank, 2005b). In telecommunications, cross-subsidization from long-distance to local service is still common. Electricity tariffs are normally below costs. Latin America's are about 75 percent of Organisation for Economic Co-operation and Development (OECD) tariff levels and do not cover full costs. Other regions' power tariffs range from a third to a half of OECD levels. Cost recovery in water supply typically lags

Table 21.3 Performance indicators vary substantially across the developing world

	Electricity (% of managers surveyed ranking this as a major constraint)	Telephone faults (per 100 mainlines)		Electric power transmission and distribution losses (% of output)	
	(2001 to 2003 whichever available)	Average 1990–95	Average 1996–2003	Average 1990–99	Average 2000–02
East Asia & Pacific	22.58	114.38	129.51	15.31	12.35
Europe & Central Asia	9.81	65.74	40.24	16.87	17.73
Latin America & Caribbean	26.23	52.62	36.49	17.50	18.07
Middle East & North Africa	–	68.62	34.02	13.75	15.27
South Asia	36.72	176.36	135.14	20.82	22.34
Sub-Saharan Africa	49.99	127.87	78.09	17.50	19.00
All developing countries	21.60	96.59	65.33	16.82	17.54

Source: World Bank (2005a).

behind power, with Latin America's water tariffs at 40 percent of OECD levels and Africa and South Asia with the weakest cost recovery. The fiscal cost of subsidies can be dramatic. In Mexico, congressionally set electricity tariffs result in a public subsidy that amounts to 1 percent of GDP.

The subsidies predominantly go to the better-off. While often advocated to reduce poverty, subsidies for infrastructure (normally from government budgets) are rarely targeted by income level and are generally available to all service users. Data from Central America suggests that in El Salvador, 70 percent of electricity subsidies benefit the non-poor; 95 percent of households in Guatemala and 85 percent in Honduras benefit from social tariffs in electricity. Attempts at targeting subsidies towards the poor by offering low prices to a first (small) block of consumption can be reasonably effective in the case of electricity, where consumption is fairly closely related to income. For water the relation to income is weaker, and effective targeting requires either connection subsidies (which favor the poor as the rich tend already to be connected) or the use of existing data to identify needy families. Connection subsidies tend also to be an effective means of

helping the poor for power and gas. Since the poor receive no consumption or service subsidy benefits when they are not connected, subsidies for connection to the network (rather than for service consumption) are likely to be much more pro-poor (Komives et al., 2005).

Improper pricing constrains service and promotes inefficient consumption. When prices do not cover costs, all customers are subsidized and service expansion increases losses to service providers. Reluctance to set prices to cover costs, or to raise prices to keep up with inflation, has been an important impediment to private investment in infrastructure, and is one of the stumbling blocks to private sector participation in many countries. It also constrains public investment because increased public investment entails larger public operating subsidies in the face of growing expenditure needs in the social sectors. Finally, when services are priced below costs, customers will overconsume the service – producing inefficiencies in consumption that can be even more costly than inefficiencies in service production. For example, economic studies indicate that an increase in the price of electricity of 10 percent reduces demand by 5 percent (Berndt and Wood, 1975). Having power prices cover only half of costs increases demand substantially and leads to additional investment in generating capacity that then does not pay for itself. This impedes financing – both private and public.

Decentralization or devolution has had advantages and disadvantages. While the extent of change varies by region, decentralization of responsibility for infrastructure has increased in most countries. In Latin America, where decentralization is most advanced, it has presented challenges for reform. Although decentralization generally improves knowledge of local needs and priorities, many municipalities lack needed technical expertise for efficient service provision. Decentralization often also leads to policy and regulatory incoherence between municipalities, particularly in services (such as water supply and sanitation) where networks are local and not national. Decentralization also increases transaction costs for private investors, who must negotiate with individual municipalities rather than with a national or state entity.

Private sector investment in developing-country infrastructure grew nearly fivefold from \$20 billion in 1993 to a peak of \$100 billion in 1997. It has since declined to around \$45 billion (World Bank, 2003). The source of private infrastructure investment is changing. In sub-Saharan Africa from 1998 to 2003, 38 percent of private investment funds were from other developing countries in the region (principally South Africa) and nearly 20 percent from domestic investors (World Bank, 2005c). This differs from earlier periods when private investment was mostly from developed countries.

Private sector investment has been concentrated in a few countries and sectors. The Latin American region received roughly half of the \$786 billion of private investment in infrastructure from 1990 to 2003, and half of these funds financed privatizations (World Bank, 2005b). Sub-Saharan Africa received less than 10 percent of this global total, and half of those funds went to one country, South Africa. The top four recipients in the African region (South Africa, Nigeria, Mozambique and Côte d'Ivoire) received 70 percent of the Africa regional total (World Bank, 2005c). Overall, the top ten recipients globally received 62 percent of total private investment in infrastructure from 1990 to 2003 (World Bank, 2003). Globally, telecoms received 42 percent of private capital flows to infrastructure in developing countries. In Africa, its share is 70 percent.

While private investment in infrastructure was growing in the 1990s, overseas development assistance (ODA) to infrastructure declined. For example, in sub-Saharan Africa annual ODA commitments for infrastructure were \$5 billion in 1989 and fell to \$2.4 billion in 2003 (World Bank, 2005c). Relative to private investment, ODA for infrastructure is more evenly distributed across countries and sectors. ODA infrastructure funding is only recovering in 2005 nearly a decade after private investment flows peaked.

The performance of private investment in infrastructure has been good, but below unrealistically high expectations. While public support for private investment in infrastructure has declined in most countries, the actual performance of private participation exceeds public perceptions (Birdsall and Nellis, 2002; Estache and Rossi, 2004). Private provision of services has expanded greatly. In Latin America in 1990, private companies provided only 3 percent of telephone and electricity connections. By 2003 their share was 86 percent for telephones and 60 percent for electricity connections (World Bank, 2005b). In Africa by 2003, half of the countries had private participation in telephone service and 40 percent in power generation. In most cases, private providers have improved efficiency, coverage and service quality. The poor have benefited mainly from increased coverage. Concessions have not made excessive profits, especially given their risks: 40 percent are unlikely to earn any profit (Sirtaine et al., 2005; Estache and Pinglo, 2004), although nearly all telephone concessions earn profits. There have been a few well-publicized failures, and a large share of concession agreements have been renegotiated. In Latin America, 30 percent of concessions have been renegotiated, including three out of four in water and sanitation (Guasch et al., 2003). Cost recovery for private providers has been a persistent problem in all sectors except telecom, and governments have tended not to enforce service payments.

Private investment did not fill the gap left by declining public investment and ODA in some regions. Reliable data across developing countries on public sector investment in infrastructure does not exist, but the available evidence shows a consistent decline since 1995 in some regions. For example, African central government expenditure on infrastructure fell from 4.2 percent of GDP in the mid-1980s to 1.6 percent in the late 1990s (World Bank, 2005c). These estimates are suggestive but only partial as they exclude state-owned enterprises and sub-national governments. Consistent estimates compiled for the Latin American region show that public investment in infrastructure declined from 3.5 percent of GDP in the mid-1980s to 1.7 percent in 2001. However, in many East Asian countries such as China, Thailand and Vietnam, total annual infrastructure investment has exceeded 7 percent of GDP, and the bulk of these funds represent public investment (Asian Development Bank, 2005). In these countries, infrastructure stocks have grown in proportion to GDP, and public investment in infrastructure has expanded over time.

Regulatory issues also remain unresolved in many countries. Independent regulators need to discipline private participants and protect consumers – and also protect investors from arbitrary intervention from governments when sunk costs are high, as is often the case with infrastructure. In fact, many regulators are insufficiently independent of governments to fulfill this latter role. Moreover, greater regulatory autonomy is associated with better quality and efficiency of service provision by private providers (Andres et al., 2006). Regulatory risk raises the cost of private capital (as much as 2 to 6 percent in Latin America (Guasch and Spiller, 1999)). While many developing countries now have regulatory bodies, their establishment often followed rather than led the initiation of private participation. Experience has shown that regulatory independence is important for effective private participation in infrastructure, and that it is difficult to attain quickly.

Current thinking on regulatory practices is now much more pragmatic. It is clear that regulatory arrangements need to be adapted to the country context. For example, price cap regulation – which limits the prices firms may charge, not their profits – may be the best model theoretically, because it promotes efficiency. But in Latin America it has not worked well. Price caps entail more risk for operators (as profits are not guaranteed), raising the cost of capital. And they have proved much more susceptible to renegotiation and to the effects of external shocks (Guasch et al., 2003). Traditional rate-of-return or hybrid regulatory regimes may therefore be preferable in Latin America and other developing regions. In addition, it is now clear that the difficulty of building independent, effective and competent regulatory agencies was seriously underestimated.

Conclusion

Private participation in infrastructure was oversold in the 1990s. Many expected private involvement in infrastructure to fill financing gaps, improve efficiency of service provision and extend service coverage. The objective evidence indicates that it achieved gains in all of those areas, but not to the extent expected. Private investment flows have subsided well below past peaks, but they are still substantial and likely to be sustainable at current levels. However, private participation (and particularly private investment) is concentrated in a few sectors: telecoms and to a lesser extent electric power, with little activity in transport. In water supply, the major involvement has been in managing concessions and not in providing investment funds. Private participation is also concentrated in a few countries, even when the share of investment is compared to the share of GDP. The sector that has seen the most private participation across countries is telecom – the best-performing sector in having tariff revenue cover costs.

Governments have a continuing but new role in infrastructure. Many have been moving away from the old model of direct provision of infrastructure services through vertically integrated state enterprises to a model with more unbundling of services and a more businesslike approach. Hopes that private investors would finance infrastructure have been replaced by the realization that services must be paid for by customers, taxpayers or foreign aid agencies – whether the service providers are public or private. The need for regulation – whether service provision is public or private – is also more widely understood. Countries are beginning to adopt a hybrid approach. This involves facilitating private participation in specific sectors such as telecom, port facilities and power generation where private interest and activity continues. And it involves more use of concessions in water supply and attempts to increase efficiency in transport through corporatization and the use of special financing mechanisms, such as road funds, that give users a voice in ensuring that revenues are efficiently spent.

The key issues going forward include restoring investor confidence in emerging markets in general and in infrastructure in particular. This will require developing appropriate risk-mitigation instruments so that each party bears that risk it is best able to manage, and improving cost recovery for operators – either by increasing prices so that the user pays for services (engendering user efficiency) or by having governments shoulder their responsibilities for social tariffs. More generally, most countries could benefit from more efficient targeting so that the poor are better protected while the cost remains affordable. Finally, it is now clear that many infrastructure services will remain publicly managed in many countries. Improved performance gains will therefore require reviving programs to improve public sector efficiency.

Note

1. The authors would like to thank Tito Yepes for his assistance on this chapter.

References

- Andres, L., V. Foster and J.L. Guasch (2006), 'The Impact of Privatization on the Performance of the Infrastructure Sector: The Case of Electricity Distribution in Latin American Countries', Policy Research Working Papers No. 3926, Washington, DC.
- Aschauer, D. (1989), 'Is Public Expenditure Productive?' *Journal of Monetary Economics*, 23:177–200.
- Asian Development Bank (2005), *Connecting East Asia: A New Framework for Infrastructure*, Asian Development Bank, World Bank, and Japan Bank for International Cooperation, Manila, The Philippines.
- Baltagi, H. and N. Pinnoi (1995), 'Public Capital Stock and State Productivity Growth: Further Evidence from an Error Components Model', *Empirical Economics*, 20: 351–9.
- Banerjee, Abhijit V. (2004), 'Who is Getting the Public Goods in India? Some Evidence and Some Speculation', in K. Basu (ed.), *India's Emerging Economy*, Cambridge, MA: MIT Press, pp. 183–214.
- Berndt, E.R. and D.O. Wood (1975), 'Technology, Prices, and the Derived Demand for Energy', *Review of Economics and Statistics*, 57: 259–68.
- Birdsall, N. and J. Nellis (2002), 'Winners and Losers: Assessing the Distributional Impact of Privatization', CGD Working Paper 6, Center for Global Development, Washington, DC.
- Brenneman, A. (2002), 'Infrastructure and Poverty Linkages, a Literature Review', World Bank, Washington, DC.
- Calderón, C. and L. Servén (2003), 'The Output Cost of Latin America's Infrastructure Gap', in W. Easterly and L. Servén (eds), *The Limits of Stabilization: Infrastructure, Public Deficits and Growth in Latin America*, Washington, DC: World Bank, pp. 95–118.
- Calderón, C. and L. Servén (2004a), 'The Effects Of Infrastructure Development On Growth And Income Distribution', Policy Research Working Paper No. 3400, World Bank, Washington, DC.
- Calderón, C. and L. Servén (2004b), 'Trends in Infrastructure in Latin America, 1980–2001', Policy Research Working Paper No. 3401, World Bank, Washington, DC.
- Canning, D. (1998), 'A Data Base of World Infrastructure Stocks, 1950–1995', Policy Research Working Paper No. 1929, World Bank, Washington, DC.
- Canning, D. (1999), 'The Contribution of Infrastructure to Aggregate Output', Policy Research Working Paper No. 2246, World Bank, Washington, DC.
- Canning, D. and E. Bennathan (2000), 'The Social Rate of Return on Infrastructure Investments', Policy Research Working Paper No. 2390, World Bank, Washington, DC.
- Cashin, P. (1995), 'Government Spending, Taxes, and Economic Growth', *IMF Staff Papers*, 42 (2): 237–69.
- Estache, A. and Pinglo, M.E. (2004), 'Are Returns on PPI in LDCs Consistent with Risk since the Asian Crisis?' Policy Research Working Paper No. 3373, World Bank, Washington, DC.
- Estache, A. and M. Rossi (2004), 'Have Consumers Benefited From The Reforms In The Electricity Distribution Sector In Latin America?' Policy Research Working Paper No. 3420, World Bank, Washington, DC.
- Fay, M., D. Leipziger, Q. Wodon and T. Yepes (2005), 'Achieving Child Health-Related Millennium Development Goals: The Role of Infrastructure', *World Development*, 33 (8): 1267–84.
- Gomez-Ibanez, Jose Antonio (2003), *Regulating Infrastructure: Monopoly, Contracts, and Discretion*, Cambridge, MA: Harvard University Press.
- Gramlich, E. (1994), 'Infrastructure Investment: A Review Essay', *Journal of Economic Literature*, 32: 1176–96.
- Guasch, Jose Luis (2005), *Granting and Renegotiating Infrastructure Concessions – Doing it Right*, WBI Development Studies, Washington, DC: World Bank.

- Guasch, J.L., J.J. Laffont and S. Straub (2003), 'Renegotiations of Concession Contracts in Latin America', World Bank Policy Research Working Paper No 3011, World Bank, Washington, DC.
- Guasch, J.L. and P. Spiller (1999), *Managing the Regulatory Process: Design, Concepts, Issues and the Latin American and Caribbean Story*, World Bank, Washington, DC.
- Holtz-Eakin, D. (1994), 'Public Sector Capital and the Productivity Puzzle', *Review of Economics and Statistics*, 76 (1): 12–21.
- Hulten, C. (1996), 'Infrastructure Capital and Economic Growth: How Well You Use It May Be More Important than How Much You Have', NBER Working Paper 5847, National Bureau of Economic Research, Cambridge, MA.
- Ingram, G.K. and Z. Liu (1999), 'Determinants of Motorization and Road Provision', in J.A. Gomez-Ibanez, W.B. Tye and C. Winston (eds), *Essays in Transportation Economics and Policy*, Washington, DC: Brookings Institution, pp. 325–58.
- International Institute for Management Development (2005), *World Competitiveness Yearbook 2005*, Lausanne, Switzerland: World Competitiveness Center.
- Jacoby, H. (2002), 'Access to Rural Markets and the Benefits of Rural Roads', *Economics Journal*, 110: 713–37.
- Kessides, I.N. (2004), *Reforming Infrastructure: Privatization, Regulation and Competition*, Oxford: Oxford University Press, and Washington, DC: World Bank.
- Komives, K., V. Foster, J. Halpern, Q. Wodon and R. Abdullah (2005), 'Who Benefits from Utility Subsidies? Water, Electricity, and the Poor', World Bank Institute Development Studies, Washington, DC.
- Krugman, P. (1994), 'Competitiveness: A Dangerous Obsession', *Foreign Affairs*, 73 (2): 28–44.
- Lee, K.S. and A. Anas (1992), 'Costs of Deficient Infrastructure: The Case of Nigerian Manufacturing', *Urban Studies*, 29 (7): 1071–92.
- Ndulu, B. (2004), 'Infrastructure, Regional Integration, and Growth in Sub-Saharan Africa', Background Paper for the Commission for Africa.
- Prud'homme, R. (2004), 'Infrastructure and Development', paper presented to Annual Bank Conference on Development Economics, World Bank, Washington, DC.
- Sirtaine, S., M.E. Pinglo, J.L. Guasch and V. Foster (2005), 'How Profitable are Infrastructure Concessions in Latin America? Empirical Evidence and Regulatory Implications', *Quarterly Review of Economics and Finance*, 45 (2–3): 380–402.
- Taylor, Lance (2001), *External Liberalization, Economic Performance, and Social Policy*, New York: Oxford University Press.
- World Bank (1994), *World Development Report 1994: Infrastructure for Development*, Washington, DC: World Bank and Oxford University Press.
- World Bank (2003), *Private Participation in Infrastructure: Trends in Developing Countries in 1990/01*, Washington, DC: World Bank and Public–Private Infrastructure Advisory Facility, Washington.
- World Bank (2004), *World Development Report 2005: A Better Investment Climate for Everyone*, Washington, DC: World Bank and Oxford University Press.
- World Bank (2005a), *World Development Indicators 2005*, Washington, DC: World Bank.
- World Bank (2005b), *Infrastructure in Latin America: Recent Developments and Key Challenges*, Washington, DC: World Bank.
- World Bank (2005c), *Infrastructure in Sub-Saharan Africa*, Washington, DC: World Bank.
- World Bank (2005d), *Doing Business in 2005*, Washington, DC: World Bank.
- World Economic Forum (2004), *The Global Competitiveness Report 2003–2004*, Geneva, Switzerland.

22 Population and development

Dennis Ahlburg and Robert Cassen

The link between population growth and economic development is among the older issues in social science, particularly because of its association with the name of Robert Malthus. His famous *Essay on Population* of 1798 argued that population growth inevitably led to poverty – fundamentally, he claimed, because it would always outstrip the means of subsistence. Populations would always increase so that the supply of labour pressed on wages, to the point where they reached subsistence level. Beyond that point population would only be held back by war, starvation or disease, but people would remain poor. While in later editions Malthus modified his views considerably, his name is mainly associated with the thesis of his original *Essay*.

This association has hardly been helpful to our subject. Malthus's early views were linked to harsh social philosophies, and in the past anyone who believed that population growth had any kind of negative impact on development was commonly labelled as 'Malthusian', and often condemned as a result. Things have changed. Today there is a growing consensus that rapid population growth in poor countries under conditions of high fertility can have negative, non-Malthusian, consequences for economic and social development. But they are not necessarily large, nor are they irremediable. And Malthus was wrong about food supplies, which have consistently outgrown population globally and in most individual countries – with some, mainly African, exceptions.

Development influences population growth and vice versa. We will treat the issues separately, though very summarily with the impact of development on population.

Development and population

Fertility

For a long period in human history populations grew very slowly, with quite high fertility but also high mortality, and episodes of extraordinary mortality such as the Black Death, which killed one-third of Europe's population in the fourteenth century. The first key to lower fertility was usually improvements in mortality. People do not want just babies, but surviving children, and when child survival is low, families will 'insure' by

having large numbers of babies. Once confident of their children's survival, parents may begin to limit the number of children they have.

The first widespread decline in fertility occurred in France in the eighteenth century. It began to decline in Europe generally in the nineteenth century, commonly – though not universally – preceded or at least accompanied by declining mortality. The fact that historical fertility decline mainly coincided with increasing prosperity, and is lower today in better-off than in poorer countries, gives the (broadly correct) impression that economic development is associated with fertility decline. But it is very far from a simple relationship. On the contrary, while the decline usually starts with better-off, urban and educated parents, it can spread to those who do not have these characteristics. In the developing countries today there is an association between falls in fertility and mortality improvements, rising education (especially female education) and the spread of contraception. But in some countries today there is significant fertility decline among the uneducated and the poor (Bhat, 2002; McNay et al., 2003). At some point the pace of fertility decline often outruns that of the spread of its correlates (Van de Kaa, 1996).

Mortality

Much the same may be said of mortality improvement. Historically mortality has responded to the gradual disappearance of major causes of death – famines, epidemics, contagious diseases – as well as to more fundamental forces such as improving nutrition, hygiene and public health measures. On the whole until relatively recently curative medicine has been a weaker force. Again broadly there is an association between increasing prosperity and mortality decline, since prosperity typically brings with it the things that reduce mortality. But it is a far from straightforward process. Infant mortality in particular can respond to very specific interventions, and throughout history there have been long periods in different countries where living standards have improved but infant mortality has not, and conversely times when infant mortality has fallen without widespread improvements in levels of living. (See for example Woods, 2000.)

Mortality in general in developing countries has often fallen rapidly due to medical and public health interventions; many countries have achieved in a few decades the kinds of declines in mortality that took a century or more in countries that are now industrialized. For this reason they have experienced rates of population growth greatly in excess of those in the history of the industrialized countries. There is still, though, high mortality in many developing countries and this is a drag on development. A number of studies have shown that healthier countries grow faster (Easterlin, 1996; Bloom and Canning, 1999), and Robert Fogel (1994)

claimed that synergies between technological and physiological improvements in health account for about one-half of the economic growth in Europe over the previous two centuries. HIV/AIDS is taking a large toll in many countries, and has still to reach its peak in several. It too has a negative impact on development, often killing men and women in their prime working years, placing huge burdens on health services, and creating millions of orphans. It has been estimated that in the 1990s, AIDS reduced per capita annual growth in Africa by 0.8 percentage points. In the worst-affected countries the reduction in growth was one to two percentage points. After two decades, the economies of these countries would be about 20 to 40 per cent smaller as a consequence of AIDS (Loewenson and Whiteside, 2001).

Fertility decline itself contributes to improvements in mortality and health. Death in childbirth is still a significant cause of female mortality in many developing countries, and high fertility is often associated with higher rates of child malnutrition in the family.

Population and development

Macroeconomics

One of the major debates in development economics has been over the macroeconomic role of population growth. In the twentieth century the seminal work was Coale and Hoover's 1958 study using the example of India. It compared two paths for the economy, one with higher fertility than the other, and reached a powerful conclusion: not only was the growth of per capita income lower under the high-fertility variant, but also even the growth of aggregate gross national product (GNP) was lower. The result derived from two assumptions in their model: one was that the burden of dependency, the ratio of non-workers to workers in the population, was greater under high fertility and led to reduced savings; and the other was that investment had to be spread over larger numbers instead of raising the amount of physical or social capital per worker. This was called 'capital widening versus capital deepening'; that is, if the population were growing more slowly, the same amount of capital would be used to improve the quality of schooling or health services received by each individual, instead of being diluted by having to extend coverage to more people; or else there would be more or better capital for each worker in the workplace.

These findings were hotly disputed; attempts to measure the burden-of-dependency effect on savings suggested it might be quite small, and similar questions were raised about the capital-dilution argument. Most attempts to measure the effect of population on economic growth (either in the aggregate or per capita) suggested the impact was small or non-existent

(Temple, 1999) and empirical estimates were fragile, dependent on model specification and data used (Levine and Renelt, 1992). Recent studies have criticized Coale and Hoover's assumption that investments in education and health did not promote economic growth – these investments have been highlighted in the 'new growth economics' literature (for example Barro, 1997) – as well as their focus on short-run impacts of population and without consideration of longer-run impacts.

Research in the last two decades has brought a swing of the pendulum in the macroeconomic discussion. Bloom and Freeman (1988) and Blanchet (1991) showed that mortality and fertility declines had different impacts on economic growth, so models that considered only aggregate population were misspecified. A series of cross-country studies by Kelley and Schmidt (1996, 2001) followed. Their initial work indicated that the positive and negative effects of population probably offset each other in the 1960s and 1970s but that a net negative effect in the 1980s was likely. In their more recent work they conclude that about 20 per cent of economic growth over the period 1960 to 1995 can be attributed to mortality and fertility declines, with the larger contribution coming from mortality.

Rapid decline in population growth, and even more the dramatic economic growth in East Asia in 1960–85, gave a further boost to the study of population and development. The keys to the relationship were thought to be decreased dependency burden (commonly known as the 'demographic bonus') leading to higher savings and more investment in education. An influential study by the World Bank (1993) argued that a large proportion of that growth was due to improvements in education, in turn made possible by lower population growth. Although the magnitude of the contribution of education to increasing economic growth has been challenged, most models of economic growth now include education as a contributing factor. Analysis of the East Asian experience also focused not just on the short-run effects of population growth, as had Coale and Hoover, but also on the intermediate effects where a rising number of young people enter the labour force, and the long-run effects that occur as they retire. Savings (and wealth) rise in the intermediate period and although they may fall in the long run, overall the demographic bonus may account for as much as one-third of the rapid economic growth in East Asia (Bloom and Canning, 1999).

It should be noted that these recent studies do not show a relationship between the rate of population growth per se and economic growth; it is rather certain demographic features and the timing of their change which may matter: fertility, the age distribution and life expectancy. Some of the models incorporate simultaneous relationships, with economic growth affecting the demographic variables and vice versa, giving rise to virtuous

or vicious circles of rapid or slow growth (Bloom and Williamson, 1997; Bloom et al., 2000; Bloom and Canning, 2001). The 'demographic bonus', if real, does not confer an automatic boost to growth. Countries can use the bonus wisely or unwisely, as evidence from Asia shows. It translates into higher economic growth if supportive policies, markets and institutions exist. If they do not, then the bonus will be squandered.

Poverty

It is widely believed that more rapid population growth increases poverty by reducing real wages. However, as noted by McNicoll (1997), the relationship with poverty is 'neither obvious nor well established'. For example, Eastwood and Lipton (2001) identify at least 60 effects of population on poverty, and a recent study has questioned the assumption that an increase in the labour force (from an increase in population) necessarily reduces wages (Ahlburg, 2002). The poverty measure in these studies is generally income poverty, rather than a broader definition such as in Sen's capability approach (Sen, 1985).

There have been surprisingly few attempts to estimate the impact of population on poverty directly, and most have been at the macro level. Three studies of Indian states found a small positive impact of population growth on income poverty (van de Walle, 1985; Evenson, 1993; Chelliah and Sudarshan, 1999). In cross-country regressions Ahlburg (1996) found no relationship between population growth and poverty. Other similar studies found that the major variables explaining cross-country variation in poverty were the rate of economic growth and the degree of income inequality. In contrast, Eastwood and Lipton (2001) did find a considerable effect of population on poverty: 'The average (developing) country in 1980 had a poverty incidence of 18.9 per cent; had it reduced its fertility by 5 per 1000 throughout the 1980s (as did many Asian countries), this figure would have been reduced to 12.6 per cent' (p. 218). This estimate must be viewed with caution for, as DeHaan and Lipton (1998) have shown, the relationship between population and poverty varies considerably across regions, countries, growth sectors and policy environments.

At the level of the household, one must take care to identify the source of population change and the timing of the measurement of the association between population and poverty. If family size increases because of a birth, poverty may rise because more mouths are trying to consume the same amount of resources. For example, in a study of 211 agricultural households from 1975 to 1983 Gaiha and Deolalikar (1992) found that larger families were more likely to be poor at any given point in time, and also more likely to experience persistent poverty. But resources may not remain constant. Members of the household may increase their labour

supply or leave the household, assets may be sold, or the family may receive income from relatives. All of these effects (and the many more noted by Eastwood and Lipton) influence the estimate of the relationship between population change and poverty. The timing of the measurement of the association is also important. Children may be the best investment the poor can make for their old age, so that increasing current household size may increase poverty in the short run with the expectation that it will reduce poverty in the long run – though this expectation may be defeated if children die early, fail to become gainfully employed, or fail to contribute income to the household. While some attention has been paid to the impact of an additional birth on poverty at the household level, much less attention has been paid to the impact of a death on poverty. The death of an adult may increase the likelihood of the family becoming poor unless there are offsetting factors, such as increased resources flowing in from relatives or increased work by other family members. The empirical evidence suggests that household income and asset ownership decline with the death of an adult. For example, it was estimated that in Botswana from the mid-1990s to the mid-2000s the poverty rate would increase by up to eight percentage points due to AIDS and average household income would fall by 10 per cent (Loewenson and Whiteside, 2001). The negative impacts of a parental death can have far-reaching implications. In a study of ten African countries, Case et al. (2004) found that orphans were less likely to be enrolled in school than non-orphans. This reduction in human capital formation raises concerns about higher poverty and lower growth in the next generation.

While the study of the direct relationship between population change and poverty has proven to be quite difficult, there have been a large number of studies of the effects of population change on aspects of well-being other than income poverty. There are quite a number of household studies which show reasonably strong correlations between measures of fertility and measures of women's and children's health and survival (Montgomery and Lloyd, 1996), and between number of surviving siblings and children's education, especially for female children (Lloyd, 1994). While these authors accept the fact that their studies can be criticized on methodological grounds, they claim that they do identify underlying causal connections.

Environment

Energy, transport and industry

The main sources of air, soil and water pollution are – apart from agricultural chemicals – energy, transport and industry. As developing economies grow, they substitute modern forms of transport and energy production for

traditional ones: cars, trucks and trains replace horse and bullock transport, and coal, gas and oil take over from vegetable matter and animal products as fuels. Manufacturing, processing, heavy industry and services become the dominant production sectors, and even agriculture uses more chemicals and mechanization. All these release chemical pollutants and particulate matter into the atmosphere, soil and water courses.

But what is the role of population? In most of these productive activities it is mainly one of derived demand, and it frequently plays a relatively modest part. The economy and the pattern of development are the powerful factors, with commercial energy use and modern forms of transport often growing at several times the rate of population growth. (Dyson et al., 2004 gives many examples for India.) At the same time, population is itself an underlying demand factor: as populations grow there are more people and goods to be transported, more demand for the products of manufacturing and the like. How these factors interact is no simple matter.

A common mistake is to take the per capita consumption for, say, energy, and multiply by population growth to project energy consumption. Per capita consumption will change as population and the economy grow together; and a key aspect is the household. Since there are economies of scale in household energy consumption, the pattern of household formation will greatly affect demand. O'Neill et al. (2001) show the very considerable difference between household-based and population-based assessments of energy demand. There are of course also important scale factors at levels beyond the household.

In addition, technology changes. Most industrial processes are subject to technological change which can reduce emissions, and the cost of reducing them falls over time (Anderson, 2001a and 2001b). Modelling the likely output of various pollutants over time, Anderson (2004) has shown that the early introduction of 'clean' technology far outweighs the influence of population growth on a variety of emissions in energy, industry and transport. The one main exception to such optimism is the small-scale sector, which can be highly polluting, and where clean technological progress is less apparent.

Undoubtedly population growth plays a part but how important this is depends upon other factors. If cleaner technologies continue to be developed, and go down in price at past rates, societies in the main can enjoy higher standards of living and modern economic growth while protecting their environments from chemical pollution. To say that they can, however, is not to say that they will. Where 'dirty' technologies are already installed and the investments have significant economic lives left, or where the costs of cleaner technologies are such that government policies are required to

ensure their introduction, but these policies are not in place, the combination of modern economic growth with rapidly growing populations can indeed be a recipe for rising pollution.

The urban environment is subject to much the same analysis. Population growth adds to demand for environmental resources, but often less slowly than the pace and pattern of economic growth. Waste is a particular problem for cities. There is much valuable experience from developing countries in waste management, which can often be a paying proposition. Municipal finances are, however, often weak, and planning capacity limited. Very few cities in the developing world have coped satisfactorily with the combination of economic growth and population growth, the latter often at very high rates, especially where natural growth is added to by inward migration from rural areas. Extremes of income inequality only add to the difficulties. The problems are not unmanageable in principle, but are frequently poorly managed in fact.

Water

Water differs from other environmental issues. The growth of demand for energy, manufactures and modern forms of transport is only partially affected by population growth, and there are commonly technological 'fixes'. The demand for water, though, is strongly influenced by population growth, and the role of technology is limited. Every additional person requires their own water supply; but more importantly for water, they have to be fed, and in economies that are self-sufficient in agriculture, about four-fifths of water demand comes from agriculture, while residential use takes up about 5 per cent, the rest being required for industry and ecological services. If such economies wish to remain self-sufficient in food, they either have to use more water, or achieve greater water efficiency. (Exporting other goods and importing food is of course an option. It is tantamount to importing water, and several countries are likely to be forced to go down this route.)

In many countries, water is already scarce, at least regionally, or seasonally, if not nationally or chronically. Water pollution and climate change may also be reducing availability. While there are potential technological means for conserving water and increasing the amount of crop yield per unit of water, they typically offer relatively small gains. It may be that desalination of seawater will become more economic in future; at present it is only economic where alternative sources are highly expensive. But that apart, and then only for areas close to the sea, there is little alternative to water regulation and pricing. These are often feasible, but politically and socially difficult. For such reasons, water is likely to be the greatest environmental challenge of population growth.

Ecology

Studies undertaken in the 1980s indicated that population growth may have had a detrimental impact upon renewable resources such as rainforests and fishing areas. Since the mid-1990s researchers have reached more conditional conclusions. While increasing numbers inevitably place pressure on such resources as forests, grazing land, animal habitats and the like, it would be wrong to 'blame' population growth for much of the diminution in these resources that has taken place all over the world. Much depends on the nature of management of the resources, development of property rights, development or adaptation of technology, land tenure relations, population mobility, and markets and other institutions and organizations. Many traditional systems of management have been successful in conserving natural resources over long periods, even with growing populations. There have been well-documented cases in Africa where increasing population density and labour availability have led to sustainable agricultural intensification, rather than degradation. Such cases have usually been where land and tree tenure have been satisfactory, good market conditions for produce have been available, and tax regimes have not punished success (Tiffen et al., 1994). But the opposite has also been the case, in Africa and elsewhere (Baland and Platteau, 1996; Jodha, 1986). Traditional systems have often been better at sharing resources equitably than in raising their productivity to match rising numbers.

Population growth is likely to lead to the degradation of resources where there is open access to the resource and real rural incomes are stagnating or falling, land tenure is insecure and there is lack of access to credit, where alternative forms of employment are lacking, and where low levels of education and skill limit labour mobility (Panayotou, 1996). This is not to say that population growth typically helps; on the contrary, *ceteris paribus* there will usually be less strain on resources if populations grow slowly or not at all.

This literature can be set in a wider account of a potential beneficial influence of population growth when it leads to agricultural intensification and improved technology, and economies of scale. Should this occur, population growth could be a positive factor even macroeconomically. Boserup made a strong claim for the importance of these effects in a much-cited study (Boserup, 1965). But while having valuable insights on such a role for population in the long sweep of history, her account did not deal with more negative experience in the shorter term, and Boserup herself changed her views in later work (Boserup, 1981). It is perfectly clear from much African experience that population growth there has commonly failed to induce countervailing technological change in agriculture.

References

- Ahlburg, D.A. (1996), 'Population Growth and Poverty', in D.A. Ahlburg, A.C. Kelley and K.O. Mason (eds), *The Impact of Population Growth on Well-being in Developing Countries*, Berlin: Springer-Verlag, pp. 219–58.
- Ahlburg, D.A. (2002), 'Does Population Matter; A Review Essay', *Population and Development Review*, 28 (2): 329–50.
- Anderson, D. (2001a), 'Technical Progress and Pollution Abatement: An Economic View of Selected Technologies and Practices', *Environment and Development Economics*, 6 (3): 283–311.
- Anderson, D. (2001b), 'Dynamic Simulation and Environment Policy Analysis: Beyond Comparative Statics and the Environmental Kuznets Curve', *Oxford Economic Papers*, 53: 721–46.
- Anderson, D. (2004), 'Modelling the Environment: the Production and Use of Energy', in T. Dyson, R. Cussen and L. Visaria (eds), *21st Century India: Population, Economy, Human Development and the Environment*, Oxford: Oxford University Press.
- Baland, J.-M. and J.-P. Platteau (1996), *Halting Degradation of Natural Resources: Is There a Role for Rural Communities?*, Rome: Food and Agriculture Organization and Oxford: Clarendon Press.
- Barro, R.J. (1997), *Determinants of Economic Growth: A Cross-Country Empirical Study*, Cambridge, MA: MIT Press.
- Bhat, Mari P.N. (2002), 'Returning a Favour: Reciprocity Between Female Education and Fertility in India', *World Development*, 30 (10): 1791–1803.
- Birdsall, N., A.C. Kelley and S.W. Sinding (eds) (2001), *Population Matters: Demographic Change, Economic Growth and Poverty in the Developing World*, Oxford: Oxford University Press.
- Blanchet, D.E. (1991), 'On Interpreting Observed Relationships Between Population Growth and Economic Growth: A Graphical Exposition', *Population and Development Review*, 17 (1): 105–14.
- Bloom, D.E. and D. Canning (1999), 'The Health and Wealth of Nations', *Science*, 287: 1207–9.
- Bloom, D.E. and D. Canning (2001), 'Cumulative Causality, Economic Growth, and the Demographic Transition', in N. Birdsall, C. Kelley and S.W. Sinding (eds), *Population Matters: Demographic Change, Economic Growth and Poverty in the Developing World*, Oxford: Oxford University Press.
- Bloom, D.E., D. Canning and P.N. Malaney (2000), 'Demographic Change and Economic Growth in Asia', *Population and Development Review*, 26 (Supplement): 257–90.
- Bloom, D.E. and R.B. Freeman (1988), 'Economic Development and the Timing and Components of Population Change', *Journal of Policy Modeling*, 10: 57–81.
- Bloom, D.E. and J.G. Williamson (1997), 'Demographic Change and Human Resource Development', in Asian Development Bank (ed.), *Emerging Asia*, Manila: Asian Development Bank, pp. 141–97.
- Boserup, E. (1965), *The Conditions of Agricultural Growth: The Economics of Agrarian Change under Population Pressure*, London: Allen & Unwin and Chicago, IL: Aldine.
- Boserup, E. (1981), *Population and Technical Change*, Chicago, IL: University of Chicago Press.
- Case, A., C. Paxson and J. Ableidinger (2004), 'Orphans in Africa: Parental Death, Poverty, and School Enrolment', *Demography*, 41 (3): 483–508.
- Chelliah, R.J. and R. Sudarshan (eds) (1999), *Income Poverty and Beyond: Human Development in India*, New Delhi: Social Science Press.
- Coale, A.J. and E.M. Hoover (1958), *Population Growth and Economic Development in Low-Income Countries*, Princeton, NJ: Princeton University Press.
- DeHaan, Arjan and Michael Lipton (1998), 'Poverty in Emerging Asia: Progress, Setbacks, and Log-jams', *Asian Development Review*, 16: 135–76.
- Dyson, T., R. Cassen and L. Visaria (eds) (2004), *21st Century India: Population, Economy, Human Development and the Environment*, Oxford: Oxford University Press.

- Easterlin, R.A. (1996), *Growth Triumphant: The Twenty-First Century in Historical Perspective*, Ann Arbor, MI: University of Michigan Press.
- Eastwood, R. and M. Lipton (2001), 'Demographic Transition and Poverty: Effects Via Economic Growth, Distribution and Conversion', in N. Birdsall, C. Kelley and S.W. Sinding (eds), *Population Matters: Demographic Change, Economic Growth and Poverty in the Developing World*, Oxford: Oxford University Press, pp. 213–59.
- Evenson, R.E. (1993), 'India: Population Pressure, Technology, Infrastructure, Capital Formation and Rural Incomes', in C.L. Jolly and B.B. Torrey (eds), *Population and Land Use in Developing Countries: Report of a Workshop*, Washington, DC: National Academy Press, pp. 70–97.
- Fogel, R.W. (1994), 'Economic Growth, Population Theory, and Physiology: The Bearing of Long-term Processes on the Making of Economic Policy', *American Economic Review*, 84: 369–95.
- Gaiha, R. and A.B. Deolalikar (1992), 'Persistent, Expected, and Innate Poverty: Estimates for Semi-arid Rural South India', *Cambridge Journal of Economics*, 17 (4): 409–21.
- Jodha, N.S. (1986), 'Common Property Resources and Rural Poor in Dry Regions of India', *Economic and Political Weekly*, 31 (27): 1169–81.
- Kelley, A.C. (1988), 'Economic Consequences of Population Change in the Third World', *Journal of Economic Literature*, 26 (4): 1685–728.
- Kelley, A.C. and R.M. Schmidt (1994), 'Population and Income Change: Recent Evidence', *World Bank Discussion Paper*, No. 249, Washington, DC: World Bank.
- Kelley, A.C. and R.M. Schmidt (1996), 'Towards a Cure for Myopia and Tunnel Vision of the Population Debate: A Dose of Historical Perspective', in D.A. Ahlburg, A.C. Kelley and K.O. Mason (eds), *The Impact of Population Growth on Well-being in Developing Countries*, Berlin: Springer, pp. 11–35.
- Kelley, A.C. and R.M. Schmidt (2001), 'Economic and Demographic Change: A Synthesis of Models, Findings, and Perspectives', in N. Birdsall, C. Kelley and S.W. Sinding (eds), *Population Matters: Demographic Change, Economic Growth and Poverty in the Developing World*, Oxford: Oxford University Press, pp. 67–105.
- Levine, R. and D. Renelt (1992), 'A Sensitivity Analysis of Cross-Country Growth Regressions', *American Economic Review*, 82 (4): 942–63.
- Lloyd, C.B. (1994), 'Investing in the Next Generation: The Implications of High Fertility at the Level of the Family', in R. Cassen (ed.), *Population and Development: Old Debates, New Conclusions*, Oxford: Transactions Publishers, pp. 181–202.
- Loewenson, R. and A. Whiteside (2001), 'HIV/AIDS Implications for Poverty Reduction', United Nations Development Programme Policy Paper, New York: UNDP Bureau for Development Policy.
- McNay, K., P. Arokiasamy and R.H. Cassen (2003), 'Why are Uneducated Women in India Using Contraception? A Multilevel Analysis', *Population Studies*, 57 (1): 21–40.
- McNicoll, Geoffrey (1997), 'Population and Poverty: A Review and Restatement', Working Paper No. 107, New York: Population Council.
- Montgomery, M. and C. Lloyd (1996), 'Fertility and Maternal and Child Health', in D.A. Ahlburg, A.C. Kelley and K.O. Mason (eds), *The Impact of Population Growth on Well-being in Developing Countries*, Berlin: Springer, pp. 37–68.
- O'Neill, B.C., L. MacKellar, and W. Lutz (eds) (2001), *Population and Climate Change*, Cambridge: Cambridge University Press.
- Panayotou, Theodore (1996), 'An Inquiry into Population, Resources and Environment', in D.A. Ahlburg, A.C. Kelley and K.O. Mason (eds), *The Impact of Population Growth on Well-being in Developing Countries*, Berlin: Springer, pp. 259–98.
- Sen, A.K. (1985), *Commodities and Capabilities*, Amsterdam: North-Holland.
- Temple, Jonathon (1999), 'The New Growth Evidence', *Journal of Economic Literature*, 37 (1): 112–56.
- Tiffen M., M. Mortimore and F. Gichuki (1994), *More People, Less Erosion: Environmental Recovery in Kenya*, Chichester: John Wiley & Sons.
- Van de Kaa, D.J. (1996), 'Anchored Narratives: the Story and Findings of Half a Century of Research into the Determinants of Fertility', *Population Studies*, 50 (3): 389–432.

- Van de Walle, D. (1985), 'Population Growth and Poverty: Another Look at the Indian Time Series Data', *Journal of Development Studies*, 21 (3): 29–39.
- Williamson, J.G. (2001), 'Demographic Change, Economic Growth, and Inequality', in N. Birdsall, C. Kelley and S.W. Sinding (eds), *Population Matters: Demographic Change, Economic Growth and Poverty in the Developing World*, Oxford: Oxford University Press, pp. 103–36.
- Woods, R. (2000), *The Demography of Victorian England and Wales*, Cambridge: Cambridge University Press.
- World Bank (1993), *The East Asian Miracle*, Washington, DC: World Bank.

23 Labor markets in developing countries

Albert Berry

It is useful to distinguish four broad levels of completeness or sophistication in the analysis of labor markets. Much of the literature stops at the first level, involving the point-of-time allocation of a fixed supply of labor to the most productive uses. The second brings in the generally trickier issues of dynamic (over time) efficiency – how possible differences in the way labor markets function can affect investment in physical capital, investment in human capital, and the rate and pattern of technological change. One aspect of dynamic efficiency involves ability to adjust to shocks, an increasingly important challenge in the light of the debt crisis and the increased volatility that characterizes more open economies. Not much analysis ventures explicitly into a third level where the objective function is not total income maximization (Pareto optimality) but rather total welfare maximization or ‘broad efficiency’. Some variables, like direct satisfaction from employment, are hard to get a quantitative handle on, though the ‘happiness literature’ does give some empirical evidence on how people’s reported satisfaction relates to various possible determinants like absolute income, relative income, stability of employment, conditions of employment, and so on (Oswald, 1997). Deciding how to weight the welfare of different people calls for some sort of third-party judgment. The fourth and final level of analysis, more the realm of sociologists and others than of economists, takes account of the way labor market functioning contributes to changes in taste over time, a relevant aspect of ‘performance’ since some preference systems (for example ones in which people cooperate easily) lend themselves to higher levels of human satisfaction than do others (for example ones in which people’s interaction with each other is mainly competitive in character, to the point where one person’s satisfaction or success makes others feel worse off).

At any of these levels of analysis, an important consideration is how the labor market is linked to the institutions of social security. One of the great human needs, which prosperous societies eventually address, is that for income security. That security is jointly determined by how the labor market works and by the complementary social and governmental institutions. The labor market must thus be judged together with those institutions in terms of how well they deal with the need for income security.

The evolution of thinking about the performance of labor markets and related policy issues

Much of the thinking on the functioning and efficiency of developing-country labor markets has its roots in the perception that something is not working well. Commonly identified warning signals of labor market malfunction leading to deadweight loss have included: gaps in the labor earnings and other conditions of employment between groups of workers whose productivity would be expected to be similar; rigidity in the movement of nominal or real wages of groups of workers; high levels of open unemployment or other indications of non-use of labor; skills mismatches, taking the form of people with more education or training than their jobs require and/or long queues of people lining up for certain types of jobs; the presence of clear market power on either the worker side (unions) or the employer side (for example monopsony), or labor legislation which interferes with competitive market forces; and direct evidence or presumption that market forces are not duly taken into account in hiring and payment decisions, with this last critique arising mainly *vis-à-vis* the public sector. Meanwhile, high levels of inequality are also sometimes blamed in part on labor market malfunctioning.

The 'labor surplus' model

Lewis (1954) posited a substantial reserve of labor that lay unutilized or underutilized in the 'traditional' sector of the economy, raising the question of whether a significant part of this reserve could be put to productive use. Early estimates of the amount of underutilized labor ranged up to 30–40 percent of the labor force (Kao et al., 1964) but, not being based on any rigorous methodology, were no doubt misleadingly high. Subsequent experience made it clear that underutilized labor might be available for some productive uses but not for others.

As economies have become more urbanized such labor surplus as exists is increasingly seen to be lodged in the large urban informal sector (Reynolds, 1969). An early concern in some countries that open urban unemployment was rising to serious levels also turned out to be overdrawn¹ and misleading on several counts. There was in fact no general upward trend in open unemployment in developing countries apart from the transitory effect that accompanied the increasing shares of workers who were young or were found in urban settings; with fairly narrow definitions of open unemployment the observed rates are seldom as high as 10 percent. A fair number of those in the open unemployment pool are from reasonably well-off families and are thus not forced quickly into self-employment or some undesirably low-paid job (Udall and Sinclair, 1982). Still, such caveats notwithstanding, open unemployment above levels interpretable as normal

'friction' (perhaps 5–6 percent) is symptomatic of problems in the labor market and is likely to imply both a deadweight loss on the output side and a higher poverty incidence.

The informal sector: another form of surplus labor or an efficient part of the economy?

Much initial speculation that a large informal sector represented serious labor misallocation was also exaggerated.² The idea that persons working there were, in general, isolated or marginalized from the formal sector (Vekemans and Silva, 1969) was fairly quickly disposed of (for example Perlman, 1976). Another interpretation had the informal sector systematically subordinate to or exploited by the formal one (Moser, 1978). At the other end of the spectrum the sector was seen in a positive light by the International Labour Organization (ILO) in its 1972 Kenya report (ILO, 1972), one of the documents that put the term 'informal' into the lexicon of development economists. The labor-intensity of most informal sector activities suggested to many (for example Liedholm and Mead, 1987) that its (narrow) efficiency levels were typically high, when calculated properly (that is, with scarcity rather than market prices as the cost of each factor of production). Presumably its 'broad efficiency' would be higher still given its importance as a source of income for poorer families.

Potential damage from labor market imperfections: debate around the pro-market perspective

Since the 1970s no major new models or frameworks to help us better understand less-developed-country (LDC) labor markets have captured widespread attention, but the neoliberal revolution in economic thinking has brought increasing and more general attention to the inefficiencies which could result from badly designed interventions in labor markets.³ Simple theory suggests that excessive protection (high minimum wages, stringent firing rules, powerful unions, and so on) can reduce employment in the formal sector where the legislation is applied, not only at the moment but also, more ominously, in the future as firms are nudged toward capital-intensive technologies. In this view the key is to let markets get on with their allocative function. Enough empirical evidence consistent with these worries has been brought forward to give them considerable weight. And it is easy to imagine a framework of labor legislation that, if fully implemented, would do a great deal of damage to economic performance. But the implementation of such legislation is notoriously incomplete in most countries. Individuals and firms both have a high incentive to reach Pareto-optimal contracts, and often do so, leading to widespread avoidance of labor regulations when these would be mutually damaging. In the light of

such considerations, it is clear that the extent of loss from bad labor market institutions is a matter that must be settled purely on the basis of empirical evidence.

Differences in wages, job security and working conditions by size of firm and by public or private employer: what do they imply?

Suspicion of inefficient allocation of labor has come to a great extent from large observed gaps in wages and other conditions of employment between groups of workers with apparently similar productive potential.

The formal sector–informal sector ‘wage and working conditions’ gap can be seen simply as an aspect of the general tendency for wages and working conditions to improve with firm size, and size–earnings gaps are important within the formal sector as well as between formal and informal sectors. Earnings differentials related to firm size can reach as high as 3:1 or 4:1 between large firms (say those with over 200 workers) and informal sector ones for workers in the same broad low-skilled category; usually they are at least 1.5:1. Public–private differentials when formality or size is held constant are much smaller and can go in either direction depending on the country and time period (Lindauer, 1991). Earnings gaps by size of firm and by type of contract have given rise to three main interpretations, both with respect to what causes them and to the extent of misallocation or other social loss to which they give rise: differential ‘protection’ of workers; the efficiency wage hypothesis, which posits that (at least) some employers will not offer a wage below a certain minimum because productivity would fall faster than the wage; and mechanisms which explain the observed earnings and working conditions differentials as natural reflections of efficient labor allocation processes.

Earnings gaps between people who differ by some personal characteristic not obviously related to productivity suggest discrimination of some sort and an associated efficiency cost. (Monthly) earnings differentials by gender tend to fall in the range 15–50 percent in most developing countries; gaps in hourly earnings are usually less and appear to have been declining in a good number of countries, including those of Latin America, where by the late 1990s the average gap was less than 25 percent in virtually all countries (Psacharopoulos and Tzannatos, 1992, pp. 5, 204). Earnings differentials by ethnic group usually fall into the same range as those by gender (Psacharopoulos and Patrinos, 1994) and are typically driven more by differences in productivity-related characteristics (like education and place of residence) than by differences in the returns to those characteristics (De Ferranti et al., 2004, p. 93). Finally, most studies of earnings gaps due to nepotistic hiring and payment practices have not detected really large effects (Psacharopoulos, 1977; Kugler, 1980). Although this phenomenon

can make quite a difference at the beginning of one's working career, the differential tends to erode fairly quickly over time.

Although not related to gaps between presumably equally productive people, earnings differentials by level of education are typically very large and account for such a high share of overall income inequality that it is natural to speculate that they may exceed productivity differences. A neo-classical 'Mincerian' reading of the evidence suggests that such gaps are likely to be efficient – both in their impact on current labor allocation and in their provision of appropriate signals to people involved in the educational and training process for upcoming generations of workers, as long as they reflect market forces. Credentialist (Dore, 1976) and screening (Arrow, 1973) interpretations, in contrast, suggest that these differentials may not reflect such an efficient process and may lead to the pursuit of more education than is socially productive, or to a situation in which both students and educational institutions focus unduly on credentials rather than the content of the education.⁴ It is clear, and hardly surprising, that these latter factors play some role in hiring and wage decisions (Strobl, 2004). The interesting question is whether their role is large enough to imply that labor market mechanisms and signals are not very effective in allocating and encouraging investment in the various skills.

The contemporary empirical evidence

Against some dramatic expressions of concern about labor market functioning, the record looks relatively good. With well-functioning labor markets, wages, unemployment and other outcomes would be expected broadly to reflect macroeconomic trends, and this has in fact been the case in all regions over the last few decades. The belief that formal wages might be downward rigid is clearly belied by the major declines observed in both Africa and Latin America since the 1970s. In Asia, the growth boom has shown up in large wage increases. And most earnings gaps look smaller on closer inspection than the simple comparisons suggest. What were or were perceived as major mismatches between supply and demand in some segments of the labor market normally recede with time. In short, the historical record suggests a good deal of capacity to adjust.

May we therefore conclude that developing-country labor markets have usually worked very well? Even against the criterion of narrow economic efficiency this conclusion would be premature. The fact that real wages do show substantial downward flexibility is, in fact, guaranteed by the constraints of aggregate productivity. So the important question is not whether such real wage declines and reallocations of labor occur, but rather whether they occur in an optimal fashion from the perspective of overall economic health.

Overblown concerns about labor market malfunction in Africa

In the late 1960s and the 1970s there was great concern that formal sector wages in Africa were out of line in relation to the rural incomes that people gave up to move to the city. But Mazumdar (2002) argues persuasively that the rural–urban earnings gap was both greatly exaggerated and misinterpreted in a number of important studies and official reports. In the 1950s and 1960s African countries had indeed followed a policy of raising wage incomes, with a view to moving from a labor system based on transient migrant labor to one based on a settled urban population stable enough to acquire needed skills. The policy did raise average earnings in the formal sector well above those in agriculture; as of the late 1980s there was a 50–60 percent gap in real median earnings between urban and rural earners, and the gap was probably substantially larger at its peak some years earlier.

Though urban formal wages in most African countries may not have been significantly out of line for long, it is less clear that job security and public sector hiring practices have been so benign in their employment and growth effects. Data from the World Bank's Research Program in Enterprise Development surveys of employer responses find high financial costs of lay-offs widely cited as a serious problem, most notably by medium-sized and larger firms. Of the elements of worker protection, minimum wages appear to be the least burdensome. Summing up this evidence, Mazumdar (2002, p. 324) notes that though employers did give importance to restrictions on hiring and firing and to relying on temporary workers their responses 'tend to confirm labor market regulations rank low in the list of obstacles to enterprise development compared to other factors, e.g. problems of finance and infrastructure'.

Latin America: an overextended informal sector? A faulty industrial relations system?

The degree of labor market intervention⁵ and the fact that the urban informal sector in Latin America is large and has grown rapidly since at least 1980 partly explain the amount of attention given to informality and to labor market issues more generally. In 1990 the average gross ratio of urban informal to formal sector income ranged from 45 percent to close to 80 percent (Berry, 2005, p. 31). Most earnings functions tend to show a moderate differential of 25–40 percent after differences in education, experience and other relevant variables are taken into account. With a modest-sized gap like this, as Maloney (1998, 2004) and others have pointed out, there is a substantial voluntary flow of people from the formal to the informal sector as well as in the opposite direction, part of the considerable labour mobility observed especially for younger workers and those at a stage of their working careers at which they may, for example, wish to go

into business (that is, small business) for themselves.⁶ Maloney (1998) concluded that the average increase in earnings of those who moved to the informal sector and the frequency of the movements were inconsistent with the idea of a marked segmentation between the sectors. There was little evidence of the rigidities that the incentives implicit in the labor code would lead one to expect, and certainly no evidence that people wanted to stay in the formal sector until retirement – at least two-thirds of those moving to the informal sector did so voluntarily, with a desire for greater independence or higher pay cited as the main motivation for doing so. Maloney argued that the very legislation that is thought to induce rigidities into the labor market might in fact stimulate turnover and encourage workers to leave salaried formal sector employment. In a rare attempt to probe different aspects of flexibility related to the labor input, Romaguera et al. (1995, p. 46) distinguish flexibility of wages, of employment, and of the tasks that a given worker can be assigned, reporting that: ‘The biggest source of rigidity seen among firms in the industrial sector is found neither in the legislation nor in the labor market, but rather in backwardness in the areas of administration and human resource planning inside each firm.’ Such a conclusion underlines the fact that too little research has been directed to figuring out exactly how labor legislation impacts upon firms and how they respond to it.

Though the informal–formal earnings gap has tended to widen since 1990 (Berry, 2005, pp. 31–32) and the share of employment in the informal sector has increased (at least until 1995) it seems unlikely that these trends are attributable in any significant degree to increasingly damaging labor market imperfections, given the general (albeit mixed – see Saavedra, 2003, p. 219) trend towards a loosening of the regulatory apparatus. Still, if labor markets need to be more flexible under more liberalized conditions, they may not have moved in this direction fast enough to accommodate the changing needs. Traditionally, shocks have been absorbed through high wage volatility, with inflation and low enforcement of labor legislation pushing the adjustment more to wages than to the level of employment. But recently unemployment has responded more sharply to macroeconomic problems, as inflation rates have been reined in (IDB, 2004, Overview).

Evidence of flexibility notwithstanding, restraints on firing and generally high worker protection may reduce job creation in the protected sector by either discouraging investment or discouraging the use of labor-intensive technologies. In that case one would expect to find larger informal sectors for a given level of development in the affected countries. Pagés’s (2004, p. 73) results suggest this is the case, based on cross-country regressions for Latin American and Organisation for Economic Co-operation and Development (OECD) countries. Maloney (2004, p. 1170) comes up with

a different result. Neither of these conflicting conclusions is definitive, among other reasons because only a few developing countries from other regions are included in the sample (those which are members of the OECD).

Definitely on the dark side of the ledger, IDB (2004, p. 12) notes that labor relations in the region are mired in conflict and distrust, which may hamper the growth of labor productivity and certainly exercises a direct negative effect on worker welfare. And the labor market may saddle workers with seriously debilitating levels of job and income insecurity. According to Latinobarometer's public opinion survey, 85 percent of Latin Americans as of 1996–97 were either unemployed or worried about losing their jobs (IDB, 2004). The comparable number for European countries was 32 percent. It is speculated that the absence of widespread social insurance in Latin America may account for the difference, since the job rotation rates are similar in the two regions. Even in Latin countries with quite low unemployment rates, this ratio is at least 65 percent while the highest figure for Europe (Spain) is under 50 percent and the lowest (Denmark) is under 15 percent.

Given the apparently very large psychological cost of worry about job security, it is arguable that the biggest institutional flaw in Latin America is not the existing regulations but the absence of a system to diminish the cost of job insecurity. Though several countries now have nascent unemployment insurance (UI) systems, their coverage remains very limited.

South and East Asia: the benefits of economic growth

Anecdotal evidence has long abounded in India and Sri Lanka (ILO, 1971) of an excess supply of job aspirants who were either overtrained or prepared for activities for which there was little demand. Such problems, especially in India, were probably a natural result of slow growth. Today India is notorious for the rapid expansion of skill-intensive information industries and education has moved in the direction of those skills. The interesting questions in this, as in many other cases, are whether labor market problems contributed to the slow growth that plagued India until the late 1970s or early 1980s⁷ or have had a role in the recent increases in inequality. More recently at least, there appears to be no strong evidence of labor market malfunction, with the earnings structure remarkably stable and not very different from that of most Latin American countries. Vasudeva-Dutta's (2004, pp. 5, 22) study of adult males in all sectors of the economy reveals an almost constant 3:1 ratio between the median weekly earnings of regular (permanent) workers and casual (temporary) ones for 1983, 1993 and 1999. Probably the human capital-corrected earnings differential between these two groups would be on the order of 2:1; that is, probably

not unlike what appears in other regions. The formal versus informal earnings gap is likely to be smaller than the regular versus casual one, since informal sectors have a good number of relatively high earners.

As one might expect, the labor market outcomes in the fast-growing East Asian countries reflect the overall economic success, with rapid increases in real wages, low levels of open unemployment and underemployment, and a generally impressive performance. Most of these countries have also been noted for strict control or repression of labor unions, and less generous fringe benefits than elsewhere. It is noteworthy that the easing of repression that accompanied the transition to democracy in Korea brought an unprecedented outburst of labor unrest and a spectacular increase in labor disputes, reflecting a significant level of previously bottled-up resentment and unhappiness (Lindauer et al., 1997). In Taiwan the positive quantitative indicators have been matched by a fairly effective procedure for resolving disputes, though there too the lifting of martial law and the establishment of independent unions in 1987 led to a bout of strikes which generated considerable wage gains, though some were resisted and there was harassment and even violence against union leaders (Galenson, 1999, p. 279–81).

With its increasing shift towards market mechanisms, China's experience with the informal sector and rural–urban migration (long dammed up by a tight system of regulation, but one gradually being freed up) takes on special interest. As in other fast-growing Asian countries, wages have been rising at a good clip, and labor shortages have been emerging with increasing frequency. The barriers to migration have created a sort of dichotomy in the urban labor force between urban 'residents' with *hukou* (household registration or permission) and 'migrants', with the former benefiting (albeit decreasingly over time) from special perks (especially the right to housing, but also other subsidies). Many of the migrants have been temporary. As migration has accelerated, it appears that the bulk of those involved have entered the previously understaffed services sector (Meng, 2001); though earnings of migrants have generally remained below those of urban residents, they are considerably above potential rural incomes. Based on a 1995 survey in the coastal area city of Jinin, Meng (2001) reports that both informal sector wage earners and the self-employed do better in terms of income and other benefits than those who work in the formal sector. The loosening of market forces in China's labor market has also been manifested in a large increase in the private returns to education, previously extremely low (Benjamin et al., 2008), one factor in the dramatic increase in income inequality suffered in China during the reform period since the late 1970s. It remains something of a puzzle why inequality is so high (with the Gini coefficient probably around 0.5 or higher) in a country where it was previously very low.

The key questions: does labor market malfunction seriously deter growth, worsen income distribution or decrease worker welfare directly?

Judging by the bulk of the empirical analysis to date, it seems unlikely that in the typical developing country a point-of-time reallocation of labor among the formal, urban informal and rural sectors could raise gross domestic product (GDP) by a significant amount – say 5 percent of GDP. Judged by their effect on wages of their members (typically an increase of 5–10 percent) unions likewise tend to create only a modest wage distortion (IDB, 2004, Overview). The real question is whether the various imperfections significantly affect investment and technology choice, and hence economic growth and income distribution. Much of the country-specific analysis involves Latin American cases. Labor institutions (especially unions) have been blamed for slow long-run growth in Argentina, and for impeding adjustment and labor mobility in Brazil and other countries, thus contributing to the painful nature of the 1980s recession and the resulting high levels of unemployment (Horton et al., 1994, p. 39). It has often been suggested (for example Birdsall and Sabot, 1997) that the lesser degree of protection and intervention in favor of the formal sector workers in the East Asian than in the Latin American countries contributed to their superior growth performance, rapid labor absorption and lower levels of inequality. But many other factors may have been at work to create these differences. Latin America's highly regulated (at least *de jure*) labor regimes have not prevented some countries (including most notably Brazil, but also Mexico, Venezuela and Costa Rica) from achieving high investment rates in the 1960s and/or 1970s, nor the region as a whole from growing fast over 1945–80.

Several factors make it difficult to assess labor market performance and policy. To begin with, the determinants of the key variables, investment and technological change, are ill understood in general, making it hard to sort out those effects which are specific to labor policy and labor market functioning. Secondly, the partial and selective application of labor legislation greatly complicates any statistical analysis since degree of implementation of laws is hard to measure. Finally, the new more open context of today may differ enough from the prior one so that analysis should not mix results from the two.

Broad conclusions

The simultaneous achievement of narrow static and dynamic efficiency in the allocation of labor together with job security and other aspects of work that matter very much to people is clearly a major challenge. Static misallocation of labor due to labor market imperfections (including bad policies) is unlikely by itself to have very severe efficiency costs.

Discouragement of physical and human investment may, but the evidence is unclear as to whether this outcome is frequent or not, since we know too little about how labour market policy affects economic performance within the range of typically observed labor force policies. The single biggest failings of many labor markets may take the form of higher inequality due to the incentive towards capital-intensive technologies, and of job insecurity. Labor market performance appears to have fallen farther short of the ideal in recent decades due to the combination of greater macroeconomic fluctuations and lower levels of inflation, with the latter reducing the capacity of labor markets to adjust.

What it is about labor institutions (including unions) that really matters clearly deserves much more research, with that research involving both greater integration between labor economics per se and industrial relations, and a much more explicit consideration of exactly what determines worker welfare and what it is about labor markets that affects firm behavior. Nearly all aspects of labor legislation and institutions are based on a valid logic; but they risk producing negative effects when those valid objectives are carried too far or somehow distorted. Unions can bring benefits in the form of job stability, a vehicle for communication with employers and a way to reconcile interests for society as a whole, or they can generate serious rigidities. It is important to recognize that there is no such thing as free labor markets in the sense of markets unaffected by institutional constraints. As Pagés (2004, p. 68) notes:

the right approach is not to discuss when or how to deregulate. Instead, the discussion must be based on which set of institutions and regulations will improve the functioning of labor markets and whether the regulations that are already in place achieve their goals or instead need to be amended.

Notes

1. As, for example, in Grant's (1971) call to arms about the looming unemployment crisis.
2. Harris and Todaro (1970) proposed a model of rural–urban migration with that implication. For a broad critique of this model, see Kannapan (1985).
3. The World Bank's 1995 *World Development Report* (World Bank, 1995) exemplified much of this thinking, as does the recent volume edited by Heckman and Pagés (2004).
4. Phelps-Brown (1977) discusses the institutional factors (such as belief about the appropriate relative pay of different types of activities, inertia, and others) that may influence the relative pay of different groups, along with their marginal productivities.
5. Riveros (1989) reports, for example, that non-wage costs have typically added 60–80 percent on top of basic wage costs in Latin American countries while the comparable ratio for Asian countries has been 20–30 percent.
6. Extensive data on labor mobility are presented by IDB (2004, Chapter 2).
7. One of a number of (controversial) hypotheses to explain the country's acceleration to fast growth since then is the idea that Prime Minister Indira Gandhi's confronting the unions and taking a more pro-business stance made the difference in unleashing a burst of investment and growth (De Long, 2003).

References

- Arrow, Kenneth (1973), 'Higher Education as a Filter', *Journal of Public Economics*, 2: 193–216.
- Benjamin, Dwayne, Loren Brandt, John Giles and Sangui Wang (2008), 'Income Inequality in the Transition Period', in Loren Brandt and Thomas Rawski (eds), *China's Economic Transition: Origins, Mechanisms, and Consequences*.
- Berry, Albert (2005), 'A Review of Literature and Evidence on The Economic and Social Effects of Economic Integration in Latin America: Some Policy Implications', paper prepared for the Inter-American Development Bank.
- Birdsall, Nancy and Richard Sabot (1997), *Pathways to Growth: Comparing East Asia and Latin America*, Washington, DC: Inter-American Development Bank.
- De Ferranti David, Guillermo E. Perry, Francisco H.G. Ferreira and Michael Walton (2004), *Inequality in Latin America: Breaking with History?* Washington, DC: World Bank.
- De Long, B. (2003), 'India Since Independence: An Analytic Growth Narrative', in D. Rodrik (ed.), *In Search of Prosperity: Analytic Narratives on Economic Growth*, Princeton, NJ: Princeton University Press.
- Dore, Ronald (1976), *The Diploma Disease: Education, Qualification, and Development*, Berkeley, CA: University of California Press.
- Galenson, Walter (1999), 'The Labor Market in Taiwan: Manpower, Earnings, and Market Institutions', in Erik Thorbecke and Henry Wan (eds), *Taiwan's Development Experience: Lessons on Roles of Government and Market*, Norwell, MA and Dordrecht, the Netherlands: Kluwer Academic Publishers.
- Grant, James P. (1971), 'Marginal Man: The Global Unemployment Crisis', *Foreign Affairs*, 50 (1): 112–24.
- Harris, John and Michael Todaro (1970), 'Migration, Unemployment and Development: A Two-Sector Analysis', *American Economic Review*, 60 (1): 126–42.
- Heckman, James and Carmen Pagés (2004), *Law and Employment: Lessons from Latin America and the Caribbean*. Chicago, IL: University of Chicago Press.
- Horton, Susan, Ravi Kanbur and Dipak Mazumdar (1994), 'Labor Markets in an Era of Adjustment: An Overview', in Susan Horton, Ravi Kanbur and Dipak Mazumdar (eds), *Labor Markets in an Era of Adjustment. Volume 2: Case Studies*, Washington, DC: World Bank.
- Inter-American Development Bank (IDB) (2004), *Good Jobs Wanted: Labor Markets in Latin America*, Washington, DC: Inter-American Development Bank.
- International Labour Office (ILO) (1971), *Matching Employment Opportunities and Expectations: A Programme of Action for Ceylon*, Geneva: ILO.
- International Labour Office (ILO) (1972), *Employment, Incomes and Equality: A Strategy for Increasing Productive Employment in Kenya*, Geneva: ILO.
- Kannapan, S. (1985), 'Employment Problems and the Urban Labor Market in Developing Nations', *Economic Development and Cultural Change*, 33: 699–730.
- Kao, C., K. Anshel and C. Eicher (1964), 'Disguised Unemployment in Agriculture: A Survey', in Carl Eicher and Lawrence Witt (eds), *Agriculture in Economic Development*, New York: McGraw Hill, pp. 129–43.
- Kugler, Bernardo (1980), 'The Influence of Education on Labor Incomes: The Colombian Case', in R. Albert Berry and Ronald Soligo (eds), *Economic Policy and Income Distribution in Colombia*, Boulder, CO: Westview Press, pp. 133–46.
- Lewis, W. Arthur (1954), 'Economic Development with Unlimited Supplies of Labour', *Manchester School of Economic and Social Studies*, 22: 139–91.
- Liedholm, Carl and Donald C. Mead (1987), 'Small Scale Industries in Developing Countries: Empirical Evidence and Policy Implications', MSU International Development Paper No. 9. East Lansing, Michigan: Department of Agricultural Economics, Michigan State University.
- Lindauer, David L. (1991), *Government Pay and Employment Policy: A Parallel Market in Labor*, San Francisco, CA: ICS Press.
- Lindauer, David L., Jong-Gie Kim, Joung-Woo Lee, Hy-Sop Lim, Jae-Young Son and Ezra F. Vogel (1997), *The Strains of Economic Growth: Labor Unrest and Social Dissatisfaction*

- in Korea*, Cambridge, MA: Harvard University Press for Harvard Institute International Development and Korea Development Institute.
- Maloney, William F. (1998), 'Are LDC Labor Markets Dualistic?' mimeo, Washington, DC: World Bank.
- Maloney, William F. (2004), 'Informality Revisited', *World Development*, 32 (7): 1159–78.
- Mazumdar, Dipak, with Ata Mazaheri (2002), *Wages and Employment in Africa*, Aldershot, UK and Burlington, VT: Ashgate Publishing.
- Meng, Xin (2001), 'The Informal Sector and Rural–Urban Migration: A Chinese Case', *Asian Economic Journal*, 15 (1): 71–89.
- Moser, Caroline (1978), 'Informal Sector or Petty Commodity Production: Dualism or Dependence in Urban Development', *World Development*, 6 (9–10): 1041–64.
- Oswald, Andrew (1997), 'Happiness and Economic Performance', *Economic Journal*, 107 (464): 334–49.
- Pagés, Carmen (2004), 'A Cost–Benefit Approach to Labor–Market Reform', *Economic Review–Federal Reserve Bank of Atlanta*, 89 (2): 67–86.
- Perlman, Janice (1976), *The Myth of Marginality: Urban Poverty and Policies in Rio de Janeiro*, Berkeley, CA: University of California Press.
- Phelps-Brown, H. (1977), *The Inequality of Pay*, Berkeley, CA: University of California Press and Oxford: Oxford University Press.
- Psacharopoulos, George (1977), 'Family Background, Education, and Achievement: A Path Model of Earnings Determinants in the UK and some Alternatives', *British Journal of Sociology*, 28: 321–35.
- Psacharopoulos, George and Harry Patrinos (eds) (1994), *Indigenous People and Poverty in Latin America: an Empirical Analysis*, Washington, DC: World Bank.
- Psacharopoulos, George and Zafiris Tzannatos (1992), *Women's Employment and Pay in Latin America: Overview and Methodology*, Washington, DC: World Bank.
- Reynolds, Lloyd G. (1969), 'Economic Development with Surplus Labor: Some Complications', *Oxford Economic Papers*, 21 (1): 89–103.
- Riveros, Luis (1989), 'International Differences in Wage and Nonwage Labor Costs', World Bank Policy, Planning and Research Working Papers (WPS) No. 188, Washington, DC.
- Romaguera, Pilar, Pablo Gonzalez, Alejandra Mizala and Cecilia Montero (1995), 'Labour Flexibility: The Case of Chile', mimeo.
- Saavedra, Jaime (2003), 'Labor Markets During the 1990s', in Pedro-Pablo Kuczynski and John Williamson (eds), *After the Washington Consensus: Restarting Growth and Reform in Latin America*, Washington, DC: Institute for International Economics.
- Strobl, Eric (2004), 'Do Employers Use Education as a Signal for Ability in Developing Countries? Evidence from Ghana', *Applied Economics Letters*, 11 (4): 259–61.
- Udall, Alan and S. Sinclair (1982), 'The "Luxury Unemployment" Hypothesis: A Review of Recent Evidence', *World Development*, 10 (1): 49–62.
- Vasudeva-Dutta, Puja (2004), 'Accounting for Wage Inequality in India, 1983–1999', paper presented at an international Seminar on Wages and Incomes in India, New Delhi, Institute for Human Development and Indira Gandhi Institute of Development Research.
- Vekemans, R. and I. Silva (1969), *Marginalidad en América Latina: Un Ensayo Diagnostico*, Barcelona: Herdes.
- World Bank (1995), *World Development Report, 1995: Workers in an Integrating World*, Washington, DC: World Bank.

24 Education and human capital

George Psacharopoulos and Harry Anthony Patrinos

History

‘A man educated at the expense of much labor and time . . . may be compared to one . . . expensive machine . . . The work which he learns to perform . . . over and above the usual wages of common labor will replace the whole expense of his education’ (Adam Smith, 1904 [1776], p. 101). Thus began the interest in education as an investment in economics. Articles on education as investment appeared sporadically in the first half of this century (for example Strumilin, 1929; Walsh, 1935). In the modern era, education and human capital entered economics in the late 1950s. The focus of the early writings was on the ‘unexplained’ residual in economic growth (Abramovitz, 1962). Schultz (1961) introduced the concept of human capital to explain Solow’s technological change (1956).

From the theoretical literature two waves can be identified. The first, roughly corresponding to the period from 1960 to the 1980s, treated education as an exogenous factor (Becker, 1964). Then, from the late 1980s to the present, education was seen as endogenous, especially in the ‘new growth’ theory literature. From the voluminous empirical literature there are two avenues: micro – largely focused on the microeconomic returns to education; and macro – with early roots in growth accounting.

There are essentially two classes of estimation methods: one that uses the internal-rate-of-return procedure, and another that approximates this procedure by means of fitting an earnings function to individual data sets (Mincer, 1974). Each of these classes is subdivided into the elaborate and short-cut methods, and the basic and extended-earnings function methods (Psacharopoulos and Mattson, 1998). The advantage of the Mincerian estimation is that it can smooth out and handle incomplete cells in an age–earnings profile matrix by level of education.

Micro estimates

The average returns to schooling are presented in Table 24.1. Clearly, the returns are higher in lower-income areas, and the global average is 10 percent. The same diminishing returns apply across countries: the more developed the country, the lower the returns to education at all levels. The high returns to education in low-income countries must be attributed to the

*Table 24.1 Mean returns to investment in education by world region
(Mincerian rate of return)*

Region	Per capita income level	Mean years of schooling	Rate of return (%) ^a
OECD	\$25 000	9.0	7.5
Europe/Middle East/North Africa	\$6 000	8.8	7.1
Asia	\$5 000	8.4	9.9
Latin America/Caribbean	\$3 000	8.2	12.0
Sub-Saharan Africa	\$1 000	7.3	11.7
World average	\$9 000	8.3	9.7

Note: a. Coefficient on years of schooling.

Source: Psacharopoulos and Patrinos (2004).

relative scarcity of human capital. Private returns are higher than social returns at all levels – a result of the public subsidization of education in most countries. The discrepancy between private and social returns is greatest at the university level – which raises issues of equity and finance.

Although the concept of the rate of return to investment in education is unassailable, empirical applications have been attacked on a number of grounds. The most important issue is that of differential ability between those who complete different levels of schooling. To put it in the extreme (Arrow, 1973), a higher education degree might be nothing else than a filter; that is, selecting the more able. There has been a stream of research on this issue. Originally, an arbitrary ‘alpha’ (for ability) coefficient equal to two-thirds was applied, in effect to reduce by one-third the earnings differentials of the more educated for unmeasured differences (Denison, 1967; Blaug, 1970). Later work by Griliches (1970, 1977; Griliches and Mason, 1972) indicated that including an IQ measure in the Mincerian earnings function reduced the rate of return to investment in education by only 10 percent.

Perhaps the ultimate test for accepting that there are returns to education is to observe directly the productivity of workers with different levels of schooling. Beyond econometric shadow pricing, or observation shadow pricing, there is an immense line of work relating education to physical farm productivity. For example, in an early review of the literature, Jamison and Lau (1982) found that, on average, the difference between zero and four years of schooling among farmers results in a 10 percent increment in production. Rosenzweig (1995) and Foster and Rosenzweig (1996) have shown that primary education has an impact on farmers adopting new high-yield varieties. In India, for example, high-yield variety use had an 18 percent

greater effect on the per-area profitability for farmers with primary schooling, compared with farmers with no schooling (Rosenzweig, 1995).

Macro estimates

The importance of education in the growth process re-emerged in the 1980s with the influential writings of Romer (1986, 1990, 1992) and Lucas (1988). Romer and Lucas start with a Solow-type (Solow, 1956, 1957) aggregate production function, augmented in two ways. First, beyond some measure of human capital that is actually used by different firms in the economy, total output also depends upon the average level of human capital. Second, human capital is endogenous, rather than exogenous, in the system; that is, human capital is produced by using resources. The dramatic theoretical implications of this formulation is that output is no longer constrained by the constant-returns-to-scale property of the production function, and that 'knowledge' becomes a kind of public good that spills over the economy as an externality, allowing output to grow beyond the measurable inputs. The empirical implication of this formulation is that different countries need not converge to a common steady-state path, as predicted by neoclassical economics. The level of per capita income between countries can diverge forever, rather than converge. Another, equally important implication of this model is that, by virtue of the average stock of human capital being available to all, there might be social underinvestment in human capital formation.

The returns to education using the macro approach are estimated either by: (1) an aggregate production function explaining GDP; or (2) an aggregate 'macro-Mincerian' earnings function where the units of observation are individual countries (Heckman and Klenow, 1998; Krueger and Lindahl, 2001). The literature on macro-level benefits is vast, complicated and controversial, leading to many different kinds of empirical estimates (Table 24.2).

Equity

Since education has such a strong bearing on individual earnings, it must also affect the distribution of income. The net effect of the expansion of schooling has been a reduction in the dispersion of earnings and hence a more equal distribution. This equitable effect, however, strongly depends on which level of schooling is expanded. The equity impact is highest for basic education, since the low earnings of otherwise illiterate workers are raised nearer to the overall mean. But if university education is expanded (and especially postgraduate education), the equity effect may be negative, in the sense that a group of workers with earnings above the mean are raised even further away from it. Taking Mexico as an example, Marin and

Table 24.2 *Examples of the contribution of education to economic growth*

Database	Main findings	Study source
Cross-section of 32 countries, 1940–85	Early literacy is threshold countries must pass to grow	Azariadis and Drazen (1990)
98 countries, 1960–85, education proxied by primary and secondary enrollments	Increase of 1% point of respective initial 1960 enrollment ratio raises 1960–85 growth rate by 0.025% points for primary and 0.035 for secondary	Barro (1991)
Cross-section/panel of 121 countries, 1960–85, education measured by % of working age population in secondary school	Coefficient of log(education) on log(GDP/worker) is 0.70; coefficient of log(education) on log(difference GDP per worker 1960–85) is 0.23	Mankiw et al. (1992)
Cross-section panel of 111 countries, 1960–90	1 year increase in average years of schooling of labor force raises output per worker by 5–15%	Topel (1999)
Panel cross-section of 110 countries, education variable is average years of schooling, fitted macro-earnings function across countries	Return to schooling equals 18–30%	Krueger and Lindahl (2001)

Psacharopoulos (1976) report that providing primary education to 10 percent of those without it would make income distribution more equal by nearly 5 percent compared with the present level of an inequality index. Giving higher education to 5 percent of those with secondary education, however, would worsen the inequality index by 2 percent. Since most university students come from the higher-income groups in any society, state subsidies for their education will boost their future earnings at the expense of the general taxpayers, who are less likely to enroll their children in higher education.

A large literature examines the benefits of education investments across the income distribution. Overall, public education expenditures are regressive, with a higher share of public spending going to groups from the highest family income categories. However, this has a lot to do with the fact that mostly individuals from high-income families enter university, which

Table 24.3 *Distribution of public education subsidy by expenditure quintiles, selected countries (%)*

Country	Year	Education level	Quintile				
			1	2	3	4	5
Indonesia	1998	Primary	25	24	21	18	13
		Junior secondary	16	20	22	22	21
		Senior secondary	10	14	19	24	34
Malawi	1990/91	Primary	20	23	21	20	16
		Secondary	9	10	16	25	39
		Tertiary	1	7	13	20	58
Ghana	1992	Primary	22	24	22	19	14
		Secondary	15	22	22	26	19
		Tertiary	6	10	19	20	45
South Africa	1993	Primary	27	21	17	16	19
		Secondary	18	18	17	21	25
		Tertiary	11	13	16	28	32

Source: Yang (2004).

is associated with a much higher expenditure per student. When the data is disaggregated by level, by and large, the poor benefit more from expenditure on primary education (see Table 24.3).

Wider human capital

Human capital includes health. Yet evidence on health impacts is not as widely available as is evidence on the effect of education. In general, it is found that the more educated a woman, the lower her fertility, with no evidence of a threshold effect. The mechanism by which this is achieved is that parental education enhances the adoption of contraceptive techniques, and most importantly that female education raises the opportunity cost of children (Becker and Lewis, 1973; Ben-Porath, 1973; Cochrane, 1979; Rosenzweig and Schultz, 1989; Barro, 1991; Appleton, 1996). Age at marriage has been rising steadily in North African countries, due largely to school attendance (Westoff, 1992). In Honduras, Indonesia, Kenya and Mexico, schooled women desire fewer children, and express this through a higher rate of contraceptive use.

Education also reduces infant mortality. For example, a ten percentage point increase in female primary education can be expected to decrease infant mortality by 4.1 deaths per 1000. Thus, in Pakistan, an extra year

of schooling for an additional 1000 girls would prevent 60 infant deaths (UNICEF, 1999, p. 7). The more educated the parents, particularly the mother, the lower is maternal mortality and the healthier is the child. Parental education is significantly associated with the health status of children (defined by a reduction in mortality or an improvement in survival risks), even after controlling for socio-economic status and for access to health services (Cleland and Wilson, 1987; Hobcraft, 1993). Rising levels of maternal education reduce the odds of the child dying before age two. This relationship holds in both urban and rural settings. As with fertility, there are no thresholds in the relationship. Child mortality falls by about 8 percent for each additional year of parental schooling.

The influence of parental schooling operates through the use of medical services (such as prenatal care and clinic visits) and changes in household health behavior (such as washing hands and boiling water). These behavioral changes may result both from perceptual and attitudinal changes and from the ability of the educated (whose incomes are higher than those of the uneducated) to afford better nutrition and better health services for their children (Caldwell, 1979).

Even before taking account of these externalities, the returns to investment in women's education exceed those to men's education for those women who obtain employment. Once the health and fertility externalities are added, the case for educating girls becomes even stronger. The benefit-cost ratio of these health and fertility externalities in Pakistan, for instance, has been estimated at about 3:1.

Non-market benefits and externalities

The benefits of education captured in the rate-of-return estimates reported above are market benefits; that is, they are based on the price more and less educated people command in the labor market. However, there is another set of benefits stemming from a host of beneficial effects of education that are not traded in the market (Duncan, 1976). Such non-market effects are often compounded with public or external effects; that is, they affect not only the recipient of education but others as well.

One of the problems in arriving at estimates of non-market and external effects is that benefits often overlap into more than one category. Table 24.4 provides a catalogue of such effects coming mainly from the United States. However, the fact that parental, especially the mother's, education lowers fertility has been well documented for developing countries (Rosenzweig and Evenson, 1977; Sathar, 1984).

As mentioned above, because of a perhaps unfortunate convention in the early 1960s literature on the economics of education, the adjective 'social'

Table 24.4 *Non-market and external benefits of education*

Benefit type	Findings	Study source
Child education	Parental education affects child's educational level and scholastic achievement.	Murnane (1981), Angrist and Lavy (1996)
Child health	Child's health positively related to parental education	Edwards and Grossman (1979), Grossman and Joyce (1989)
Fertility	Mother's education lowers daughter's births	Sandefur and McLanahan (1990), Rosenzweig and Evenson (1977), Sathar (1984)
Own health	More education increases life expectancy	Feldman et al. (1989), Robins (1984)
Spouse's health	More schooling improves spouse's health and lowers mortality	Auster et al. (1969), Grossman (1975)
Job search efficiency	More schooling reduces cost of search, increases mobility	Greenwood (1975), DaVanzo (1983)
Desired family size	More schooling improves contraceptive efficiency	Michael and Willis (1976), Rosenzweig and Schultz (1989)
Technological change	Schooling helps R&D and diffusion	Nelson (1972), Wozniac (1987)
Social cohesion	Schooling increases voting and reduces alienation	Gintis (1971), Comer (1988)
Crime	Education reduces criminal activity	Yamada et al. (1991), Ehrlich (1975)

Source: Based on and adapted from Wolfe and Zuvekas (1997).

attached to a rate-of-return calculation really meant 'adjusted for the full cost of education, whether paid by the individual or the state'. Because of the universal public subsidization of education, by arithmetic definition the social rate of return is lower than the private rate. However, if one were to include difficult-to-measure spillover effects of education (say, in reducing crime) not realized by the individual, then the externalities-inclusive social rate of return might well be above the private one. The problem is that it is

very hard to measure the spillover effects of education and add them up to the conventional (wage-based) benefits.

The issue has also been raised as to what level of schooling is associated with most externalities relative to the other levels, in order to correct the hierarchy of the returns to education. For example, Birdsall (1996) argued that the returns to higher education are probably underestimated, given the assumed externalities university graduates bring to the economy. On the other hand, Psacharopoulos (1996) counter-argued that if externalities by level of schooling should be considered, then probably primary education has the highest externalities. This (untested) result is achieved by weighing the probability of a university graduate inventing a new vaccine, against the social costs imposed onto the rest of society by the illiterate masses.

In the case of farmers adopting new varieties, better-schooled farmers are the first ones to use them and act as a source of information to others on the benefits of the new seeds. Based on farm surveys in India and the Philippines, Rosenzweig (1996, p. 28) reports that the profits of a farmer were 4 percent higher if his 'representative' neighbor in the village had completed primary schooling compared to his profitability when the neighbor had no schooling. Beyond the effect on neighbors (the classic geographic-proximity example of an externality), Basu (1998) has carried the concept within the family, identifying intra-household externalities arising from the presence of a literate member.

There might be a threshold in terms of human capital accumulation before a country can reap growth benefits. Azariadis and Drazen (1990) were the first to suggest this in the growth literature, while educators (Bowman and Anderson, 1963) and economic historians (Easterlin, 1981) had been suggesting it for a long time. Once the stock of knowledge surpasses certain critical values, aggregate production possibilities may expand especially rapidly (Azariadis and Drazen, 1990). In a back-of-the-envelope empirical testing of this theory, they found that the threshold might be early literacy.

Using data from Brazil, Lau et al. (1996) found a threshold effect of education on output, namely an interval over which the effects are convex, between three and four years of average education. In other words, a country must have a critical mass of basic education before the returns to education manifest themselves. Or, there are increasing returns to the average level of education. This finding is consistent with Romer's (1986) hypothesis that there exist increasing returns to intangible capital.

The Mincerian earnings function was used in a country cross-section to decompose the effect of education on growth into: (1) an effect of the changed returns to education over time; (2) an indirect effect of schooling's positive effect on schooling growth; and (3) a direct effect of education

raising income, holding education growth constant (Glaeser, 1994). The indirect, schooling-to-schooling effect had the greatest impact in the decomposition. This finding is in the spirit of Becker and Murphy (1992) suggesting that earlier human capital creates later human capital, and the new growth literature on increasing returns to scale. Several other studies have found that parental education is a strong determinant of children's school participation and eventual educational attainment (see, for example, Birdsall, 1985). But this is practically all there is in terms of empirical evidence. As noted by Schultz (1994, p. 45), there is little concrete guidance in this literature on where precisely to look for this externality.

Quality versus quantity

A standard criticism of empirical estimates of the returns to education is that such returns refer to the quantity of schooling, saying nothing about quality. Several studies have shown the importance of school quality in determining earnings (Behrman and Birdsall, 1983; Solmon, 1985; Psacharopoulos and Velez, 1993; Card and Krueger, 1996; Bedi, 1997). This is not really a critique of the rate-of-return literature – rather it is pointing to an omission because of the difficulty of obtaining information on learning outcomes.

Yet a counter-argument could be that rates of return to investment in education, as conventionally estimated, by definition refer to the average level of quality across all schools in the sample. So, if school quality is important in determining earnings, improving school quality must yield even higher returns to education.

Most of the evidence on the developmental effects of education refers to the extensive margin; that is, to the number of years of schooling of the labor force. Evidence on the intensive margin – the quality of education provided – is scarce (Behrman and Birdsall, 1983). The reason is that, in developing countries, longitudinal data sets that follow the student from school to adult life and measure economic performance are rare. Furthermore educational quality means different things to different people. First, there is the traditional input definition, by which higher expenditure per pupil or lower repetition rates are indicators of good quality. But throwing money at schools does not necessarily mean that such money will be used efficiently, and automatically promoting everyone in a class does not mean that graduates will (at least) have been made literate. Second, there is the output definition of educational quality, based on the students' learning achievement. But because so many factors other than schooling (for example, prior cognitive knowledge and family background correlate with cognitive achievement in a cross-section), it is difficult to isolate the particular effects of education.

Policy implications

In education, as in any other field, universal policy prescriptions simply do not exist. The strategy and tactics of education depend upon the initial conditions in a particular country, which means that whereas policy A is suitable for country X, policy B may be more suitable for country Y. Given this qualification, the accumulated evidence in the economics of education in the past 30 years permits some broad policy generalizations. The list which follows is conservative, in the sense that, unless the initial conditions in a given country dictate otherwise, the propositions may be applicable to a large number of countries.

Emphasis on primary education in developing countries

Human capital theory holds that investment in human resources results in improved productivity, and that both the costs of the investments and the benefits of improved productivity can be used to calculate an economic rate of return. Human capital investments generally take the form of education or training and may include health care as well. An important distinction is made between private and social rates of return. Private rates of return accrue to families from human capital investments. Social rates of return include private returns, but also consider positive externalities such as improved public health, diffusion of democratic values and practices, and more freedoms for individuals in society. The existence of social returns provides a rationale for public investment in primary education. The World Bank policy paper on *Primary Education* and subsequent education policy papers (World Bank, 1990, 1995, 1999) embraced human capital theory, observing that education, particularly at the primary level, increases the productivity of the workforce through improved literacy, numeracy and health status. Other international public agencies, governments and academics have substantially agreed with the general interpretation of the human capital justification for public investment in primary education.

Emphasis on general over specific skills

Manpower planning models were debunked as a planning tool for a dynamic market economy (Psacharopoulos et al., 1983). Cost-benefit analysis was used to show that, not only were these investments generally unable to match the demand and supply of skills well, but they also suffered low returns because of their small benefits and high cost (see for example, Psacharopoulos and Loxley, 1985).

Emphasis on cost recovery in higher education

At the highest level of education, cost recovery is the most promising policy for both efficiency and equity reasons. Too much of a typical education

budget is devoted to the university level, which typically has the lowest social rate of return, and where a disproportionate number of students come from the more affluent parts of society (World Bank, 1986). Some sons and daughters of poor farmers make it to university, but they are the exceptions that prove the rule. Yet attendance at the university is typically free, and students may even receive a cash allowance. If students pay at least part of the cost of their education, they are more likely to make better choices on whether to enroll and what to study. For the talented poor, selective scholarships or loans can be provided. Along with cost recovery, universities could adopt more traditional efficiency measures, such as the consolidation of dispersed campuses into larger units. Economies of scale apply as much to university campuses as to industrial plants. The average cost per student declines sharply once enrollment exceeds 500 (Psacharopoulos, 1982).

Conclusions

The concept of human capital has a long history in the economics literature. Decades of writings have established firmly that spending on education is an investment with an economic return. Firm conclusions about education's contributions to productivity have been established. The empirical literature counts hundreds of studies that have estimated the economic return to investments in education, as well as other forms of human capital.

Still, research on the subject is ongoing, given that important theoretical and empirical questions remain unanswered. New methodological tools enable researchers to estimate the causal impact of education on earnings and the heterogeneity of returns to schooling across population subgroups. The impact of the quality of human capital – in addition to the quantity – is gaining more attention in the empirical literature.

The link to growth has been especially critical in recent years. Data limitations are partly to blame for a lack of consensus among researchers, but much more needs to be done to reconcile for example the high and robust returns to schoolings at the individual level, with the mixed signals at the macro or cross-country level. Given that the degree of model sophistication is not matched by the data used in empirical applications, then one may want to rely more confidently on the micro evidence.

Other forms of human capital produce returns, and often these other forms interact with education. Among the established relationships, for example, is that the more educated a woman, the lower her fertility. Education also reduces infant mortality. Also, the more educated the parents, particularly the mother, the lower is maternal mortality and the healthier is the child. Parental education is significantly associated with the health status of children, even after controlling for socio-economic status and for access to health services.

References

- Abramovitz, M. (1962), 'Economic Growth in the United States', *American Economic Review*, 52: 762–82.
- Angrist, J. and V. Lavy (1996), 'The Effect of Teen Childbearing and Single Parenthood on Childhood Disabilities and Progress in School', National Bureau of Economic Research Working Paper No. 5807.
- Appleton, S. (1996), 'How Does Female Education Affect Fertility? A Structural Model for the Cote D'Ivoire', *Oxford Bulletin of Economics and Statistics*, 58 (1): 139–66.
- Arrow, K. (1973), 'Higher Education as a Filter', *Journal of Public Economics*, 2: 193–216.
- Auster, R., I. Leveson and D. Sarachek (1969), 'The Production of Health: An Exploratory Study', *Journal of Human Resources*, 4: 411–36.
- Azariadis, C. and A. Drazen (1990), 'Threshold Externalities in Economic Development', *Quarterly Journal of Economics*, 105 (2): 501–26.
- Barro, R.J. (1991), 'Economic Growth in a Cross-Section of Countries', *Quarterly Journal of Economics*, 106 (2): 407–44.
- Basu, K. (1998), 'On Measuring Literacy', *Economic Journal*, 108: 1733–49.
- Becker, G.S. (1964), *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*, 3rd edn, Chicago, IL and London: University of Chicago Press.
- Becker, G.S. and H.G. Lewis (1973), 'On the Interaction between Quantity and Quality of Children', *Journal of Political Economy*, 81: 279–88.
- Becker, G.S. and K.M. Murphy (1992), 'The Division of Labor, Coordination Costs and Knowledge', *Quarterly Journal of Economics*, 107: 1137–60.
- Bedi, A.S. (1997), 'The Importance of School Quality as Determinant of Earnings in a Developing Country: Evidence from Honduras', *International Journal of Educational Development*, 17 (4): 427–37.
- Behrman, J.R. and N. Birdsall (1983), 'The Quality of Schooling: Quantity Alone Is Misleading', *American Economic Review*, 73 (5): 928–46.
- Ben-Porath, Y. (1973), 'Economic Analysis of Fertility in Israel: Point and Counterpoint', *Journal of Political Economy*, 81: 202–33.
- Birdsall, N. (1985), 'Public Inputs and Child Schooling in Brazil', *Journal of Development Economics*, 18 (1): 67–86.
- Birdsall, N. (1996), 'Public Spending on Higher Education in Developing Countries: Too Much or too Little?' *Economics of Education Review*, 15 (4): 407–19.
- Blaug, M. (1970), *Economics of Education*, London: Penguin.
- Bowman, M.J. and C.A. Anderson (1963), 'Concerning the Role of Education in Development', in C. Geertz (ed.), *Old Societies and New States*, Glencoe: Free Press of Glencoe, pp. 245–79.
- Caldwell, J. (1979), 'Education as a Factor in Mortality Decline: An Examination of Nigerian Data', *Population Studies*, 33 (2): 3995–413.
- Card, D. and A. Krueger (1996), 'Labor Market Effects of School Quality: Theory and Evidence', in G. Burtless (ed.), *The Link Between Schools, Student Achievement and Adult Success*, Washington, DC: Brookings Institution, pp. 97–140.
- Cleland, J. and C. Wilson (1987), 'Demand Theories of the Fertility Transition: An Iconoclastic View', *Population Studies*, 41 (1): 5–30.
- Cochrane, S. (1979), 'Fertility and Education: What Do We Really Know?', World Bank Staff Occasional Paper No. 26.
- Comer, J.P. (1988), 'Educating Poor Minority Children', *Scientific American*, 29 (5): 42–48.
- DaVanzo, J. (1983), 'Repeat Migration in the US: Who Moves Back and Who Moves On?' *Review of Economics and Statistics*, 65: 552–9.
- Denison, E.F. (1967), *Why Growth Rates Differ?*, Washington, DC: Brookings Institution.
- Duncan, G.J. (1976), 'Earnings Functions and Nonpecuniary Benefits', *Journal of Human Resources*, 11: 464–83.
- Easterlin, R. (1981), 'Why Isn't the Whole World Developed?', *Journal of Economic History*, 41 (1): 1–19.

- Edwards, L.N and M. Grossman (1979), 'The Relationship between Children's Health and Intellectual Development', in S. Mushkin (ed.), *Health: What is it Worth?* New York: Pergamon Press, pp. 273–314.
- Ehrlich, I. (1975), 'On the Relation between Education and Crime', in F.T. Juster (ed.), *Education, Income and Human Behavior*, New York: McGraw-Hill, pp. 313–38.
- Feldman, J., D. Mkuc, J. Kleinmman and J. Cornoni-Huntly (1989), 'National Trends in Educational Differences in Mortality', *American Journal of Epidemiology*, 129 (5): 919–33.
- Foster, A.D. and M.R. Rosenzweig (1996), 'Technical Change and Human Capital Returns and Investments: Evidence from the Green Revolution', *American Economic Review*, 86: 931–53.
- Gintis, H. (1971), 'Education, Technology and the Characteristics of Worker Productivity', *American Economic Review*, 61: 266–79.
- Glaeser, E.L. (1994), 'Why Does Schooling Generate Economic Growth?' *Economics Letters*, 44 (3): 333–37.
- Greenwood, M.J. (1975), 'Research on Internal Migration in the US: A Survey', *Journal of Economic Literature*, 13: 266–79.
- Griliches, Z. (1970), 'Notes on the Role of Education in Production Functions and Growth Accounting', in W.L. Hansen (ed.), *Education, Income and Human Capital*, New York: Columbia University Press, pp. 71–127.
- Griliches, Z. (1977), 'Estimating the Returns to Schooling: Some Econometric Problems', *Econometrica*, 45 (1): 1–22.
- Griliches, Z. and W.M. Mason (1972), 'Education, Income and Ability', *Journal of Political Economy*, 80 (3), Part II: S74–103.
- Grossman, M. (1975), 'The Correlation between Health and Schooling', in N.E. Terleckyj (ed.), *Household Production and Consumption*, New York: National Bureau of Economic Research, pp. 147–223.
- Grossman, M. and T. Joyce (1989), 'Socio-economic Status and Health: A Personal Research Perspective', in J. Bunker, D. Genby and B. Kehrer (eds), *Pathways to Health: The Role of Social Factors*, Menlo Park: Kaiser Foundation, pp. 139–63.
- Heckman, J. and P. Klenow (1998), 'Human Capital Policy', Working Paper, University of Chicago.
- Hobcraft, J. (1993), 'Women's Education, Child Welfare and Child Survival: A Review of Evidence', *Health Transition Review*, 3: 159–90.
- Jamison, D.T. and L. Lau (1982), *Farmer Education and Farm Efficiency*, Baltimore, MD: Johns Hopkins University Press.
- Krueger, A.B. and M. Lindahl (2001), 'Education for Growth: Why and for Whom?' *Journal of Economic Literature*, 39 (4): 1101–36.
- Lau, L.J., D.T. Jamison, S. Liu and S. Rivkin (1996), 'Education and Economic Growth: Some Cross-sectional Evidence', in N. Birdsall and R.H. Sabot (eds), *Education in Brazil*. Washington, DC: Inter-American Development Bank, distributed by Johns Hopkins University Press, Baltimore, MD, pp. 83–116.
- Lucas, R. (1988), 'On the Mechanics of Economic Development', *Journal of Monetary Economics*, 22 (1): 3–42.
- Mankiw, N.G., D. Romer and D.M. Weil (1992), 'A Contribution to the Empirics of Economic Growth', *Quarterly Journal of Economics*, 107 (2): 407–37.
- Marin, A. and G. Psacharopoulos (1976), 'Schooling and Income Distribution', *Review of Economics and Statistics*, 58 (3): 332–8.
- Michael, R.T. and R.J. Willis (1976), 'Contraception and Fertility: Household Production and Uncertainty', in N.E. Terleckyj (ed.), *Household Production and Consumption*, New York: National Bureau of Economic Research, pp. 99–138.
- Mincer, J. (1974), *Schooling, Experience and Earnings*, New York: Columbia University Press.
- Murnane, R.J. (1981), 'New Evidence on the Relationship between Mother's Education and Children's Cognitive Skills', *Economics of Education Review*, 1: 245–52.
- Nelson, R. (1972), 'Recent Exercises in Growth Accounting: New Understanding or Dead End?' *American Economic Review*, 63: 462–68.
- Psacharopoulos, G. (1982), 'The Economics of Higher Education in Developing Countries', *Comparative Education Review*, 26 (2): 139–59.

- Psacharopoulos, G. (1996), 'Public Spending on Higher Education: Too Much Rather Than Too Little', *Economics of Education Review* 15 (4): 421–2.
- Psacharopoulos, G. and W. Loxley (1985), *Diversified Secondary Education and Development: Evidence from Colombia and Tanzania*, Baltimore, MD: Johns Hopkins University Press for the World Bank.
- Psacharopoulos, G. and R. Mattson (1998), 'Estimating the Returns to Education: A Sensitivity Analysis of Methods and Sample Size', *Journal of Educational Development and Administration*, 12 (3): 271–87.
- Psacharopoulos, G. and H.A. Patrinos (2004), 'Returns to Investment in Education: A Further Update', *Education Economics*, 12 (2): 111–34.
- Psacharopoulos, G. and E. Velez (1993), 'Educational Quality and Labor Market Outcomes: Evidence from Bogota, Colombia', *Sociology of Education*, 66: 130–45.
- Psacharopoulos, G., K. Hinchliffe, C. Dougherty and R. Hollister (1983), 'Manpower Issues in Educational Investment: a Consideration of Planning Process and Techniques', World Bank Staff Working Paper No. 624, Washington, DC.
- Robins, L.N. (1984), 'Lifetime Prevalence of Specific Psychiatric Disorders in Three Sites', *Archives of General Psychiatry*, 51: 949–58.
- Romer, P. (1986), 'Increasing Returns and Long-run Growth', *Journal of Political Economy*, 94: 1002–37.
- Romer, P. (1990), 'Endogenous Technological Change', *Journal of Political Economy*, 89 (5), Part 2: S71–102.
- Romer, P. (1992), 'Two Strategies for Economic Development: Using Ideas and Producing Ideas', *Proceedings of the World Bank Annual Conference on Development Economics*, Washington, DC: World Bank, pp. 63–91.
- Rosenzweig, M.R. (1995), 'Why Are There Returns to Schooling?' *American Economic Review*, 85 (2): 153–8.
- Rosenzweig, M.R. (1996), 'When Investing in Education Matters and When it Does Not', *Challenge* (March–April): 22–9.
- Rosenzweig, M.R. and R. Evenson (1977), 'Fertility, Schooling, and the Economic Contribution of Children in Rural India: An Economic Analysis', *Econometrica*, 45 (5): 1065–79.
- Rosenzweig, M.R. and T.P. Schultz (1989), 'Schooling, Information and Non-market Productivity: Contraceptive Use and its Effectiveness', *International Economic Review*, 30 (2): 457–77.
- Sandefur, G. and S. McLanahan (1990), 'Family Background, Race and Ethnicity, and Early Family Formation', Institute of Research on Poverty Discussion Paper No. 911–90.
- Sathar, Z.A. (1984), 'Does Female Education Affect Fertility Behaviour in Pakistan?', *Pakistan Development Review*, 23 (4): 573–90.
- Schultz, T.P. (1994), 'Human Capital Investment in Women and Men', International Center for Economic Growth Occasional Papers No. 44.
- Schultz, T.W. (1961), 'Investment in Human Capital', *American Economic Review*, 51 (1): 1–17.
- Smith, A. (1904), *The Wealth of Nations*, 5th edn, ed. Edwin Cannan, London: Methuen & Co. first published (1776).
- Solmon, L. (1985), 'Quality of Education and Economic Growth', *Economics of Education Review*, 4: 273–90.
- Solow, R. (1956), 'A Contribution to the Theory of Economic Growth', *Quarterly Journal of Economics*, 70: 65–94.
- Solow, R. (1957), 'Technical Change and the Aggregate Production Function', *Review of Economics and Statistics*, 39: 312–20.
- Strumilin, S.G. (1929), 'The Economic Significance of National Education: The Planned Economy'; reprinted in UNESCO, *Readings in the Economics of Education*, Paris: UNESCO, pp. 413–50.
- Topel, R. (1999), 'Labor Markets and Economic Growth', in O. Ashenfelter and D. Card (eds), *Handbook of Labor Economics*, Amsterdam: North-Holland, pp. 2943–84.
- UNICEF (1999), *The State of the World's Children 1999*, Oxford: Oxford University Press.

- Walsh, J.R. (1935), 'Capital Concept Applied to Man', *Quarterly Journal of Economics*, 49 (1): 255–85.
- Westoff, C.F. (1992), 'Age at Marriage, Age at first Birth, and Fertility in Africa', World Bank Technical Papers Series No. 169.
- Wolfe, B. and S. Zuvekas (1997), 'Non-Market Effects of Education', *International Journal of Education Research*, 27 (6): 491–502.
- World Bank (1986), *The Financing of Education in Developing Countries: An Exploration of Policy Options*, Washington, DC: World Bank.
- World Bank (1990), *Primary Education: A World Bank Policy Paper*, Washington, DC: World Bank.
- World Bank (1995), *Priorities and Strategies for Education: A World Bank Review*, Washington, DC: World Bank.
- World Bank (1999), *Education Sector Strategy*, Washington, DC: World Bank.
- Wozniac, G. (1987), 'Human Capital, Information, and Early Adoption of New Technology', *Journal of Human Resources*, 22: 101–12.
- Yamada, T., T. Yamada and J. Kang (1991), 'Crime Rates versus Labor Market Conditions: Theory and Time-related Evidence', National Bureau of Economic Research Working Paper No. 3810.
- Yang, P. (2004), 'Incidence of Public Spending for Public and Private Education', Human Development Department, World Bank.

25 Health and nutrition and economic development

Harold Alderman, Jere R. Behrman and John Hoddinott

Introduction

Improvements in health and nutrition are themselves central indicators for concepts of development that focus on fulfilling human potentialities and capabilities. Sharp declines in infant and childhood mortality, reductions in proportions of populations that are malnourished, and increased life expectancies are also associated with the process of development as measured by per capita income. Figure 25.1 gives an example for the inverse cross-country associations between malnutrition rates and gross national product (GNP) per capita. Also associated with development is a shift in the disease composition with communicable diseases (for example, diarrhea, respiratory infection, malaria, parasites) becoming relatively less important and non-communicable diseases (for example, cancers, cardiovascular diseases, diabetes) becoming more prominent, despite the rapid expansion of HIV/AIDS in the former group. Further, it is often argued that a healthier, better-nourished population is more productive; indeed Leibenstein's (1957) well-known concept of efficiency wages is predicated on this notion.

However, associations such as these are not very informative about causality. They leave unanswered key questions about to what extent improved health and nutrition cause economic development and to what extent economic development and related policies improve health and nutrition. The primary focus of this chapter, therefore, is on setting out a framework to guide analyses of links among health and nutrition, economic development and policies. We contend that persuasive answers to such questions require careful empirical micro studies grounded in explicit models of behaviors with data that enable control for endogenous behavioral choices that respond to individual, household, community and policy conditions in the presence of unobserved characteristics. Aggregate associations between indicators of health, nutrition and development are not likely to be very informative.

Analytical framework

Good analysis of determinants and impacts of health and nutrition has tripartite foundations: data, modeling and estimation. Policy analysis

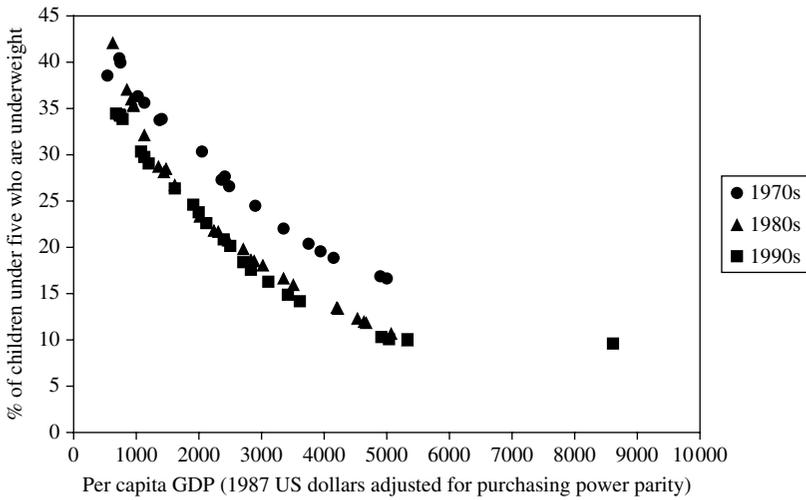


Figure 25.1 *Fitted relationship between child malnutrition rate and per capita GDP in developing countries, 1970s, 1980s and 1990s*

requires sensitivity to basic policy motivations and integration of information about costs and impacts into indicators that permit comparisons among policy options. We consider components of analysis and policy considerations in turn.

Data

Health and nutrition are separate, though related, concepts. While health can be thought of as a state of complete physical, mental and social well-being (World Health Organization, 2005), it is often measured in terms of the presence of a disease or infirmity. Another approach is to consider life expectancy on the presumption that longer lives are likely to be healthier lives. An alternative approach is to track the incidence of diseases. Such data are sometimes aggregated into ‘disability-adjusted life years’ (DALYs) that estimate the disease burden by combining estimates of the impact of an illness on premature mortality and disability (World Bank, 1993). At the individual level, data on health can come from respondent or clinical reports on disease histories, respondent reports on capabilities for undertaking certain activities or tests for doing so (sometimes referred to as ‘activities for daily living’), or respondent self-assessments. While some of these indicators may be good measures of particular disease conditions, they are not necessarily good measures of what people mean by good health or of what aspects of health are most important for productivity. In

other cases, health is proxied by health-related inputs rather than outputs, for example, nutrients, curative health care visits and preventative health measures such as vaccinations.

Some of these indicators may have systematic measurement problems, which can confound attempts to examine relations between economic development and health. If, for example, poorer individuals report less sickness for the same objective health conditions than richer individuals (perhaps poorer individuals perceive a degree of sickness as being 'normal' or because they do not receive sick leave when ill), health appears to improve less than it actually does as income rises.

Nutritional outcomes are the consequence of interactions between food consumption (both quantity and quality), activity levels, illnesses and health behaviors. Nutritional outcomes manifest themselves in terms of body size, body composition or body function reflecting single or multiple nutrient deficiencies. For example, in pre-school and school-age children, nutritional status is often assessed in terms of anthropometry. A particularly useful measure is height-for-age as this reflects the cumulative impact of events affecting nutritional status that result in stunting. Micronutrient status (deficiencies in intake and biological absorption of vitamins and minerals, with Vitamin A, iodine, iron and zinc being of particular concern in developing countries) is expressed in terms of clinical and serological measures.

Basic model

Households and the individuals in them are the proximate sources of demands for health and nutrition, given their preferences, predetermined assets (physical, financial and human, including endowments such as genetically determined innate health) and production technologies. They can be considered to maximize expected utility subject to production functions for health and nutrition as well as expected impacts of health and nutrition on productivity. These decisions take into account current and expected prices for inputs used in the production of health and nutrition and for outcomes of the investments. Policies may enter directly or indirectly through a number of channels ranging from the accessibility and quality of public and private services to the functioning of capital markets for financing health or nutrition investments and the functioning of markets in which these investments are expected to have returns.

Becker (1967 [1975]) provides a well-known simple and useful framework for investments in human resources that captures many critical aspects of investments in health and nutrition. Human resource investments, under risk neutrality, equate expected marginal private benefits and expected marginal private costs (both in present discounted terms) for investments

in a given individual. Marginal private benefits depend importantly, *inter alia*, on expected private gains in productivity arising from improvements in health or nutrition. The marginal private benefit curve is downward-sloping for sufficiently high investment levels, in part because of diminishing returns to health and nutrition given genetic and other endowments. Marginal private costs may increase with health or nutrition because of higher opportunity costs of more time devoted to such investments and because of increasing marginal private costs of borrowing on financial markets. This simple stylized representation can be extended to a dynamic perspective over the life cycle. Marginal private benefit and costs curves vary across individuals because of differences in observed and unobserved individual, family, household and community characteristics, the latter in part related to policies and to markets. Changes in any of these factors can shift these curves and thus equilibrium investment levels. This simple framework systematizes four critical points for investigating dimensions of the determinants and the effects of health and nutrition – and how these relate to policy choices.

First, the impacts of changes in policies may be hard to predict by policy-makers and analysts. If households or other entities face policy or market changes, they can adjust all of their behaviors in response, with cross-effects on other outcomes, not only on the outcome to which the policy is directed.

Second, marginal benefits and costs of investments in an individual differ depending upon the point of view from which they are evaluated: (1) there may be externalities or capital or insurance market imperfections so that social returns differ from private returns (see below); and (2) there may be a difference between who makes the investment decision (for example, parents) and in whom the investment is made (for example, infants and small children). The effectiveness of policies is likely to depend crucially on perceived private effects by private decision-makers, and these may differ from social effects of interest to policy-makers and the broader society.

Third, health and nutrition are determined by a number of individual, family, community, (actual or potential) employer, market and policy characteristics, only a subset of which are observed in available data sets. To identify the impact of the observed characteristics on health and nutrition or the impact of health – nutrition, it is important to control for correlated unobserved characteristics.

Fourth, empirically estimated determinants of, and effects of, health and nutrition are relevant only for a given macroeconomic, market, policy, schooling and regulatory environment in which there may be feedback both at the local and at a broader level.

Estimation

To assess the impact of a particular intervention designed to improve health or nutrition, such as increasing use of iron supplements or of health fees, the ideal is a double-blind experiment with random assignment to treatment and control groups over a long enough time to assess the effects of interest. If there were available data from well-designed and well-implemented experiments (see below), associations between observed health and nutrition and observed outcomes would reveal the underlying causality directly. Experiments have been conducted to evaluate a relatively small number of policies related to health and nutrition in developing countries (see the next section for some examples).

But possibilities for using experiments for policy evaluation are limited. First, most such experiments cannot be double-blind. That the treated know that they are treated may create better performance. That those not treated know that they are not treated may create incentives to obtain treatment through migration, political pressure, market purchases or other means, or to drop out of evaluation samples. Second, if control group members expect that they will eventually be affected by the program and if they can transfer resources over time, they should immediately adjust their behavior to reflect their changed command over resources due to expected eventual future direct program benefits. If so, comparisons between treatment and control groups probably underestimate program impacts. Third, many experiments cannot be conducted because randomized design of particular treatments is considered unethical. Fourth, even for the policies for which good experiments can be conducted at a reasonable (resource, ethical and political) cost, they reveal only the gross changes induced by the experimental treatment conditional on a particular situation, not what would happen in somewhat different circumstances.¹ Fifth, there may be insufficient time to observe effects of interest to policy-makers. This is a particular problem for investments in nutritional or health status of infants or children, since many of the desired outcomes are expected to occur much later in life. For such reasons, though it is desirable to increase experimental evaluation of policies and to assure that experiments that are undertaken are of high quality, there are limits on what policies can be evaluated by experimental means. Nevertheless, the experimental design is an important benchmark against which other means of evaluation should be compared and judged.

In the absence of experimental data, econometric methods can be used. These estimation procedures should be grounded firmly in the analytical framework used for the analysis, control for data problems such as measurement errors in variables and right-side variables that reflect current or past choices that are correlated with unobserved preferences and individual

or household factors such as innate ability, address sample selection (for example, clinic health test results only from those who go to health clinics and have the tests) and control for program selection, when, for example, programs are purposively placed on the basis of unobserved community factors that may correlate with outcomes (Rosenzweig and Wolpin, 1986).

Policy motives

Policies designed to improve health or nutrition are often rationalized for their intrinsic value – better health or nutrition is considered ‘a good thing’ – as well as their instrumental value; it is presumed that better health will contribute to other outcomes such as productivity, incomes and thus economic development. By themselves, however, such motivations provide little in the way of guidelines for choosing among policies.

Policies should be chosen to maximize social welfare. However, given difficulties (both theoretically and in political economy terms) of determining social welfare functions, a more practical approach is to consider the two standard economic justifications for policies: (1) to increase efficiency or productivity; and (2) to redistribute resources. The distributional justification includes as a special case poverty reduction as well as intra-household and intergenerational considerations.

An investment is efficient if marginal social benefits of the last unit of that investment just equal marginal social costs. Private maximizing behavior leads to investments at the level at which marginal private benefits equal marginal private costs under the assumption that, given information available to them and constraints that they face, individuals act in what they perceive to be their best interests.

Private incentives for investments may differ from social incentives for such investments. For example, marginal social benefits may exceed marginal private benefits for health and nutrition for the following reasons. Firstly, there are negative externalities such as second-hand smoking and contagious diseases such as HIV/AIDS. Secondly, similarly, there are positive externalities such as private investments achieving some reduction of infectious disease transmission. Thirdly, privately held information may misrepresent private rates of return to these investments because it is incomplete or incorrect (for example, youth may not know about contraceptives or about risks associated with sexually transmitted diseases). The ‘public good’ nature of information leads to underproduction of information from a social point of view by private markets. Fourthly, the combination of uncertainty, risk aversion and imperfect insurance markets may result in private incentives to underinvest in health and nutrition (and other) assets because from a social point of view the risks are pooled. Fifthly, the social discount rate may be lower than private discount rates

because individuals value future outcomes more collectively than they do individually.

Marginal social costs may be less than marginal private costs because: (1) there may be capital market imperfections for these investments; (2) the sectors that provide some related services may produce inefficiently because institutional arrangements do not induce efficient production of an efficient basket of commodities, or regulations preclude efficient production of an efficient basket of commodities (for example, regulations that limit hours during which clinics are open, or that limit provision of services to public providers).

Policy choices to increase efficiency and to improve distribution

If all other markets are operating efficiently and there are differences between marginal private and social incentives in a given market related to health and nutrition, policies that induce investments at the socially efficient levels increase efficiency. That still does not indicate what policies would be best to induce health and nutrition at the socially desirable levels. There is a large set of possibilities, including governmental fiat, governmental provision of services at subsidized prices, price incentives in markets related to health and nutrition, price incentives in other markets, and changing institutional arrangements. To choose among alternatives based on efficiency alone, there are two important considerations.

First, policies have costs: direct costs of implementing and monitoring policies and distortionary costs due to encouragement of socially inefficient behavior such as rent-seeking. An efficiency policy hierarchy that gives the preferential ordering of policies in terms of efficiency can be defined, in which alternative policies to attain the same improvement in efficiency are ranked according to their social marginal costs.

Second, there are substantial information problems regarding exactly what effects policies have. This is an argument in favor of policies that are as transparent as possible, which generally means higher in the efficiency policy hierarchy with regard at least to distortionary costs because more direct policies are likely to be more transparent. Information problems also provide an argument for price policies (taxes or subsidies) because if there are shifts in the underlying demand and supply relations they are likely to be more visible in a more timely fashion to policy-makers if they have impact on governmental budgets than if they only change the distortions faced by private entities as tends to happen with quantitative policies.² Finally, information problems in the presence of heterogeneities across communities point to the possible desirability of decentralization and empowerment of users of health and nutrition-related services to increase the efficacy of the provision of those services, though such considerations

must be balanced against possible economies of scale, higher staff quality and possibly lower levels of corruption at more centralized levels, as well as intercommunity distributional concerns.

Thus, for efficiency or productivity reasons, there is an argument for choosing policies as high as possible in the efficiency policy hierarchy – and thereby using interventions that are focused as directly on the problem as possible. But good efficiency reasons for public support for health and nutrition does not imply that the best way to provide that support is through governmental provision of the relevant services. In some circumstances, subsidies or taxes that create incentives for the efficient provision of these services, whether the actual providers are public, private or some mixture of the two, may be higher in the efficiency policy hierarchy.

Now consider distribution. Generally subsidization of specific goods and services is not a very efficient way of lessening distributional problems (for example, Coady et al., 2004). Instead, it is generally more efficient to redistribute income to consumers, allowing them to allocate the income in ways that lead to efficient patterns of consumption. Nevertheless, there are cases in which subsidization of selected goods and services may be defensible to attain distributional objectives. For example, in cases where it is difficult to target poor households or poor types of individuals within households, subsidizing certain goods and services that are mainly consumed by the poor may be relatively high in the efficiency policy hierarchy.

Ranking policy alternatives

A challenge for policy-related analyses is how to measure and rank possible investments designed to improve health and nutrition, or to use health and nutrition-related investments to increase productivity or to improve distribution. One common approach, ‘cost-effectiveness analysis’ (CEA) consists of ranking related investments according to their costs per unit of effectiveness.³ The measure of ‘effectiveness’ is clearly defined and as narrow as practical, for example, cost per life saved or cost per DALY. However, CEA has several shortcomings: it requires a single effectiveness measure, which is impractical for many health and nutrition investments because they involve a wide range of possible outcomes; it does not provide any basis for comparing health-related interventions to other investments such as economic infrastructure that may also yield health benefits; and CEA does not address the efficiency motive for policies comprehensively.

Cost–benefit analysis (CBA) is an alternative approach. If the benefits to improved health or nutrition outcomes can be expressed in monetary terms, CBA generates results (benefit–cost ratios or internal rates of return) that readily permit comparisons with alternative investments. However,

expressing benefits in monetary terms is not straightforward: for example, how should one value a life saved? In addition, because this is a partial equilibrium approach, important market feedbacks, particularly within relatively closed economies, may be missed. Also, information with which to assess private versus social rates of return, and therefore, the efficiency motive, is relatively rare.

Empirical evidence on the links between health and nutrition, economic development and policies

Better health and nutrition are associated with economic development. Figure 25.1 gives an illustration of the inverse cross-country association between one indicator of health and nutrition, the child malnutrition rate, and GNP per capita (Haddad et al., 2003). This inverse association is consistent with a cross-country elasticity of pre-school underweight rates with respect to per capita income for 1980–96 of -0.5 . While income growth is apparently associated with improved nutrition, the association is modest. Figure 25.1 also illustrates that this aggregate cross-country curve shifted over time, perhaps due to declining food prices and dissemination of simple health care technology (for example, immunization and oral rehydration treatment for diarrhea). Thus increased per capita income may be an important, if modest, factor in improving health and nutrition but other factors may be important as well. There are similar patterns of inverse associations between other indicators of health and nutrition and per capita income, also with substantial variance after controlling for income, both on an aggregate and a micro level. However, identifying causality from such associations, as noted in the previous section, is very difficult. To make confident assessments of what determines health and nutrition in the developing countries and to what extent better health and nutrition cause development, careful micro studies are needed that have good foundations in data, modeling, estimation and policy concerns. This section summarizes selected recent careful studies and their strengths and weaknesses in the light of the discussion of the previous section.⁴

What might improve health and nutrition in the developing countries?

Income growth is one possibility that is suggested by the association between per capita income and health and nutrition. Haddad et al. (2003), in addition to the aggregate associations summarized in Figure 25.1, estimate reduced-form demand relations for child weight-for-age Z scores (that is, in terms of standard deviations of distributions for an international reference population) as dependent on per capita income, household access to piped water and sewage, parental schooling attainment, mother's height, ethnicity, household size and demographic composition, and community

fixed effects using micro household data sets from 12 developing countries for which nationally representative samples with the necessary information from the 1990s are available and that represent a variety of locations and conditions. Their estimation strategy is sensitive to measurement and endogeneity problems for income with both Ordinary Least Squares (OLS) and Instrumental Variables (IV) estimates (using assets as instruments) and a number of specification tests. The median of their preferred estimates indicates that doubling income would increase weight-for-height by 0.47 standard deviations. But the range is from 0.14 to 1.20, indicating that average cross-country estimates might be very misleading for the average impact of income in a specific country. Questions can be raised about some aspects of this investigation: for example, cross-sectional data precludes controlling for individual and household endowments that some other studies have found to be important, and household demographic characteristics including fertility are assumed to be exogenous rather than behavioral choices. But all in all this is a systematic investigation that is sensitive to most of the concerns discussed in the previous section and that probably provides the best available evidence on the impact of income on this important nutritional indicator. The substantive results indicate that income importantly affects child nutrition, but that such effects generally are modest and vary across economies due to differential effects of other factors. Therefore per capita income growth is likely to reduce child malnutrition significantly, but not fast enough to attain targets such as the Millennium Development Goal of halving child malnutrition by 2015.

Policies that are directed at health and nutrition may also be important, beyond general development policies, in improving health and nutrition in developing countries. Improved access to safe water is thought to be one possibly important such policy, and there is an ongoing debate as to whether this is best accomplished via public or private service delivery. It is within this context that Galiani et al. (2005) evaluate the impact of privatization of water supplies in the 1990s in about 30 percent of Argentinian municipalities with about 60 percent of the population. Their analysis uses the fact that local governments were responsible for delivering water services and made different decisions regarding whether and when they privatized. They use a generalized difference-in-difference matching estimator that conditions on fixed effects that may have led to selective privatization. They find that child mortality fell 8 percent in areas that privatized and that the effect was largest (26 percent) in the poorest areas. As a further check on their results, they look at cause of death and find that privatization affected child mortality from water-borne diseases but not other diseases. Thus this study advances knowledge on an important policy issue critically related to health and nutrition by using data, models and estimation

methods to control for what otherwise would have been confounding problems, particularly selection into privatization.

There are also some studies that investigate the impact of policies directed towards improving health and nutrition by using experimental data, as suggested in the previous section, the ‘gold standard’ or benchmark approach for evaluating a given program impact. Behrman and Hoddinott (2005) is an example that also illustrates some possible limitations of experimental data. They investigate the impact on child nutrition of PROGRESA, a large Mexican rural anti-poverty program that had an evaluation sample in which overall treatment was randomly assigned to 326 communities but not to 180 others. While this randomization should have been sufficient to address potentially confounding effects of unobservable heterogeneity on assessment of program effectiveness, a shortage in the availability of one component of this intervention – a nutritional supplement provided to pre-school children – appears to have led local administrators to exercise discretion in deliveries of this intervention, systematically favoring children with poorer nutritional status. When Behrman and Hoddinott examine the impact of PROGRESA based on the presumption of randomized allocations, they find that PROGRESA had no, or even a negative, impact on child nutrition. However, not all children designated to receive nutritional supplements actually did. Their preferred estimates – child fixed-effects estimates that control for unobserved heterogeneity correlated with access to the supplement – indicate a significantly positive and fairly substantial program effect of nutritional supplements on children aged 12–36 months, an increase of about a sixth in mean growth per year. Thus even with experimental data, to understand program impact care must be taken to incorporate the details of how programs actually function. This study also illustrates the use of a distributional program (an income transfer) to achieve improvements in efficiency, in this case child nutrition.

Does better health and nutrition contribute to economic development?

Better health and nutrition can contribute to economic development by:

- reducing deaths, thereby saving costs associated with premature mortality;
- saving resources that would be otherwise used to treat illnesses;
- allowing individuals to be more productive; and
- transmitting these benefits to future generations.

Some recent studies use special data to investigate aspects of these pathways within the considerations outlined in the previous section. Two

interesting recent examples focus on impacts on productivity through increasing education using experimental data.

Miguel and Kremer (2004) evaluate a Kenyan program of school-based mass treatment with inexpensive deworming drugs. Intestinal helminths (for example, hookworm) affect a quarter of the world's population, particularly school-age children. Seventy-five schools were phased into the program in random order. Health and school participation improved not only at program schools, but also at nearby schools, due to reduced disease transmission. Absenteeism in treatment schools was 25 percent (or seven percentage points) lower than in comparison schools. Including spillover effects, schooling increased by 0.15 years per person treated. The collection of data on all children in these schools allowed the assessment of externalities that relate to the efficiency motive for policy discussed in the previous section, but for which empirical estimates are rare. Despite the qualifications due to the selected school-based sample, these estimates are suggestive of an important way in which improved health increases development prospects through education with both efficiency and distributional gains.

Maluccio et al. (2005) evaluate impacts on education of an early-childhood nutritional intervention in rural Guatemala. They advance beyond the previous literature by using longitudinal data from an early-childhood nutritional experiment with educational measures for school-age through prime adult years, avoiding confounding estimates by excluding potentially endogenous right-side variables, using estimators that allow for non-normal distributions, and testing for robustness of results to calculation of standard errors due to clustering and control for sample attrition – all within the general framework of the previous section. Their substantive results indicate significantly positive, and fairly substantial, effects of the nutritional intervention: increased grade attainment by women – via increased likelihood of entry and of completing at least some secondary school; speedier grade progression by women; higher scores on cognitive tests; and higher scores on educational achievement tests by men. Studies such as are summarized in this section provide important inputs into evaluating policy alternatives. But they need to be combined with information on costs, on how society values all of the relevant pathways noted above, and on time paths and discount rates in order to permit comparisons among alternative possible interventions through CBA as discussed in the previous section. Some examples for health and nutrition are given in Behrman et al. (2004) and Mills and Shillcut (2004). These studies suggest that benefit–cost ratios for a number of interventions in health and nutrition are considerable – suggesting substantial possible contributions to economic development, particularly through increasing productivity but also

through lessening resources devoted to illnesses. But these studies also suggest such estimates can be very sensitive to assumptions regarding factors such as how to value averted mortality and the discount rate.

There are a number of other interventions that appear to have potentially high benefit–cost ratios in low-income settings when these services are regularly provided; the challenge is to find delivery mechanisms that are sustainable in such locales at reasonable costs. For example, a range of randomized trials have shown that when children one to five years of age take vitamin A prophylaxis every six months, child mortality is reduced by 25–35 percent (Beaton et al., 1993). However, many countries have failed to find a means to deliver this inexpensive nutrient on a regular basis to the populations most in need of such supplementation without undermining the health care infrastructure that is necessary for other routine services.

One issue in financial sustainability is the willingness to pay for services. This is illustrated in Miguel and Kremer (2004). Despite the benefits they documented, few parents invested in this medicine in the absence of subsidies. Similarly, Alaii et al. (2003) reported that private returns in the form of reduced medical expenses and increased labor participation from the use of insecticide-treated bednets exceeded costs in Kenya, yet few beneficiaries expressed a willingness to purchase the nets or to have them re-treated. In both these studies there was a documented externality; treating one individual reduced the probability that another would be infected. This justifies a level of subsidy on efficiency grounds. Still, it remains a puzzle why private investment in these preventative services was less than expected.

Conclusions

A number of themes run through this chapter. First, there are links both from health and nutrition to economic development, and from economic development to health and nutrition. Second, interpreting these links must be done cautiously. Good analysis of what causes household and individual health and nutrition and what effects such investments have is difficult, and requires much more systematic approaches than simply looking at associations among observed variables. Explicit economic models permit exploring systematically the determinants and the impacts of health and nutrition, point to what data are needed for such explorations, facilitate interpretations of empirical findings, and help to identify probable estimation issues that should be addressed given the data used. In the absence of experimental data, estimates based on behavioral data should use an explicit model of the underlying behaviors, though far too often in the literature the models used are not explicit. Those who are not clear about their framework of analysis may think they are revealing underlying truths unconstrained by such frameworks, but they are instead usually making

implicit assumptions that may upon examination not be plausible. Even if experimental data are available, care must be paid to the details and context of the experiments to interpret the results. And while experiments are to be encouraged, they have limitations, including not providing information about counterfactual policies, so they do not eliminate the need for careful economic modeling. Third, while there is persuasive evidence on some aspects of relations between health and nutrition and development that suggest, for example, high returns to some interventions, there are many important issues remaining that warrant better data and analysis in order to understand better links of health and nutrition with development and the policy implications.

Notes

1. Structural models, in contrast, can be used to evaluate counterfactual situations and policies, including with respect to their duration (for example, Todd and Wolpin, 2004).
2. Nevertheless there are likely to be some cases, such as providing information regarding qualities of goods and services related to health and nutrition, for which quantitative regulations may be higher in the efficiency policy hierarchy than price policies because of the nature of information requirements.
3. Identifying costs, however, is tricky. Direct governmental budgetary expenditures for particular investments may be intermingled with other expenditures in budgets, and therefore be difficult to identify or they may be spread among various budgets at various administrative levels. A further problem is that budgetary expenditures do not reflect the true costs because of distortions in market prices, perhaps created by policies (for example, salary scale limitations for public sector health personnel). Finally, the costs of a program to the private sector may be considerable. For example, many programs require considerable time of private individuals that has opportunity costs in the form of other uses, but such costs are often ignored in cost–benefit analysis.
4. Given space constraints, we review only a few studies. For surveys of health and nutrition and development, see Strauss and Thomas (1998) and the more recent though somewhat more specialized updates in Alderman et al. (2005), Behrman et al. (2004) and Mills and Shillcutt (2004).

References

- Alaii, Jane A., H.W. van den Borne, S. Patrick Kachur, Halima Mwenesi, John M. Vulule, William A. Hawley, Martin I. Meltzer, Bernard L. Nahlen and Penelope A. Phillips-Howard (2003), 'Perceptions of Bed Nets and Malaria Prevention Before and After Randomized Controlled Trial of Permethrin-Treated Bed Nets in Western Kenya', *American Journal of Tropical Medicine and Hygiene*, 68 (Supplement): 142–48.
- Alderman, Harold, Jere R. Behrman and John Hoddinott (2005), 'Nutrition, Malnutrition and Economic Growth', in Guillem López-Casasnovas, Berta Rivera and Luis Currais (eds), *Health and Economic Growth: Findings and Policy Implications*, Cambridge, MA: MIT Press, pp. 169–94.
- Beaton, G., R. Martorell and K. Aronson (1993), *Effectiveness of Vitamin A Supplementation in the Control of Young Child Morbidity and Mortality in Developing Countries*, New York: United Nations.
- Becker, Gary S. (1967), 'Human Capital and the Personal Distribution of Income: An Analytical Approach' Ann Arbor, MI: University of Michigan, Woytinsky Lecture; republished in Gary S. Becker (1975), *Human Capital*, 2nd edn, New York: NBER, pp. 94–117.

- Behrman, Jere R. and John Hoddinott (2005), 'Program Evaluation with Unobserved Heterogeneity and Selective Implementation: The Mexican PROGRESA Impact on Child Nutrition', *Oxford Bulletin of Economics and Statistics*, 67 (4): 547–69.
- Behrman, Jere R., Harold Alderman and John Hoddinott (2004), 'Hunger and Malnutrition', in Bjorn Lomborg (ed.), *Global Crises, Global Solutions*, Cambridge: Cambridge University Press, pp. 363–420.
- Coady, David, Margaret Grosh and John Hoddinott (2004), 'Targeting Outcomes *Redux*', *World Bank Research Observer*, 19 (1): 61–85.
- Galiani, Sebastian, Paul Gertler and Ernesto Schargrodsky (2005), 'Water for Life: The Impact of the Privatization of Water Services on Child Mortality', *Journal of Political Economy*, 113 (1): 83–120.
- Haddad, Lawrence, Harold Alderman, Simon Appleton, Lina Song and Yisehac Yohannes (2003), 'Reducing Child Malnutrition: How Far Does Income Growth Take Us?' *World Bank Economic Review*, 17 (1): 107–31
- Leibenstein, Harvey (1957), *Economic Backwardness and Economic Growth*, New York: John Wiley.
- Maluccio, John A., John Hoddinott, Jere R. Behrman, Agnes Quisumbing, Reynaldo Martorell and Aryeh D. Stein (2005), 'The Impact of an Experimental Nutritional Intervention on Education into Adulthood in Rural Guatemala', Middlebury College Economics Discussion Paper no. 06–14, Middlebury College.
- Miguel, Edward and Michael Kremer (2004), 'Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities', *Econometrica*, 72 (1): 159–217.
- Mills, Anne and Sam Shillcutt (2004), 'Communicable Diseases', in Bjorn Lomborg (ed.), *Global Crises, Global Solutions*, Cambridge: Cambridge University Press, pp. 62–128.
- Rosenzweig, Mark R. and Kenneth I. Wolpin (1986), 'Evaluating the Effects of Optimally Distributed Public Programs', *American Economic Review*, 76 (3): 470–87.
- Strauss, John and Duncan Thomas (1998), 'Health, Nutrition and Economic Development', *Journal of Economic Literature*, 36 (2): 766–817.
- Todd, Petra and Kenneth I. Wolpin (2004), 'Using a Social Experiment to Validate a Dynamic Behavioral Model of Child Schooling and Fertility: Assessing the Impact of a School Subsidy Program in Mexico', Philadelphia, PA: University of Pennsylvania, mimeo.
- World Bank (1993), *World Development Report 1993: Investing in Health*, Washington, DC: World Bank and Oxford University Press.
- World Health Organization (2005), 'WHO definition of health', <http://www.who.int/about/definition/en/>, accessed 1 September 2005.

26 Entrepreneurship and development

E. Wayne Nafziger

Introduction

Economic theory and development economics neglect the entrepreneur. He or she is not needed in the neoclassical model of the firm which analyzes the optimum in well-defined problems with variables clearly specified (Baumol 1968). Leibenstein (1968, p. 72) contends:

If all inputs are marketed and their prices are known, and if all outputs are marketed and their prices are known and if there is a definite production function that relates inputs to outputs in a determinate way, then we can always predict the profits for any activity that transforms inputs into outputs. If net profits are positive, then this should serve as a signal for entry into this market. The problems of marshaling resources and turning them into outputs appear to be a trivial activity. From this point of view it is hard to see why there should ever be a deficiency of entrepreneurship. The answer is that the standard competitive model hides the vital function of the entrepreneur.

This chapter examines concepts of the entrepreneurs, their location, determinants of entrepreneurship and technological transfer, whether there is a shortage in less-developed countries (LDCs), the supply of and demand for entrepreneurs, the role of the state and policy implications.

Theory and concept of the entrepreneur

Decision-maker

The entrepreneur can be viewed as: (1) decision-maker under uncertainty; (2) gap-filler and input completer; or (3) innovator (Schumpeter, 1961).

To Knight, the firm's ultimate control lies with the risk bearer and not the hired manager. The crucial decision is selecting people to make decisions; any other decision is routine (Knight, 1921, pp. 97–111, 297–8). Knight makes profit a return to the entrepreneur, who is the capitalist.

Gap-filler

Entrepreneurship involves activities where markets are poorly established or production functions are not completely known. Coase (1937) identifies two major coordinators: the entrepreneur, who organizes the firm, and the price mechanism, which coordinates decisions between firms. The choice between firm or market ('make or buy') is determined by transactions costs.

An entrepreneur (individual or group) makes up for market deficiencies. Many firms operate with slack (Leibenstein, 1966). The LDC entrepreneur may need to engineer a hard-to-get machine, supervise production workers, or manage, since he or she may not be able to hire someone.

The entrepreneur must be an 'input completer', making up inputs less than that required. Growth depends on increased labor and capital and entrepreneurial gap-filling. No fixed input-output relationship exists, because entrepreneurship cannot be quantified, predicted or controlled.

Innovator

The rapid growth of the West and Japan since the nineteenth century is largely a story of how novel and improved ways of satisfying wants were discovered and adopted. This story is not just of inventions, as many inventions were unneeded, unsponsored or lacking markets. The early twentieth-century Stanley Steamer probably failed not because of inferiority to the internal combustion engine but because the Stanley brothers did not mass produce. To explain growth, we must emphasize innovation, the embodiment in commercial practice of some new idea or invention.

Schumpeter (1961, 1939) links innovation, the source of private profits and growth, to the entrepreneur, who carries out new economic combinations, which include introducing new products or new production functions reducing inputs per output, opening new markets, exploiting new sources of materials and reorganizing an industry.

The model begins with a stationary state, an unchanging economic process reproducing constantly. This model assumes perfect competition, full employment, and no savings or technical change. No entrepreneurship is needed, since workers can routinely repeat orders and operations. However, into this stationary process, a profit-motivated entrepreneur introduces an innovation raising resource productivity. Eventually such innovation means new plants, firms and leadership.

The stationary economy may have high earnings for management, monopoly gains, windfalls or speculative gains, but no profits, the premium for innovation. Innovation sets up a temporary monopoly gain that is wiped out by imitation. Innovators must keep a step ahead of rivals for profits to continue. Profits result from the entrepreneur, who may not always receive them.

New bank credit finances innovation, which is imitated by competitors. Innovations are not isolated but arise in clusters from reduced risk. Eventually entrepreneurial waves force out old firms while exhausting gains from innovation. As borrowing diminishes and loans are repaid, entrepreneurship slackens and finally ceases. Innovation, saving, credit creation and

imitation explain growth, while their ebb and flow determine the business cycle.

Where are Schumpeterian entrepreneurs?

The entrepreneur's contribution, although important, cannot be precisely measured. Can we find a proxy for entrepreneurship? Schumpeter (1949) attributes novel ways of doing things (innovations responsible for technical change) to the entrepreneur. Thus, a proxy might be technical progress or growth in total factor productivity (TFP), output per combined factor input.

Solow (1957), Abramovitz (1956), Massell (1960) and Denison (1962) find that the major sources of growth per worker are not inputs but a residual, technical progress. Econometric studies observe that whereas most growth per worker in developed countries (DCs) is explained by technical progress (innovation), most LDC growth is attributed to increased capital per worker. The World Bank's (2004a, p. 44) decomposition of gross domestic product (GDP) growth, 1990–2000, reinforces earlier empirical studies, while finding negative TFP for sub-Saharan Africa.

These findings are consistent with a world divided into technological innovators (Schumpeterian entrepreneurs), technological adapters and the technologically excluded (Sachs, 2000; p. 81). Schumpeter (1961, pp. 229–30; 1939, Vol. 1, pp. 131–7) contrasts the innovator with the imitator. Addison (2003, p. 5) finds that LDCs imitating DCs, boosted by increased education, is the major contributor to increased TFP. Sayigh (1962, pp. 12–50) and Nafziger (1977a, p. 21) broaden Schumpeter's entrepreneur to those who imitate, derive or adapt existing innovations (or technological adopters). Techniques developed outside the firm must be modified, especially when relative factor prices differ.

Where are entrepreneurs? Fifteen percent of the world, with about 80 percent of GDP, most Organisation for Economic Co-operation and Development (OECD) countries (including South Korea) plus Taiwan, are technological innovators. About 50 percent of the world, with more than 15 percent of GDP (and at least 2 percent of GDP high-tech exports), are technological adopters. Adopters include northern Mexico, Costa Rica, Argentina, Chile, Tunisia, South Africa, Israel, India (except the Ganges valley), Singapore, Malaysia, Indonesia, Thailand, coastal China, the Baltic states, Russia (near St Petersburg), plus OECD countries New Zealand, Spain, north-eastern Greece, Poland, the Czech Republic, the Slovak Republic, Hungary, Slovenia, Romania and Bulgaria. The remaining 35 percent, with no more than 4 percent of GDP, are technologically excluded (Sachs, 2000, p. 81).

What determines technological transfer?

How do entrepreneurs in technological adapting economies differ from those in innovating economies? Technological adopting entrepreneurs tend to be in economies that produce standardized goods at late stages of the product cycle or that participate in global production networks (GPNs).

Comparative advantage may be based on a technological advantage, a new product or production process that provides a temporary world monopoly until other countries imitate. The product-cycle model states that a good that requires highly skilled labor, with standardization, can be mass-produced by LDCs with less skill (Vernon, 1966).

England specialized in cotton textiles from the mid-eighteenth to late nineteenth centuries. In the 1880s and 1890s, Japan substituted indigenous production with British machines for imports from Britain. By 1921–39, Japan had a comparative advantage in textiles, invading Western markets. In the 1960s, Japan imported cotton textiles from the newly industrialized countries (NICs) of South Korea, Taiwan, Hong Kong and Singapore, many of which used Japanese investment and technology to compete even in Japan. Although China combined existing technology and cheap labor to expand textile exports, 1980–2007, Africa gained little from this phase.

How do LDCs adopt foreign technology? Late nineteenth-century Japan hired foreigners, bought foreign machinery and learned from foreign buyers' standards, eventually displacing foreigners as technology standardized. In the twentieth century, newly industrializing countries' (NICs') borrowing emphasized foreign investment and technology, competing with DCs. For contemporary LDC adopters, a major strategy is: 'participation in the global production networks established by multinational enterprises (MNEs) [that] provide developing countries [access to] global know-how and expanding their integration into world markets' (World Bank, 1997, p. 2).

Among emerging nations, Mexico, Thailand, Malaysia, China and Korea comprise 78 percent of sales of parts and components to DCs. The link between emerging countries and DCs can involve ownership, arm's-length transactions and supplier–purchaser relationships. To maintain control, foreign direct investment (FDI) in a subsidiary is preferred, reflected in the increasing share of intra-firm exports in MNEs' parents' exports, especially in Japan (World Bank, 2003a, pp. 59–62).

Beyond these five emerging nations, low-income countries' (LICs') manufactures as a percentage of exports rose from about 20 percent in 1981 to almost 80 percent in 2001, while middle-income countries' (MICs') percentage rose from 24 to 70. Even without China and India, the rise in export manufactures is substantial (World Bank, 2004a, pp. xiv, 65). Might LDCs' increased manufacturing exports reflect only the increasing value-added

steps? No. Indeed, LICs and MICs groupings have a higher percentage of value added in industry than DCs do (World Bank, 2003a, pp. 41–7, 189–92).

Much international trade and FDI has shifted from the exchange of consumer goods to the production and exchange of parts and components, blurring the nationality of the products. In 1998, the USA comprised only 37 percent of its cars' value added (World Bank, 2003a, p. 55; 2004a, pp. 66–9), in contrast to Henry Ford's 1921 assembly-line-produced Model T and the US factory of the 1950s. GPNs, organized by MNEs, with LDCs participating in lower value-added steps, expanded rapidly the last quarter of the twentieth century. The share of imported to total intermediate inputs in manufacturing in DCs and their value added from imported inputs as a share of exports increased substantially from the 1970s to the 1990s (World Bank, 2003a, p. 56).

Technological progress in transport, communications, electronics, and data processing has increased GPNs and FDI. Cheaper and faster telephone, fax, internet, cargo connections, and improved data analysis, using electronic interchange, have facilitated GPNs. Lower tariffs also contributed to increased imported inputs to total sales of US affiliates (World Bank, 2003a, pp. 57–8).

Since the late 1980s and early 1990s, MNEs increased international outsourcing, offshoring services (especially in ICT – information and communications technology), importing components from low-cost producers and exporting to overseas processors. In 1998, almost half of ICT imports by OECD countries were from non-OECD (mostly Asian) countries (Nafziger, 2006, pp. 324–5). US, Japanese, other OECD and even South African firms have organized GPNs utilizing LICs in a new division of labor. Trans-Pacific or Asian borderless economies may be organized like flying geese, with technically advanced economies in the lead; NICs specializing in sophisticated research and development (R&D)-intensive and technology-intensive industries; and LICs undertaking less sophisticated, labor-intensive, low value-added assembly. However, recently countries such as India and China have climbed the ladder toward providing the skills and organization for late-stage ICT processing.

GPNs enable the breaking up of production into discrete, specialized stages. LDCs may produce a low-tech component of a high-tech good, improving productivity and climbing stages through learning by doing. Both China and India doubled GPN output from 1980 to 1998, primarily from increased duty-free access of imported intermediates for exports. Moreover, both countries, particularly China, have negotiated favorable joint ventures and technology transfers, enabling learning gains (World Bank Group, 2004, pp. 69–77).

The technologically excluded – tropical Africa, Bangladesh, Burma, Laos, Cambodia and Haiti – with poor investment climate, are not in GPNs. Although FDI is not a panacea, improved institutions and infrastructure, good governance, sound macroeconomic policies and investment incentives would facilitate GPN gains.

Is there a shortage of entrepreneurs?

How do supply and demand differ between technologically innovative (I) and excluded or underdeveloped (U) economies?

Demand

The demand for entrepreneurship is a function of its marginal net revenue productivity. Marginal product depends on other production factors, capital (including infrastructure), skilled and unskilled labor, natural resources and the state of the art.

For Schatz (1963), lack of LDC entrepreneurship suffers from a poor economic environment, such as inadequate technology and complementary resources (demand) rather than insufficient entrepreneurial capacity (supply). LDCs with entrepreneurial ability comparable to DCs would make little profit because of a poor environment. The evidence is the success of LDC entrepreneurs (from India before 1991) abroad.

Demand is dampened by poor credit markets and insufficient wealth for collateral to borrow start-up capital from banks and moneylenders (Ray, 1998, pp. 229–36). LICs had an adjusted net savings–GDP of 7.8 percent compared to 13–14 percent for MICs and DCs (World Bank, 2003b, p. 17). Inadequate infrastructure, slow and irregular deliveries, and unsatisfactory servicing of equipment and supplies reduce demand. Efficiency wages (wage divided by labor productivity) above equilibrium, resulting from minimum-wage pressure, may mean LDC labor is not cheap, discouraging demand. Another impediment is that skills (marketing, purchasing, labor relations, political administration, finance, production, engineering) frequently cannot be bought in the marketplace (Kilby, 1971, pp. 1–40). Access to technology is limited because of unfavorable FDI policies, limited R&D, insufficient skills to transfer technology and small market size. Moreover, rent-seeking, unproductive activity to obtain private benefit from public action (frequent illegal bribes and coercion), reduces demand.

Nafziger and Sudarsana Rao (1996, pp. 90–103) reinterviewed a universe of 1971 industrial entrepreneurs from Visakhapatnam, India (population 1 million) whose firms still survived in 1993. For the authors, who identified one entrepreneur for each of 50 firms, innovators were the 16 who carried out ‘new combinations’ in Visakhapatnam. Examples included utilizing previously flared natural gas, safety regulators for gas cylinders, adapting

anodized products and aluminum architectural fittings, tire-retreading after official limitation on tire supply, introducing steel barges and motor launches to the port, the first daily delivery of fresh-wrapped bread, and the city's first commercial printing press (Nafziger, 1978, pp. 149–51). Surprisingly, as innovation rose, firm survival fell (coefficient insignificant). India's licensing policy before the 1990s enticed entrepreneurs to make profits from long-term government-granted monopoly rents rather than innovation. Many entrepreneurs could not acquire licenses essential for business; India's 1991 delicensing came too late. Pervasive rent-seeking occurs in technologically excluded economies where the state is soft and lacks clear business rules of law (Myrdal, 1968, Vol. 2), preventing returns to innovation because of the premium on arbitrary license grants (no explicit criteria for allocation).

The demand for entrepreneurs is limited by poorly developed institutions (constitutions, law, regulations, enforcement, informal constraints) and poor governance. Institution-building takes time, evolving by trial and error (North, 1997, p. 2).

A barrier to LDC and post-socialist innovation is insecure property rights. De Soto (2000) attributes Western success to legally enforceable property titling, based on painstaking accrual of legislation consistent with the social contract. Although LDC governments may provide credit and industrial estates for start-up firms, insufficient property rights limit growth, illustrating de Soto's dead capital, inaccessible as collateral for borrowing or bonds. Formal credit markets are non-existent for most LDC businesses. If De Soto's estimate of the world's dead capital is overstated by 70 percent, as Woodruff (2001) suggests, dead capital is still \$2.8 trillion, \$1250 for every LDC citizen.

Will Chinese capitalists invest and innovate when land use rights are insecure? Past experience suggests China's property-rights regime is discontinuous, raising questions about 99-year leases.

Education and occupation as supply factors

Literacy and education The supply of entrepreneurs is hampered by low adult literacy (LICs 63 percent) and education (LIC primary completion 74 percent) (World Bank, 2004b, pp. 82, 86). Yet ironically, education can limit entrepreneurship by increasing occupational choices. In the early 1960s, when Nigerians were replacing colonials in government, the few university graduates were lured by its salary, security and perquisites rather than by entrepreneurship. By contrast, university graduates in excess supply, as pre-1990 south India, may choose small business to avoid unemployment or blue-collar jobs.

Occupational background In most LDCs, numerous young people apprentice to learn baking, shoemaking, tinsmithing, blacksmithing, tanning or dressmaking from a relative or other artisan. Where crafts require a lengthy apprenticeship, education may be negatively related to success. Time and money spent on education represents relinquished opportunities in training closely related to entrepreneurship (Nafziger, 1977a).

Artisans trained this way may become manufacturers during early industrialization, as in England's Industrial Revolution and today's LICs. The enterprise's scale may gradually expand. Even so, few artisans make the leap to manufacturing. However, artisans and apprentices benefit from innovation, training and extension. In Meiji Japan (1868–1912), apprentice systems improved with new techniques.

Studies of LDC industrial entrepreneurs indicate trade as a former occupation. Trade provides market knowledge, management experience, sales contacts and capital. For traders, an industrial venture may await government protection and industrial policy, technical and management training, and loans.

Most successful industrialists have had management responsibility – in manufacturing, handicrafts, trade, transport or contracting – before becoming entrepreneurs. Landowners, with high propensities to consume and little experience in secondary labor relations, and government employees, who are risk-averse, rarely become entrepreneurs.

Nafziger's survey (2006, pp. 398–407) of empirical studies shows few entrepreneurs were previously blue-collar workers. They tend to be 'pushed' into business from poor options such as unemployment, rather than 'pulled' by rapidly expanding markets.

Does self-employment pay as well as paid employment? Hamilton (2000) finds that the present value of income to the median entrepreneur of a long-lasting business is substantially less than that of a paid job with zero tenure. Moreover, most businesses, in both DCs and LDCs, fail within four years of establishment (Nafziger, 1968). Might prospective entrepreneurs face 'push' factors of few options or inflated expectations of 'striking it rich'? No, Hamilton (2000, p. 628) shows that entrepreneurship offers significant non-monetary benefits, such as 'being your own boss'. Perhaps this motive, encountered among scores of LDC entrepreneurs I interviewed, is most important in spurring entrepreneurship.

Samurai, the formerly feudal warrior class, were disproportionately represented among industrialists and bankers during Japan's Meiji Restoration. Hirschmeier (1964) stresses the community-centered samurai, sacrificing profits for national economic progress. But Yamamura (1968) indicates the blurring of samurai status from purchase by farmers and merchants during the late Tokugawa period. Profit, not nationalism, provided

the impetus for business. Samurai benefited from pensions, government contracts, scarce capital, subsidies, debt payments from financing the 1867–68 coup, and the purchase of state enterprises at low prices from the Meiji emperor (1868–1912), whose bureaucracy was controlled by lower-ranking samurai.

Socio-economic factors affecting supply

Family as entrepreneur I think that the extended family is not an obstacle to development. True, family members may demand entrepreneurs share savings, diverting capital. However, if family members need education, training or start-up funds, the larger family may pool risks to support them.

Poorly developed credit markets may hamper entrepreneurs from borrowing from banks. Most LDC entrepreneurs in firms with less than 100 employed receive most start-up funds from the extended family.

Family entrepreneurship can mobilize resources, make quick, unified decisions, hire trustworthy managers and limit irresponsibility. Among Nigeria's Igbo, families guarantee debt payments, and their solidarity prevents default, a blemish on family reputation. The family frequently funds apprentice training and initial capitalization, although hindering firm expansion by diverting resources to consumption (Nafziger, 1977a, pp. 187–93).

Some LDC industrial conglomerates are family owned. India's largest private manufacturers are usually members of old mercantile families, who control several companies and assign specialized roles to family members.

In India, the business family is usually methodical in investing in its children. The family may use its wealth for its youths' entrepreneurial development, and may diversify children's educations. Each son and (recently) daughter is moved within the family's firms, gradually increasing responsibility. Moreover, families sometimes arrange marriages to further alliances with other business families. Family entrepreneurship, however, may be conservative about taking risks, innovating, delegating authority and hiring professional managers (Nafziger, 2006, pp. 398–402).

Achievement motivation and self-assessment Childhood in traditional societies produces an authoritarian personality with a low need for achievement (urge to improve) and a high need for submission. When parents think children cannot manage their world, they treat them oversolicitously and prevent initiative. The child avoids anxiety by obeying powerful people. Society requires changes in child rearing to stress independence and creativity (Hagen, 1962).

McClelland (1961) contends that a society with high need for achievement produces more energetic entrepreneurs, who bring about faster

economic development. He supports achievement motivation training for entrepreneurship development programs.

Jovanovic (1982, pp. 649–70) finds that differences in entrepreneurial ability, learned over time, determine business entry or exit. From business experience, people estimate their ability more precisely, expanding output as they revise estimates upward, and contracting with downward revisions.

Religion and ethnicity Weber, in *The Protestant Ethic* (1958), tried to explain why continuous capitalist development originated in sixteenth-century Western Europe. He noted that European businessmen were overwhelmingly Protestant and that capitalism was most advanced in England, Holland and other Protestant countries.

Protestantism, like Catholicism, was ascetic, systematically regulating the Christian's conduct. For Weber, the Protestant ethic's 'inner-worldly' asceticism meant vigorous activity in a secular vocation or calling (in contrast to the Catholic monastery's 'other-worldly' asceticism). The Protestant ethic fostered hard work, frugality, sobriety and efficiency, components of the capitalist spirit. Protestantism's calling nurtured systematic organization of free labor and gave religious justification for unstinting work at low wages in God's (and the employer's) service. These attitudes resulted in savings, hard work and economic progress. Calvinists and other puritans comprised ascetic Protestantism.

Critics argue that Protestantism, although correlated with the rise of capitalism, may not be causal. A third factor, the fall of Catholic all-encompassing power, may have contributed to both. Alternatively Protestantism may have accommodated a rising capitalism. Or the secularization, ethical relativism and social realism of Protestantism may have been as important as 'this-worldly' asceticism in contributing to development.

Although scholars found non-Western ideological sources of capitalism, Weber's work stimulated scholars to ask how entrepreneurship is affected by religious, ethnic and linguistic communities.

Social origins and mobility The dominant American folk hero goes from rags to riches through business. One highly celebrated was steel magnate Andrew Carnegie (1835–1919), an uneducated immigrant, a working man's son, forced to find work as a boy. Through cleverness and hard work, he rose from yarn spinner to messenger to assistant railroad superintendent to industrial leader. For Carnegie (1902, p. 189): 'The millionaires who are in active control started as poor boys and were trained in the sternest but most efficient of all schools – poverty.' Even so, he is atypical. The nineteenth-century rags-to-riches stories by US popular author Horatio Alger paint a false picture. The typical industrial leader in the late nineteenth and early

twentieth centuries was usually American by birth, English in national origin, urban, educated through high school, and born and bred in an atmosphere in which business and a relatively high status were associated with family life (Miller, 1962).

Industrialists outside the United States have a similar sociological profile. Innovators during the English Industrial Revolution were primarily sons of men in comfortable circumstances (Hagen, 1962). Industrial entrepreneurs from Greece, Nigeria, Pakistan, India and the Philippines had an occupational and family status substantially higher than the population generally. Industrial corporate managers, mostly from families having the funds to pay for a university education, generally have an even higher socio-economic status than entrepreneurs. Nafziger (2006, p. 405) found a high socio-economic status among business people, mostly public sector executives, in Stalinist Russia and Maoist China.

Advantages of wealth and privilege Entrepreneurs frequently profit from monopoly advantage, usually the result of greater opportunities, such as: (1) access to more economic information than competitors; (2) superior training and education; (3) a lower discount of future earnings; and (4) agreements to restrict entry or output. All four are facilitated by wealth or position (Dobb, 1926).

What effect does wealth inequality have on entrepreneurship? Ray (1998, pp. 229–37) thinks inequality reduces entrepreneurship because fewer persons attain a critical level of wealth (or collateral) for bank loans. However, in LICs, this critical level may be so in excess of mean income that higher inequality increases those qualified for loans.

In India, high castes upper classes and large business families use monopoly advantages to become industrial entrepreneurs in disproportionate numbers. In Visakhapatnam, 52 percent of entrepreneurs (but 11 percent of blue-collar workers) were from high Hindu castes, which comprise 26 percent of the population. No entrepreneur, but a disproportionate share of blue-collar workers, was from low-caste backgrounds (Dalits and Protestant or Roman Catholic Christians). This lopsided distribution of business, which reflects differences in opportunities between the privileged and less privileged, is typical of many other countries too (Nafziger, 1977b).

Entrepreneurship is a means of moving up economically. Research indicates that the socio-economic status of entrepreneurs is higher than that of their parents, which is substantially higher than that of the general population (ibid.).

Gender In the United States, there are relatively few women in business – because of sex discrimination and female socialization. Feminists charge

that girls are raised to aspire to be secretaries, nurses, dancers and kindergarten teachers rather than to start businesses.

In many LDCs, the percentage of businesswomen is lower than in the USA. Despite exceptions, such as female traders in large West African open-air marketplaces, few large-scale LDC entrepreneurs are women.

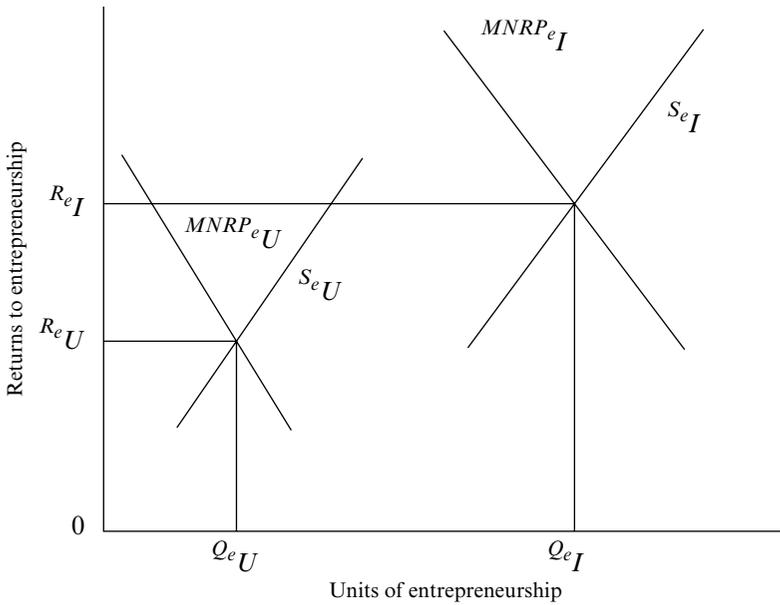
Most LDCs have cultural norms dictating how males and females behave at work. Frequently a woman's physical mobility and social contact are restricted. Lessinger (1980) states that Indian women are not allowed to deal directly with strange men, since it is assumed that all unmonitored contact between unrelated men and women must be sexual. Furthermore, according to Lessinger, Indian women are viewed as naturally weaker, more emotional, less socially adept and less rational than men. These views have been used to limit business competition between women and men, and sometimes to justify a woman's restriction to the household.

Moreover the culture may view the requisite shrewdness, quick judgment, gregariousness and force of personality of the successful entrepreneur as inconsistent with a good woman. Where a woman is determined to be an entrepreneur, she is daily reminded that she is going against the norm: sexual harassment is likely if she steps beyond accepted behavior. Although a woman can circumvent these restrictions by surrounding herself with people who can vouch for her good behavior, this is cumbersome for the entrepreneur, who must be mobile. Also, bankers and suppliers may refuse the businesswoman credit. These attitudes toward female entrepreneurs are prevalent in LICs.

Since the 1980s, however, LDC governments and non-governmental organizations have established institutions to lend to women. The most prominent is group lending, similar to Bangladesh's Grameen Bank, established in 1988, which avoids subsidizing credit. Peer borrowing groups of five or so women with joint liability approve loans to other members to substitute for the bank's screening. The members discuss loans, scrutinize the borrower's plan, and save part of the loan, which remains on deposit. Failure to repay by any member jeopardizes the group's credit. Grameen has more than 1170 branch offices, has served 2 million clients (94 percent women), and has a repayment rate of 92 percent (Yunus, 2003). Other LDC credit programs, in the Philippines, the Dominican Republic and Indonesia, provide training and technical aid for the urban poor, especially women, in microenterprises.

Demand and supply: a summary

What determines the demand for and supply of entrepreneurship? Demand is determined by complementary factors (infrastructure, credit markets,



Note:

- $MNRPe$ – Marginal net revenue productivity of (demand for) entrepreneurship
- Se – Supply of entrepreneurship
- Q_e – Equilibrium quantity of entrepreneurship
- R_e – Equilibrium returns to entrepreneurship
- U – A technologically underdeveloped economy
- I – A technologically innovative economy (with the same population)

Figure 26.1 The market for entrepreneurs

labor quality, high-level skills), market size, institutions (such as property rights), the state and its policies (investment climate and freedom from rent-seeking) and technology (dependent on policies toward technological diffusion). Supply is determined by education, occupational choice and socio-psychological factors shaped by ethnicity, religion and experience. Schumpeter contends that entrepreneurs are not distributed randomly like singers. Figure 26.1 shows how supply and demand, based partly on cross-national differences from dissimilar histories, affect the quantity of entrepreneurs (innovators and adopters).

The relatively low remuneration and quantity of entrepreneurs in the technologically underdeveloped economy, U, suggests a lack of supply and weak demand. By contrast, the relatively high quantity and remuneration of entrepreneurs in the technologically innovative country, I, indicates stronger supply and demand.

The state

Gerschenkron (1982) attributes the strategies of latecomers (nineteenth-century France, Germany, Russia and Italy) to the advantages of relative backwardness: adopting the leaders' (Britain and the USA) modern technologies, using powerful ideological medicines and state intervention.

Japan's gross national product (GNP) per capita was less than that of Western countries in the late nineteenth century. However, since the 1867 abolition of feudal property, Japan's growth was the fastest in the world. The early Meiji period (1868–1900), which lacked a capitalist class, relied on government to start factories and support private ventures. The Western threat spurred 'guided capitalism' to 'enrich the nation to strengthen the army'. The state invested in infrastructure (posts, telegraphs, transport, electricity, gas and technical research), institutionalized corporations and enterprise freedom, exhibited and borrowed abroad, aided exports, imported machines cheaply, established trading companies, set marketing standards, organized banking (central Bank of Japan, 1872), educated students and officials abroad and (in the absence of foreign aid) hired thousands of foreigners to adapt technology (Nafziger, 1995).

Franko (1983, p. 23) contends:

The Japanese are without doubt the world's champion importers of 'other people's' technology . . . Japan has continuously sent its sons to be educated abroad and then to live or travel abroad to search out ways of catching up with or surpassing the West.

The nineteenth-century Meiji government initiated half the investment outside agriculture but sold most industrial properties, often at bargain prices, to private businesspeople. Additionally government aided industry through low-wage policies, low taxes on business and high incomes, a favorable legal climate, destruction of fiefs' economic barriers, lucrative purchase contracts, tax rebates, loans and subsidies. The Meiji exploited agriculture, relying on a land tax for revenue for industrial investment and assistance. State-assisted entrepreneurs formed financial cliques (*zaibatsu*), dominating industry until World War II's devastation.

The Ministry of International Trade and Industry's (MITI, now METI) policy contributed to rapid growth and entrepreneurial development in 1950–85. MITI created infrastructure; provided loans, subsidies and tax incentives; restructured industries; subsidized R&D; negotiated technological transfers; and protected industries (heavy machinery, color televisions, color film and integrated circuits).

From 1960, South Korea's intervention to spur industrial entrepreneurship was more systematic than Japan's. The state instigated 'every major shift in industrial diversification', in 1960–80, including import substitution

in cement, fertilizers, oil refining and synthetic fibers (Amsden, 1989, p. 80). Korea initially emphasized producing standardized goods with a low-wage comparative advantage, shifting to a dynamic advantage by innovating through learning by doing. Although the state initiated entrepreneurship, over time, with private sector experience, entrepreneurship increasingly became joint public-private ventures. State strategies included multiple pricing (subsidies or favorable exchange rates for imports) or negative real interest rates by nationally controlled banks for favored entrepreneurs. Contrary to Schumpeter, Korea relied on existing firms, frequently state-supported diversified business groups (*chaebols*), for new ventures (Amsden, 1989).

In the 1990s Japan exhausted four decades of catch-up growth from scale economies, learning by doing and adapting foreign technology. Japan's distorted prices, banking bad debts, cartels, *keiretsu* and entrenched interests, less costly while catching up, failed to spur entrepreneurship while a slow-growing technological leader. Entrepreneurs in other latecoming nations could suffer a similar fate.

Policy implications

Facilitating innovative and adapting entrepreneurship requires a well-governed state with well-developed institutions, infrastructure investment and clear business rules of law, including secure, legally enforceable property rights. The state needs the competence to prevent pervasive rent-seeking that allocates licenses, subsidies and monopoly grants capriciously or corruptly, thus reducing returns to innovation.

LICs and MICs can learn from the success of latecomers that took advantage of relative backwardness to catch up with DCs. Latecomers' state initiatives included: infrastructure investment; direct or joint investment in key sectors; education (especially in business, science and engineering) from technological leaders; extension, training and exhibitions; aiding exports; providing credit; and enabling entrepreneurs and bureaucrats to learn and borrow from DCs.

Developing entrepreneurship requires the state facilitating technological transfer by GPN participation and encouraging foreign investment and technology to replace DCs when the product cycle favors cheap labor. Hiring experienced foreigners, buying foreign machinery and technology, and learning foreigners' exacting standards can contribute to displacing foreign competitors as technology becomes standardized. To do so, LDCs need a strong investment climate, with advances in transport, communications, electronics and data processing, and low tariffs contributing to GPN participation. Being a part of a North American or Asian borderless economy can foster learning by doing, eventually climbing the

value-added ladder toward the skills and organization for late-stage processing.

Schumpeter correctly emphasizes innovation, while overstating the ingenious entrepreneur. The crucial factor is a facilitating state that refrains from hindering innovation and adaptation.

References

- Abramovitz, Moses (1956), 'Resources and Output Trends in the United States', *American Economic Review*, 44 (2): 5–23.
- Addison, Douglas (2003), 'Productivity Growth and Product Variety', World Bank Working Paper 3023.
- Amsden, Alice (1989), *Asia's Next Giant: South Korea and Late Industrialization*, New York: Oxford University Press.
- Baumol, William (1968), 'Entrepreneurship in Economic Theory', *American Economic Review*, 58 (2): 64–71.
- Carnegie, Andrew (1902), *The Empire of Business*, New York: Doubleday, Page & Co.
- Coase, Ronald H. (1937), 'The Nature of the Firm', *Economica*, New Series, 4: 386–405.
- Denison, Edward (1962), *The Sources of Economic Growth in the United States and the Alternatives Before Us*, New York: Committee for Economic Development.
- De Soto, Hernando (2000), *The Mystery of Capitalism: Why Capitalism Triumphs in the West and Fails Everywhere Else*, New York: Basic Books.
- Dobb, Maurice (1926), *Capitalist Enterprise and Social Progress*, London: George Routledge & Sons.
- Franko, Lawrence G. (1983), *The Threat of Japanese Multinationals: How the West can Respond*, Chichester: Wiley.
- Gerschenkron, Alexander (1982), *Economic Backwardness in Historical Perspective: A Book of Essays*, Cambridge, MA: Belknap Press.
- Hagen, Everett E. (1962), *On the Theory of Social Change: How Economic Growth Beings*, Homewood, IL: Dorsey Press.
- Hamilton, Barton H. (2000), 'Does Entrepreneurship Pay?' *Journal of Political Economy*, 108 (3): 604–32.
- Hirschmeier, Johannes (1964), *The Origins of Entrepreneurship in Meiji Japan*, Cambridge, MA: Harvard University Press.
- Jovanovic, Boyan (1982), 'Selection and Evolution of Industry', *Econometrica*, 50: 649–70.
- Kilby, Peter (ed.) (1971), *Entrepreneurship and Economic Development*, New York: Free Press.
- Knight, Frank (1921), *Risk, Uncertainty, and Profit*, Boston, MA: Houghton Mifflin.
- Leibenstein, Harvey (1966), 'Allocative Efficiency vs. "X-Efficiency"', *American Economic Review*, 56 (3): 392–415.
- Leibenstein, Harvey (1968), 'Entrepreneurship and Development', *American Economic Review*, 58 (2): 72–83.
- Lessinger, Johanna (1980), 'Women Traders in Madras City', New York: Barnard College.
- Massell, Benton (1960), 'Capital Formation and Technological Change in the United States Manufacturing', *Review of Economics and Statistics*, 42: 182–8.
- McClelland, David C. (1961), *The Achieving Society*, Princeton, NJ: D. Van Nostrand.
- Miller, William (ed.) (1962), *Men in Business: Essays on the Historical Role of the Entrepreneur*, New York: Harper & Row.
- Myrdal, Gunnar (1968), *Asian Drama: An Inquiry into the Poverty of Nations*, vol. 2, Harmondsworth: Penguin.
- Nafziger, E. Wayne (1968), 'A Reconsideration of "Capital Surplus" in Nigeria', *Nigerian Journal of Economic and Social Studies*, 10 (1): 111–16.
- Nafziger, E. Wayne (1977a), *African Capitalism: A Case Study in Nigerian Entrepreneurship*, Stanford, CA: Hoover Institution.

- Nafziger, E. Wayne (1977b), 'Entrepreneurship, Social Mobility, and Income Redistribution in South India', *American Economic Review*, 67 (1): 76–80.
- Nafziger, E. Wayne (1978), *Class, Caste, and Entrepreneurship: A Study of Indian Industrialists*, Honolulu: University Press of Hawaii.
- Nafziger, E. Wayne (1995), *Learning from the Japanese: Japan's Pre-War Development and the Third World*, Armonk, NY: M.E. Sharpe.
- Nafziger, E. Wayne (2006), *Economic Development*, 4th edn, New York: Cambridge University Press.
- Nafziger, E. Wayne and R. Sudarsana Rao (1996), 'Industrial Entrepreneurship and Innovation under Licensing and Liberalization in India', *Indian Economic Journal*, 44 (4): 90–103.
- North, Douglass C. (1997), 'The Process of Economic Change', UNU/WIDER Working Paper 128.
- Ray, Debraj (1998), *Development Economics*, Princeton, NJ: Princeton University Press.
- Sachs, Jeffrey (2000), 'A New Map of the World', *Economist*, 24 June, pp. 81–3.
- Sayigh, Yusif A. (1962), *Entrepreneurs of Lebanon: The Role of the Business Leader in a Developing Economy*, Cambridge, MA: Harvard University Press.
- Schumpeter, Joseph (1939), *Business Cycles*, 2 vols, New York: McGraw-Hill.
- Schumpeter, Joseph (1949), 'Economic Theory and Entrepreneurial History', in *Change and the Entrepreneur*, Cambridge, MA: Harvard University Press, pp. 63–84.
- Schumpeter, Joseph (1961), *The Theory of Economic Development*, Cambridge, MA: Harvard University Press.
- Schatz, Sayre (1963), 'Economic Environment and Private Enterprise in West Africa', *Economic Bulletin*, 7 (4): 42–56.
- Solow, Robert (1957), 'Technical Change and the Aggregate Production Function', *Review of Economics and Statistics*, 39: 312–20.
- Vernon, Raymond (1966), 'International Investment and International Trade in the Product Cycle', *Quarterly Journal of Economics*, 80 (2): 190–207.
- Weber, Max (1958), *The Protestant Ethic and the Spirit of Capitalism*, trans. Talcott Parsons (1930), New York: Charles Scribner's Sons.
- Woodruff, Christopher (2001), 'Review of de Soto's *The Mystery of Capital*', *Journal of Economic Literature*, 39 (4): 1215–23.
- World Bank (1997), *Global Economic Prospects 1997*, Washington, DC: World Bank.
- World Bank (2003a), *Global Economic Prospects 2003*, Washington, DC: World Bank.
- World Bank (2003b), *World Development Report 2003*, New York: Oxford University Press.
- World Bank (2004a), *Global Economic Prospects 2004*, Washington, DC: World Bank.
- World Bank (2004b), *World Development Indicators 2004*, Washington, DC: World Bank.
- World Bank Group (2004), *World Bank Poverty-focused Activities*, Washington, DC: World Bank.
- Yamamura, Kozo (1968), 'A Re-examination of Entrepreneurship in Meiji Japan (1868–1912)', *Economic History Review*, 21 (1): 148–58.
- Yunus, Muhammad (2003), 'Halving Poverty by 2015: We Can Actually Make it Happen', London: Commonwealth Institute.

27 Natural resources and development

*Richard Auty*¹

After a brief discussion of the minimal role that natural resources have been assigned in most mainstream growth models, this chapter focuses upon the performance of the developing countries in recent decades, during which time the impact of natural resources upon economic growth appears to have become highly significant, an outcome that has generated much interest in the so-called 'resource curse'. This recent outcome is paradoxical because a rich endowment of natural resources should help boost the rate of economic growth. This is because the natural resource rent provides additional revenues above those required by an efficient producer with which to raise the rate of investment and it also generates extra foreign exchange with which to import the capital goods required to build the infrastructure of a modern economy. Certainly, by 1960, the median income of the resource-rich developing countries was 50 percent higher than that of the resource-poor countries (Auty, 2001, p. 5). Yet a generation later, the median income of the resource-poor countries exceeded that of the resource-rich countries, many of which were struggling to recover from a protracted growth collapse. This reversal of fortune strongly suggests that the resource curse is not a deterministic phenomenon and that it may have its roots in policy failure.

Natural resources in mainstream models of economic growth: an overview

Economists have long tended to regard natural resources as generally far less important to economic growth than capital and labor. In the early nineteenth century, however, classical economists voiced concern that natural resources, notably land, might constitute a limit to per capita gross domestic product (GDP) growth. Yet by the close of that century most economists believed that society could surmount the Malthusian population trap and the law of diminishing returns, so that sustained economic and social progress was likely, if not certain. This optimism was supported by the development of agriculture in new territories overseas, the continued discovery of minerals, and rapid technological advances in extracting and refining ores of lower grades. Mainstream economists came to believe that increased capital and technological progress would prevent natural resources from ever constraining global economic growth (Auty and Mikesell, 1998).

Several growth models were formulated in the mid-twentieth century, based on production functions like the Cobb-Douglas, such as the Harrod-Domar models in which natural resources played little part. Capital accumulation was believed to drive economic growth, while technological advances served to increase the productivity of produced and human capital. These models suggested that low domestic investment was responsible for keeping developing-country income at low levels, and that foreign investment, technical assistance and government planning could remove this capital constraint (Lewis, 1954; Myrdal, 1957; Rostow, 1956). In an open (national) economy shortages of natural resources would not constrain economic growth because countries could purchase resource-intensive products on world markets with revenue earned from exporting labor-intensive goods. In a closed economy like the global economy, natural resources appeared to be plentiful and self-evidently beneficial.

In 1972, the publication of the report of the Club of Rome (Meadows et al., 1972) ignited sharp exchanges about the degree to which natural resource scarcity constrained growth. The report focused on the scarcity of minerals and arable land through the medium term, and on the risk of degrading life-supporting environmental assets over the longer term. The report met fierce criticism (Cole et al., 1973; Beckerman, 1974). In more measured tones, Solow (1974) argued that continued technological progress is necessary to sustain a positive consumption flow from finite resources: consequently, with unlimited technological progress and an elasticity of substitution between natural resources and labor- and capital-intensive goods no less than unity, finite resources can support continual consumption with additions to the stock of reproducible capital (Solow, 1974, p. 41).

Subsequent work on environmental and natural resource accounting concluded that produced and human capital could substitute for diminishing natural capital, so that sustainable development from finite resources required that the current generation should pass to future generations either the same total stock of capital or a larger stock (Pearce et al., 1996). Nordhaus (1992) contests even this relaxed conclusion, arguing that if gains in technology offset the rate of contraction in the capital stock, then a shrinking capital stock can sustain consumption indefinitely. Many economists also remain skeptical of claims about the long-term risks from world pollution, such as global warming. They argue that incremental adjustment, based upon expanding knowledge and improving technology, can resolve such problems at a much lower cost to society than by rapid adoption of physical limits on emissions (Manne, 2004).

Many economists identify the 'Dutch disease' effects as the principal cause of the recent underperformance of the resource-rich countries. This perspective is examined in the next section and is shown to offer only a

partial explanation. The third section therefore reviews the literature on export base theory because it explains why natural resource abundance should be advantageous, but it also shows why some natural resources may contribute less to development than others. The fourth section synthesizes recent research to produce a set of resource-driven models of economic development that explain why growth collapsed in many resource-rich countries in recent decades. These models expressly include government motivation and policy responses, issues that most mainstream economic growth models considered to be beyond their remit.

Dutch disease effects and growth collapses

The Dutch disease effect is attributed to the booming resource sector keeping the value of the currency so high that other tradeables sectors cannot compete internationally. Corden and Neary (1982) explain the mechanism of the Dutch disease in terms of a three-sector model comprising a resource sector, a sector of other tradeables, usually manufacturing and agriculture, and a non-tradeables sector. They posit that a boom in the resource sector has two basic effects: namely a spending effect that alters relative prices, and a resource movement effect.

First, expenditure of the increased export revenues boosts demand for tradeables and non-tradeables, but global competition precludes price rises for tradeables so the spending effect shifts relative prices in favor of the non-tradeables sector. This brings a real appreciation of the currency, which reduces the competitiveness of non-booming tradeable activity. Whereas domestic prices of tradeable goods are moderated by import competition, the domestic prices of non-tradeables rise due to increased demand, being unaffected by either the currency appreciation or by competitive imports. Second, capital and labor move from tradeables to non-tradeables in response to changed relative prices, which reduces exports and expands imports. It also lowers capital accumulation if the non-tradeable sector is more labor-intensive than the tradeable sector, because such movements in favor of the non-tradeable sector raise wages and lower returns to capital. Furthermore, if resource booms cause manufacturing to shrink and manufacturing is especially beneficial to growth (due, for instance, to the gains from learning by doing), the resource-abundant economy can experience slower long-term growth than it would if it had no resources (Matsuyama, 1992). Krugman (1987) identifies the conditions under which temporary resource booms can lead to an enduring loss of competitiveness and a lower level of per capita income than would have been the case in the absence of the resource boom.

However, Neary and van Wijnbergen (1986, p. 40–41) point out that some deindustrialization may be a symptom of the economy's adjustment

to a new equilibrium rather than a symptom of a disease. Provided the rate of absorption of boom-sector revenues is matched to domestic absorptive capacity (typically by sterilizing a fraction of the revenue in capital funds held offshore) and unsustainable patterns of domestic consumption are not established, a well-managed economy can adjust smoothly to the changing composition of sectoral activity during commodity booms (Gelb and Associates, 1988). Consequently, those like Sachs (1999) who argue that Dutch disease effects largely explain the growth collapses in resource-abundant economies neglect the importance of policy in mediating outcomes. Export base theory provides a subtler explanation by showing how natural resource booms can sustain economic diversification, and also why the socio-economic linkages of some commodities further this outcome more than others.

Export base theory, economic diversification and development

Innes (1920), North (1955) and Watkins (1963) developed export base theory (or staple theory) in the context of the 'unsettled' regions of North America. The theory was formulated in order to explain the growth of diversified, prosperous, regional economies based upon the export of primary products, rather than upon industrialization. The theory conceives the economy as comprising two complementary sectors: a 'basic' sector that is export-oriented and attracts 'new' revenue into a region or country, whereas the 'service' or 'residential' sector recirculates such expenditure through the domestic economy. The mechanism of the export base model can be described in terms of four principal sets of linkages or socio-economic stimuli, which after Hirschman (1977) are:

- Backward linkage (the establishment of firms to provide inputs to the export commodity).
- Forward linkage (the establishment of firms to process the commodity prior to its export).
- Fiscal linkage (the spending of government taxes levied on the commodity).
- Final demand linkage (the activities set up in response to the local spending of wages and profits by labor and the owners of capital).

Export base theory explains the diversification of mono-product commodity-exporting regions in terms of a five-stage sequence (Watkins, 1963), which begins when a farmer or company identifies a primary product with potential comparative advantage and begins to export it. The second stage sees production expand, yielding both internal economies of scale and external economies (such as improved shipping facilities) that

lower average production costs and further boost the commodity's competitiveness. In this second stage, investment remains mostly channeled into the exported primary product, but in the third stage productive linkages are triggered in the form of investment in local supplies to the primary sector (backward linkage), which displace hitherto imported inputs, and/or in investing in processing (forward linkage) prior to export, as with a refinery or fabricating plant. In the fourth stage of the export base model, capital overflows from the primary sector into the non-export sector to supply growing household demand as well as the needs of firms supplying inputs to the lead sector. Typical examples include brewing, furniture production and business services in addition to production inputs, any of which may eventually enter export markets. Meanwhile, fiscal linkage takes the form of government spending on economy-wide physical infrastructure and human capital and, perhaps, finances policies designed to speed economic diversification. The region reaches its fifth and final 'mature' stage when the economy is sufficiently diversified into additional commodity exports, services and/or manufacturing that the initial primary export can no longer be identified.

Examples where natural resource exports have been the catalyst for the emergence of diversified economic regions include the Pacific North-West, South-East Brazil, the Witwatersrand in South Africa and Western Australia. But different primary exports yield different patterns of linkage, some of which are more conducive to competitive diversification than others. Baldwin (1956) shows why by examining how the commodity production function impacts upon economic growth with reference to contrasting regions in nineteenth century America. A capital-intensive production function typical of mines and plantations imposes sizeable barriers to entry and concentrates the linkages on a small number of economic agents, including large companies and the government. The heavy sunk investment causes mine and plantation owners to respond to lower prices by cutting costs (in line with the Prebisch, 1964, critique), and in some cases by seeking to stifle local activity that competes for land and labor, thereby aborting the diversification of the regional economy predicted by export base theory. Moreover, a capital-intensive production function limits both final demand linkages and production linkages. This is because the specialized equipment for mines and plantation crop processing is often most cheaply supplied as imports from overseas producers who can capture localization economies. In addition, market-oriented processing of resources is often more competitive than raw material-oriented processing. Moreover, capital-intensive activities employ relatively few workers so even if those workers are highly paid, their aggregate final consumption expenditure is relatively low. This often leaves fiscal linkage as the principal

conduit for the revenues from capital-intensive commodities, and such revenue is all too easily siphoned into the national treasury and/or into the personal bank accounts of politicians, and dissipated.

Engerman and Sokoloff (1997) contrast this pattern, which they term 'point source' linkages, with the 'diffuse' linkages of commodities like peasant cash crops, whose production function is more flexible because it offers few barriers to entry and also funnels revenue across many more economic agents. This more flexible production function responds to small additions to investment, which boost productivity and incomes, so that final demand linkage keeps rising, which stimulates a wide range of production linkages to supply farm inputs and basic consumer goods. A further source of flexibility comes from low sunk costs, which facilitate economic diversification by allowing producers to respond to falling prices by switching from low-growth to high-growth commodities (Duncan, 1993). In addition, farmers support fiscal linkage because they benefit directly from public expenditure on farm roads, rural education and health care, and therefore accept appropriate levels of taxation. According to Baldwin (1956), small farming also produces a stream of entrepreneurs (younger sons) to establish the local businesses that diversify the economy. Provided that central governments do not convert these diffuse linkages into point-source linkages by imposing swingeing taxes (through commodity marketing boards, for example, that allow them to siphon away crop rent, and more, Osei, 2001) fiscal linkage rarely dominates diffuse linkages, leaving more expenditure than in the case of point-source linkages with local economic agents that exhibit a higher propensity to save and invest effectively than many governments do (Bevan et al., 1987).

Despite the potential economic benefits from diffuse linkages identified by Baldwin (1956), fashionable policies during the immediate post-World War II decades encouraged developing-country governments to increase the scale of their intervention in the economy in order to force industrialization and reduce the economy's dependence on volatile primary products with allegedly declining terms of trade (Prebisch, 1964). Krueger et al. (1992) present empirical evidence of the scale of the impact of the resulting policy distortion. They show that direct agricultural taxation in sub-Saharan Africa averaged 25 percent of revenue, some four times that of Latin America and ten times Asian rates. The sub-Saharan African figure doubles when indirect taxes, such as overvalued exchange rates, are added (Krueger et al., 1992). In fact, rates of taxation reached levels well above those required to maximize the tax take.

Easterly and Levine (1997) suggests that these growth-repressing characteristics of sub-Saharan Africa may be rooted in the region's high level of ethnic fractionalization, whose adverse impact manifests itself through

maladroit policies. Contests for rents between ethnic groups impair stabilization efforts (Alesina and Drazen, 1991), reduce outlays on public goods below appropriate levels and boost levels of corruption (Mauro, 1995). A marked feature of ethnically driven rent-seeking activity is a lack of coordination (Shleifer and Vishny, 1993), which leads to imperfect information about the complex interplay of the numerous rent-related economic distortions. In addition, rent-seeking governments in fractional societies tend to exhibit short time horizons. They recognize that many crops, notably tree crops, and most mines have high fixed costs. This encourages government officials to set tax rates in relation to the variable costs of production and, in effect, to capture the returns to the farmers' investment in planting (McMillan, 1997). Farmers respond by withdrawing from planting the next time round, so that the government revenue base shrinks, creating even greater pressure to squeeze revenues from elsewhere in the system.

The basic contrasts in socio-economic linkages can be synthesized with several strands in recent research in development economics to derive resource-driven models that help to explain the political economy behind the growth collapses that have beset many resource-rich developing countries in recent decades.

Models of resource-driven development trajectories

The central insight from resource-driven development models is that the smaller the natural resource rent relative to GDP and the more diffusely it is spread across economic agents, the higher the probability of engendering a developmental political state that sustains rapid growth in per capita income, which in turn strengthens sanctions against anti-social governance. A developmental political state is defined here, after Lal and Myint (1996), as one that has sufficient autonomy to pursue a coherent economic policy and the aim of raising social welfare. The corollary is that the larger the natural resource rent relative to GDP and the more concentrated it is on a handful of economic agents, the greater the probability of engendering a non-developmental political state that distorts the economy, presides over a growth collapse and represses political accountability. The oil-exporting economies (excluding the Gulf states with favorable reserve–population ratios) tend to exhibit the latter set of features to a heightened degree because their natural resource rent has been both large relative to GDP, even after the 1985 oil price collapse (Table 27.1), and channeled overwhelmingly into fiscal linkage, which concentrates rent on the state.

The competitive industrialization trajectory traced by resource-poor countries

The competitive industrialization model is strongly associated with resource-poor economies. It predicts that the political state is more likely to

Table 27.1 Share of rents in GDP 1994 and GDP growth 1985–97, by natural resource endowment

Resource endowment	PCGDP growth 1985–97 (%)	Total rent (% GDP)	Pasture and cropland rent (% GDP)	Mineral rent (% GDP)
<i>Resource-poor</i> ^{1,2}				
Large	4.7	10.56	7.34	3.22
Small	2.4	9.86	5.41	4.45
<i>Resource-rich</i>				
Large	1.9	12.65	5.83	6.86
Small, non-mineral	0.9	15.42	12.89	2.53
Small, hard mineral	-0.4	17.51	9.62	7.89
Small, oil exporter	-0.7	21.22	2.18	19.04
<i>All countries</i>		15.03	8.78	6.25

Notes:

¹ Resource-poor = 1970 cropland/head < 0.3 hectares

² Large = 1970 GDP > \$7 billion

Source: Auty and Gelb (2001, p. 131). (Comprehensive data on rents available for 1994 only.)

be developmental in the presence of low natural resource rents. This is because the government must rely heavily for income on the taxation of productive activity rather than upon capturing rent. Therefore the government has a strong incentive to expand wealth. It does so by providing public goods and maintaining efficiency incentives. In consequence, the economy adheres to its comparative advantage, which for a low-rent country lies initially in labor-intensive manufactured goods. This initiates a virtuous development trajectory for the political economy that is explained by the competitive industrialization model.

The necessarily brief initial dependence of resource-poor countries on primary product exports calls for industrialization at a relatively low per capita income, which must be competitive and is therefore labor-intensive. This in turn triggers beneficial economic and social linkages. Taking the economic circuit first: early industrialization brings early urbanization, which accelerates passage through the demographic cycle so that the dependency–worker ratio falls quickly and the share of saving and investment in GDP rises and causes GDP growth to accelerate (Bloom and Williamson, 1998). Labor-intensive industrialization also quickly absorbs surplus rural labor. This boosts wage costs, which must be offset by raising worker

productivity through the acquisition of skills, a process that drives diversification into capital-intensive and skill-intensive manufacturing, which increases the capacity of the economy to cope with economic shocks. Finally, exposure to global competition sustains investment efficiency so that high investment doubles per capita income every decade and the transition to developed status can occur in less than two generations. This compares with more than five generations for the developing countries on average, according to Syrquin (1986, p. 232). In this way, competitive industrialization rapidly accumulates produced capital (Auty and Kiiski, 2001), human capital (Birdsall et al., 2001) and social capital (Woolcock et al., 2001). It also sharply raises the genuine saving coefficient, implying that the development trajectory is strongly sustainable (Hamilton, 2001).

As for the virtuous social circuit in low-rent countries: the sustained rapid increases in per capita income nurture an equitable income distribution because the elimination of surplus labor puts a floor under the wages of the poor, while the rapid accumulation of skills puts a ceiling on the skill premium. Rapid GDP growth also strengthens three key sanctions against anti-social governance to foster an endogenous democratization that is incremental. More specifically, as per capita income rises, then: (1) taxation shifts from exports towards value-added, income and profits taxes and thereby intensifies pressure for political accountability (Ross, 2001); (2) rapid urbanization builds bridging social capital that empowers civic associations (Woolcock et al., 2001); and (3) expanding private enterprise promotes demands for stronger property rights within the rule of law (Li et al., 2000).

Resource-abundance and the staple trap trajectory

In contrast, the scale of the resource rent in resource-abundant countries, which typically ranged between 13–21 percent of GDP in the mid-1990s (Table 27.1), diverts government effort from promoting wealth creation and into rent capture and distribution. This tends to nurture a non-developmental political state. Resource-abundance also prolongs the period of reliance on primary product exports, which delays competitive industrialization and constrains employment creation (Lal and Myint, 1996). Governments therefore deploy rent to expand jobs that competitive markets would not support.

The second basic characteristic of development with high rent is a longer reliance on primary product exports, which under a non-developmental political state does not retard competitive industrialization, but postpones it indefinitely. More specifically, the absence of labor-intensive industrialization perpetuates surplus rural labor in resource-abundant countries and feeds rising income inequality, which prompts governments to use the rent to boost employment directly and inefficiently instead of indirectly by

providing incentives for competitive wealth creation. The result is an over-expanded government bureaucracy and/or forced industrialization via import protection, which is not only inefficient but also, ironically, capital-intensive (Auty and Gelb, 2001). Consequently, far from achieving competitive diversification of the economy, the staple trap trajectory expands a parasitic sector that depends on subsidies from the natural resource rent. Human capital accumulates more slowly (Birdsall et al., 2001) than is the case in the competitive industrialization model and social capital is less resilient because it is dependent, being shaped by political patronage rather than mutually beneficial market transactions. Moreover, the efficiency of capital investment declines as the distortion of the economy increases.

When the rents inevitably shrink relative to GDP, because of either ongoing structural change or falling prices for the leading commodity, economic reform is blocked by the rent recipients, which form a powerful vested interest. Therefore, even as the parasitic sector increases its share of GDP and depresses the economy-wide efficiency of investment so that GDP growth slows, so the government finds it politically attractive to sustain transfers by extracting the returns to capital from the commodity sector as well as the rent. But as Mcmillan (1997) notes, this depresses incentives in the commodity sector and erodes its competitiveness. This is the essence of the staple trap model: a non-developmental political state uses rent to subsidize employment, which spawns a parasitic sector that aborts competitive diversification and renders the economy increasingly vulnerable to shocks and a growth collapse. A growth collapse retards the demographic transition so that population growth remains relatively high (Auty, 2001, p. 24) even as all forms of capital run down. Recovery from a growth collapse may therefore take a generation or more.

These negative features of the staple trap trajectory are amplified in the case of the oil-exporting countries because the rents are unusually large relative to GDP and/or 'point source' in nature. The resulting pattern of rent extraction and deployment is closely associated with failing political states, including those in Algeria, Angola, Azerbaijan, Egypt, Iran, Iraq, Kazakhstan, Nigeria, Saudi Arabia and Venezuela among the oil-exporting countries. The governments tend to extract and deploy rent in ways that weaken all three key sanctions against anti-social governance by applying rent to: (1) reduce personal taxation, which saps demands for political accountability (Ross, 2001); (2) create a dependent form of social capital based on political patronage; and (3) repress competition, thereby easing pressure from the private sector for institutional safeguards like property rights and the rule of law (Li et al., 2000; Woolcock et al., 2001).

Rent-deploying governments also lean towards authoritarianism, but although they appear strong when oil revenue is rising they become weak,

sometimes abruptly, when falling oil revenue collides with increasing social entitlements. The resulting 'bunker' states (Henry and Springborg, 2001) repress dissent and will reform only reluctantly, if at all.

The growth collapses that resulted from the cumulative distortion of the economy in resource-rich countries may not be the 'trigger' for civil strife but they provide the conditions in which such triggers can all too easily emerge. This interpretation is consistent with Collier (2000), who suggests that civil strife is strongly positively linked not only to primary exports, but also to economic decline (that is, a growth collapse). The result is a relatively large young male population (because a growth collapse retards passage through the demographic cycle into the low-growth stage) with little education, for whom conflict offers immediate financial gain. Le Billon (2001) extends the explanation for the link between natural resources and civil strife further: some resources yield revenues that are more lootable than others. The high value-weight ratios of alluvial diamonds and narcotics, for example, render them especially attractive because of their mobility, while production sites that are remote from the seat of government and close to porous borders are additional accommodating factors. Even so, bulkier commodities yielding high rent, notably oil, can also engender civil strife: Ross (2001) shows that oil-exporting governments are more vulnerable than other developing-country governments to violent overthrow.

Not all resource-rich countries experienced a growth collapse, however, even among the oil exporters. For example, some oil exporters with high ratios of reserves to population such as the paternalistic political states of the Persian Gulf faced less pressure than larger and less favorably endowed oil-exporting countries for rapid rent absorption. Their governments adopted cautious policies: they accumulated a fraction of the rent overseas, relied heavily on external advisors and made liberal use of immigrant labor to adjust to skill gaps and fluctuating demands for workers (Table 27.2). They also expanded domestic absorptive capacity by building up the economic infrastructure and human capital with which to diversify the economy away from hydrocarbon dependence. Kuwait had almost reached the point on the eve of the Iraqi invasion in 1991 of generating more income from its overseas investments than from current oil revenues, but like Saudi Arabia, the cushion of large hydrocarbon reserves also delayed reforms to stimulate private enterprise so that diversification, and also political liberalization, proceeds slower than is desirable.

Even some countries with less favorable oil-population ratios that faced strong pressures to absorb the oil revenues rapidly, like Indonesia and Malaysia, managed to boost per capita income, reduce poverty and competitively diversify their economies (Table 27.2, Memo item). In both cases,

Table 27.2 High-income oil-exporters in the Gulf and Brunei, some basic indices 2002

Country	Population (millions)	PCGNI (PPP US\$)	Hydrocarbon reserves (bbl/oe) ^a	Hydrocarbon exports (% total)	Reserves/population (000 bls/head)	Expatriot workers (% total) ^b
Bahrain	0.650	16 190	1.0	60	1.5	64
Kuwait	2.300	17 780	105.9	93	46.0	81
Oman	2.400	13 000	10.7	77	4.5	55
Qatar	0.720	31 400	83.3	79	114.3	90
Saudi Arabia	23.000	12 660	299.8	89	13.0	55
United Arab Emirates	3.300	24 030	135.5	92	41.1	90
Brunei	0.351	18 000	3.9	90	11.1	65
Memo items						
Indonesia	212.000	3 070	18.1	24	0.1	n.a.
Malaysia	24.000	8 500	18.7	9	0.8	10

Source: World Bank (2004) except: ^a BP (2002) and ^b Rodenbeck (2002).

the government gave priority to sound macroeconomic management and controlled rent-seeking activity (which nevertheless still absorbed a sizeable fraction of GDP). The governments also ensured that rural incomes rose by using rent to extend the agricultural frontier, boost irrigation and provide rural roads to link farmers to markets (Timmer, 2004). Importantly, despite protectionist policies during the 1974–78 and 1979–81 oil booms, Indonesia developed a dualistic manufacturing sector in which some firms took the rent in terms of reduced levels of efficiency while others, mostly Chinese-owned, strove for high levels of efficiency and ploughed the rents back into further industrial investment. When oil prices collapsed in 1985, the government stabilized the economy, sustained a 60 percent depreciation of the real exchange rate and began dismantling import protection, which allowed the efficient firms to export. Manufactured goods rose to account for more than half of Indonesian exports within a decade as the economy switched from a staple trap trajectory to competitive industrialization. This rapid economic restructuring was echoed in Malaysia, which had begun to establish export-processing zones from 1971 in response to disappointing results with earlier efforts to industrialize through import substitution.

Conclusion

The growth collapses that affected many resource-rich developing countries during the period of heightened commodity price volatility of 1973–85 had their roots in flawed post-World War II policies, which responded to pessimistic assessments of resource-driven development by promoting increased state intervention to force industrialization. The advocates of these policies underestimated the capacity of the discretionary deployment of rent to divert effort from wealth creation into policies for wealth redistribution, especially in sub-Saharan Africa, that deflected the economy away from its comparative advantage, rendering it vulnerable to a growth collapse.

A caveat is in order, however, where rent-rich governments enjoy high ratios of rents to population or are sensitive to the needs of a relatively large poor rural constituency, as with Indonesia, Malaysia and Botswana. Here the political state may remain developmental. Consequently, the competitive diversification of the economy is retarded, rather than aborted. The incentive to boost rural welfare was weak in most rent-abundant countries during the immediate post-World War II decades, however, so that various forms of non-developmental political state have been the norm.

A critical policy implication from this recent manifestation of the resource curse is that economic policies to revive collapsed economies may take decades due to the extreme degradation of all forms of capital, and

they must also incorporate incentives for governments to channel rent-seeking away from rent redistribution to the favored few and into wealth creation to the benefit of society at large.

Note

1. The author gratefully acknowledges helpful suggestions by the editors.

References

- Alesina, A. and A. Drazen (1991), 'Why are Stabilizations Delayed?' *American Economic Review*, 81: 1170–88.
- Auty, R.M. (2001), *Resource Abundance and Economic Development*, Oxford: Clarendon Press.
- Auty, R.M. and A.H. Gelb (2001), 'Political Economy of Resource-Abundant States', in R.M. Auty (ed.), *Resource Abundance and Economic Development*, Oxford: Oxford University Press, pp. 126–44.
- Auty, R.M. and S. Kiiski (2001), 'Natural Resources, Capital Accumulation and Welfare', in R.M. Auty (ed.), *Resource Abundance and Economic Development*, Oxford: Oxford University Press, pp. 19–35.
- Auty, R.M. and R.F. Mikesell (1998), *Sustainable Development in Mineral Economies*, Oxford: Clarendon Press.
- Baldwin, R.E. (1956), 'Patterns of Development in Newly Settled Regions', *Manchester School of Social and Economic Studies*, 24: 161–79.
- Beckerman, W. (1974), *In Defence of Economic Growth*, London: Jonathan Cape.
- Bevan, D., P. Collier and J.W. Gunning (1987), 'Consequences of a Commodity Boom in a Controlled Economy: Accumulation and Redistribution in Kenya', *World Bank Economic Review*, 1: 489–513.
- Birdsall, N., T. Pinckney and R. Sabot (2001), 'Natural Resources, Human Capital and Growth', in R.M. Auty (ed.), *Resource Abundance and Economic Development*, Oxford: Oxford University Press, pp. 57–75.
- Bloom, D.E. and J.G. Williamson (1998), 'Demographic Transitions and Economic Miracles in Emerging Asia', *World Bank Economic Review*, 12: 419–55.
- BP (2002), *Statistical Review of World Energy 2002*, London: British Petroleum.
- Cole, H.S.D., C. Freeman, M. Jahoda and K.L.R. Pavitt (1973), *Models of Doom: A Critique of the Limits to Growth*, New York: Universe Books.
- Collier, P. (2000), 'Doing Well out of War', in M. Berdal and D.M. Malone (eds), *Greed and Grievance: Economic Agendas in Civil Wars*, London: Lynne Rienner, pp. 91–111.
- Corden, M. and J.P. Neary (1982), 'Booming Sector and Dutch Disease Economics: A Survey', *Economic Journal*, 92: 826–44.
- Duncan, R.C. (1993), 'Agricultural Export Prospects for Sub-Saharan Africa', *Development Policy Review*, 11: 31–45.
- Easterly, W. and R. Levine (1997), 'Africa's Growth Tragedy: Policies and Ethnic Divisions', *Quarterly Journal of Economics*, 112: 1203–50.
- Engerman, S. and K. Sokoloff (1997), 'Factor Endowments, Institutions, and Differential Paths of Growth Among New World Economies: A View From Economic Historians of the United States', in S. Haber (ed.), *Economic Growth and Latin American Economic Historiography*, Cambridge, MA: MIT Press, pp. 260–304.
- Gelb, A.H. and Associates (1988), *Oil Windfalls: Blessing or Curse?*, New York: Oxford University Press.
- Hamilton, K. (2001), 'The Sustainability of Extractive Industries', in R.M. Auty (ed.), *Resource Abundance and Economic Development*, Oxford: Oxford University Press, pp. 36–56.
- Henry, C.M. and R. Springborg (2001), *Globalization and the Politics of Development in the Middle East*, Cambridge: Cambridge University Press.

- Hirschman, A.O. (1977), 'A Generalised Linkage Approach to Development with Special Reference to Staples', *Essays on Economic Development and Cultural Change in Honor of Bert F. Hoselitz*, Chicago, IL: Chicago University Press, pp. 67–98.
- Innes, H. (1920), *The Fur Trade in Canada*, New Haven, CT: Yale University Press.
- Krueger, A.O., M. Schiff and A. Valdes (1992), *Political Economy of Agricultural Pricing Policies*, Baltimore, MD: Johns Hopkins University Press.
- Krugman, P.R. (1987), 'The Narrow Band, the Dutch Disease and the Competitive Consequences of Mrs Thatcher', *Journal of Development Economics*, 27: 41–55.
- Lal, D. and H. Myint (1996), *The Political Economy of Poverty, Equity and Growth*, Oxford: Clarendon Press.
- Le Billon, P. (2001), 'The Political Ecology of War: Natural Resources and Armed Conflicts', *Political Geography*, 20: 561–81.
- Lewis, W.A. (1954), 'Economic Development With Unlimited Supplies of Labour', *Manchester School of Economics and Social Studies*, 22 (2), May: 139–91.
- Li, S., S. Li and W. Zhang (2000), 'The road to Capitalism: Competition and Institutional Change in China', *Journal of Comparative Economics*, 28: 269–92.
- Manne, A.S. (2004), 'Perspective Paper on Cline', in B. Lomborg (ed.), *Global Crises: Global Solutions*, Cambridge: Cambridge University Press, pp. 49–55.
- Matsuyama, K. (1992), 'Intercultural Productivity, Comparative Advantage, and Economic Growth', *Journal of Economic Theory*, 58: 317–34.
- Mauro, P. (1995), 'Corruption and Growth', *Quarterly Journal of Economics*, 110: 681–712.
- McMillan, M. (1997), 'A Dynamic Theory of Primary Export Taxation: Evidence from Sub-Saharan Africa', Working Paper, New York: Columbia University.
- Meadows, D.H., D. Meadows, J. Randers and W. Behrens (1972), *The Limits to Growth*, New York: University Books.
- Myrdal, G. (1957), *Economic Theory and Underdeveloped Regions*, London: Methuen.
- Neary, P.J. and S.N. van Wijnbergen (1986), *Natural Resources and the Macro Economy*, Cambridge, MA: MIT Press.
- Nordhaus, W.D. (1992), 'Is Growth Sustainable? Reflections on the Concept of Sustainable Growth', mimeo, paper prepared for the International Economic Association, Varenna.
- North, D.C. (1955), 'Location theory and regional economic development', *Journal of Political Economy*, 63: 248–58.
- Osei, R. (2001), 'A growth collapse with diffuse resources', in R.M. Auty (ed.), *Resource Abundance and Economic Development*, Oxford: Oxford University Press, pp. 165–78.
- Pearce, D., K. Hamilton and G. Atkinson (1996), 'Measuring Sustainable Development: Progress on Indicators', *Environment and Development Economics*, 1: 85–101.
- Prebisch, R. (1964), *Proceedings of the United Nations Conference on Trade and Development: Towards a New Trade Policy for Development*, Vols 1–8, New York: United Nations.
- Rodenbeck, M. (2002), 'Time Travelers: A Survey of the Gulf', *The Economist*, 23 March, centre supplement.
- Ross, M. (2001), 'Does Oil Hinder Democracy?', *World Politics*, 53 (3): 325–61.
- Rostow, W.W. (1956), 'The Take-Off into Self-Sustained Growth', *Economic Journal*, 66: 25–48.
- Sachs, J.D. (1999), 'Resource Endowments and the Real Exchange Rate: A Comparison of Latin America and East Asia', in T. Ito and A.O. Krueger (eds), *Changes in Exchange Rates in Rapidly Developing Countries*, Chicago IL: University of Chicago Press, pp. 133–53.
- Shleifer, A. and R. Vishny (1993), 'Corruption and Economic Growth', *Quarterly Journal of Economics*, 108: 599–617.
- Solow, R.M. (1974), 'Intergenerational Equity and Exhaustible Resources', *Review of Economic Studies*, 41: 29–45.
- Syrquin, M. (1986), 'Productivity Growth and Factor Reallocation', in H. Chenery, S. Robinson and M. Syrquin (eds), *Industrialization and Growth: A Comparative Study*, New York: Oxford University Press, pp. 228–62.
- Timmer, C.T. (2004), 'Operationalizing Pro-Poor Growth: Indonesia', Washington, DC: World Bank.

- Watkins, M. (1963), 'A Staple Theory of Economic Growth', *Canadian Journal of Economics and Political Science*, 29: 141–58.
- Woolcock, M., J. Isham and L. Pritchett (2001), 'The Social Foundations of Poor Economic Growth in Resource-Rich Countries', in R.M. Auty (ed.), *Resource Abundance and Economic Development*, Oxford: Oxford University Press, pp. 76–92.
- World Bank (2004), *Economic Development Indicators 2004*, Washington, DC: World Bank.

28 The environment and development

John McPeak

Introduction

There are a variety of reasons for an explicit focus on the environment in development economics. At a conceptual level, development and environmental change are linked dynamic processes where decisions made today influence the choices that are available to us in the future. As economic and natural systems jointly evolve, there is a need for linked analysis of these two systems. At a practical level, it is clear that the state of the environment plays a critical role in the process of development. One aspect of this is that many residents of developing countries are critically reliant on the natural resource base, reflected for example in the fact that over half of the labor force in developing countries is involved in agriculture.¹ Another aspect of this recognizes the distinct environmental challenges posed by rapid urbanization in developing countries (World Bank, 2003). Finally, sudden environmental shocks and long-term environmental change can both impact upon the prospects for development (World Bank, 2006).

In this chapter, the environment is viewed as consisting of both natural resource stocks and flows of goods and services from these stocks. Examples of natural resource stocks are things like arable soil, clean water and clean air. Examples of flows of goods and services from these natural resource stocks are: crop production and soil formation, waste assimilation and improved human health (see Costanza et al., 1997 for a more detailed list). Efforts to improve human welfare impact, positively or negatively, upon the environment as just defined when they lead to changes in either the stock of natural capital or the ability of natural capital to provide a flow of goods and services. Likewise, changes in these environmental stocks or flows can impact upon the prospects of development as defined by improving human welfare. This chapter is organized into what I see as the six major themes in the current literature on environment and development.²

The relationship between economic growth and environmental change

One important topic in the literature on environment–development interactions considers sustainable development. The Bruntland Commission Report (WCED, 1987), in what is probably the most commonly cited definition of sustainable development, defines the concept as: ‘development that meets the needs of the present without compromising the ability

of future generations to meet their own needs' (p. 43). That said, there is a vast spectrum of alternative definitions of this concept. Pezzy, in a 1989 paper on sustainable development, was able to find enough definitions to fill an eight-page appendix with alternative definitions, and the list has only become longer since then.

Pezzy and Toman (2002) provide a review of the journal articles relating to sustainability. They trace the origins of the concept to a set of papers by Dasgupta and Heal (1974), Stiglitz (1974), and Solow (1974) that were written in response to concerns raised in Meadows et al.'s 1972 book *Limits to Growth*. These papers incorporated the use of natural resource stocks into models based on standard economic growth theory and derived implications for welfare and resource trajectories using present value (PV) optimality as the efficiency criterion. Pezzy and Toman summarize these models in a general mathematical form based on maximizing the utility of consumption over time. Consumption is specified as $C_t = K_t^\alpha R_t^\beta e^{\tau t} - \dot{K} - \delta K_t - \xi R_t$, where t denotes the time period, C is consumption flow, K is (human-made) capital stock, R is the rate of natural resource change defined as a function of the resource stock S according to the equation $R = -S + G(S)$ with $G(\cdot)$ representing renewable resource growth, τ is exogenous technical progress, δ is the rate of capital depreciation, and ξ is the per unit cost of resource extraction (pp. 5–6). Note that this specification assumes K and R are substitutes in the production of the consumption flow, and the inclusion of R in the production function for consumption differs from the basic 'Solow' growth model.

Pezzy and Toman report that a 'key finding from Dasgupta and Heal's 1974 analysis was that the PV-optimal outcome is grim for far-distant generations' (2002, p. 6). The practice of discounting future values through use of a discount rate was a key factor that called the application of conventional economic tools into question when considering sustainable development. Stiglitz identified the potential for technological progress as one possible way of resolving this problem and Solow suggested that adding other constraints into the model that captured issues of fairness and a moral obligation to future generations could be considered. This led to a series of studies that investigated the interplay between intergenerational rights and equity, dynamic efficiency and sustainability (Howarth and Norgaard, 1990; Stavins et al., 2002 as examples), with many new insights coming from adopting overlapping-generations dynamic models.

A key issue identified by Pezzy and Toman (2002) is the distinction between 'weak sustainability' and 'strong sustainability' in the literature. The difference revolves around the question of how the analyst views substitution between human-made capital and natural resources. If it is unlimited, as in the consumption flow equation specified above, the analyst is

considering what can be called weak sustainability. When such substitution is viewed as very limited or not allowed at all, such as when the analyst argues that there are fundamental differences in the characteristics of human-made capital and natural resources, leading them to be complements rather than substitutes, the analyst is considering what can be called strong sustainability. From the point of view of weak sustainability it is acceptable to return to future generations manufactured capital of a given value in place of a commensurate level of natural capital. Strong sustainability, as exemplified by Daly (1990), does not find this acceptable and places emphasis on maintaining critical levels of natural resource stocks to pass on to future generations.

The sustainability literature is related to the literature on 'green national accounting' (Repetto et al., 1989) as seen in Pearce and Atkinson's 1993 effort to estimate empirically what countries were meeting a 'weak sustainability' requirement that the savings rate has to be greater than the sum of the physical capital and the natural capital depreciation rates to be deemed 'sustainable'. Debate, however, continues about how to value natural capital (Nordhaus and Kokkelenberg, 1999). Recently there have been efforts to clarify the distinction between the value of natural resource commodities and the value of ecosystem services. Costanza et al. (1997) estimate the annual value of ecosystem services and find it to be almost twice as large as annual global gross national product (GNP). Boyd and Banzhaf (2006) call for increased precision of the definition of 'ecosystem services'. They develop a definition of what is and is not to be included in a measure of what they call environmental service units.³ The valuation issue is part of an important and ongoing debate. Krautkraemer (2005) argues that the evidence suggests that technological progress has ameliorated problems arising from increased scarcity of resource commodities, but there is reason to be concerned that technological progress is not sufficient to address the growing problem of increasingly scarce resource amenities.

Another major topic in the literature on growth – environment interactions – focuses on what has been termed the 'environmental Kuznets curve' (World Bank, 1992; Grossman and Kreuger, 1995; special issue of *Environment and Development Economics*, 1997, special issue of *Ecological Economics*, 1998 to name but a few). The core idea is based on the original Kuznets curve that predicted an inverted 'U' relationship between a measure of average national income, such as gross national income (GNI) per capita, and the degree of income inequality within the nation as measured by a Gini coefficient. The argument is adapted to predict that environmental harm will be minimal at low levels of income, will increase over the middle range, and will decline as incomes reach higher levels. The explanation for the eventual decline in environmental 'bads' is a combination of

a demand-side story where environmental quality is a luxury good that we demand as incomes increase, and a supply-side story that focuses on the higher capacity to pay for effective regulatory institutions as incomes increase (Gangadharan and Valenzuela, 2001; Dasgupta et al., 2002). One message such a relationship would send is that developing countries should '[g]row first, then clean up' (Dasgupta et al., 2002, p. 147).

Given the shape of the curve, many of the empirical studies on this issue have focused on identifying the level at which the turning point of the Kuznets curve is reached. The use of cross sectional data, capturing variation across space at a given point in time, to estimate a relationship that hypothesized to take place for a given space over time is problematic, as is well known from the literature on the original Kuznets Curve. However, panel data are increasingly available as a means to address this problem. The empirical evidence is decidedly mixed concerning the existence of the proposed relationship and it is not always consistent in the turning point identified. In addition, debate about the methodology used in empirical studies continues. It appears at this point that the findings consistent with the environmental Kuznets curve hold for a subset of environmental issues, such as SO₂ and particulate matter, but do not hold for others, such as energy use or municipal waste generation per capita.

An alternative perspective on the relationship between growth and development is found in the literature on the 'poverty–environment nexus' (Duraiappah, 1998, Dasgupta et al., 2005). Poverty and degradation are seen as linked processes. Poverty leads people to degrade the environment (cultivation on highly erosive hillsides, extracting resources such as fish or fuelwood from commons at a harmful rate, unhygienic waste disposal in urban slums), and degraded environments impact upon the poor the most severely (declining revenue due to erosion or overexploitation, health impacts from pollution) thus exacerbating their poverty. As argued in the influential *World Development Report* of 1992, 'economic development and sound environmental management are complementary aspects of the same agenda' (World Bank, 1992, p. 25). Arguing that one must choose one or the other is a 'false dichotomy'. As the World Bank became more focused on poverty reduction as an objective of development (see the *World Development Report 2000/2001*, World Bank, 2000/2001), research increasingly considered how poverty reduction and positive environmental change could be mutually compatible outcomes. One particularly attractive aspect of this work is that it identifies when multiple desirable objectives can be mutually compatible (reduce poverty, prevent environmental degradation) or when there may be trade-offs (increase agricultural production, decrease water quality) – a quest well captured by the title of and many of the contributions in Lee and Barrett's 2001 volume *Tradeoffs or Synergies*. In

addition, there are efforts to add additional objectives such as improving social capital, local participation or some other measure of the well-being of a collection of individuals. The questions then revolve around what policy or combination of policies, if any, will lead to 'win-win-win' outcomes that increase economic welfare, environmental conditions and social cohesion. And if no policy or set of policies allow such an outcome, what is the nature and magnitude of the trade-offs facing decision-makers given the policy options available?

A growing literature on economic growth and environmental change attempts to answer these questions by modeling dynamic interactions between economic and environmental systems when there are feedback effects. This approach is usually taken when analyzing an identifiable ecological system, such as a watershed or a forest-farmland zone. These studies use bioeconomic modeling, where the dynamics of a biophysical system and the dynamics of human production and consumption activities are jointly modeled (special section of *Environment and Development Economics*, 2005; Fisher et al., 2005; Grafton et al., 2005; Brown, 2000, to note but a few). These studies often attempt to identify the implications of how human behavior will interact with the natural system in terms of long-run human welfare and environmental condition in response to a given policy. Such studies harness rapidly increasing computing power and improved modeling methodology to embrace greater nuance and complexity in both human and natural systems and their evolution over time.

Agriculture and rural issues

Another theme in the literature on the environment and development focuses on the critical role of agricultural intensification. As noted above, agriculture remains a major source of employment in developing countries, thus making agricultural intensification a key component of the development process. In addition, concerns about population growth outstripping food production have been expressed from the time of Malthus to more recent publications such as Brown and Kane's *Full House* (1994). Further, concerns about the environmental impact of 'Green Revolution' technologies that led to rapid increases in agricultural yields in the past have been expressed. Finally, agricultural intensification is viewed as needed as demand for agricultural products will inevitably grow in the future both in response to increases in population and incomes and there are limits to the amount of land available for cultivation.

In a policy forum in *Environment and Development Economics*, Barrett et al. (2001) identify three distinct technological approaches that can be taken to allow agricultural intensification. The first is renewed effort to sustain the 'Green Revolution' methods that were used to increase

agricultural yields in the past through increased use of irrigation and inorganic fertilizer; the second is built around agroecological methods such as intercropping and agroforestry; and the third focuses on genetic modifications to cultivars that increase productivity and resistance to pests and disease. Whether the impact of a particular approach or combination of approaches on human welfare and the environment will be positive or negative is debated, as evident by comparing the essays in the policy forum.

Lee et al. (2006) identify four main issues to consider surrounding agricultural intensification in developing economies when the goal is to improve human welfare without causing environmental damage. One is the role of labor in the intensification process. This focuses on the supply and demand for labor, the nature of the labor constraints, and whether labor is a substitute or a complement for other inputs used in production. A second issue is whether the proposed intensification can be implemented without imposing an environmental cost, or even in such a way so as to lead to positive environmental outcomes. A third issue focuses on the processes of adoption and innovation: what determines when innovation will occur and what influences adoption of the products or techniques that were created by this innovation? A fourth issue revolves around the policy framework that will lead to the most positive outcome in terms of both improved human welfare and environmental quality. What can be done by decision-makers to make agricultural intensification sustainable?

Urban issues

While it is true that agriculture remains critically important in the development process, we also need to recognize that an increasing share of the world's population lives in urban areas, and developing countries are no exception to this trend. The share of the population of low- and middle-income countries living in urban areas has increased from 24 percent of the total in 1960 to 43 percent of the total in 2004.⁴ Todaro and Smith (2006) estimate that 'urban centers of the developing world will absorb over 80% of future increases in world population' (p. 487). This would indicate that urban environmental issues will become increasingly prominent in the analysis of environment and development interactions.

The United Nations Environment Programme's Division of Technology, Industry, and Economics⁵ lists priority urban environmental problems concerning water supply, sewage, solid waste, energy, loss of green and natural spaces, urban sprawl, land contamination, traffic, transport, air pollution and noise as key areas where urban growth can pose environmental problems. UN-HABITAT in its 2005 annual report describes how environmental action plans can be designed by residents of urban areas in developing countries to deal with the problems associated with rapid growth.

As a growing share of the population in developing countries inhabits urban areas, it will be increasingly important to address these environmental issues, particularly as a failure to address them will have severely negative consequences for the health of the urban population, and impose negative externalities on those who may not be urban residents as well. One hopes, given concerted policy effort, it will be possible that these problems can be addressed in ways which make the urbanization process a contributor rather than a detriment to the overall development effort (Hardoy et al., 1992; World Bank, 2003, Chapter 6).

Non-market aspects

As noted above, there are issues related to the valuation of environmental goods and services that need to be considered in the context of development. As many environmental goods and services never enter the market, special techniques must be used to arrive at estimates of these values.

One theme in this literature focuses on the application of the toolkit of environmental economics to issues in developing countries. These techniques are particularly useful in settings where the environmental good or service to be valued is a public good; that is one that is characterized by non-rivalry and non-excludability. The UN Food and Agriculture Organization (FAO) (2000) published an overview of applications of the contingent valuation method in developing countries, which provides a summary of 20 studies that have used this method. Almost all the studies discussed have a focus on environmental issues, with half focusing on water quality issues. A different method used in the literature is the travel cost method. With this approach, the implicit price of a good is revealed by the implicit expenditure that is revealed by calculating the cost of obtaining the good (Choe et al., 1996; Hegan et al., 2003). A third element of the environmental economist's toolkit that is applied to developing-country topics is the hedonic method. This method takes an observed market price and decomposes this price by regressing it on identifiable characteristics of the good. This leads to an implicit price of a non-market good that is embedded in the marketed good (Humavindu and Stage, 2003; special issue of *Ecological Economics*, 2003). A final method is to value an environmental good or service through a measure of averting expenditures. In these studies, the implicit value of a good or service is represented by the expenditure on a substitute for this good or service (McConnell and Rosado, 2000).

A different aspect of the non-market question arises when we consider programs to pay residents of developing countries for the provision or protection of environmental goods and services. Zbinden and Lee (2004) present a discussion of a program in Costa Rica that pays farmers and

forest owners for reforestation, forest conservation and sustainable forest management activities. Pagiola et al. (2005) consider the potential impact of payments for environmental services on the objective of poverty reduction. There is also a growing interest in payments to residents of developing countries for conservation that leads to carbon sequestration in response to concerns about global warming. These payment mechanisms have arisen as a result of the Kyoto Protocol (Olszewski and Benitez, 2003), as it is hoped that such payments will play a role in addressing global climate change.⁶ Intriguing issues arise as such payments are towards a global public good, bringing up issues of bilateral and multilateral development aid and poverty reduction in developing countries.

Finally, there has been a great deal of effort at developing integrated conservation–development projects. One of the best known and documented examples is Zimbabwe’s CAMPFIRE project (Fisher et al., 2005; Alexander and McGregor, 2000 as examples). These types of projects attempt to offer populations living in or near particular natural resources that have value but do not enter the market (wild animals in the case of CAMPFIRE) an economic incentive to protect these resources by providing them with a share of the benefits generated by the resources and/or income-generating opportunities that do not negatively impact upon the resource. Barrett and Arcese (1995) discuss some of the problematic issues that arise when designing such projects.

Common property

Distinct issues arise when considering open access or common property regimes. In such settings, consumption of a given good is rivalrous, yet excludability is partial (common property) or not possible (open access). Examples in the literature have predominantly focused on fisheries, forests, rangelands and water resources. The fundamental issue is that users of the resource make decisions that may impose a cost on other users, but may not take this cost into account when making the decision.⁷ Perhaps the most widely known version of this argument is Hardin’s (1968) ‘Tragedy of the Commons’.

A variety of authors have taken issue with the basic argument of the tragedy of the commons proposed by Hardin. Baland and Platteau (1996) link issues of common property management to local management, and identify determinants of success or failure of common property management regimes. Bromley (1992) provides a useful collection of case studies that illustrate how different societies have addressed issues of common property management. Ostrom (1990) also devotes considerable attention to case studies, and also explore issues of common property management using experimental economics (Ostrom et al., 1994). Taken together, these

studies suggest that there is a social and political context within which commonly held resources are used that must be considered when analyzing such systems.

Some other themes emerge from recent research on common property systems. One, as illustrated by Nugent and Sanchez (1998), is that common property systems may serve as a risk-minimizing form of property ownership. The case they consider is common property management of rangelands in Sudan as a response to spatially variable rainfall. A second theme is the interaction between conflict management and common property management (Haro et al., 2005; Castro and Nielsen, 2003). Another notable theme in the literature focuses on patterns of exploitation in a spatial heterogeneous commons (McPeak, 2003; Sanchirico and Wilen, 1999). Heterogeneity of users in terms of inequality in assets or income and use of the commons is an additional area of active research (Narain et al., 2005; Baland and Platteau, 1996; Jodha, 1986).

Prediction of and vulnerability to environmental shocks and change

An emerging theme in the literature is the role of severe shocks and uncertain change on the environment and prospects for development. One focus of this research considers *ex ante* aspects of managing shocks that are environmentally driven. Perhaps most prominent in this literature are attempts to grapple with the implications of climate change for developing countries (see for example the special issue of *Disasters*, 2006, devoted to the overall implications of climate change and natural disasters). As noted by a recent World Bank study (2006), natural resources in the form of agricultural land and pastureland make up a large share of developing-country wealth, and this wealth is threatened by climate change. These studies suggest we need to plan now to ensure we have minimized the impact of the risk of climate change. In addition, as noted above in the section on paying the poor for environmental services, there are questions about who should be compensated for the damages associated with climate change and who should be paying this compensation.

A related literature focuses on the information dissemination aspect of environmentally driven shocks. How do we best deliver information about expected shocks to residents of developing countries and how will they use this information in formulating their own plans (Lybbert et al., 2007; Luseno et al., 2003; Phillips et al., 2002; Ingram et al., 2002)? These studies focus on how information about potential climate-driven risks are processed, and investigate the role that better information delivery systems can have on efforts to manage risk exposure.

Another theme in this literature investigates the *ex post* impact of natural disasters on income and assets, and how individuals manage such risk.

Households can find the assets they have accumulated over many years suddenly swept away by environmental events such as hurricanes, tsunamis, droughts, floods, earthquakes, mudslides or other related phenomena. McPeak (2004) contrasts short-term shocks to income with long-term shocks to assets in livestock economies as a result of a drought. Carter et al. (2004) investigate the long-term impact of environmental disasters on households in Ethiopia and Honduras, finding that shocks can lead to long-term poverty traps.

Concern over the humanitarian impact of environmentally driven shocks appears to be increasing due to recent vivid events such as Hurricane Katrina and the Asian tsunami, combined with growing concerns about global warming. One would expect that research on predicting, planning, information delivery and coping with environmentally driven shocks and change in developing countries will continue to grow.

Notes

1. Data from FAOSTAT available from www.fao.org, accessed by the author May 2006.
2. In the end, I am sure there are some issues and contributions I have overlooked in my attempt to summarize this huge and growing field of environmental and development economics, but I hope to have arrived at a reasonably comprehensive assessment of the main themes in the literature.
3. Many of the valuation issues involved here will involve some of the non-market valuation methods outlined later in this chapter, as most environmental services occur outside markets.
4. Figures obtained from the World Development Indicators Online of the World Bank, accessed May 2006 by the author.
5. <http://www.unep.or.jp/ietc/Issues/Urban.asp>.
6. See for example <http://earthtrends.wri.org/features/index.php?theme=3>.
7. Ostrom (1990) provides a useful distinction between the types of externalities that can be imposed. One type of externality takes place in the current period. She calls this an appropriation externality, and it reflects the rival nature of consumption – only one user can consume the resource to which access is shared. The second type of externality is a dynamic one. In this setting, decisions made today impose a cost on the future provision of the resource, thus they are called provision externalities.

References

- Alexander, J. and J. McGregor (2000), 'Wildlife and Politics: CAMPFIRE in Zimbabwe', *Development and Change*, 31 (3): 605–27.
- Baland, J.-M. and J.-P. Platteau (1996), *Halting Degradation of Natural Resources: Is there a Role for Rural Communities?*, New York: Oxford University Press.
- Baland, J.-M. and J.-P. Platteau (1999), 'The Ambiguous Impact of Inequality on Local Resource Management', *World Development*, 27 (5): 773–88.
- Barrett, C. and P. Arcese (1995), 'Are Integrated Conservation–Development Projects (ICDPs) Sustainable? On the Conservation of Large Mammals in Sub-Saharan Africa', *World Development*, 23 (7): 1073–84.
- Barrett, C.B., E. Barbier and T. Reardon (2001), 'Introduction: Farming the Garden of Eden', *Environment and Development Economics*, 6: 503–31.
- Boyd, J. and S. Banzhaf (2006), 'What Are Ecosystem Services? The Need for Standardized Environmental Accounting Units', RFF Discussion Paper 06-02, Resources for the Future, Washington, DC.

- Bromley, D. (ed.) (1992), *Making the Commons Work: Theory, Practice, and Policy*, San Francisco, CA: ICGS Press.
- Brown, D. (2000), 'A Review of Bio-Economic Models', paper prepared for the Cornell African Food Security and Natural Resource Management (AFSNRM) Program, Cornell University, Ithaca, NY.
- Brown, L. and H. Kane (1994), *Full House: Reassessing the Earth's Population Carrying Capacity*, New York: W.W. Norton & Company.
- Carter, M., Peter D. Little, Tewodaj Mogues and Workneh Negatu (2004), 'Shocks, Sensitivity and Resilience: Tracking the Economic Impacts of Environmental Disaster on Assets in Ethiopia and Honduras', BASIS-CRSP working paper, University of Wisconsin, Madison.
- Castro, P. and E. Nielsen (eds) (2003), *Natural Resource Conflict Management Case Studies: An Analysis of Power, Participation, and Protected Areas*, Rome: FAO.
- Choe, KyeongAe, Dale Whittington and Donald T. Lauria (1996), 'The Economic Benefits of Surface Water Quality Improvements in Developing Countries: A Case Study of Davao, Philippines', *Land Economics*, 72 (4): 519–37.
- Costanza, Robert, Ralph D'Arge, Rudolf de Groot, Stephen Farber, Monica Grasso, Bruce Hannon, Karin Limburg, Shahid Naeem, Robert V. O'Neill, Jose Paruelo, Robert G. Raskin, Paul Sutton and Marjan van den Belt (1997), 'The Value of the World's Ecosystem Services and Natural Capital', *Nature*, 387: 253–60.
- Daly, H. (1990), 'Toward Some Operational Principles of Sustainable Development', *Ecological Economics*, 2 (1): 1–6.
- Dasgupta, P. and G.M. Heal (1974), 'The Optimal Depletion of Exhaustible Resources', *Review of Economic Studies*, 41: Symposium on the Economics of Exhaustible Resources, 3–28.
- Dasgupta, S., Benoit Laplante, Hua Want and David Wheeler (2002), 'Confronting the Environmental Kuznets Curve', *Journal of Environmental Perspectives*, 16 (1): 147–68.
- Dasgupta, S., Uwe Deichmann, Craig Meisner and David Wheeler (2005), 'Where is the Poverty–Environment Nexus? Evidence from Cambodia, Lao PDR, and Vietnam', *World Development*, 33 (4): 617–38.
- Disasters* (2006), Special Issue on Climate-Related Disasters, 30 (1).
- Duraiappah, A. (1998), 'Poverty and Environmental Degradation: A Review and Analysis of the Nexus', *World Development*, 26 (12): 2169–79.
- Ecological Economics* (1998), The Environmental Kuznets Curve, 25 (2).
- Ecological Economics* (2003), Special Issue on Animal Genetic Resources, 45 (3).
- Environment and Development Economics* (1997), Special Issue on the Environmental Kuznets Curve, 2 (4).
- Environment and Development Economics* (2001), Policy Forum, 6 (4): 503–31.
- Environment and Development Economics* (2005), Special section on the Dynamics of Coupled Human and Natural Systems, 11 (1): 9–75.
- FAO (2000), 'Applications of the Contingent Valuation Method in Developing Countries', FAO Document Repository, www.fao.org.
- Fisher, C. et al. (2005), 'Bioeconomic Model of Community Incentives for Wildlife Management Before and After CAMPFIRE', RFF Discussion Paper 05-06, Resources for the Future, Washington, DC.
- Gangadharan, L. and M.R. Valenzuela (2001), 'Interrelationships between Income, Health and the Environment: Extending the Environmental Kuznets Curve Hypothesis', *Ecological Economics*, 36: 513–31.
- Grafton, R. Quentin, Tom Kompas and Viktoria Schneider (2005), 'The Bioeconomics of Marine Reserves: A Selected Review with Policy Implications', *Journal of Bioeconomics*, 7: 161–78.
- Grossman and Kreuger (1995), 'Economic Growth and the Environment', *Quarterly Journal of Economics*, 112: 353–77.
- Hardin (1968), 'The Tragedy of the Commons', *Science*, 162: 1243–8.
- Hardoy, Jorge E., Diana Mitlin and David Satterthwaite (1992), *Environmental Problems in Third World Cities*, London: Earthscan.

- Haro, Guyo O., Godana J. Doyo and John G. McPeak (2005), 'Linkages Between Community, Environmental, and Conflict Management: Experiences from Northern Kenya', *World Development*, 33 (2): 285–99.
- Hegan, R. et al. (2003), 'Is the Tragedy of the Commons Likely? Factors Preventing the Dissipation of Fuelwood Rents in Zimbabwe', *Land Economics*, 79 (2): 181–97.
- Howarth, R. and R. Norgaard (1990), 'Intergenerational Resource Rights, Efficiency, and Social Optimality', *Land Economics*, 66 (1): 1–11.
- Humavindua, Michael and Jesper Stage (2003), 'Hedonic Pricing in Windhoek Townships', *Environment and Development Economics*, 8 (2): 391–404.
- Ingram, K., M.C. Roncoli and P.H. Kirshen (2002), 'Opportunities and Constraints for Farmers of West Africa to Use Seasonal Precipitation Forecasts with Burkina Faso as a Case Study', *Agricultural Systems*, 74: 331–49.
- Jodha, N.S. (1986), 'Common Property Resources and Rural Poor in Dry Regions of India', *Economic and Political Weekly*, 21 (27): 1169–82.
- Krautkraemer, J. (2005), 'Economics of Natural Resource Scarcity: The State of the Debate', RFF Discussion Paper 05-14, Resources for the Future, Washington, DC.
- Lee, D.R. and C.B. Barrett (2001), *Tradeoffs or Synergies? Agricultural Intensification, Economic Development, and the Environment*, New York: CABI Publishing.
- Lee, D.R., C.B. Barrett and J.G. McPeak (2006), 'Policy, Technology and Management Strategies for Achieving Sustainable Agricultural Intensification', *Agricultural Economics*, 34: 123–7.
- Luseno, W. et al. (2003), 'Assessing the Value of Climate Forecast Information for Pastoralists: Evidence from Southern Ethiopia and Northern Kenya', *World Development*, 31 (9): 1477–94.
- Lybbert, T., C.B. Barrett, J.G. McPeak and W.K. Luseno (2007), 'Bayesian Herders: Updating of Rainfall Beliefs in Response To External Forecasts', *World Development*, 35 (3): 480–97.
- McConnell, K. and M. Rosado (2000), 'Valuing Discrete Improvements in Drinking Water Quality through Revealed Preference', *Water Resources Research*, 36 (6): 1575–82.
- McPeak, J. (2003), 'Analyzing and Addressing Localized Degradation in the Commons', *Land Economics*, 79 (4): 515–36.
- McPeak, J. (2004), 'Contrasting Income Shocks with Asset Shocks: Livestock Sales in Northern Kenya', *Oxford Economic Papers*, 56: 263–84.
- Meadows, D., J. Randers, D.L. Meadows and W.W. Behrens (1972), *The Limits to Growth: A Report for the Club of Rome's Project on the Predicament of Mankind*, New York: Universe Books.
- Narain, U., S. Gupta and K. van 't Veld (2005), 'Poverty and the Environment: Exploring the Relationship between Household Incomes, Private Assets, and Natural Assets', RFF Discussion Paper 05-18, Resources for the Future, Washington, DC.
- Nordhaus, W. and E. Kokkelenberg (eds) (1999), *Natures Numbers: Expanding the National Economic Accounts to Include the Environment*, Washington, DC: National Academy Press.
- Nugent, J. and N. Sanchez (1998), 'Common Property Rights as an Endogenous Response to Risk', *American Journal of Agricultural Economics*, 80 (3): 651–57.
- Olshewski, R. and P.C. Benitez (2003), 'Secondary Forests as Temporary Carbon Sinks? The Economic Impact of Accounting Methods on Reforestation Projects in the Tropics', *Ecological Economics*, 55: 380–94.
- Ostrom, E. (1990), *Governing the Commons: The Evolution of Institutions for Collective Action*, Cambridge: Cambridge University Press.
- Ostrom, E., R. Gardner and James Walker (1994), *Rules, Games, and Common-Pool Resources*, Ann Arbor, MI: University of Michigan Press.
- Pagiola, S., A. Arcenas and G. Platais (2005), 'Can Payments for Environmental Services Help Reduce Poverty? An Exploration of the Issues and Evidence to Date from Latin America', *World Development*, 33 (2): 237–53.
- Pearce, D. and G. Atkinson (1993), 'Capital Theory and the Measure of Sustainable Development: An Indicator of "Weak" Sustainability', *Ecological Economics*, 8 (2): 103–8.
- Pezzy, J. (1989), 'Economic Analysis of Sustainable Growth and Sustainable Development', World Bank Environment Department Working Paper 15, Washington, DC.

- Pezzy, J. and M. Toman (2002), 'The Economics of Sustainability: A Review of Journal Articles', RFF Discussion Paper 02-03, Resources for the Future, Washington, DC.
- Phillips, J.G., D. Deane, L. Uganai and A. Chimeli (2002), 'Implications of Farm-Level Response to Seasonal Climate Forecasts for Aggregate Grain Production in Zimbabwe', *Agricultural Systems*, 74 (3): 351–69.
- Repetto, R. et al. (1989), *Wasting Assets: Natural Resources and the National Income Accounts*, Washington, DC: World Resources Institute.
- Sanchirico, J. and J. Wilen (1999), 'Bioeconomics of Spatial Exploitation in a Patchy Environment', *Journal of Environmental Economics and Management*, 37 (2): 129–50.
- Solow, R. (1974), 'Intergenerational Equity and Exhaustible Resources', *Review of Economic Studies*, 41: Symposium on the Economics of Exhaustible Resources, 29–45.
- Stavins, R., A.F. Wagner and G. Wagner (2002), 'Interpreting Sustainability in Economic Terms: Dynamic Efficiency Plus Intergenerational Equity', RFF Discussion Paper 02-29, Resources for the Future, Washington, DC.
- Stiglitz, J. (1974), 'Growth with Exhaustible Natural Resources: Efficient and Optimal Growth Paths', *Review of Economic Studies*, 41: Symposium on the Economics of Exhaustible Resources, 123–37.
- Todaro, M. and S. Smith (2006), *Economic Development*, 9th edn, New York: Pearson, Addison Wesley.
- UN-HABITAT (2005), *Annual Report*, Nairobi, Kenya: United Nations Human Settlement Programme.
- World Commission on Environment and Development (WCED) (1987), *Our Common Future*, Oxford: Oxford University Press.
- World Bank (1992), *World Development Report 1992, Development and the Environment*, Washington, DC: World Bank.
- World Bank (2000/2001), *World Development Report 2000/2001, Attacking Poverty*, Washington, DC: World Bank.
- World Bank (2003), *World Development Report 2003, Sustainable Development in a Dynamic World*, Washington, DC: World Bank.
- World Bank (2006), *Where is the Wealth of Nations*, Washington, DC: World Bank.
- Zbinden, S. and D.R. Lee (2004), 'Paying for Environmental Services: An Analysis of Participation in Costa Rica's PSA Program', *World Development*, 33 (2): 255–72.

29 Technical choice and technological change in development

Howard Pack

Technological issues entered the analysis of developing countries relatively early, primarily in response to the perceived need to create employment in the industrial sector. Early analysis was concerned with whether the elasticity of substitution between unskilled labor and capital was sufficiently great to generate adequate growth in employment opportunities in urban areas. A considerable empirical literature investigated this question, relying on the estimation of production functions but also on engineering analyses and visits to firms that allowed insights that were not obtainable solely from econometric investigation. Soon afterwards technology in the agricultural sector became an important area of research largely in response to the effects of the Green Revolution that increased yields several-fold.

By the mid-1980s attention switched from static factor choice to total factor productivity (TFP) levels and rates of growth (TFPG). Correct choice of technology that was not accompanied by high levels of productivity would not contribute to high levels of income per capita. And low rates of TFPG could imply that the cost advantage stemming from the initial correct choice of factor proportions could be undermined by firms that adopted initially inappropriate technology but succeeded in improving TFP levels.

A large number of papers calculating economy-wide TFPG were motivated in part by the desire to establish whether specific development strategies such as import substituting industrialization (ISI) were in fact dynamically inferior to export orientation. The spectacular success of a small number of Asian countries generated another round of research in the 1980s and 1990s whose focus was on the relative contribution of capital accumulation and TFP growth to their exceptional growth. Measurement of TFPG has been a fraught topic and a number of controversies have arisen in its calculation and interpretation. There has been only a limited effort to go behind the estimates in order to understand its determinants at the firm level, although there have been numerous cross-country analyses attempting to identify the sources of differences in performance at the national level. Since 1996 a number of analyses, some using firm-level panel data, have investigated whether exports lead to greater productivity growth

than domestic sales. More recently the domain of technology issues has been extended to include the role of industrial clusters and the abilities of firms to deal with information technology and the technical requirements of large purchasers of final goods in the Organisation for Economic Co-operation and Development (OECD) nations.

The first section analyzes the static issues of technology choice and TFP levels. The second and third sections discuss the dynamic ones relating to productivity growth. A final section presents conclusions.

Static technology questions

Choice of technology

The classic article of W. Arthur Lewis (1954) that outlined a model whose main feature was the shift in the center of economic activity from agriculture to manufacturing implicitly set the technology research agenda for a considerable period. The absorption of job seekers leaving an overpopulated rural area to seek employment in urban areas depended on an outward shift in the demand for labor in the urban sector. Additional demand for labor depends on both the magnitude of investment and its technical characteristics, in particular whether the technology allows significant possibilities for labor-intensive production or whether the technology offers limited employment. A large literature demonstrated that the elasticity of substitution, σ , was high in most of the industries in which less-developed countries (LDCs) had a potential comparative advantage.¹ This implied that if relative factor prices reflected factor scarcities, expansion in the industrial sector would occur with a low capital–labor ratio and be accompanied by a growth of employment that would allow absorption of the influx of unskilled workers from the rural areas.

Given values of σ considerably in excess of zero, the focus of discussion then shifted to one of the earliest analyses in the development literature of political economy, namely, the political forces that underlay a continuing distortion of relative factor prices. In particular, it was argued that premature introduction of Western notions of minimum wages, ‘excessive’ power of unions, government limitations on the interest rate that could be charged by financial institutions, and an overvalued exchange rate that reduced the local currency cost of imported advanced equipment all militated in favor of capital-intensive technology choices. Many of these distortions reflected a political equilibrium, for example, employers and employees in sectors benefiting from protection against imports had sufficient power to prevent liberalization. One consequence of import-substituting industrialization was the artificial overvaluation of local currencies. Despite prompting by international financial institutions, bilateral aid agencies, and local

and foreign economists, the factor market distortions were only slowly removed. In the interim, there was a loss of both output and employment that was technically achievable. With growing trade and financial liberalization some of these distortions have been removed and the expansion of output according to comparative advantage should, on the margin, lead to a greater labor intensity in production.

Since the 1980s analyses of technology choice have largely disappeared from the literature as attention has shifted to the sources of productivity growth. This shift in emphasis occurred despite the continuing high rates of open and disguised unemployment in most developing countries, particularly those in sub-Saharan Africa and the Middle East. Yet even in China and India, which have experienced rapid growth in industrial output, the correct choice of technology remains important as large numbers of poorly educated rural labor force members continue to migrate to urban areas. In China, the absorption problem has been partly addressed by encouraging foreign direct investment (FDI) in special economic zones. Foreign firms have invested in sectors such as clothing and toy production that are intrinsically more labor-intensive and have chosen technologies that are appropriate to the factor prices they face. In addition construction, which is often more labor-intensive than manufacturing, has provided large numbers of jobs.

TFP levels

One source of low per capita income in developing countries is their low TFP levels for a given capital-labor ratio. A number of studies at the firm level have calculated relative TFP in LDCs with advanced country comparators. Using plant-level data Pack (1987) found that in Kenya and the Philippines most textile factories were characterized by low TFP levels relative to British firms using identical equipment. McKinsey & Company comparing similar firms in Korea and Brazil with those in the USA obtained similar results (McKinsey Global Institute, 1998a, 1998b). Relative TFP levels (A_{ldc}/A_{dc}) range from 0.4 to 0.6. The results for Korea, obtained in the late 1990s, are surprising insofar as Korea already had an advanced industrial sector. These results confirmed a study of Pilat (1994) that calculated TFP levels at the sectoral rather than the firm level for Korean manufacturing in 1988 and found that most sectors exhibited values of (A_{ldc}/A_{dc}) that were quite low. Not only is (A_{ldc}/A_{dc}) low, but so is the average level of efficiency in individual countries relative to the most efficient firms within the countries.²

There have been several economy-wide analyses of A_{ldc}/A_{dc} . Parente and Prescott (2000) explain differences in income per capita levels within a Solow neoclassical model in which the equilibrium level of income is:

$$y^* = A [s/(g + d)]^{\alpha/(1-\alpha)} \quad (29.1)$$

where y^* is the level of per capita income, A is the level of total factor productivity, s is the national saving rate, g the rate of growth of Harrod-neutral technical change, d is the rate of depreciation and α is the share of capital. All countries are assumed to be able to move along an *ex ante* common internationally available production function so that A and $\alpha/(1-\alpha)$ are identical. In their simulations, differences in saving rates do not explain much of the very large variations in per capita income levels, nor does an expansion of the equation to reflect differences in human capital, time devoted to education and intangible capital. Their explanation of the cross-country variation in y^* is that A varies enormously due to barriers imposed by governments to adopting the internationally available technology, often embodied in new machinery, and from opposition by key factor suppliers, particularly labor unions.

While obstacles to adoption undoubtedly have explanatory value, much of the development literature has been devoted to alternate though complementary explanations. The international production function may not be costlessly accessible. Access to technology occurs through the purchase of new equipment, licensing of existing production processes, and through the location of foreign firms (foreign direct investment – FDI). Each of these has a cost that may discourage LDC firms from moving towards a potentially available international best-practice production function. These costs include licensing charges, consultants' fees and possible excess profits earned by foreign firms. Moreover, successful identification and assimilation of newer technologies requires domestic absorptive capacity, particularly technically competent educated labor (Nelson and Phelps, 1966; Borenszstein et al., 1998). To take an example from the agricultural sector, Foster and Rosenzweig (1995) found that more educated farmers were earlier and more efficient adopters of the Green Revolution in India, a major TFP-increasing technological change given the relative importance of the sector in India.

Bad policies affect TFP levels (and growth rates). A huge literature investigates the harmful allocative effects of import-substituting industrialization which show up in reduced TFP levels as a result of the absence of competitive pressure and the resulting diminished incentive to undertake costly efforts to improve TFP. Short production runs for a purely domestic market also have adverse effects where scale economies are present. In agriculture, high taxes on export and low prices paid to farmers who must sell part of their output to government monopsonies have generated very large losses in output for a given commitment of land and labor. One measure of foregone production is the increase of 50 percent in net farm output in

China in five years after agricultural liberalization in the late 1970s, implying that any measure of the TFP levels in 1977 would have conflated the low state of agricultural technology with the effects of policy.

The main conclusion I draw from the literature is that TFP levels are indeed lower in LDCs than in the OECD countries. Partly, this is a technological issue reflecting the failure to take effective advantage of the large backlog of advanced technologies that are highly productive, many of which are appropriate for LDCs. To a major extent it also reflects the failure of LDCs to utilize their existing technologies effectively – even the best firms achieve lower TFP than peer firms in developed countries, and most firms within countries operate with considerably less efficiency than the frontier firms in the same country.

Measurement of growth in total factor productivity

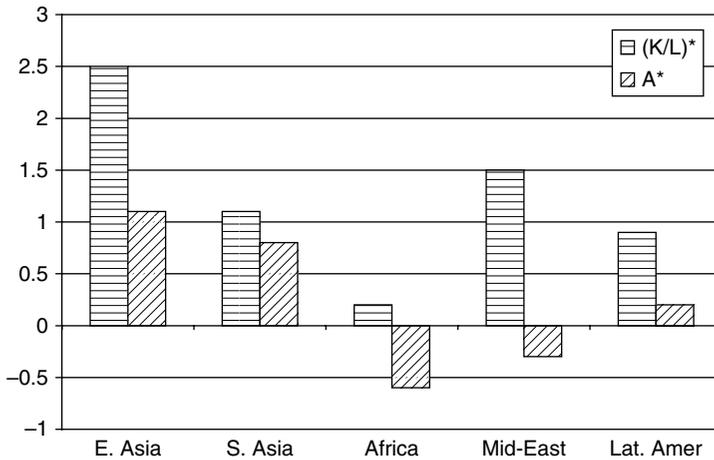
Since the late 1980s there has been growing attention to rates of TFP growth in LDCs. Part of this reflects a general interest in applying these techniques to developing countries, but much of this literature had a policy-driven interest, namely, analyzing the effectiveness of various development strategies, particularly import-substituting industrialization versus export promotion. A difficulty with this approach is that most of the studies have examined economy-wide data, whereas the specific strategy being tested was limited to manufacturing. Insofar as the aggregate data reflect the performance of other sectors including services, construction, agriculture and utilities, the various estimates do not provide a focused test of the strategy with respect to manufacturing unless it is assumed that there are significant spillover effects to other sectors.

The rate of growth of TFP assumed a particularly important place in analyses of the Asian Miracle (World Bank, 1993). The absence of unusual rates of TFP growth compared to countries in other regions (Young, 1992, 1995; Krugman, 1994) was understood to imply that capital accumulation was the critical source of growth and thus the important policy was the ability to suppress consumption.

Impact of development strategies

Two forms of formal analysis have been utilized to examine the contribution of factor accumulation and TFP to aggregate growth, namely, growth accounting and econometric estimation of production functions. Before examining these estimates I briefly examine the stylized facts.

Figure 29.1 shows growth rates for the capital–labor ratio and total factor productivity by regions for the period 1960–94 using data from Collins and Bosworth (1996).³ The East Asian group had higher rates of capital deepening and higher rates of TFP growth than other regions for



Source: Collins and Bosworth (1996).

Figure 29.1 Growth rates of capital-labor ratio and TFP, 1960–94

the period 1960–94. The study using consistent adjustments across countries suggests that there was something different about the ability of the export-oriented newly industrialized countries (NICs) of Asia to absorb capital productively. While their TFP growth rates were not ‘miraculous’ they were higher than those in other regions that had relied on import substituting industrialization (ISI). Similar results are obtained by Nehru and Dhareshwar (1994). The observed differences contribute to sorting out the ISI versus export debate, but at a deeper level the difference could be due to factors such as better macroeconomic management. The (relatively) high TFP growth rates may have prevented diminishing returns that could have caused a decline in the very high rates of saving in the NICs, the latter being a phenomenon for which no completely adequate account has been adumbrated (Deaton and Paxson, 1994).

Growth accounting

Growth accounting employs observed factor shares from the national accounts to estimate partial output elasticities.⁴ The change in the aggregate amount of inputs is calculated using the Tornqvist index:

$$T = \sum_i [1/2(S_{i,t} + S_{i,t-1}) (\ln x_{i,t} - \ln x_{i,t-1})] \quad (29.2)$$

where $S_{i,t}$ is the observed share of factor x_i in period t . This is subtracted from the log difference in output to obtain TFP growth:

$$A^* = \log(Y_t/Y_{t-1}) - T \quad (29.3)$$

There are many assumptions built into the calculation.⁵ Consider one: namely, the measurement of the $S_{i,t}$. The $S_{i,t}$ within a country may be endogenous, reflecting structural change, for example, the growing importance of large relative to small firms (Nelson and Pack, 1999). Existing estimates make very strong assumptions about the nature of technological change: for example, growth accounting exercises such as Young (1992, 1995) and Collins and Bosworth (1996) assume Hicks-neutral technical change. But this assumption cannot be supported by independent production function estimates for one country given the impossibility theorem of Diamond et al. (1972) which shows that for a general neoclassical production function, the elasticity of substitution and the bias of technical change cannot be estimated simultaneously.⁶

Nelson and Pack (1999) show that the $S_{i,t}$ are endogenous by assuming a neoclassical production function $Q = f(K, mL)$ in which m represents Harrod-neutral, (labor-augmenting) technological advance. The rate of change of factor shares $S_{i,t}$ is a function of the elasticity of substitution, σ , and m , or:

$$S_K^* = [S_L^0 (1 - \sigma)/\sigma] (m - k^*) \quad (29.4)$$

$$S_L^* = [S_K^0 (1 - \sigma)/\sigma] (k^* - m) \quad (29.5)$$

where k^* is the growth rate of the capital-labor ratio. Equations (29.4) and (29.5) show that the factor shares utilized in calculating the Tornqvist index are affected by both technical change, in this case labor-augmenting, and changes in capital intensity. If σ is high, close to unity, a high value of k^* does not decrease the share of capital even if m is small. If σ is low, a high value of m could prevent a fall in S_k . In growth accounting exercises the $S_{i,t}$ are assumed to measure the elasticity of output with respect to factor inputs. But the $S_{i,t}$ are 'uncorrupted' measures only if the assumed underlying production function exhibits Hicks-neutral technical change. If technical change was, in fact, labor-augmenting as in (29.4) and (29.5), $S_{k,t}$ used in (29.2) would have been lower, hence the calculated value of T would have been smaller (as k^* was > 0), and the calculated value of A^* would have been greater.

Given that rapid rate of growth of capital weighted by $S_{K,t}$ is employed in the calculations attempting to demonstrate the absence of high productivity growth, the precise assumptions about the nature of technical change are critical. Unless there is a strong basis for assuming the existence of Hicks-neutral technical change, calculations of TFP growth using

Tornqvist indices provide estimates that may be biased. On theoretical grounds Hicks-neutral technical change is problematic as steady-state growth in neoclassical models can occur only if technical change is Harrod-neutral (labor-augmenting).

There are many other problems. For example, if input markets are distorted due to the suppression of labor unions, the factor shares may not yield good estimates of the elasticity of output with respect to the factor in question; if output markets are not competitive due to high rates of effective protection or high concentration, mark-up pricing could give rise to distorted values of the $S_{i,t}$. It cannot be assumed that factor shares represent competitive imputations derived from Euler's theorem.

Production function estimates

Production function estimates of TFPG are not subject to the same problems. A paper by Kim and Lau (1994) finds that there has been no TFP growth and that technical change has been capital-augmenting. Their results do not suffer from the many problems of growth accounting such as the assumption of constant returns to scale and no bias in technical change. Yet their approach raises questions of interpretation. In particular, their primary assumption is that LDCs can move freely along an international 'meta' production function, much as Parente and Prescott (2000) assume.⁷ But this abstracts from a number of empirical facts: (1) production knowledge is imperfectly available and requires large amounts of tacit knowledge that is not in the possession of LDCs;⁸ (2) fear of generating future competitors makes some industrial-country firms reluctant to provide technology; (3) because of information asymmetries, contracts for existing technology may not be consummated (Arrow, 1969); (4) much of the successful use of knowledge requires production experience (Rosenberg, 1994) and domestic absorptive efforts; (5) much learning, particularly in manufacturing, is specific to a particular locale (Evenson and Westphal, 1995) and as firms move away from their existing capital-labor ratios, their technical efficiency may decline (Atkinson and Stiglitz, 1969);⁹ (6) knowledge is rarely transferred in non-traded goods, particularly services and construction.

Given the many problems associated with both growth accounting and production function estimation, description of the experience of countries or regions by estimates of TFPG have to be recognized as providing gross measures of differences in performance. Referring back to the data in Figure 29.1, no change in assumptions is likely to alter the ordinal ranking of regions. But more fine-grained conclusions about the relative contribution of capital accumulation and TFPG are less certain.

Sources of TFP growth

Determinants of TFP growth

Cross-country regression analyses of the sources of economy-wide TFP growth have recently concentrated on 'fundamental' sources such as geography, institutions and economic policies (or their determinants). While these are facilitating factors, a technology focus suggests looking at the level of foreign direct investment, technology licensing, the use of consultants, imports of equipment and other major vectors of imported technology that permit the technology gap between poor and rich countries to be closed. In addition purely domestic sources of productivity improvement such as research and development (R&D) and education need to be considered.

Countries need not employ all of the potential vectors of new technology but they need to utilize some. During the 1950s Japan relied heavily on technology licensing while discouraging FDI (Ozawa, 1974; Nagaoka, 1989). In the 1960s and 1970s Korea also largely excluded FDI but used technology licensing, consultants and imported equipment and intermediates as sources of technical advance (Hobday, 1995; Enos and Pak, 1987). The primary orientation was an openness to foreign ideas, many embodied in physical inputs. Coe et al. (1996) provide evidence that countries with high levels of imports of goods benefiting from R&D in their home country increase TFP growth in the importing country.¹⁰

The success of the Asian countries depended, *inter alia*, on their ability to improve their technological levels and a considerable amount of firm-level evidence documents the process (Pack, 2001), which was facilitated by the interaction between highly skilled labor and inflows of technology. The precise mechanism of this interaction is not captured in standard cross-country regressions and the measures of 'institutions' typically do not include openness to technology. Almost four decades ago, Nelson and Phelps, (1966) presented a model that provides a plausible alternative to growth accounting explanations that assume growth is a function simply of factor accumulation that has no complex interactions among the factors.

The intuition of their model is that new technology can be a major source of growth but its successful absorption depends on the presence of high skills. Education will have its greatest impact when there is rapid technological change. If the basic technology (a loom used in weaving) is largely unchanged over time, the production process becomes routine and the ability to deal with change is not germane – high or growing education results in only limited productivity gains. In contrast, where technology is rapidly evolving, learning about the existence of new processes, learning to use them when they are deployed and staying abreast of new developments requires the adaptability provided by formal education. This intuition may

help to explain the puzzling low returns to education in some developing countries (Pritchett, 1996). Even in sub-Saharan Africa, which is generally thought to suffer from a shortage of skilled labor, the presence of educated managers has little pay-off in improved productivity in the face of a very low inflow of new technology (Pack and Paxson, 2001).

Do exports contribute to TFP growth rates?

The rapid growth of many export-oriented countries and the slower growth of countries following ISI has led to conjectures that the act of exporting has some particular beneficial effects in generating greater TFPG. One possibility is that the need to compete in foreign markets forces firms to seek means of improving their productivity. An alternate view is that firms that are initially more productive self-select into exporting. Clerides et al. (1998) establish the latter result using panel data of firms in Chile, Colombia and Mexico, and find that contingent on the initial export decision, there is no evidence of faster learning in exporting firms than in those that do not export. However, research for other countries does not support this. One paper examining eight African countries (Bigsten et al., 1997) finds evidence of learning from exporting as does a study by Kraay (1997) of China. One problem with empirically examining the issue is that some critical micro detail is not available when estimation with large numbers of firms occurs. For example, Kim (1997) describes the considerable effort by a Korean firm to improve both the quality of microwave ovens and the productivity with which they were manufactured, after the firm received an export order but two years before the exports were realized. This implies that TFPG grew before exports were realized.¹¹

Do clusters affect growth?

A number of analyses have noted the potential importance of agglomeration economies for firms in LDCs. Such observations date back to Marshall's *Principles* (1870) – this emphasis has received new impetus from the success of high-technology enclaves in Silicon Valley and Bangalore, and industrial clusters in the special economic zones in China. The presumed benefits of clusters on the level of TFP stem from real external economies such as pooling specialized labor skills, and the pecuniary economies resulting from the competition among component makers, as well as a greater variety of readily available intermediate inputs. Most of the evidence on the benefits of agglomeration relies on the result that labor (not total factor) productivity is greater in urban areas, but the possibility that this simply reflects greater capital intensity is not tested given the general lack of availability of capital stock by region (Henderson, 2002).

Recent research (Yusuf, 2003) argues that clusters may be more innovative and thus lead to higher TFPG. There is little evidence of agglomerations being the impetus to innovation rather than an *ex post* result of the entry of new firms following an initial spurt of innovation and growth. Hewlett-Packard was innovative and grew rapidly before the many firms now constituting Silicon Valley could be identified as a cluster. Yusuf (pp. 239–45) notes that several science clusters in Japan and Korea have not been particularly innovative. The recent experience of India's software sector and the Hsinchu Science Park in Taiwan suggest that international linkages may be of equal or greater importance than the presence of nearby firms. The development of the Indian software industry resulted from the efforts of domestic Indian entrepreneurs and a generation of expatriates who had migrated to the USA and were prominent in Silicon Valley (Saxenian, 1999, 2001). The role of clusters in improving the level and rate of growth of productivity is potentially important, but the empirical work in this area is as of now too limited to allow strong conclusions.

Is high TFPG enough?

Higher rates of TFPG and the implied lowering of unit costs in manufacturing were critical to the industrial success of a number of Asian countries in the 1960s and 1970s. Four decades later such growth is no longer a sure recipe for success for those countries with limited domestic markets that need to export to earn foreign exchange, and even for larger countries that expose their domestic firms to competition from imports. Increasingly, world trade in intermediates and in many final goods occurs through international production networks.

Two types of organization have evolved: (1) international production networks (IPNs), in which a producing firm organizes large numbers of suppliers in a number of locations; (2) buyer-led networks in which large retail chains provide specifications for the desired final product and encourage suppliers in developing countries to organize their own production system that most often includes large numbers of local subcontractors.¹² These networks have become increasingly important, and are dominant in clothing and electronics and are growing in importance in other products such as automotive components. In East Asia in recent years components 'constitute at least a fifth of manufacturing exports and . . . have typically grown 4–5 percent faster than overall trade in East Asia' (Yusuf, 2003, p. 272).

One effect of the growing importance of IPNs is their efficiency at organizing production and continuously reducing costs so that the global price that non-member firms must compete with shifts down rapidly. Infant firms undergoing learning face other hurdles: rapidly improving quality,

changing characteristics of existing products (Ernst, 2002) and an array of new goods that compete with existing ones. For firms attempting to enter export markets or defend home markets, it cannot be assumed that simply improving TFP levels and achieving low cost is sufficient to realize foreign sales. There is no guarantee that lead firms will be able to identify one or two firms in a small African nation that is otherwise unremarkable in manufacturing.

New technological skills are required, for example, LDC firms in the network must: 'label track, respond to product orders in real time on the basis of style, color, fabric, and size; exchange information on an . . . electronic basis, provide goods to a retailer's distribution center that can be efficiently moved to stores . . . including containers with bar codes concerning contents' (Yusuf, 2003, p. 283). Such technological requirements for successful participation in globalization impose much greater technological sophistication than that required of the early Asian NICs, including considerable abilities in information technology. Without such skills, greater TFPG may have limited pay-off.¹³

Conclusions

There is no off-the-shelf recipe for either choosing the correct technology or generating accelerated TFP growth. Even acknowledging that the various measures of TFPG have some degree of uncertainty, all studies indicate that the Asian countries have performed better than countries in other regions. Part of their success undoubtedly reflects better macroeconomic management. Another represents the greater competitive pressures that firms face as a result of lowered import duties and export incentives. But part also stems from the concentration on acquiring and improving technology which itself reflects competitive pressures. Firms were open to acquiring foreign knowledge and were not precluded by inimical policies such as limits on royalty payments for technology licenses. The potential high productivity of foreign technology was enhanced by an educated domestic labor force. It is worth emphasizing that the skills that are relevant are science and technology training rather than law or the liberal arts. Argentina in 1960 had one of the highest average levels of education of any country, but its industry suffered from a relative shortage of technicians, compounded by adverse incentives. Indeed, a large supply of lawyers may encourage the type of counterproductive surveillance of foreign technology inflows characteristic of the Andean Pact nations.

Conversely, the establishment of science and technology institutes, even those devoted to facilitating industrialization, may not be necessary or as productive as it initially appears to outsiders. For example, the Korean Institute of Science and Technology (KIST) and the Industrial Technology

Research Institute (ITRI) in Taiwan have been widely viewed as critical to the rapid growth of newer, higher-technology industries. Yet firm-level studies of the development of competitiveness in these industries in both Korea and Taiwan rarely find that these institutions were critical to the success of firms (Hobday, 1995; Kim, 1997). Rather, the ability of firms to obtain knowledge from abroad, through licenses, equipment purchases, strategic alliances and returning nationals, is a recurrent theme for firms in newer industries. Moreover, this knowledge was improved upon by engineers on the firms' own staffs (Kim, 1997). While the KIST and the ITRI played an important role in a few technological areas, it was the initiative of entrepreneurial firms and their own internal R&D which improved upon foreign technology that was critical. Publicly supported research consortia, whether SEMATECH in the USA or the fifth-generation computer project in Japan, do not have a compelling record of economic success. A few examples from Japan in large-scale integrated circuits are offset, even in that country, by a number of failed efforts.

Finally, efforts to improve productivity levels need to be complemented by initiatives to insure that local firms are on the radar screen of international producers' networks, a task at least as complex as providing conditions conducive to productivity growth.

Notes

1. Pack (1982) provides a survey of the issues and an evaluation of their quantitative importance.
2. Tybout (2000) provides a thorough review of a large number of studies in LDCs that have used firm-level data to calculate frontier production function and the dispersion of TFP levels and the inter-firm variation of TFP. He notes that this dispersion is no greater than that in industrialized countries. However, given that the TFP level of firms on the LDC frontier is typically low relative to DC counterparts, the dispersion implies a low efficiency of resource use for the typical industrial sector.
3. Although their calculations are open to many of the same criticisms of growth accounting that I present below, it is unlikely that the qualitative pattern of their findings would be reversed with revisions of their procedure.
4. Some analysts such as Young (1995) adjust the shares to correct various deficiencies such as those dealing with the remuneration of unpaid family members.
5. (a) all countries on the same production function; (b) the $S_{i,t}$ are unaffected by technical change as it is Hicks neutral; (c) input markets are not distorted and all factors receive their marginal value product; (d) Y_t and $x_{i,t}$ are measured correctly. The implications of using a cost function rather than a production function are considered in Pack (2001). For full discussion of these issues see Pack (2001).
6. Kim and Lau (1994) estimate production functions and solve this problem by using several countries in their pooled cross-section time-series analysis.
7. Hayami and Ruttan (1985) postulated the meta production function (MPF) for agriculture and argued that as factor prices changed secularly, induced innovation would occur, reducing the demand for more expensive factors. The process was envisioned as one in which new isoquants would be developed by research rather than be chosen from an existing menu available across the world. Moreover, even in agriculture it is known that the successful adoption of new technologies such as the Green Revolution required long

and expensive domestic efforts (Evenson and Westphal, 1995). Farms in India could not move toward the world frontier without considerable research in each region of India. In non-agricultural activities, the world frontier may be even more difficult to achieve, lessening the plausibility of the MPF metaphor.

8. On these and other questions related to interpretation of the production function see Nelson and Winter (1982). For evidence that firms in some LDCs do not possess the same production knowledge see Pack (1987).
9. These factors also underlie the considerable diversity in productivity among firms in the same industry in developed countries. Estimates of frontier production functions have demonstrated the very large range of productivity achieved within developed countries among firms in which relatively similar equipment is employed (Caves and Bailey, 1992). The existence of such divergences within DCs is one more reason for doubting that all firms throughout the developing and developed world produce along the same production function.
10. A good survey is provided by Saggi (2002). Evenson and Westphal (1995) and Gisselquist et al. (2002), provide interesting evidence on agricultural technology transfer.
11. Westphal (2002) has a thorough examination of much of the research on the export–TFPG link.
12. A good description of these alternatives and evidence on their quantitative importance is given in Gereffi (1999). Sturgeon and Lester (2002), provide evidence on the empirical importance of the IPNs.
13. Interesting questions also arise about the extent to which participation in networks constrains the extent of learning. For a good discussion see Humphrey and Schmitz (2002).

References

- Arrow, Kenneth A. (1969), 'Classificatory Notes on the Production and Transmission of Technological Knowledge', *American Economic Review*, 59 (2): 29–35.
- Atkinson, A.B. and J.E. Stiglitz (1969), 'A New View of Technological Change', *Economic Journal*, 79: 573–8.
- Bigsten, Arne, Paul Collier, Stefan Dercon, Marcel Fafchamps, Bernard Gauthier, Jan Willem Gunning, Jean Habarurema, Abena Oduro, Remco Oostendorp, Catherine Patillo, Måns Söderbom, Francis Teal and Albert Zeufack (1997), 'Exports and Firm Level Efficiency in the African Manufacturing Sector', School of Business, University of Montreal.
- Borensztein, E., J. De Gregorio and J.W. Lee (1998), 'How Does Foreign Direct Investment Affect Economic Growth?', *Journal of International Economics*, 45: 115–35.
- Caves, Richard and Sheryl Bailey (eds) (1992), *Industrial Efficiency in Six Nations*, Cambridge, MA: MIT Press.
- Clerides, Sofronis, Saul Lach and James R. Tybout (1998), 'Is Learning by Exporting Important? Microdynamic Evidence from Columbia, Mexico, and Morocco', *Quarterly Journal of Economics*, 113: 903–47.
- Coe, David T., Elhanan Helpman and Alexander Hoffmaister (1996), 'North–South R&D Spillovers', *Economic Journal*, 107 (440): 134–49.
- Collins, Susan M. and Barry P. Bosworth (1996), 'Economic Growth in East Asia: Accumulation vs. Assimilation', *Brookings Papers on Economic Activity*, 21: 135–91.
- Deaton, Angus and Christina Paxson (1994), 'The Effects of Economics and Population Growth on National Saving', *Demography*, 34: 97–114.
- Diamond, Peter, Daniel MacFadden and Michel Rodriguez (1972), 'Identification of the Elasticity of Substitution and the Bias of Technical Change', in Daniel MacFadden (ed.), *An Econometric Approach to Production Theory*, Amsterdam: North-Holland, pp. 125–47.
- Enos, John L. and Uhui Pak (1987), *The Adoption and Diffusion of Imported Technology: The Case of Korea*, London: Croom Helm.
- Ernst, Dieter (2002), 'Global Production Networks in East Asia's Electronics Industry', East-West Center, Honolulu, Hawaii.

- Evenson, Robert E. and Larry E. Westphal (1995), 'Technological Change and Technology Strategy', in Jere Behrman and T.N. Srinivasan (eds), *Handbook of Development Economics 3a*, Amsterdam: North-Holland, pp. 2209–2300.
- Foster, Andrew D. and Mark R. Rosenzweig (1995), 'Technical Change and Human Capital Returns and Investments: Evidence from the Green Revolution', *American Economic Review*, 86: 931–53.
- Gereffi, Gary (1999), 'International Trade and Industrial Upgrading in the Apparel Commodity Chain', *Journal of International Economics*, 48: 37–70.
- Gisselquist, David, John Nash and Carl Pray (2002), 'Deregulating the Transfer of Agricultural Technology: Lessons from Bangladesh, India, Turkey, and Zimbabwe', *World Bank Research Observer*, 17: 237–66.
- Hayami, Yujiro and Vernon Ruttan (1985), *Agricultural Development: An International Perspective*, Baltimore, MD: Johns Hopkins University Press.
- Henderson, Vernon (2002), 'Urbanization in Developing Countries', *World Bank Research Observer*, 17: 89–112.
- Hobday, Mike (1995), *Innovation in East Asia: The Challenge to Japan*, Aldershot, UK and Brookfield, US: Edward Elgar.
- Humphrey, John and Hubert Schmitz (2002), 'Governance and Upgrading: Linking Industrial Cluster and Global Value Change Research', Institute for Development Studies, University of Sussex, Working Paper 120.
- Kim, J.I. and L.J. Lau (1994), 'The Sources of Economic Growth in the East Asian Newly Industrialized Countries', *Journal of Japanese and International Economics*, 8, 235–71.
- Kim, Linsu (1997), *From Imitation to Innovation: Dynamics of Korea's Technological Learning*, Boston, MA: Harvard Business School Press.
- Kraay, Art (1997), 'Exports and Economic Performance: Evidence from a Panel of Chinese Enterprises', World Bank, Development Economics Department.
- Krugman, Paul (1994), 'The Myth of Asia's Miracle', *Foreign Affairs*, December: 62–78.
- Lewis, W. Arthur (1954), 'Economic Development with Unlimited Supplies of Labor', *Manchester School*, 22: 139–91.
- McKinsey Global Institute (1998a), *Productivity Led Growth in Korea*, Washington, DC and Seoul: McKinsey & Co.
- McKinsey Global Institute (1998b), *Productivity: The Key to an Accelerated Development Path for Brazil*, Washington, DC and Sao Paulo: McKinsey & Co.
- Marshall, Alfred (1870), *Principles of Economics*, Cambridge: MacMillan.
- Nagoaka, Sadao (1989), 'Overview of Japanese Industrial Technology Development', Industry Series Paper 6, World Bank Industry and Energy Department, Washington, DC.
- Nehru, Vikram and Ashok Dhareshwar (1994), 'New Estimates of Total Factor Productivity Growth for Developing and Industrial Countries', Washington, DC, World Bank.
- Nelson, Richard R. and Howard Pack (1999), 'The Asian Growth Miracle and Modern Growth Theory', *Economic Journal*, July: 416–36.
- Nelson, Richard R. and Edmund Phelps (1966), 'Investment in Humans, Technological Diffusion, and Economic Growth', *American Economic Review*, May (56): 69–75.
- Nelson, Richard R. and Sidney Winter (1982), *An Evolutionary Theory of Economic Change*, Cambridge, MA: Harvard University Press.
- Ozawa, Terutomo (1974), *Japan's Technological Challenge to the West, 1950–74: Motivation and Accomplishment*, Cambridge, MA: MIT Press.
- Pack, Howard (1982), 'Aggregate Implications of Factor Substitution in Industrial Processes', *Journal of Development Economics*, November: 1–37.
- Pack, Howard (1987), *Productivity, Technology, and Industrial Development*, New York: Oxford University Press.
- Pack, Howard (2001), 'Technological Change and Growth in East Asia: Macro and Micro Perspectives', in J. Stiglitz and S. Yusuf (eds), *The Asian Miracle Revisited*, New York: Oxford University Press, pp. 95–142.
- Pack, Howard and Christina Paxson (2001), 'Is African Manufacturing Skill Constrained?', in A. Szirmai (ed.), *The Industrial Experience of Africa*, London: Macmillan.
- Parente, Stephen and Edward Prescott (2000), *Barriers to Riches*, Cambridge, MA: MIT Press.

- Pilat, Dirk (1994), *The Economics of Rapid Growth, the Experience of Japan and Korea*, Aldershot, UK and Brookfield, US: Edward Elgar.
- Pritchett, Lant (1996), 'Where Has All the Education Gone?' World Bank, Policy Research Working Paper No. 1581, March.
- Rosenberg, Nathan (1994), 'Uncertainty and Technological Advance', manuscript, Department of Economics, Stanford University, Stanford, CA.
- Saggi, Kamal (2002), 'Trade, Foreign Direct Investment, and International Technology Transfer', *World Bank Research Observer*, 17: 191–236.
- Saxenian, AnnaLee (1999), *Silicon Valley's New Immigrant Entrepreneurs*, San Francisco, CA: Public Policy Institute of California.
- Saxenian, AnnaLee (2001), 'Taiwan's Hsinchu Region: Imitator and Partner for Silicon Valley', Stanford Institute for Economic Policy Research, Stanford, CA.
- Sturgeon, Timothy and Richard Lester (2002), 'Upgrading East Asian Industries: New Challenges for Local Suppliers', Cambridge, MA: Industrial Performance Center, MIT.
- Tybout, James R. (2000), 'Manufacturing Firms in Developing Countries: How Well Do They Do, and Why?' *Journal of Economic Literature*, 38: 11–44.
- Westphal, Larry E. (2002), 'Technology Strategies for Economic Development in a Fast Changing Global Economy', *Economics of Innovation and New Technology*, 11: 275–320.
- World Bank (1993), *The East Asian Miracle*, Washington, DC: World Bank.
- Young, Alwyn (1992), 'Tale of Two Cities: Factor Accumulation and Technical Change in Hong Kong and Singapore', in Oliver Blanchard and Stanley Fischer (eds), *NBER Macroeconomics Annual*, Cambridge, MA: MIT Press, pp. 13–53.
- Young, Alwyn (1995), 'The Tyranny of Numbers: Confronting the Statistical Realities of the East Asian Growth Experience', *Quarterly Journal of Economics*, 110: 641–680.
- Yusuf, Shahid (2003), *Innovative East Asia: The Future of Growth*, Washington, DC: World Bank.

PART V

SECTORS IN DEVELOPMENT

30 Factor market imperfections in poor agrarian economies

Parikshit Ghosh and Ashok Kotwal¹

Introduction

There has been a voluminous literature since the 1970s analyzing factor markets in poor agrarian economies. There are two sources of interest in this area: a curiosity about various contractual arrangements that seem quite puzzling at first glance, and a search for a policy perspective on a sector in which the majority of the world's poor make a living. Many agrarian economies are characterized by highly skewed distributions of land ownership, and consequently, large variations in factor endowment ratios across households. There are rich families with an inadequate amount of family labor to cultivate their land, while there are many poor families who own little or no resources other than their labor. Unless production functions are highly non-homothetic, this state of affairs will place an enormous burden on the markets for factor services (rental, labor, credit, water, and so on) to achieve overall efficiency in cultivation.

It is puzzling that asset markets for land are conspicuous by their near absence in such an environment. If poor families could acquire ownership rights on land through arrangements (for example, mortgage) similar to what we have in developed countries to enable home ownership, it would bring about an efficient organization of production by equalizing the marginal product of land across owners. But instead, we observe an almost exclusive reliance on the markets for factor services. Moreover, unlike the anonymous exchange and uniform pricing characteristic of developed societies, these transactions are in the form of highly personalized and informal contractual arrangements. Why? What is different about these poor agrarian economies except that they are poor? Or, is it the poverty itself that inhibits the evolution of asset markets, or (possibly) distorts the functioning of input markets? On the policy front, we need to know how to reform the agricultural sectors in less-developed countries in order to make a significant dent in poverty. Even in fast-growing economies like India (since the 1980s), poverty reduction has been slower in areas where rural markets are less developed, suggesting the existence of local institutional bottlenecks that hinder growth. What should governments do? Should they subsidize credit to poor farmers? Should they undertake a comprehensive land

reform? Would land reform promote equality at the cost of aggregate output? The goal of this chapter is to present a framework that will allow us to deal with these questions.

In order to understand the distinctive characteristics of factor markets in poor agrarian economies, we must first understand the distinctive characteristics of agricultural production in such an environment. The first such characteristic is weather uncertainty. In the absence of well-developed irrigation, the output (and hence incomes) fluctuates with the vagaries of weather.² A bad year can jeopardize the survival of a cultivator who lacks savings to stem the fall in the consumption of his household. Cultural norms and institutions have evolved to deal with this very common problem which afflicts a large segment of the rural population. The rich are socially obligated to ensure the survival of their clients – those who borrow money or lease land from them. This kind of ‘limited liability’ is a characteristic feature of land and credit contracts in such an environment.

Moral hazard arising from the unobservability of effort is another common problem. In labor-intensive agriculture, the output is especially sensitive to the quality of effort. Small mistakes can cause huge losses. For example, the wrong fertilizer mix can kill the whole crop. At the same time, production activities are spread over a large area rather than confined to a shop floor, making supervision difficult. In addition, the final output (that is, crop) is available only at the end of the growing season, while labor inputs are spread over the entire season. This, coupled with the difficulty of disentangling the contributions of negligence and random elements like the weather, makes the issue of moral hazard in cultivation even more important than in many manufacturing activities.

Since weather uncertainty coupled with extreme poverty is the defining characteristic of a poor agrarian economy and gives rise to land and credit contracts subject to limited liability, we will focus on the moral hazard of cultivators in this context. We will show that poverty itself can be a source of inefficiencies in factor markets. A major implication of our analysis is that some form of mandated land redistribution may be essential for increasing productivity. In the absence of such measures, other policies such as credit extension may be ineffective or, in some situations, even counterproductive. Our models help explain the puzzling thinness of land markets. The analysis borrows the framework of a very insightful paper by Dilip Mookherjee (1997).

The model

The baseline model consists of two agents, labeled a lender (L) and a farmer (F), and a plot of land whose size is normalized to one. Cultivation requires

land to be combined with labor and a purchased input (seeds) to yield output. Land is perfectly divisible, and we assume a fixed coefficient linear technology: one unit of land must be used with one unit of labor and one unit of seeds (market price = 1).

We assume L has no household labor available for cultivation, but has unlimited funds to buy seeds. On the other hand, F has enough household labor (at zero opportunity cost), but total cash reserves of only $w < 1$ for the purpose of buying seeds. This implies neither agent is self-sufficient for cultivation up to the maximum scale, and must rely on labor or credit relationships to try and bridge shortfalls. The land may be owned either by L or F. The former case further creates the scope for either rental or asset transactions in land, and we will consider both cases separately.

To introduce moral hazard in the simplest possible way, assume output is stochastic, taking either a high (q) or a low (\tilde{q}) value ($q > \tilde{q}$). Let p denote the probability of obtaining the higher output. We assume p is determined by the farmer's unobservable effort and is a monotonically increasing function of p . For convenience, we will refer to p as effort. If the farmer cultivates a fraction λ of the land (keeping the rest fallow) and wants to achieve the higher output with probability p , he faces an 'effort cost' of $\lambda c(p)$, where $c(\cdot)$ is a differentiable, increasing and strictly convex function, satisfying the Inada end-point conditions: $c'(0) = 0$, $c'(1) = \text{infinity}$. This formulation ensures that the technology displays no scale economies with respect to observable (land, labor, seeds) and unobservable (effort) inputs taken together. The unobservability of effort (that is, p) implies it cannot be contracted upon, and will be chosen to maximize the farmer's expected earnings under all scenarios considered.

Both agents are risk-neutral and the farmer has no alternative use for his labor. We model the interaction within a principal-agent framework, tacitly giving L unlimited power to allocate the surplus in the relationship. This seems like a reasonable approximation of the conditions prevailing in a labor-surplus economy with substantial inequality in land ownership. The last crucial assumption is the introduction of a limited liability constraint – the contract offered by L must give F at least amount s of consumption in each possible state. s can be thought of as a 'subsistence income' that must be left with the farmer for mere survival.

Before analyzing the inefficiencies arising from agency problems under different scenarios, let us define the first-best as a benchmark for comparisons. Consider for a moment the effort level that would have been chosen (represented by p^*) if the cultivator owned all necessary inputs and could keep the entire output for himself. Mathematically:

$$(\lambda^*; p^*) = \arg \max_{\lambda, p} \lambda [pq + (1 - p)\tilde{q} - c(p) - 1]$$

that is, we solve for the optimal scale of cultivation and the optimal level of effort for a self-sufficient owner-cultivator. Since the objective function is linear in λ , λ^* is either 0 or 1. We will focus only on the interesting case of $\lambda^* = 1$. The first-best effort level is then captured by the first-order condition:

$$c'(p^*) = q - \tilde{q} \quad (30.1)$$

Factor markets and moral hazard

Sharecropping has been a pervasive system of land leasing all over the world. Historically, it was popular in the US antebellum south and many parts of medieval Europe, and continues to be a popular arrangement in much of the developing world today. Under sharecropping, the cultivator gets only a fraction of his marginal product, and the share accruing to the landlord works as a tax on the tenant, lowering the incentive to apply inputs. This criticism of sharecropping as an inefficient form of cultivation dates back to Marshall (1920). Indeed, Binswanger and Rosenzweig (1986) and Shaban (1987) find that controlling for land characteristics and comparing against the productivity of owned land, farmers who both own and lease generate 30 percent lower yields on their sharecropped land. Shaban (1987) also finds that owner-cultivated land has a higher price per hectare, indicating that leasing discourages quality-improving investments. Thus, poor incentives seem to matter over the short as well as the long run. There has been an enormous amount of theoretical literature on sharecropping, motivated by a desire to understand why such an institution, with obviously poor incentive properties, has emerged in many diverse areas of the world at a certain stage of development. This literature represents one of the earliest applications of agency theory in economics.

A fixed rental contract (where the rent is independent of crop output) makes the cultivator the residual claimant and would solve the incentive problem. However, this places all the risk on the tenant, and it has been suggested that since tenants are typically poorer and more risk-averse than landlords, sharecropping has emerged as an institution that balances risk-sharing motives with the need to provide incentives to the cultivator (Stiglitz, 1974). A second reason (see Basu, 1992) why rents may need to be made state dependent is 'limited liability' – the physical and social constraints on the landlord in extracting payment when crop output is low. This kind of state dependence of obligations, again, acts as a tax on incremental production and distorts the incentive to apply costly inputs. Moreover, since the root of limited liability lies in the cultivator being close to subsistence, it is an endemic feature of poor economies, and affects other kinds of transactions (such as credit)³ as well, and in similar ways. In this chapter, we take an integrated approach – by evaluating the performance

of various factor markets based on common features of the environment (moral hazard and limited liability), we are able to compare agricultural performance under certain alternative scenarios (say, alternative patterns of land ownership or alternative assumptions about access to credit). This enables us to evaluate the effects of various policy measures, like land reform or provision of subsidized credit.

In the next two subsections, we analyze how the pattern of land ownership affects incentives, efficiency and agricultural productivity. We compare outcomes under two alternative assumptions of ownership: one where the farmer is landless and depends on the other party for both land and credit, and the other where the farmer owns the land and borrows to overcome liquidity constraints. We show that under sufficient pressure of poverty (that is, if the farmer's borrowing needs are high, or his ability to pay in bad states very low), inefficiency will arise in both cases, in the sense that the effort level and expected output will be less than first-best. We also show that the effort level is weakly higher when the farmer owns the land, and under some conditions, strictly so. The last result indicates that effective land reform may increase agricultural productivity.

Land-owning cultivators: pure credit

Suppose the land belongs to F. However, since F does not have enough financial capital to cultivate the entire land, he may want to borrow from L up to the amount $1 - w$. F's reservation utility (\tilde{u}) is the expected income he can generate by himself, without depending on L for credit. Note that due to the linearity of the technology, F will choose the first-best effort p^* when he is not indebted, even if liquidity constraints curtail the scale of cultivation. Hence:

$$\tilde{u} = w [p^*q + (1 - p^*)\tilde{q}] - c(p^*) - 1 \tag{30.2}$$

The lender makes a take-it-or-leave-it offer to F, advancing x amount of credit ($0 \leq x \leq 1 - w$) and asking for a (possibly state-contingent) interest rate. As above, it is easy to see that the choice of scale of cultivation always has a corner solution, that is, if F cultivates at all, he will always cultivate as much land as his liquidity (his own plus any borrowed funds) permits. Let r be total obligation per dollar of loan (that is, principal plus the interest rate) when output is high, and \tilde{r} when output is low. L chooses x , r and \tilde{r} to maximize his expected profit:

$$\max x; r; \tilde{r} \ x [pr + (1 - p)\tilde{r} - 1] \tag{30.3}$$

subject to certain constraints, which we will now specify. As long as the farmer undertakes some debt, his optimal effort choice is given by:

$$\begin{aligned}
& \arg \max_p p [(w+x)q - rx] + (1-p) [(w+x)\tilde{q} - \tilde{r}x] \\
& \quad - (w+x)c(p) - w \\
& = \arg \max_p (w+x) [pq + (1-p)\tilde{q} - c(p)] - x [pr + (1-p)\tilde{r}] - w
\end{aligned} \tag{30.4}$$

which yields the first-order condition

$$c'(p) = (q - \tilde{q}) - (r - \tilde{r}) (x / (w + x)) \tag{30.5}$$

This is the incentive constraint: that is, in computing expected profit, L must take into account the fact that F's effort choice will be determined by the interest rates as described in (30.5).

The terms of the credit contract must be such that F will prefer to accept it instead of being self-reliant. This leads us to the participation constraint:

$$\max_p (w+x) [pq + (1-p)\tilde{q} - c(p)] - x [pr + (1-p)\tilde{r}] - w > u \tag{30.6}$$

Finally, limited liability imposes upper bounds on the interest rates that can be charged:

$$r \leq ((w+x)/x)q - s/x = q + ((wq-s)/x) \tag{30.7}$$

$$\tilde{r} \leq ((w+x)/x)\tilde{q} - s/x = \tilde{q} + ((w\tilde{q}-s)/x) \tag{30.8}$$

L's problem boils down to maximizing (30.3), subject to the constraints (30.5), (30.6), (30.7) and (30.8).

The optimal choice of x is either 0 or $1-w$, that is, L will either refuse to lend or offer enough credit to allow cultivation of the entire land.⁴ To make the problem interesting, we will assume parameter values are such that L finds it optimal to lend.

Limited liability has a crucial effect on the outcome of transactions like this, and may reduce the level of effort below first best. To understand this, let us consider the case where (30.7) and (30.8) do not bind (this is true in the extreme when $s = \text{negative infinity}$ and will be generally true if s is low enough). Then, the optimum contract has the property that $r = \tilde{r}$, that is, L will charge an interest rate which is not contingent on the outcome.⁵ Using this in (30.5) and comparing against (30.1), we find that the effort level under debt without limited liability will be the first best effort p^* . The intuition is fairly straightforward. If L charges a fixed interest rate independent of the state of the world, this leaves F's incentives unaffected, since at the

margin, he retains all of the increase in surplus. By choosing the interest appropriately high, L can capture as much of this maximized surplus as is possible without violating F's participation constraint.

Next, we analyze cases when at least one of the limited liability constraints is binding when the contract is optimal. This will be true if and only if⁶ the parameters satisfy the condition:

$$(1 - w) [p^*q + (1 - p^*)\tilde{q} - c(p^*) - 1] + s \geq \tilde{q} \quad (30.9)$$

Whenever (30.9) holds, using the fact that $x = 1 - w$, we have: $\tilde{r} = (\tilde{q} - s) / (1 - w)$. Moreover, if (30.9) holds strictly, then in the optimal contract, $r > \tilde{r}$. Finally, observe from (30.5) that the effort level is less than p^* whenever this is the case.

Limited liability combined with the lender's profit-extraction motive reduces efficiency by inducing a level of effort below the first-best. Limited liability restricts the interest rate that can be imposed in bad states, when output is low and the farmer does not have enough liquidity to service a large debt. Under some circumstances, the lender will find it profitable to extract a greater repayment in good states, when the farmer is cash rich. However, this implies an implicit tax, that is, part of the extra output in a good state is captured by the creditor, correspondingly reducing the farmer's incentive to increase the likelihood of such an event through his effort. The lender will trade off extra profit in the good state against the disincentive effect on the borrower, but the solution will be interior in many cases. Farmers' indebtedness may have a negative impact on agricultural productivity.

Equation (30.9) shows that such inefficiencies are more likely to arise in economies where many farming households live under poverty. One consequence of poverty is that farmers lack resources of their own and have high borrowing needs (w is low), which makes (30.9) easy to bind with equality. Another manifestation may be the limited ability to service debt in the event of a poor harvest (which may require liquidating assets, drawing on savings or being insured by relatives), which is roughly captured in a relatively high value of s .

Landless cultivators: tenancy-credit interlinkage

Suppose the land is owned by L instead. L may offer a contractual arrangement involving either labor hiring, or a combination of tenancy and credit. These alternatives yield identical results in the model, so we will cast our analysis in terms of the latter. It is also not hard to see that due to the linear technology, corner solutions will obtain exactly as in the previous section, that is, L will either supply the entire land and all the credit needed for

full-scale cultivation, or none at all. We will skip to that assumption straight away. Any payment obligations of the farmer can obviously be decomposed into rent and interest in arbitrary ways, so the distinction is meaningless within this model. Therefore, without loss of generality, we set the rent to be zero in all states. We assume in this subsection that the farmer has no alternative source of credit, and must enter into an interlinked contract with L for both inputs.

For expositional simplicity, we will make a few additional assumptions before proceeding further. These do not affect the main results to follow, but reduce clutter and allow the intuition to stand out in sharper relief. Specifically, assume $\tilde{q} = s = 0$. This implies that the limited liability constraints in the bad state will always bind. The landlord or creditor can only extract payments in the good state. It is also obvious that he must leave the farmer with some surplus in the good state. Otherwise the farmer will be left with no incentive to provide any effort at all ($p = 0$). In other words, the limited liability constraints in the good state will always be slack at the optimum, and can therefore be ignored. Incorporating these simplifications, and defining the return (principal + interest) as r , the landlord-creditor's problem (that is, the variant where L owns the land) can be rewritten as:

$$\max_r (1 - w)[pr - 1] \quad (30.10)$$

subject to the incentive constraint

$$c'(p) = q - (1 - w)r \quad (30.11)$$

and the participation constraint

$$\max_p p [q - (1 - w)r] - c(p) - w \geq 0 \quad (30.12)$$

Note that once the simplifying assumptions are incorporated, the problem is identical to the previous one except in one crucial aspect. Since the farmer has no land of his own, his reservation utility is 0 instead of u , giving the landlord-creditor greater effective bargaining power in the interaction. The trade-off for the landlord-creditor is analogous to the one faced by a monopolist while choosing price. If he increases r , he adds directly to his revenues but it lowers the effort level and p , the probability of a 'good' state occurring. This is evident from the incentive constraint above. Incorporating (30.11) directly into the objective function and the participation constraint, the solution to the problem is the same as the solution to the following problem:

$$\max_p p [q - c'(p)] \quad (30.13)$$

subject to

$$pc'(p) - c(p) \geq w \quad (30.14)$$

Let p_0 be the solution to the unconstrained problem, characterized by the first-order condition:

$$c'(p_0) + p_0 c''(p_0) = q$$

The creditor would like to choose p_0 , but it may not satisfy the farmer's participation constraint. If so, the creditor would lower the rent r and thus raise p so that the farmer would be willing to participate. Let p_t denote the lowest value of p that satisfies the participation constraint (30.14). Assuming the objective function (30.13) is concave in p ,⁷ the solution to the problem is the higher of the two values, p_0 and p_t . Note that since p_t is increasing in w , it is the solution when w is high enough. It is worth noting that the lower the w (that is, the poorer the farmer) the lower is the expected output. In other words, the poverty of cultivators is partly responsible for low agricultural production. When the poverty recedes, we can expect the output to increase.

It is instructive to see how the surplus is divided between the creditor and the farmer:

- The total surplus from the agricultural production is: $pq - c(p) - 1$.
- The surplus accruing to the creditor is: $pq - pc'(p) - (1 - w)$.
- The surplus accruing to the farmer is: $pc'(p) - c(p) - w$, and this increases with p .

Comparison: landowning versus landless cultivators

Agriculture in many poor labor-abundant countries has displayed an inverse relationship between farm size and agricultural yield (for example, Deininger et al., 2003). Some authors (Eswaran and Kotwal, 1986) have suggested that compared to small family farms, large commercial farms face higher labor costs due to the added supervision requirements. This may reduce input usage per acre on large farms, accounting for lower yields. Large landowners could seek to improve efficiency by leasing, but tenancy has distortionary effects similar to hired labor, and will fail to eliminate inefficiencies. It is tempting to conclude from this that transfer of ownership from large landowners to small ones will increase overall output by reducing the reliance on factor markets.

However, the picture is more complex if one considers the whole gamut of farm inputs. Poorer households may face lower labor costs due to the availability of surplus family labor relative to their landholdings, but these households are also more likely to face financial constraints that may impede the application of purchased inputs (seeds, fertilizers, pesticides, energy, water, and so on) and capital goods (bullocks, pumps, and so on), acting as a countervailing factor to cheap labor. This disadvantage would, of course, be neutralized if there was a well-functioning, accessible credit market. However, rural credit markets are often characterized by high interest rates and rationing, because they are beset with the same moral hazard and default possibilities that plague tenancy and labor hiring. Nevertheless, one can see a number of theoretical reasons why the transfer of ownership to landless tenants may increase their credit access. First, it provides them with a collateralizable asset. Second, by improving their outside opportunities and self-sufficiency, it may enhance their bargaining position with local moneylenders, making credit cheaper. We will illustrate this last channel using the model sketched above, by comparing the solutions to the contractual problem when ownership is reversed.

We return briefly to the case where F owns the land, and incorporate our simplifying assumptions ($\bar{q}=s=0$). The problem (when F is the landowner) has the same objective function and incentive constraint as in the previous section (where L is the landowner), but is subject to a tighter participation constraint, which in this simplified case, boils down to:

$$\max_p p [q - (1 - w)r] - c(p) - w \geq u \quad (30.15)$$

On incorporating the incentive constraint (30.11) and (30.2), this becomes:

$$pc'(p) - c(p) \geq w [p^*q - c(p^*)] \quad (30.16)$$

Note that the right-hand side is greater than w , because the profitability of cultivation implies $p^*q - c(p^*) - 1 > 0$. Denote by p_d the lowest value of p that will satisfy (30.16). Just as in the previous case, the solution to the landlord-tenant's problem is the higher of the two values p_0 and p_d , and p_d is obtained if w is high enough.

Compare p_d and p_r . When the opportunity utility of the farmer is lower, the maximum r (and thus the minimum p) that the creditor can extract out of the tenant is higher (lower). The rent is higher and the probability of the good state is lower when the farmer's participation constraint is more relaxed. Therefore, when w is sufficiently high so that the participation constraint binds in both problems, $p_d > p_r$, by comparing (30.16) and (30.14), and noting that the left-hand side (or, the surplus accruing to the farmer)

is increasing in p . We can see clearly that the land ownership results in an improvement in the incentives for the farmer to work hard. It raises the expected output as well as making him better off. Land reform under these conditions improves the output while making the distribution more equal. In fact, it enhances output by increasing the share of the surplus accruing to the poor.

To understand the intuition behind this result, observe the tension between revenue extraction and efficiency. Since the farmer must default in the bad state, the only way to earn any profit is by imposing payments in the good state. However, this acts as a tax on output and reduces incentives. If the land does not belong to the farmer, even if his credit requirements are minimal, his reservation utility is very low. This will allow the landlord to negotiate a very high interest rate (or rent). However, if the farmer owns the land and most of his working capital requirements, the lender is prevented from pushing the interest rate too high, since the farmer can achieve a high level of utility from self-sufficiency. When the farmer acquires land, it is the lender's weakened bargaining position and consequent inability to inflate rents or interest rates which leads to improvements in efficiency in this model.

The absence of land markets

According to the Coase theorem, an economy without transaction costs will always achieve productive efficiency regardless of the distribution of property rights and endowments. In the real world, of course, moral hazard problems introduce transactions costs, which can be substantial in the context of underdevelopment. In our analysis so far, these transactions costs arise from moral hazard and limited liability.

Since differences in factor endowments, coupled with the inefficient functioning of the markets for factor services, are the root cause of inefficiency, it is a natural question why agents do not try to overcome it by trading in the factors themselves. In particular, it might seem that the asset market for land (as opposed to the lease market) could play an important role in reducing skewed factor endowments across households and improving productivity. One of the puzzling features of underdevelopment is the extreme thinness of land markets (Kumar, 1975; Rosenzweig and Wolpin, 1993; Moll, 1988; Deininger et al., 2003) in most developing as well as some developed countries. However, one must recognize that most poor households lack the funds to make out-of-pocket purchases, and must rely on credit to acquire land. Unlike consumption credit or working capital loans, mortgage has the advantage of a natural collateral – the asset itself. However, indebtedness raises the same issues with moral hazard that we have seen in the previous section, by taxing the fruits of labor on the mortgaged land. Our framework

can be used to understand why the problem can be serious enough to prevent the emergence of asset and mortgage markets.

Return to the case of the landlord-lender, and suppose there is a third party (the government, a bank, a non-governmental organization – NGO) prepared to give loans to F at zero profit (necessary for the viability of such programs in the long run). Let T be the selling price of land set by L. Suppose, after learning land prices, the bank charges an interest rate τ on its loans, which can be used for purchasing both land and other inputs. Finally, F decides whether to borrow at going interest rates to buy land at the offered sale price. We look at the perfect equilibrium of the sequential move game between the bank, the landlord and the farmer. L chooses T to maximize profit, correctly anticipating the bank's choice of τ , and the bank chooses τ with the objective of breaking even in expected terms, while correctly anticipating the effort level p that will be forthcoming from F if a financed land sale goes through on the terms (τ, T) .

The bank's break-even condition implies:

$$p\tau = 1$$

F, after learning (τ, T) , chooses the level of effort by solving:

$$\max_p p [q - \tau (T + 1 - w)] - c(p) - w$$

On rearranging the first-order condition for this problem and using the breakeven condition, we get:

$$T = p [q - c'(p)] - (1 - w) \quad (30.17)$$

If land transactions are to take place at all, it must be incentive compatible for F to buy at the offer price and interest rates. For the farmer to be a willing buyer, it must be true that:

$$pc'(p) - c(p) - w \geq 0 \quad (30.18)$$

The landowner, exploiting his assumed monopoly power, should choose the highest sustainable land price, that is, choose the maximum value of the expression in (30.17), without violating (30.18). Note that this problem is identical to the one faced by a landlord-creditor, that is, maximizing (30.13) subject to (30.14). Therefore, the solution coincides with that of tenancy, that is, even if land sales take place, the land price will be set so high and the buyer accumulate so much debt that it will be impossible to get any improvement in effort levels in the end. It can also be checked that \tilde{T} is

exactly enough to compensate the seller for his lost profit from rental and lending, that is, it is his reservation price. Another way to think about this is that when the farmer's participation constraint is binding, his choice of effort level and hence the probability of the good state is determined by the participation constraint. Since the participation constraints are identical in the two cases – 'tenancy' and 'mortgaged land' – so are the effort levels and hence p in the two cases. But note that the price T is given by $[pq - pc'(p) - (1 - w)]$, which is exactly the same (in terms of p) as the surplus gathered by the landlord-creditor in the tenancy case. This means that T which allows the farmer to be at the same level of utility gives the landlord the same surplus as under the tenancy arrangement. Land sale will not be a Pareto-improving arrangement over tenancy. There is no reason for a land market to evolve since there are no mutual gains from a possible transaction. If there is the slightest administrative cost in mortgage transactions, a land transaction will result in a net loss. Curiously, a coercive land redistribution is efficiency-enhancing; voluntary land sales are not. This is once again a consequence of poverty. If farmers were rich enough, limited liability, and the consequent moral hazard resulting from debt, would not be an issue and there would be no obstacles to land markets emerging.

Other sources of credit

Our analysis so far assumes that the farmer has a single source for all his input requirements – the landlord-cum-creditor. How are things affected if he can secure access to credit from alternative sources? Will the reduced dependence on the landlord shift bargaining power in favor of the tenant, reducing rents and the resultant inefficiencies? The question has both explanatory and policy relevance in the context of LDCs. Development economists have long puzzled over the phenomenon of interlinkage commonly observed in many parts of the developing world. Unlike the specialized markets which characterize industrialized market economies, farmers in poor countries often do one-stop shopping for all their farming-related transactions and even beyond. Typically, a farmer will rent land, take loans, sell his crops to and maybe even buy household supplies from the same individual. Since such multilateral transactions are extremely rare in the developed world, one wonders whether the institution has developed in order to meet certain needs of the environment (and if so, what the economic implications of this are). In recent years, there has also been concerted effort by governments as well as NGOs to extend credit at affordable rates, and without stringent collateral requirements, into the rural sector. This has been attempted by setting up outreach programs administered by government lending agencies, imposing statutory lending requirements on private banks, bank nationalization or subsidizing rural loans disbursed by

non-government organizations. In recent decades, the micro-credit movement⁸ has spanned many countries and has generated a great deal of interest among researchers and policy-makers alike. Underlying a movement of this kind is the belief that credit is a primary bottleneck that arrests agricultural development, and its removal may provide a significant boost to productivity increases and poverty reduction. We will argue that such conclusions should be treated with caution. Infusion of cheap rural credit may be ineffective (and in some situations, counterproductive) if not accompanied by complementary policies like land reform.

Suppose F has obtained all the credit he needs (that is, $1 - w$) from some other source (a specialized private moneylender, a government agency or an NGO). Suppose his debt to this creditor is R_0 , which he must service in the good state (in the bad state, he is still protected from all obligations by limited liability). L will choose the rent R on land to solve:

$$\max_R pR$$

subject to the incentive constraint:

$$c'(p) = q - R_0 - R$$

and the participation constraint:

$$p(q - R_0 - R) - c(p) \geq 0$$

where the incentive constraint is the first-order condition of maximizing the left-hand side expression above, which is the farmer's expected income. If the participation constraint binds at the optimum, we can combine the two constraints to write $pc'(p) - c(p) = 0$, which solves for some value of p , call it \tilde{p} . When this is the case, the effort level is independent of R_0 , the debt owed to third parties. However, consider the case where the participation constraint is not binding at the optimum. Then incorporating the incentive constraint directly into objective function, we get an unconstrained problem:

$$\max_R p[q - R_0 - c'(p)]$$

The solution is described by the first-order condition

$$c'(p) + pc''(p) = q - R_0$$

The second-order condition for a maxima implies that the left-hand side above must be locally increasing in p . This immediately yields the following

comparative static property: the induced effort level p is a decreasing function of R_0 , the third-party debt borne by the farmer.

This simple observation sheds light on a number of different scenarios. If interlinked transactions were replaced by specialized markets, R_0 would be determined by a zero-profit condition if the credit market was perfectly competitive. Alternatively, if credit was supplied by a monopolist money-lender, we would have to close the model by solving for the Nash equilibrium in which the landlord monopolist's choice of rent (R), and the lender monopolist's choice of interest (R_0) are best responses to each other. Regardless, it is clear that since R_0 must be positive, the effort level will be lower than in the case where a single landlord-lender was solving an interlinked problem ($R_0 = 0$). Therefore, interlinkage is socially (as well as privately)⁹ more efficient than separate markets, which explains its prevalence. The same effect would arise if R_0 is the debt owed to a government agency or a micro-credit organization, underscoring the need for caution and careful consideration of the general equilibrium effects before infusing cheap credit into a rural economy. The phenomenon discussed here is analogous to the tragedy of the commons: overall indebtedness worsens the tenant's incentive, but the landlord does not internalize the negative effect of raising his rent on others' expected returns.

Empirical evidence

It is natural to ask whether the theoretical effects we have talked about are also relevant to the environment of the present day LDCs. We believe that the set-up is appropriate for several less-developed countries today. The questions posed in this chapter and the results of the theoretical model echo the findings of some recent empirical literature.

First, there is little doubt that poor farmers in even a country as rapidly growing as India depend significantly on informal local credit markets, and that credit constraints are pervasive. Sahu et al. (2004) present evidence that the access to formal credit is limited in rural India (the State of Orissa), and a high degree of credit rationing is present. Dasgupta (1989) reports results from case studies carried out in two south Indian states. Professional moneylenders supply 45.6 percent of the credit and the average annual interest rate is about 52 percent. Ghate (1992) reports a study from Thailand where the average interest rates range between 80 and 125 percent. The high interest rates indicate that the demand for credit is high and that credit obligations loom large in the decisions of the borrowing farmers.

In support of the claim that land ownership has a big impact on a cultivator's bargaining position, there are several studies that can be cited. Banerjee et al. (2002) find that increasing the tenant's bargaining power and

security of tenure in West Bengal, India resulted in a substantial increase in the productivity of land. Otsuka (1991) finds similar effects in Thailand. In a study of rural Mexico, Finan et al. (2005) demonstrate that access to even a small plot of land improves the welfare of agricultural workers by a great deal.

Deininger et al. (2003) examine the performance of land rental and sales markets in Nicaragua during 1995–98. Despite a strong inverse size–productivity relationship and large differences in land productivity between large and small producers, land markets have been found never to have led to an equalization of returns. Their conclusion is that credit market imperfections impeded the working of land sales markets in Nicaragua.

Anderson (1990) investigates the effects of the introduction of a credit regulation requiring banks to lend a specified volume of credit to small farmers. Her conclusions cast doubt on the effectiveness of such a policy in improving their access to credit.

Conclusion

We believe the way all factor markets work in poor agrarian economies is a function of a few salient features of the environment. Extreme poverty of cultivators and weather-dependent agricultural production translate into all contracts being subject to limited liability. This creates moral hazard on the part of indebted farmers or renters. In an economy where land and hence wealth is unequally distributed, agricultural production takes place inefficiently. The result is low growth rates and the persistence of poverty. Poverty begets poverty. Land markets may not evolve in response to differences in the marginal product of land, because an asset transaction is not Pareto-improving as long as the buyer has to go into debt to acquire the land.

More generally speaking, high transaction costs imply that aggregate output is not independent of the initial distribution. A land reform that improves the bargaining position of the cultivator increases expected output. Of course, land reforms are hardly easy to implement politically; yet there are few substitutes when landlords enjoy monopoly power that can be used to extract the surplus generated by the cheap availability of other inputs. Government provision of credit to tenants may only end up benefiting the landlords, even at the detriment of incentives and productivity. However, any reform that strengthens the bargaining position of the tenants is good for both equity and efficiency.

Notes

1. We thank Nishant Chadha for excellent research assistance on this chapter.
2. Rosenzweig and Binswanger (1993) estimate that in the villages in South India that they studied, a one standard deviation increase in the coefficient of variation of rainfall leads to a 35 percent reduction in the profits of poor farmers.
3. The incentive effects of limited liability were first analyzed in a model of the credit market by Stiglitz and Weiss (1981), to explain endogenous rigidity of interest rates and credit rationing. Of course credit markets are prone to a wide array of opportunistic behavior and informational problems. For a survey of the theoretical issues, see Ghosh et al. (2001).
4. Suppose the optimal contract involves incomplete funding ($0 < x < 1 - w$). Increase x by some Δx , while keeping the farmer's consumption in each state, that is, $[(x + w)q - rx - w]$ and $[(x + w)\bar{q} - \bar{r}x - w]$, constant. This would mean p remains unchanged and consequently the left hand side of the participation constraint remains unchanged. Since the farmer's income in either state is held constant, the creditor's expected profit will change by exactly the amount of the change in social surplus, that is, $\Delta x[pq + (1 - p)\bar{q} - 1] > 0$ as long as cultivation is profitable in the first place. Thus, the creditor has an incentive to give either the full amount of credit (that is, $x = (1 - w)$) or none at all.
5. Suppose $r > \bar{r}$, inducing an effort level p . If, instead, L charges the expected interest rate $pr + (1 - p)\bar{r}$ in both states, either agent's expected income will be unaffected as long as F does not change his choice of p , implying the participation constraint must continue to be satisfied. However, examination of (30.5) reveals that F's optimum effort is now strictly higher, implying his expected income must increase strictly, leaving the participation constraint non-binding. L can then increase the (state-independent) interest rate further, thereby increasing his profit.
6. If the limited liability constraints are non-binding, the participation constraint must be binding, because otherwise there is room for L to increase the interest rate. Taking this to be the case and treating (30.6) as an equality, we get $r = \bar{r} = p^*q + (1 - p^*)\bar{q} - c(p^*) - 1$. This is the optimal solution, therefore, as long as it does not violate (30.8), the tighter of the two limited liability constraints. The constraint binds when it (weakly) violates (30.8), which gives us condition (30.9).
7. What is necessary for the characterization presented is the objective function be single-peaked, that is, decline monotonically on either side of p_0 . This would guarantee that it has no other local optima, and the solution is the closest point to p_0 permitted by the participation constraint. Concavity, which depends on the third derivative properties of $c(\cdot)$ (true if $c'''(\cdot) \geq 0$, for example) is sufficient but not necessary for this to obtain.
8. See Ghatak and Guinnane (1999) and Morduch (1999) for surveys of the literature.
9. If the landlord and the creditor merged into one entity, he could earn the sum of the profits if he kept interest rates and rents unchanged. Generally, however, he can do even better by adjusting these rates, taking into account their negative externality on each other via the farmer's effort.

References

- Anderson, J. (1990), 'Does Regulation Improve Small Farmers' Access to Brazilian Rural Sector?', *Journal of Development Economics*, 33 (1): 67–87.
- Banerjee, A., P. Gertler and M. Ghatak (2002), 'Empowerment and Efficiency: Tenancy Reform in West Bengal', *Journal of Political Economy*, 110 (2): 239–80.
- Basu, K. (1992), 'Limited Liability and the Existence of Share Tenancy', *Journal of Development Economics*, 38: 203–20.
- Binswanger, H. and M. Rosenzweig (1986), 'Behavioural and Material Determinants of Production Relations in Agriculture', *Journal of Development Studies*, 22 (3): 503–39.
- Dasgupta, A. (1989), 'Reports on Informal Credit Markets in India: Summary', New Delhi: National Institute of Public Finance and Policy.
- Deininger, K., E. Zegarra and I. Lavadenz (2003), 'Determinants and Impacts of Rural Land Market Activity: Evidence from Nicaragua', *World Development*, 31 (8): 1385–1404.

- Eswaran, M. and A. Kotwal (1986), 'Access to Capital and Agrarian Production Organization', *Economic Journal*, 96 (382): 482–98.
- Finan, F., E. Sadoulet and A. de Janvry (2005), 'Measuring the Poverty Reduction Potential of Land in Rural Mexico', *Journal of Political Economy*, 77 (1): 27–51.
- Ghatak, M. and T. Guinnane (1999), 'The Economics of Lending with Joint Liability: Theory and Practice', *Journal of Development Economics*, 60 (1): 195–228.
- Ghate, P. (1992), *Informal Finance: Some Findings from Asia*, Oxford, New York, Toronto and Hong Kong: Oxford University Press for the Asian Development Bank.
- Ghosh, P., D. Mookherjee and D. Ray (2001), 'Credit Rationing in Developing Countries: An Overview of the Theory', in D. Mookherjee and D. Ray (eds), *Readings in the Theory of Economic Development*, London: Blackwell, pp. 283–301.
- Kumar, D. (1975), 'Landownership and Inequality in Madras Presidency, 1853–54 to 1946–47', *Indian Economic and Social History Review*, 12: 229–61.
- Marshall, A. (1920), *Principles of Economics*, 8th edn, London: Macmillan.
- Moll, P. (1988), 'Transition to Freehold in the South African Reserves', *World Development*, 16: 349–60.
- Mookherjee, D. (1997), 'Informational Rents and Property Rights in Land', in John Roemer (ed.), *Property Relations, Incentives and Welfare*, London: Macmillan Press and New York: St Martin's Press.
- Morduch, J. (1999), 'The Micro finance Promise', *Journal of Economic Literature*, 37 (4): 1569–1614.
- Otsuka, K. (1991), 'Determinants and Consequences of Land Reform Implementation in the Philippines', *Journal of Development Economics*, 35 (2): 339–55.
- Rosenzweig, M. and H. Binswanger (1993), 'Wealth, Weather Risk and the Competition and Profitability of Agricultural Investments', *Economic Journal*, 103 (416): 56–78.
- Rosenzweig, M. and K. Wolpin (1993), 'Credit Market Constraints, Consumption Smoothing, and the Accumulation of Durable Production Assets in Low Income Countries', *Journal of Political Economy*, 101 (21): 223–44.
- Sahu, O., S. Madheswaran and D. Rajashekhar (2004), 'Credit Constraints and Distress Sales in Rural India: Evidence from Kalahandi District, Orissa', *Journal of Peasant Studies*, 31 (2), pp. 210–41.
- Shaban, R. (1987), 'Testing between Competing Models of Sharecropping', *Journal of Political Economy*, 95 (5), pp. 893–920.
- Stiglitz, J.E. (1974), 'Incentives and Risk Sharing in Sharecropping', *Review of Economic Studies*, 41: 219–55.
- Stiglitz, J.E. and A. Weiss (1981), 'Credit Rationing in Markets with Imperfect Information', *American Economic Review*, 71: 393–410.

31 The Green Revolution

Robert E. Evenson

Introduction

The Green Revolution is the term used to describe the rapid adoption of Green Revolution Modern Varieties (GRMVs) in 1964–65 of rice and wheat in Asia, the Middle East, sub-Saharan Africa and Latin America.¹ Some observers of the Green Revolution suggest that by 1990, the Green Revolution was over.² But this was certainly not the case. The production of GRMVs has been steadily increasing each decade since 1960.

Many developing countries, including many countries in sub-Saharan Africa and Pakistan in South Asia, experienced a tripling of population between 1950 and 2000. For these countries, many of which were already cultivating high proportions of their arable land and many of which were experiencing varying degrees of malnutrition and high rates of child mortality, prospects were grim indeed. Many observers applied the ‘carrying capacity’ model favored by many biologists and concluded that food consumption per capita would inevitably decline.

But because of the Green Revolution, food production per capita actually increased in most developing countries (even countries not realizing a Green Revolution; see Figure 31.1 and Table 31.1 below). In Pakistan, calories consumed per capita (calculated by the Food and Agriculture Organization – FAO) actually rose by 40 percent from 1960 to 2000 as a result of the Green Revolution.

The Green Revolution had its origins in The Monetary and Financial Conference at Bretton Woods, New Hampshire in June 1944. The World Bank and the International Monetary Fund were set up at Bretton Woods, and many UN agencies were established at the same time. The General Agreement on Tariffs and Trade (GATT) soon followed as did many bilateral aid agencies (the United States Agency for International Development – USAID, the Canadian International Development Research Centre – IDRC, the German Gesellschaft für Technische Zusammenarbeit – GTZ, and so on). The age of globalization was well established before the first International Agricultural Research Center (IARC) was built.

By the mid-1950s, the nature of the ‘population explosion’ was apparent. With a fall in death rates preceding a fall in birth rates, population growth was inevitable. Many countries introduced programs to distribute contraceptives by the late 1950s in an attempt to reduce birth rates. These

< 2%	2–10%	10–20%	20–30%	30–40%	40–50%	50–65%	> 65%
Afghanistan	Burkina Faso	Bolivia	Colombia	Cuba	Dominican Republic	Algeria	Argentina
Angola	Cambodia	Benin	Costa Rica	Egypt	Iran	Bangladesh	Chile
Burundi	Chad	Botswana	Ecuador	Mexico	Kenya	Brazil	China
Central African Republic	El Salvador	Cameroon	Ghana	Namibia	Morocco	Myanmar	India
Congo (B)	Gabon	Congo (Z)	Laos	Paraguay	Nepal	Tunisia	Indonesia
Gambia	Guatemala	Côte d'Ivoire	Madagascar	Peru	Thailand		Malaysia
Guinea Bissau	Guinea	Ethiopia	Mali	Saudi Arabia	Turkey		Pakistan
Mauritania	Haiti	Liberia	Sierra Leone	South Africa			Philippines
Mongolia	Jamaica	Honduras		Syria			Sri Lanka
Niger	Libya	Mauritius					Vietnam
Somalia	Malawi	Nicaragua					
Yemen	Mozambique	Nigeria					
	Panama	Rwanda					
	Senegal	Sudan					
	Swaziland	Tanzania					
	Togo	Uruguay					
	Uganda	Venezuela					
	Zambia	Zimbabwe					

Figure 31.1 Green Revolution clusters by GRMV adoption level in 2000

Table 31.1 Green Revolution Cluster Indicators

Economic indicators

Clusters By GRMV Adoption	Crop value per Ha (dollars)	Fertilizer per hectare (kg/ha)	Crop TFP growth (1961–2000)	Scientists per million Ha cropland		Extension work per million Ha		Industrial competitiveness (UNIDO)	
				1960	2000	1960	2000	1985	1998
<2%	78	2	0.09	0.019	0.030	0.230	0.461	0.002	0.002
2–10%	128	22	0.72	0.018	0.093	0.392	0.402	0.020	0.028
10–20%	94	6	1.07	0.013	0.033	0.149	0.220	0.028	0.029
20–30%	112	12	0.87	0.033	0.076	0.245	0.416	0.037	0.051
30–40%	180	40	1.30	0.033	0.179	0.070	0.371	0.050	0.076
40–50%	227	52	0.96	0.023	0.063	0.287	0.827	0.038	0.072
50–60%	300	68	1.36	0.050	0.063	0.070	0.140	0.060	0.080
>65%	488	166	1.56	0.079	0.120	0.150	0.442	0.047	0.111

Table 31.1 (continued)

Social indicators

Clusters By GRMV Adoption	Countries in class	Population in 2000 (millions)	Population (millions)		Birth rates (millions)		Child mortality rates (millions)		Dietary energy sufficiency		GDP per capita	
			1960	2000	1960	2000	1960	2000	1960	2000	1960	2000
<2%	12	75	2.2	6.1	47	41	293	160	2029	2192	361	388
2-10%	18	153	3.1	8.5	45	36	236	118	2074	2387	815	1291
10-20%	18	385	7.0	21.4	44	36	214	134	1983	2282	866	1295
20-30%	8	115	9.0	14.3	46	32	238	124	2070	2384	695	1156
30-40%	9	337	14.3	37.4	42	26	156	27	2050	2574	1169	3514
40-50%	2	284	15.5	40.3	46	26	221	61	2084	2506	805	1660
50-60%	5	385	34.9	76.7	46	23	240	50	2038	2391	1096	2153
>65%	10	2886	135.1	288.6	39	22	165	43	2100	2719	1049	2305

programs were successful, but as a practical matter most developing countries experienced rapid demographic transitions, and birth rates have now been reduced in most, if not all, developing countries.

In the mid-1950s many international policymakers evaluated the National Agricultural Research System (NARS) programs, then in place in most developing countries in Asia, the Middle East and North Africa, and Latin America, few were in place in sub-Saharan Africa because former colonies in Africa did not obtain independence until after 1960. This has been a major factor in explaining why many countries in sub-Saharan Africa are 'failed states', unable to deliver basic public goods to their populations. Most policy-makers concluded that NARS programs needed to be supplemented by a system of International Agricultural Research Centers (IARCs).

The first nominal IARC was the International Rice Research Institute, established in Los Banos, Laguna in the Philippines in 1959. But the de facto first IARC was based on the wheat breeding program established in 1943 by the Rockefeller Foundation in Mexico and led by Norman Borlaug. In 1952 Borlaug acquired the semi-dwarf wheats from Oregon State University that were to become the basis for the first 'modern' wheat varieties.³ Accordingly, the International Wheat and Maize Improvement Center was de facto the first IARC. Several other IARCs followed the establishment of the first two IARCs.⁴ Note that all of these IARCs are located in developing countries, an important factor in their success.

Expenditures in NARS and IARCs

Table 31.2 reports expenditure in constant US dollars by developed and developing countries. Data for the International Agricultural Research Centers are reported in Table 31.2. Data for 1995 are reported for private sector research and development (R&D) in agriculture. In 1995, private sector R&D in developed countries was roughly equal to public sector R&D and now exceeds it in developed countries. Private sector R&D is unimportant in developing countries.

Table 31.3 expresses private and public sector expenditure as a share of agricultural gross domestic product (GDP) and on a per capita basis. It is clear that developed countries have considerably higher expenditure as a share of agricultural GDP than do developing countries.

Green Revolution Modern Variety Adoption

The Green Revolution was primarily a 'crop yield' revolution.

GRMV adoption rates

Table 31.4 reports Green Revolution Modern Variety (GRMV) adoption rates for 11 Green Revolution crops as of 2000.⁵

Table 31.2 Global expenditure on agricultural research in 1995 (millions 2001 US dollars)

	1965	1976	1985	1995
Public sector agricultural research				
Developed countries	6532	8270	10192	11900
Developing countries				
China	377	709	1396	2036
Other Asia	441	1321	2453	4619
Middle East & North Africa	360	582	981	1521
Latin America & Caribbean	562	1087	1583	1947
Sub-Saharan Africa	472	993	1181	1270
International agricultural research centers	12	163	315	400
Private sector R&D in agriculture				
Developed countries				10829
Developing countries				672

Source: Pardey and Beintema (2001) and Boyce and Evenson (1975).

Table 31.3 Public agricultural research intensities

	Expenditure as a share of agricultural GDP			Expenditure per capita		
	1976	1985	1995	1976	1985	1995
Developed countries	1.53	2.13	2.64	9.6	11.0	12.0
Developing countries	0.44	0.53	0.62	1.5	2.0	2.5
China	0.41	0.42	0.43	0.7	1.3	1.7
Other Asia	0.31	0.44	0.63	1.1	1.7	2.6
Latin America and Caribbean	0.55	0.72	0.98	3.4	4.0	4.6
Sub-Saharan Africa	0.91	0.95	0.85	3.5	3.0	2.0

Source: Pardey and Beintema (2001), Evenson Estimates for sub-Saharan Africa.

The development of GRMVs

There is general agreement that GRMV development passes through the following stages:

- MV1: This stage is a very challenging stage, because it entails the development of a higher-yielding plant type for a region. This has

Table 31.4 Green Revolution Modern Variety (GRMV) adoption rates as of 2000

	Region			
	Latin America & Caribbean	Asia	Middle East & North Africa	Sub-Saharan Africa
Wheat	90	86	66	66
Rice	65	74	–	40
Maize	45	82	–	52
Sorghum	–	70	–	26
Pearl Millet	–	78	–	19
Barley	–	–	49	–
Beans	20	–	23	15
Lentils	–	–	23	–
Groundnuts	–	5	–	49
Potatoes	84	90	–	78
Cassava	5	10	–	16

now been achieved for all Green Revolution crops, but not for all production regions. In the case of rice and wheat, dwarfing genes played a role because the new plant type did not lodge when fertilizer was applied. For other crops, dwarfing genes were not important.

- MV2: Virtually all MV1 varieties were susceptible to insect pests and to plant diseases. In most crops ‘landraces’ in the cultivated species provided both insect resistance and disease resistance. But for some crops, interspecific hybridization was required.⁶ For example, for rice the only source of resistance to Grassy Stunt (a virus disease) was the uncultivated species, *Oryza Nivara*. Interspecific hybridization techniques have now been developed for all major crop species.
- MV3: At a relatively advanced stage of breeding, plants can be bred to withstand ‘abiotic stresses’ (drought, excess rainfall, and so on).

All Green Revolution crops have proceeded through the MV2 stage and several have entered the MV3 stage.

Of all GRMVs developed it may be noted that:

- No GRMVs were developed by non-governmental organizations (NGOs).
- Approximately 5 percent of GRMVs were developed by private sector hybrid seed companies. But private firms pursued hybridization

programs only after IARCs developed improved open-pollinated varieties (OPVs).⁷

- 36 percent of GRMVs were the result of an IARC cross.
- 59 percent of GRMVs were the result of a NARS cross. Of the NARS-crossed varieties, one-third had an IARC-crossed parent.

Economic impacts of GRMVs

GRMV–traditional variety conversion rates

Economic impacts require that the ratio of GRMV yields to traditional variety yields be computed. Table 31.5 reports these conversion rates.

Growth rates of food production, area and yield by region and period

Table 31.6 reports a growth accounting exercise by major production region. This is done for two periods: the early Green Revolution, 1961–80 and the late Green Revolution, 1981–2000. The calculations include the eleven Green Revolution crops.

Since production is equal to area times yield, the growth in production (in tonnes) can be expressed as the growth in area cropped plus the growth in yield, and since the Crop Genetic Improvement (CGI) component can be obtained from Table 31.6, the growth in yield can be decomposed into a CGI component and an ‘other inputs’ component.

A striking feature of Table 31.6 is that for both periods, production growth exceeded population growth, except for sub-Saharan Africa in the early period (recall that this period includes the Nigerian food production decline associated with high oil prices). A second feature of note is that the CGI contribution to production was higher in the late Green Revolution

Table 31.5 GRMV/Traditional variety conversion rates

Crop	Conversion rate
Wheat	0.45
Rice	0.47
Maize	0.50
Sorghum	0.45
Pearl millet	0.45
Barley	0.41
Lentils	0.41
Groundnuts	0.40
Beans	0.25
Potatoes	0.35
Cassava	0.48

Table 31.6 Growth rates of food production, area, yield, and yield components, by region and period

	Early Green Revolution 1961–80	Late Green Revolution 1981–2000
<i>Latin America</i>		
Production	3.083	1.631
Area	1.473	-0.512
Yield	1.587	2.154
MV contributions to yield	0.463	0.772
Other input/Ha	1.124	1.382
<i>Asia</i>		
Production	3.649	2.107
Area	0.513	0.020
Yield	3.120	2.087
MV contributions to yield	0.682	0.968
Other input/Ha	2.439	1.119
<i>Middle East – North Africa</i>		
Production	2.529	2.121
Area	0.953	0.607
Yield	1.561	1.505
MV contributions to yield	0.173	0.783
Other input/Ha	1.389	0.722
<i>Sub-Saharan Africa</i>		
Production	1.697	3.189
Area	0.524	2.818
Yield	1.166	0.361
MV contributions to yield	0.097	0.471
Other input/Ha	1.069	-0.110
<i>All developing countries</i>		
Production	3.200	2.192
Area	0.683	0.386
Yield	2.502	1.805
MV contributions to yield	0.523	0.857
Other input/Ha	1.979	0.948

Notes: Data on food crop production and area harvested are taken from FAOSTAT data, revised 2003 (<http://apps.fao.org/page/collections?subset=agriculture>) on total cereals, total roots and tubers, and total pulses. Asia consists of 'Developing Asia' less the countries of the 'Near East in Asia'. Africa consists of 'Developing Africa' less the countries of the 'Near East in Africa' and the countries of 'North-West Africa'. The Middle East-North Africa consists of 'Near East in Africa', 'Near East in Asia', and 'North-West Africa'. Latin America includes Latin America and the Caribbean. Crop production is aggregated for each region using area weights from 1981. Estimates of production increases due to MVs are from (4). Growth rates of other inputs are taken as a residual. Growth rates are compound and are computed by regressing time series data on a constant and trend variable. The totals for 'All developing countries' are derived by weighting the regional figures by 1981 area shares.

than in the early Green Revolution in all regions. The Green Revolution has not run its course by any means. A further feature of Table 31.6 is the disparity in CGI contributions between regions. Asia and Latin America realized large CGI gains over both periods. The Middle East and North Africa region realized significant gains in the late period. But sub-Saharan Africa realized miniscule gains in the early period and only modest gains in the late period.

The area contributions to production are also of interest. By the late period, Latin American countries were reducing the area planted to food crops, as were all developed Organisation for Economic Co-operation and Development (OECD) countries. Asia had virtually ceased expanding cropped area in the late period. The Middle East and North Africa region reduced the area contribution, but it remains high. Sub-Saharan Africa, by contrast, realized most of its production growth from expanded crop area. Furthermore, the Green Revolution MVs in sub-Saharan Africa in the late period were not accompanied by increased input use.

Thus, even though these data show that sub-Saharan Africa is finally realizing some Green Revolution gains, the nature of these gains is disquieting. Most production gains will soon be exhausted. Perhaps more relevantly, the absence of increased fertilizer use in sub-Saharan Africa suggests that GRMVs bred for Africa were not fertilizer-responsive.

Returns to investment in IARC and NARS programs

In the classic work of Griliches (1957) on hybrid maize, a benefit–cost analysis was performed. This requires a cost series $\{c_t\}$ over time and a benefit series $\{b_t\}$ over time. It is possible to construct a series for each region from International Service for National Agricultural Research (ISNAR) data on research expenditure and estimates of the CGI share of the expenditures. This cost series can be constructed for the 1950–2000 time period. The data on GRMV adoption and impact can be used to construct the benefits series, $\{b_t\}$. The cost and benefit series can then be utilized to calculate the following:

- PVB: the present value of the benefits stream computed at a specific interest rate (I use 6 percent)
- PVC: the present value of the costs stream computed at the same specific interest rate.
- $B/C = PVB/PVC$, the benefit–cost ratio.
- IRR: the rate of interest at which $PVB = PVC$.

Table 31.7 reports the Internal Rate(s) of Return (IRRs) for both NARS crop improvement programs and IARC crop improvement programs, by

Table 31.7 *Estimated B/C ratios and internal rates of return from Green Revolution contributions*

Region	NARS B/C	NARS IRRs	IARC B/C	IARC IRRs
Latin America	56	31	34	39
Asia	115	33	104	115
West Asia–North Africa	54	22	147	165
Sub-Saharan Africa	4	9	57	68

Source: Evenson calculations.

region. Note that these estimates include long periods of costs where few benefits are achieved. For example, for sub-Saharan Africa, benefits exceeded costs almost 15 years later than was the case for Latin America and Asia.

The IARC program Benefit/Cost (B/C) ratios and IRRs are very high. These high IRRs appear to be very real and they reflect the ‘leveraging’ associated with the high production of IARC crosses and the high volume of IARC germplasm delivered to NARS programs. NARS B/C ratios and IRRs are high except in sub-Saharan Africa where long periods of investment without benefits occurred.

The uneven delivery of the Green Revolution

Figure 31.1 lists 87 countries classified according to aggregate Green Revolution Modern Variety (GRMV) adoption rates in 2000. The 12 countries in the first column report negligible GRMV adoption in the year 2000. All other classes are based on area-weighted GRMV adoption rates for the 11 crops included in the GRMV study.

Table 31.1 lists indicators by Green Revolution cluster. The clusters can be roughly categorized as Non-Performing (Cluster 1), Underperforming (Clusters 2 and 3) and Performing (Clusters 4, 5, 6, 7 and 8). Economic and social indicators by cluster are reported in Table 31.7. The economic indicators show:

1. Crop value (in US dollars) per hectare is very low for countries not realizing a Green Revolution and rises to high levels for countries realizing the highest levels of GRMV adoption.
2. Fertilizer application per hectare is negligible for the first four clusters and significant for the highest GRMV clusters.
3. Crop TFP growth is negligible for countries not realizing a Green Revolution and highest for countries with the highest levels of GRMV adoption.⁸

4. Countries without a Green Revolution did have both agricultural scientists and extension workers. Scientists per million hectares of cropland rise with higher levels of GRMV adoption.
5. Extension workers per million hectares of cropland are roughly 20 times as great as scientists per million hectares of cropland. The number of extension workers increased in every cluster. No correlation between extension workers per million hectares of cropland and GRMV adoption exists.
6. None of the countries without a Green Revolution has industrial competitiveness. A UNIDO index of 0.05 or greater indicates industrial competitiveness. Only countries in 30–40 percent GRMV clusters and above have industrial competitiveness. Improvement in industrial competitiveness is greatest for the highest GRMV clusters.⁹

The social indicators show the following:

1. Sixty-three percent of the 4.65 billion people living in developing countries are located in the ten countries in the highest Green Revolution cluster. Eighty-four percent live in performing clusters (countries classified according to aggregate Green Revolution Modern Variety (GRMV) adoption rates). Countries without a Green Revolution make up less than 2 percent of the total population in developing countries.
2. The average population of countries in 1960 and 2000 rises as GRMV adoption rises. This suggests a strong bias against small countries.
3. In 1960, birth rates were similar across GRMV clusters. By 2000, birth rates had declined in all GRMV clusters, with highest declines in the highest GRMV clusters.
4. Child mortality rates in 1960 were similar in most GRMV clusters. By 2000, they had declined in all GRMV clusters with highest declines in the highest GRMV clusters. In the top two GRMV clusters, child mortality rates in 2000 were only 24 percent of their 1960 levels.
5. Dietary Energy Sufficiency (DES) was similar for all GRMV clusters in 1960. By 2000, improvements were achieved in all clusters with highest improvements in highest GRMV clusters. DES improvement is highly correlated with child mortality reduction.
6. Gross domestic product (GDP) per capita (using exchange rate conversion to dollars, Atlas Method) was lowest in countries without a Green Revolution in 1960 and did not improve in 2000. GDP per capita for the next three GRMV clusters rose by 56 percent from 1960 to 2000. GDP per capita for the highest four GRMV clusters rose by 140 percent from 1960 to 2000.

NARS programs in specific countries bear the ultimate responsibility for failing to deliver GRMVs to their farmers. But IARC programs are not immune from criticism. There are three IARCs located in Africa: the World Agroforestry Centre (ICRAF) in Kenya, the International Livestock Research Institute (ILRI) in Ethiopia and Kenya, and the International Institute of Tropical Agriculture (IITA) in Nigeria. ICRAF has had little impact because agroforestry generates little income for farmers. ILRI has also had little impact although it does not deal with crops. IITA has had an impact only after developing breeding programs with the Centro Internacional de Mejoramiento de Maiz y Trigo (CIMMYT) for maize and with the Centro Internacional de Agricultura Tropical (CIAT) for cassava. Similarly, the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) had little impact until sorghum, millet and groundnut breeding programs were developed in Africa.

Why did 12 countries fail to produce a Green Revolution? A closer examination suggests three explanations. The first is the 'failed state' explanation. The second is the 'small state' explanation. The third is the 'civil conflict' explanation. Most or all of the countries failing to deliver a Green Revolution to their farmers are effectively failed states. They cannot manage to 'deliver the mail', much less produce a Green Revolution. But they are also small states with an average population of 2.5 million people in 1960 (Angola and Yemen had 5 million people in 1960). None have universities to train agricultural scientists. All have been in civil conflict for much of the past 40 years. Given low GDP per capita, limited taxing power and civil conflict, it is not surprising that they did not produce Green Revolutions.

The second GRMV cluster did have a small Green Revolution, but they too are small countries (Mozambique and Uganda being the largest, with populations around 7 million in 1960). Most of these countries have also been in civil conflict. Few have universities to train agricultural scientists, but they did manage a small Green Revolution.

Figure 31.2 depicts 'real' prices for the 1960 to 2000 period (a five-year moving average; IFPRI). The prices of rice, wheat and maize in 2000 were approximately 45 percent of their 1960 levels (35 percent of their 1950 levels). The real prices of the world's major cereal grains have been declining by more than 1 percent per year for the past 50 years.

In the OECD developed countries, it is estimated that total factor productivity (TFP) rates (a measure of cost reduction in agriculture) have been roughly 1 percent per year higher than in the rest of the economy. For developing countries, crop TFP growth rates have been high except for countries in the lowest GRMV clusters. A few of the industrially competitive countries have had industrial TFP growth rates that are higher than agricultural TFP growth rates.

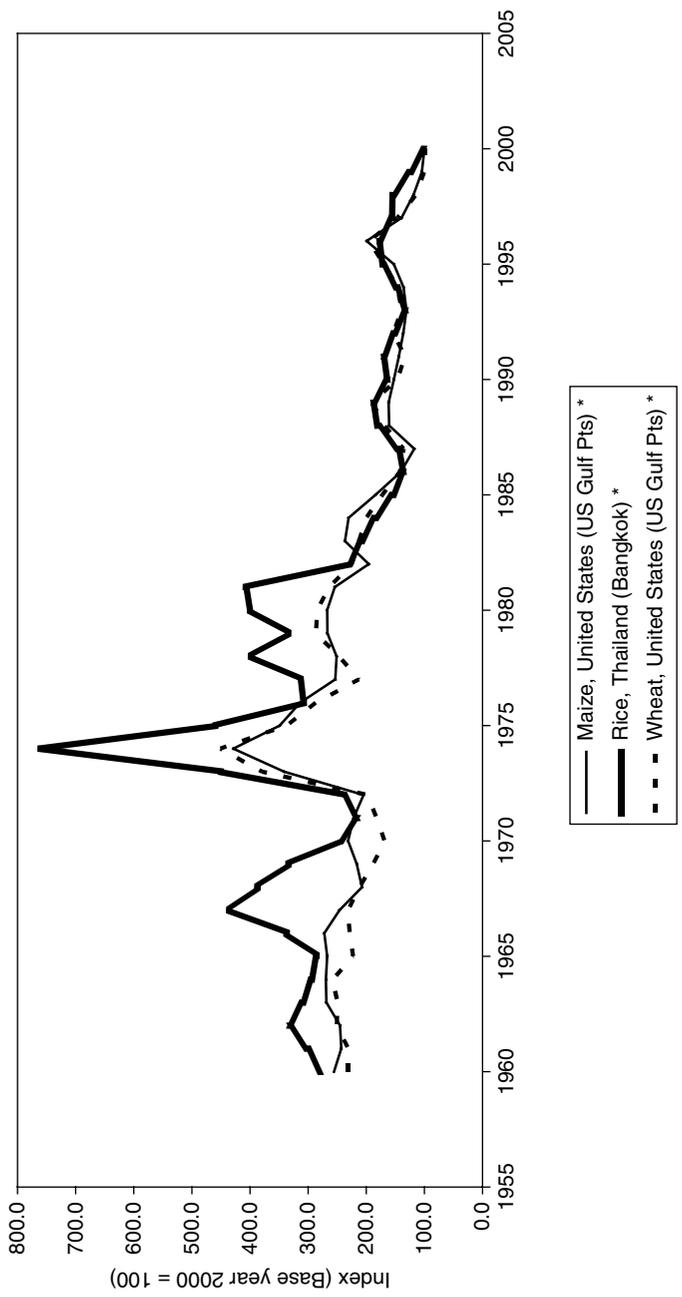


Figure 31.2 World grain prices, 1960–2000

Why then do we have ‘hunger in a world awash with grain’? For this we need only look at crop value per hectare in Table 31.7. With low crop yields, crop value per hectare is low. The highest GRMV cluster produces more than six times as much crop value per hectare as does the lowest cluster. At 1960 prices, farmers in sub-Saharan Africa with 1.2 hectares could earn \$2 per day per capita. At 2000 prices with 0.8 hectares, farmers in sub-Saharan Africa can earn only \$1 per day per capita. Farmers in a number of countries have been delivered price declines without cost declines, and many have moved from mass poverty to extreme poverty.

Critics of the Green Revolution

There were two sets of criticisms of the Green Revolution. The first was traditional Marxist–Leninist criticism. The second was environmental criticism. Curiously, neither form of criticism dwelt on the uneven delivery of the Green Revolution.

The Marxist-Leninist critics made the following points:

1. New technology is monopolized by large farms.
2. Small farms are unable to take advantage of GRMVs because they are credit-constrained.
3. Large farms are likely to expand farm size.
4. Large farms are likely to purchase large-scale farm machinery.

Ruttan effectively shows that these criticisms were misplaced.¹⁰ He noted that small farmers adopted GRMVs as rapidly as large farmers and that most farms in Asia are small farms. He further noted that large farms actually use more labor than small farms to produce GRMVs. He further noted that large farms adopted tractors before the introduction of GRMVs and that mechanization was determined by the relative price of labor and machines. And, finally, farm incomes improved when GRMVs were adopted.

The second set of criticisms was by environmentalists. This criticism was based on ‘crop intensification’, particularly of rice and wheat GRMVs. Because both rice and wheat GRMVs were based on a ‘plant type’ designed to use more fertilizer, fertilizer use increased (see Table 31.7, where fertilizer use is low in countries not achieving a Green Revolution and high in countries with high levels of GRMV adoption). Yes, increased fertilizer use does raise the possibility of ‘fertilizer run-off’, but many countries (for example, the Netherlands) have fertilizer application rates more than double those of the highest Green Revolution clusters.

There was also a concern that herbicide use would be increased, but herbicide use is determined by relative wages, and most developing countries in Asia weed with hand labor rather than by using herbicides.

The Green Revolution: a summary

There is little doubt that the Green Revolution made a huge difference. The populations of developing countries increased by factors of 2.5 to more than 3 from 1950 to 2000. This was of major concern to Malthusianists. But because of increased food production, food consumption per capita increased in almost all countries. Even countries importing food grains could import more because the real prices of food grains in world markets declined.

But not all countries realized Green Revolutions, and the consequences for these countries realizing no Green Revolution or minor Green Revolutions have been severe. In effect, these countries have been falling behind the successful developing countries and as a result, income disparities in the developing world have been widening.

Notes

1. Ultimately, the Green Revolution expanded to other crops, including maize, sorghum, millets, barley, beans, groundnuts, lentils, potatoes and cassava.
2. Developed countries realized Green Revolutions in the first half of the twentieth century.
3. Both the first-generation rice and wheat GRMVs utilized dwarfing genes to create new plant types.
4. These include the International Potato Center (CIP) in Lima, Peru; the International Livestock Research Institute (ILRI) in Addis Ababa, Ethiopia and Nairobi, Kenya; the International Center for Tropical Agriculture (CIAT) in Cali, Colombia; the West African Rice Development Association (WARDA) in Bouaké, Côte d'Ivoire (now forced to relocate by civil violence). The International Institute for Tropical Agriculture (IITA) in Ibaden, Nigeria; The International Crop Research Institute for the Semi-Arid Tropics (ICRISAT) in Hyderabad, India; and the International Center for Agriculture in the Dry Areas (ICARDA) in Aleppo, Syria.
5. GRMV Adoption rates by country are available from the author.
6. Species are defined by 'breeding barriers', that is, a cross between two species will not produce progeny, but for closely related species (species in a genus) it is possible to combine species by 'embryo rescue' techniques (the embryo is shifted to another medium shortly after it is formed) and other techniques.
7. The varieties developed by private sector hybrid seed companies should not be confused with varieties developed by interspecific hybridization techniques. Hybridization of maize, sorghum, millets and rice is achieved to take advantage of the heterosis effect. This requires a sequence of 'selfing' (crossing lines with themselves) and outcrossing. Most hybrids, including rice, now developed are based on 'male sterile' genes. (See Evenson and Kislew, 1975 and Kortum, 1997 for the search model applied to GRMVs.)
8. Crop TFP growth is reported in Avila and Evenson (2004).
9. None of the countries without a Green Revolution reported investing in industrial R&D in 1970. The Central African Republic reported industrial R&D in 1990. Of the 18 countries in the 2–10 percent cluster, five reported industrial R&D in 1970, 12 reported industrial R&D in 1990.
10. See Ruttan (2004).

Bibliography

- Avila, A.F.D. and R.E. Evenson (2004), 'Total Factor Productivity Growth in Agriculture: The Role of Technological Capital', in R.E. Evenson and P. Pingali (eds), *Handbook of Agricultural Economics*, Vol. 4, Amsterdam: Elsevier-North Holland.

- Boyce, J.K. and R.E. Evenson (1975), *Agricultural Research Extension Programs*, New York: Agricultural Development Council, Inc.
- Evenson, R.E. and D. Gollin (eds) (2003a), *Crop Variety Improvement and its Effect on Productivity: The Impact of International Agricultural Research*, Wallingford: CABI Publishing.
- Evenson, R.E. and D. Gollin (2003b), 'Assessing the Green Revolution, 1960 to 2000', *Science*, 300: 758–62.
- Evenson, R.E. and Y. Kislev (1976), 'A Stochastic Model of Applied Research', *Journal of Political Economy*, 84 (2): 256–81.
- Griliches, Z. (1957), 'Hybrid Corn: An Exploration in the Economics of Technological Change', *Econometrica*, 25: 501–22.
- Kortum, S. (1997), 'Research Patenting and Technological Change', *Econometrica*, 65: 1389–419.
- Pardey, P.G. and N.M. Beintema (2001), 'Slow Magic: Agricultural R&D a Century after Mendel', *IFPRI Food Policy Report*, Washington, DC: International Food Policy Research Institute.
- Ruttan, V.W. (2004), 'Controversy about Agricultural Technology: Lessons from the Green Revolution', *International Journal of Biotechnology*, 6 (1): 43–54.

32 Industry and industrial policy

Helen Shapiro

This chapter will highlight how the rationales, objectives and instruments of industrial policy, along with the criteria for success, have changed over time. It will also make an explicit comparison between the strategies of East Asia and Latin America. These are the most industrialized regions within the developing world, accounting for 80 percent of manufacturing value added (Lall et al., 2004). East Asia's income and manufacturing growth rates have surpassed those of Latin America, and much of the literature on industrial policy engages in explicit comparison between the regions and offers explanations for their diverging performance.

Why industrial policy?¹

In their arguments promoting government intervention, many early development economists focused on a 'missing factor' – capital, technology, entrepreneurship – which was unlikely to emerge from market forces alone. Therefore, different methods were required to elicit these missing ingredients for growth. Imperfect capital markets, for example, were unlikely either to generate sufficient savings or allocate them efficiently without some form of market intervention. Technological and pecuniary externalities lead to underinvestment; Nurkse (1953), Rosenstein-Rodan (1943) and Scitovsky (1954) argued that government needed to coordinate investment decisions and promote a 'Big Push'. In addition, investors' expectations were often based on past experience, requiring some kind of 'inducement' mechanism to elicit investment in new industrial activities (Hirschman, 1958, 1967).

There was broad consensus around the basic assumption that development required non-marginal change that market forces alone could not generate. There were two other implicit, but ultimately questionable, assumptions that experience would later make apparent. The first had to do with the nature of technological change. The development process was typically portrayed as one of factor accumulation, and technology, like labor and capital, was viewed as just another missing factor. Embodied in capital, it could be imported and, assuming fixed-technology production functions, applied in the same methods as in the country of origin. The second had to do with the state and technocratic omniscience. State planners, armed with input–output tables from industrialized countries like the United States,

and given the assumptions about technology, could simply allocate resources accordingly and leapfrog into the modern industrial era.

The first neoclassical reaction to state-guided industrialization aimed to show that industrial policies were inefficient and that 'distorted' policy regimes were correlated with poor economic performance.² Their claims that dynamic gains could be had from free trade were bolstered by the success of export-oriented countries such as South Korea and Taiwan. In the 1980s, a second wave of critique attacked the early development economists' implicit belief in the efficacy of government intervention. Various models of the interaction between the state and private actors pointed to the possibility that 'bureaucratic failure' could be worse than 'market failure'. Krueger (1974) argued how quantitative restrictions on imports led firms to compete for import licenses and their attached rents, thereby squandering resources in unproductive, rent-seeking activities.

Empirical findings and emerging consensus

Although early in this debate, some claimed that the East Asian newly industrialized countries (NICs) had relatively free trade and non-interventionist governments, it soon became clear that the governments were extremely interventionist. Subsequently, a huge literature has documented how all late-industrializing countries followed quite similar strategies and relied on the same policy instruments to kick-start industrialization in the 1950s and 1960s. They all tried to substitute imports with domestic production and used government planning to target priority sectors. They used selective protection (tariffs, quotas, import licensing and foreign exchange rationing), domestic content requirements and subsidized credit. The public sector had a large role in capital formation that diminished over time; what is more, each country targeted a similar group of basic and mid-technology industries.³

A key difference among these countries was how fast and how extensively they moved into manufacturing exports. While some inward-oriented countries such as Mexico and Brazil grew at fast rates during the 1960s and 1970s, the East Asian export-oriented countries grew even faster.⁴ Although Latin American manufactured exports also grew in the 1970s, they were a much smaller share of both total manufacturing value added and of gross domestic product (GDP). They also failed to keep up with imports, as the region entered into the balance-of-payments crises of the late 1970s and early 1980s.

Based on the conclusion that East Asian success was due to its outward orientation, and in the wake of the debt crisis in Latin America, countries were encouraged by the World Bank and others to liberalize trade. The assumption was that the anti-export bias of import-substitution policies, along with the lack of domestic competition, discouraged innovation and

encouraged rent-seeking behaviors. These micro inefficiencies, in turn, had led to macro imbalances and slower growth rates.

Subsequent work by Rodrik (1995a) and others point out how these assumptions about the gains from trade are open to question. The static efficiency costs of import substitution turn out to be relatively small and cannot explain slower growth. The dynamic learning effects from trade in East Asia are also open to dispute. Rodrik argues that in the case of South Korea's innovative firms, causation may have been from efficiency to exporting, rather than the other way around.

Additional work by Amsden (1989, 1994), Fishlow et al. (1994) and Wade (1990) also countered the emphasis on outward orientation and focused on the efficacy of East Asia's selective interventions. In this framework, exports are a reflection of their governments' superior 'reciprocal control mechanisms'. All these governments required some kind of performance targets in exchange for special favors – ranging from exports to domestic content, research and development (R&D) spending or financial arrangements – but they were not as extensive or effective.

The conclusion that selective industrial policies led to East Asian success is by no means universally accepted.⁵ However, to the extent that their contributions are seen as consequential, the conditions that allowed for their efficacy are seen as non-replicable. The capacity of governments elsewhere to enforce reciprocity commitments is questioned; markets are therefore required to enforce this discipline on firms. The focus in policy-making once again shifted to state, rather than market, failure, just when the theoretical development literature began to move in the opposite direction.

Theory and practice divide

In contrast to the 1960s, a kind of schizophrenia began to emerge in the 1980s and 1990s as theory and practice moved in opposite directions. Governments in Latin America and elsewhere, often encouraged by multilateral institutions, weakened or dismantled the public institutions associated with state-led industrialization and liberalized trade. Pressure mounted on East Asian countries to do the same, although they moved more slowly in this direction.

Just as these reforms were being vigorously promoted, their theoretical foundations were being undermined. Many of the underlying assumptions about market failure which motivated industrial policies of the 1960s have made a comeback in development economic theory. In addition, new approaches to technical change and innovation have challenged previous assumptions about firm behavior. Together, they have generated a huge literature documenting how market forces will not produce optimal

results and that some kind of state intervention is necessary to promote industrialization.

Although repackaged in formalized models, the arguments behind coordinating investment or a 'Big Push' have changed little since first proposed by Rosenstein-Rodan, Nurkse, and Scitovsky more than 50 years ago. The notion that countries can be stuck in a low-level equilibrium trap has also made a comeback, as it has been shown that multiple equilibria can exist in the face of pecuniary externalities driven by increasing returns. What this work suggests, in contrast to traditional models of comparative advantage, is that a country's specialization pattern determines its rate of growth. As Ros (2000, 2001) explains, specializing in sectors with increasing returns allows for a higher return on capital and subsequently, a higher investment rate. This literature also offers new explanations for the success of East Asia and the relative failure of Latin America that have focused not on prices or exports but on coordinated investment programs (Murphy et al., 1989) and/or policy interventions which sped up the transition from one pattern of production to another (Ros, 2001).

The acknowledgement that sectors are not all equal in a world of differential returns to labor and capital reflects the insights from the literature on firm strategy and competitiveness. In contrast to the passive price-taking firms assumed in comparative static analysis, this literature portrays successful firms as those that create and maintain barriers to entry and the rents associated with them. By exploiting 'competitive' advantages based on innovation, firms are then not dependent on unsustainable cost advantages such as low wages or exchange rates (Porter, 1980). By extension, a 'competitive' nation does not specialize in sectors dependent on low wages or exchange rates, either. In explicit contrast to theories of comparative advantage, a country's competitive advantage is determined by innovation rather than factor endowments (Porter, 1990).

A related literature on late-developing countries also puts firms and their technological capacity at the heart of development.⁶ These countries' ability to shift away from primary resources to knowledge-based capabilities so that they can produce things at lower cost determines their long-term growth. In contrast to the standard emphasis on getting the macro right, the starting point is the firm.

The treatment of technology also distinguishes this work from both current and early development economists. Rather than a missing factor akin to capital or labor, knowledge or technology is portrayed as a learning process. In a world of imperfect information and technology rents, the firm is not a competitive price-taker implicit in most macro approaches. Moreover, public support is crucial to help build firms' technological capabilities. In recent decades, research and development capabilities have become even

more central as the competitive pressures to be near the technological frontier have increased with the fall in trade barriers and in transportation, communication and information costs (Amsden, 2001; Lall, 2003).

This work on the firm, with its assumption of imperfect information and information externalities, particularly with respect to technology, has challenged what has been the dominant view of rents since Krueger's classic 1974 article. Since then, rent-seeking was portrayed as the main scourge of development and the trump card against any selective state interventions, even in the presence of market failure. Now, the acknowledgement that rents are at the heart of technological change and not simply politically derived is common in the theoretical and empirical literature that focuses on the microfoundations of development. Free trade, rather than forcing firms to innovate, may simply force them out of business if the productivity gap with foreign competitors is too large.⁷ Using the findings from endogenous growth models (Romer, 1986, 1990, 1994), this work provides a new twist to old infant industry arguments. For example, Traca (2002) argues that temporary protection is warranted for firms if they are far from the technological frontier. Otherwise, they would not be able to maintain market share and returns necessary to sustain the costs of R&D necessary to become internationally competitive.

Hausmann and Rodrik (2003), Rodrik (2004) and Klinger and Lederman (2004) also argue that firms will invest in risky non-traditional activities only with the assurance that their rents will not dissipate from foreign or domestic competition. Individual entrepreneurs take on the risk of investing in new activities; they appropriate the full costs of failure but not necessarily the gains from successfully demonstrating that a good can be produced domestically at low cost. Under conditions of unlimited entry, these information externalities will reduce the level of domestic innovation or 'self-discovery' (Hausman and Rodrik, 2003).

These works are helpful in explaining the divergent performances of regions since liberalization. During the 1980s and 1990s, Latin America's total and per capita growth rates did not compare favorably with either East Asia or its 1950–80 performance.⁸ Its performance in manufacturing was also relatively weak. As a result, Latin America's share of the developing world's manufacturing value added (MVA) fell from 48 percent to 22 percent, while East Asia's rose from 29 percent to 58 percent. On a per capita basis, the Latin American region is still the most industrialized, but that lead is diminishing. Indeed, manufacturing is no longer the engine of growth in the region, as its share of GDP has been falling.⁹

In contrast to its lagging performance in manufacturing, Latin America and the Caribbean did shift to exports at a fast rate. The region's manufacturing exports grew faster than MVA during the period from 1981–2000, as

did global manufacturing exports. Still, the region's share of developing country manufactured exports fell from 25 percent to 19 percent, while East Asia's share rose from 52 percent to 69 percent.

The sectoral breakdown of manufacturing also diverged between the regions. In many Latin American countries, such as Brazil, Chile, Argentina, Colombia and Peru, the fastest-growing industries are those that process natural resources. In Mexico and Central America, there has been a shift towards labor-intensive assembly operations, mostly for export. Generally, labor-intensive sectors geared for the domestic market fared poorly, as did capital goods and consumer durables. This rise in resource-based activities is in contrast to trends in global manufacturing, where the share of resource-based and low-technology activities in total manufacturing fell, as that of medium- and high-tech activities grew.

Explanations for divergent paths

A variety of explanations has been put forth to explain the different rates of growth across regions, most having to do with the need for further reforms. The literature that emphasizes industrialization and firm capabilities suggests different causal factors. One is the relative decline in R&D spending in Latin America. According to a variety of indicators, the gap in technological capacity between Latin America and countries such as South Korea, China, Taiwan and India is growing (Amsden 2001; Lall et al., 2004). The reasons behind this growing gap are hard to specify.¹⁰ One may be the fact that governments in countries such as India, Korea, China and Taiwan have historically promoted R&D and technology to a greater extent than those in Latin America, and have continued to do so. These programs have both supported capabilities for domestic firms and pressured foreign companies to invest in local R&D and to maximize spillovers.¹¹

Latin America's R&D gap may also reflect how sectors with a relatively high level of technological content were hit hard by the combination of free trade and overvalued exchange rates. Those industries, which had spent more intensively on research and development, have had difficulty competing with imports from more industrialized advanced countries (Katz and Stumpo, 2001). Brazil is an exception to this regional trend, and may have been more successful in retaining industries with high engineering content precisely because it reduced its trade barriers relatively late.¹²

In her survey of late developing countries, Amsden comes to a similar conclusion about how the timing of liberalization matters, particularly with respect to the relative strength of domestic and transnational firms. Countries outside of Latin America that opened relatively late and had supported domestic firms were more likely to retain medium- and high-tech industries. In the recent phase of mergers and acquisitions that has taken

place in all of the late-developing countries to enhance scale economies, national firms in Taiwan, China, Korea and India were more likely to have national firms strong enough to survive and/or to be viable as joint venture partners.

This raises the question of whether the greater role of foreign firms in manufacturing in Latin America has any implications for its relatively weak performance compared to East Asia. Interestingly, the theoretical literature cited above – on the need to coordinate investment or to protect firms until they reach the technological frontier or generate adequate returns – fails to mention ownership, implicitly assuming that the firms are independent and nationally owned. Much of the literature on competitiveness makes similar assumptions, and does not consider the ramifications of transnational firms' global strategies on national industrial development.¹³

In Latin America, foreign firms have dominated the most dynamic manufacturing sectors since their inception, and their control has increased since liberalization (Garrido and Peres, 1998). Even large national conglomerates which held dominant positions in their local markets found themselves poorly positioned to confront trade liberalization. Evidence suggests that transnationals invest virtually nothing in local R&D in developing countries (Amsden, 2001, p. 207), putting even successful sectors at risk (Lall et al., 2004).

Transnational firms also have the option of confronting new competitive pressures by integrating their subsidiaries into their global production networks. This can involve limiting national production to particular product lines and complementing them with imports, or importing parts and components for final assembly. As a result, intermediate and supplier industries are drastically shrinking. Given the importance attributed to these sectors, the potential consequences for future development are dire. For example, Porter and others who have focused on the role of geographic agglomeration have emphasized the importance of strong supplier linkages for innovative firms.¹⁴ Ciccone and Matsuyama's (1996) work suggests that new sources of innovation may be concentrated at the intermediate, rather than the final, output stage of production, and sees 'the proliferation of intermediate inputs and producer services as the essential part of economic development and growth' (1996, p. 57).

This pattern has also led to balance-of-payments concerns. Numerous studies have shown that transnational firms in Latin America are leading an 'import-intensive' or 'deficit-prone' industrialization process. While exports of natural resource processing industries, foodstuffs and primary commodities have grown fast, imports of capital goods and labor-intensive products have been growing even faster, so the manufacturing trade balance is increasingly negative. Economic concentration has increased, as transnational

subsidiaries and large national firms are in a better position to take advantage of the new environment; small and medium-sized firms are losing out, many of which had been suppliers to big foreign and domestic firms and are now being replaced by imports. Many have concluded that the costs of adjustment were high and growth in output slower as a result of vertical deintegration and the increased dependence on imports.¹⁵

In sum, in the context of a favorable international climate, domestic liberalization and macroeconomic stability, and rapid export growth, Latin America's GDP growth rates have been disappointing. They have not matched earlier growth rates or those in East Asia. Moreover, the few successes in manufacturing cannot be attributed to liberalization per se. With the exception of maquila industries, all of these sectors were established under import-substitution regimes. In Latin America, natural resource processing industries received state support. This came in the form of financial and technical support to non-traditional agriculture and forestry, or as subsidies in the 1970s and 1980s to help firms invest in state-of-the art, capital-intensive processing plants.¹⁶ The auto industry remains subject to special sectoral policies throughout the region.¹⁷

In both East Asia and Latin America, exports were based on the productive capacity and expertise developed during import substitution.¹⁸ Indeed, the logic behind import substitution policies was to force firms to make large investments that were not easily reversible. Once these investments were made, firms were subsequently forced to consider the need to protect access to these markets and their past investments, which they did not treat as sunk costs (Shapiro, 1994). In addition, even if non-traditional exports were distinct from products initially produced for the domestic market, and were therefore not the outcome of import-substitution policies per se, they were usually produced by the same firms that did mature under the import substitution industrialization (ISI) regime. To the extent that managerial and technological capabilities at the firm level are key to development, then acknowledging this continuity of major firms is critical. Work by Roberts and Tybout (1995) on Colombian exports and Maloney and Azevedo (1995) on Mexico reinforces this point. Costs associated with entering export markets lead to path-dependence, in that firms already exporting are more likely to continue doing so.

Finally, there is a peculiar 'back to the future' quality with respect to Latin America's situation, similar to the trends in the theoretical literature. Liberalization was expected to increase efficiency at a micro level, which in turn would help address its macro balance-of-payments problems. Similarly, ISI was adopted in part to overcome the region's chronic external imbalances by reducing its dependence on raw material exports and manufactured imports. As first noted by Diaz-Alejandro (1965), ISI paradoxically made

countries even more dependent on imports, at least in the short run, and therefore growth more vulnerable to an import constraint. Likewise, although exports have increased under liberalization, imports have risen even more, in part due to the vertical deintegration of the manufacturing sector. As discussed above, many observers today are concerned about an ever more binding balance-of-payments constraint.¹⁹

Recent characteristics are disturbingly reminiscent of an earlier phase. In the 1950s and 1960s, Latin America was concerned about specializing in raw materials with low income and price elasticity of demand; today it still finds itself at the low-growth, commodity end of the industrial spectrum. While its export industries are no longer the raw material export enclaves of the past, they have become increasingly delinked from the domestic economy as they move towards the assembly of imported parts and components while the design- and technology-intensive activities are done elsewhere.

Conclusion

In many ways, theories of industrialization have come full circle. In the 1950s and 1960s, the reigning paradigm considered market failure to be endemic. After years of being discredited or ignored, many of the assumptions behind this paradigm have made a comeback. The policy implications of these theories, however, have not been similarly resurrected. In contrast to their predecessors, contemporary theorists of market failure have been reticent about policy recommendations. Given the acknowledged limitations of import substitution policies, skepticism about government capacity, and a very different global economy, this is not surprising. Moreover, the challenge facing the more developed countries – making the existing industrial infrastructure more competitive, or upgrading technological capabilities – requires different approaches to that of kick-starting industrialization.

The default policy recommendation is still the market.²⁰ The emphasis of reform has switched to institutions that will allow the market to perform more efficiently. Given the weakening theoretical and empirical foundation for market-based solutions, the assumption that state failure is worse than market failure needs to be reconsidered.

Notes

1. This section is based on Shapiro and Taylor (1990).
2. See Little et al. (1970) and Balassa (1982).
3. See Amsden (2001), Bruton (1998), Lall (2003) and Westphal (2002).
4. The comparative performance figures on industrialization and growth have been well documented. See World Bank (1993).
5. See Nolan and Pack (2003).
6. See Nelson and Winter (1982), Best (2001), Lall (2001), Paus (2005), Katz (1996), Peres (1998) and Amsden (2001).

7. Work by political scientists on Latin America also documents how economic liberalization does not eliminate incentives for rent-seeking but generates different ones. See Shamis (1999).
8. See Lall (2003) and Lall et al. (2004) for comparative data.
9. This trend started in the 1970s, but accelerated in the 1980s and 1990s. See Benavente et al. (1996).
10. Lall et al. (2004) suggest, but do not analyze, possible explanations for this gap.
11. For details on these programs, see Wade (1990), Rodrik (1996), Amsden (2001) and Lall (2003).
12. In the UNDP's *Human Development Report 2003*, Stiglitz also points out that East Asia was slower to reduce trade barriers and liberalize capital accounts, and still used selective policies. Lall (2003, p. 9) points out that India also liberalized more slowly and selectively, and performed better in terms of growth in manufactured value added.
13. Porter's (1990) *The Competitive Advantage of Nations*, based primarily on firms in advanced, industrialized countries, deals almost exclusively with national firms. For a discussion of related works on developing countries, see Shapiro (2003).
14. Porter (1990).
15. See Kosacoff (2000), Miranda (2000), Moreno Brid (2000) and Ocampo (2004–05).
16. See Meller (1995) and French-Davis (1997) on support to Chilean agriculture; see Bisang et al. (1995) and Stumpo (1995) on capital-intensive processing plants.
17. See Katz and Stumpo (2001) for the role of industrial policy in revitalizing the Latin American auto industry in recent decades.
18. See Shapiro (2003). For a discussion on Turkey's export 'miracle' of the 1980s, which was also based on a pre-existing industrial base created during import substitution, see Boratav (1988).
19. Katz and Stumpo (2001) also note the similarities to the debate over balance of payments in the 1950s.
20. See World Bank (2002) and Nolan and Pack (2003).

Bibliography

- Amsden, Alice H. (1989), *Asia's Next Giant: South Korea and Late Industrialization*, New York: Oxford University Press.
- Amsden, Alice H. (1994), 'Why Isn't the Whole World Experimenting with the East Asian Model to Develop? Review of the East Asian Miracle', *World Development*, 22 (4): 627–33.
- Amsden, Alice H. (2001), *The Rise of the Rest: Challenges to the West from Late-Industrializing Economies*, Oxford: Oxford University Press.
- Balassa, Bela (1982), *Development Strategies in Semi-Industrial Economics*, Baltimore, MD: Johns Hopkins University Press.
- Benavente, Jose Miguel, Gustavo Crespi, Jorge Katz and Giovanni Stumpo (1996), 'Changes in the Industrial Development of Latin America', *Cepal Review*, 60.
- Best, Michael (2001), *The New Competitive Advantage*, Oxford: Oxford University Press.
- Bielschowsky, Ricardo A. and Giovanni Stumpo (1995), 'Transnational Corporations and Structural Changes in Industry in Argentina, Brazil, Chile, and Mexico', *Cepal Review*, 55 (April): 143–69.
- Bisang, Roberto, Gustavo Burachik and Jorge Katz (eds) (1995), *Hacia un nuevo modelo de organizacion industrial. El sector industrial argentino en los anos 90*, Buenos Aires: CEPAL y Alianza Editorial.
- Boratav, Korkut (1988), 'Turkey', Stabilization and Adjustment Policies and Programmes Country Study No. 5, Helsinki: WIDER.
- Bruton, Henry J. (1998), 'A Reconsideration of Import Substitution', *Journal of Economic Literature*, 36 (2): 903–36.
- Chenery, H.B., S. Robinson and M. Syrquin (1986), *Industrialization and Growth*, New York: Oxford University Press.
- Ciccone, Antonio and Kiminori Matsuyama (1996), 'Start-up costs and pecuniary externalities as barriers to economic development', *Journal of Development Economics*, 49 (1): 33–59.

- Diaz-Alejandro, Carlos F. (1965), 'On the Intensity of Import Substitution', *Kyklos*, 18 (3): 495–511.
- Dussel Peters, Enrique (1996), 'From Export-Oriented to Import-Oriented Industrialization: Changes in Mexico's Manufacturing Sector 1988–1994', in Gerardo Otero (ed.), *Neoliberalism Revisited*, Boulder, CO: Westview Press.
- Ellis, Howard S. (1958), 'Accelerated Investment as a Force in Economic Development', *Quarterly Journal of Economics*, 72 (4) (November): 485–95.
- Fishlow, Albert, Catherine Gwin, Stephan Haggard and Dani Rodrik (1994), *Miracle or design? Lessons from the East Asian Experience*, Washington, DC: Overseas Development Council.
- French-Davis, R. (1997), 'La Industrialización Chilena Durante el Proteccionismo y Despues (1940–95)', Washington, DC: Inter-American Development Bank.
- Garrido, Celso and Wilson Peres (1998), 'Las Grandes Empresa y Grupos Industriales Latinoamericanos en Los Anos Noventa', in W. Peres (ed.), *Grandes Empresas y Grupos Industriales Latinoamericanos*, Mexico City: Siglo Veintiuno Editores and Cepal.
- Gerschenkron, Alexander (1966), *Economic Backwardness in Historical Perspective*, Cambridge, MA: Harvard University Press.
- Hausmann, Ricardo and Dani Rodrik (2003), 'Economic Development as Self-Discovery', *Journal of Development Economics*, 72: 603–33.
- Hirschman, Albert O. (1958), *The Strategy of Economic Development*, New Haven, CT: Yale University Press.
- Hirschman, Albert O. (1967), 'The Political Economy of Import Substituting Industrialization in Latin America', *Quarterly Journal of Economics*, 82 (1968): 1–32.
- Hofman, Andre (2000), 'Economic Growth and Performance in Latin America', Serie Reformas Economicas 53, Santiago: ECLAC.
- Katz, Jorge and Giovanni Stumpo (2001), 'Sectoral Regimes, Productivity and International Competitiveness', *Cepal Review*, 75 (December).
- Katz, Jorge M. (ed.) (1996), *Estabilización macroeconómica, reforma estructural y comportamiento industrial*, Santiago de Chile and Buenos Aires: Cepal and Alianza Editorial S.A.
- Klinger, Bailey and Daniel Lederman (2004), 'Discovery and Development: An Empirical Exploration of "New" Products', World Bank, August.
- Kosacoff, Bernardo (ed.) (2000), *Corporate Strategies under Structural Adjustment in Argentina*, New York: St Martin's Press.
- Krueger, Anne (1974), 'The Political Economy of the Rent-Seeking Society', *American Economic Review*, 64 (3): 291–303.
- Lall, Sanjaya (2001), *Competitiveness, Technology and Skills*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.
- Lall, Sanjaya (2003), 'Reinventing Industrial Strategy: The Role of Government Policy in Building Industrial Competitiveness', paper prepared for Intergovernmental Group on Monetary Affairs and Development (G24), September.
- Lall, Sanjaya, Manuel Albaladejo and Mauricio Mesquita Moreira (2004), 'Latin American Industrial Competitiveness and the Challenge of Globalization', INTAL-ITD Occasional Paper SITI-05, Inter-American Development Bank, June.
- Lewis, W. Arthur (1955), *The Theory of Economic Growth*, Homewood, IL: R.D. Irwin.
- Little, Ian M.D., Tibor Scitovsky and Maurice Scott (1970), *Industry and Trade in Some Developing Countries: A Comparative Study*, London: Oxford University Press.
- Maloney, William and Rodrigo Azevedo (1995), 'Trade Reform, Uncertainty and Export Promotion: Mexico 1982–1988', *Journal of Development Economics*, 48 (1): 67–89.
- Meller, Patricio (1995), 'Chilean Export Growth, 1970–90: An Assessment', in G.K. Helleiner (ed.), *Manufacturing for Export in the Developing World*, London: Routledge, pp. 21–53.
- Miranda, J.C. (2000), 'Abertura Comercial, Reestruturacao Industrial Exportacoes Brasileiras. Avaliacao de Potencial Exportador das 500 Maiores Empresas Industrias Brasileiras', Proyecto CLA/99-1671, Rio de Janeiro, Universidade Geral de Rio de Janeiro, September.
- Moreno Brid, Juan Carlos (2000), 'Essays on the Balance of Payments Constrain with Special Reference to Mexico', PhD dissertation, Faculty of Economics and Politics, University of Cambridge, October.

- Murphy, Kevin M., Andrei Shleifer and Robert W. Vishny (1989), 'Industrialization and the Big Push', *Journal of Political Economy*, 97 (5): 1003–26.
- Nelson, R.R. and S.J. Winter (1982), *An Evolutionary Theory of Economic Change*, Cambridge, MA: Harvard University Press.
- Nolan, Marcus and Howard Pack (2003), *Industrial Policy in an Era of Globalization: Lessons from Asia*, Washington, DC: Institute for International Economics.
- Nurkse, Ragnar (1953), *Problems of Capital Formation in Underdeveloped Countries*, Oxford: Oxford University Press.
- Ocampo, Jose Antonio (2004–05), 'Beyond the Washington Consensus: What Do We Mean?' *Journal of Post Keynesian Economics*, 27 (2): 293–314.
- Paus, Eva (2005), *Foreign Investment, Development, and Globalization: Can Costa Rica Become Ireland?* New York: Palgrave Macmillan.
- Peres, Wilson (ed.) (1998), *Grandes empresas y grupos industriales latinoamericanos*, Mexico City: Siglo veintiuno editors and Cepal.
- Porter, Michael (1980), *Competitive Strategy*, New York: Free Press.
- Porter, Michael (1990), *The Competitive Advantage of Nations*, New York: Free Press.
- Roberts, Mark J. and James R. Tybout (1995), 'An Empirical Model of Sunk Costs and the Decision to Export', Policy Research Working Paper 1436, World Bank International Economics Department, International Trade Division, March.
- Rodrik, Dani (1995a), 'Getting Interventions Right: How South Korea and Taiwan Grew Rich', *Economic Policy*, 10 (20): 53–107.
- Rodrik, Dani (1995b), 'Trade and Industrial Policy Reform', in J. Behrman and T.N. Srinivasan, *Handbook of Development Economics*, Volume 3, Cambridge, MA: Elsevier Science.
- Rodrik, Dani (2004), 'Industrial Policy for the Twenty-First Century', John F. Kennedy School of Government Faculty Research Working Papers Series, Harvard University, November RWP04-047.
- Romer, Paul (1986), 'Increasing Returns and Long-Run Growth', *Journal of Political Economy*, 94 (5): 1002–37.
- Romer, Paul (1990), 'Endogenous Technological Change', *Journal of Political Economy*, 98 (5): 71–102.
- Romer, Paul (1994), 'New Goods, Old Theory, and the Welfare Costs of Trade Restrictions', *Journal of Development Economics*, 43: 5–38.
- Ros, J. (2000), *Development Theory and the Economics of Growth*, Ann Arbor, Michigan: University of Michigan Press.
- Ros, J. (2001), 'Industrial Policy, Comparative Advantages, and Growth', *Cepal Review*, 73 (April).
- Rosenstein-Rodan, Paul N. (1943), 'Problems of Industrialization of Eastern and South-Eastern Europe', *Economic Journal*, June–September.
- Scitovsky, Tibor (1954), 'Two Concepts of External Economics', *Journal of Political Economy*, 62: 143–51.
- Shamis, Hector E. (1999), 'Distributional Coalitions and the Politics of Economic Reforms in Latin America', *World Politics*, 51 (2): 236–68.
- Shapiro, Helen (1993), 'Automobiles: From Import Substitution to Export Promotion in Brazil and Mexico', in David Yoffie (ed.), *Beyond Free Trade: Firms, Governments and Global Competition*, Boston, MA: Harvard Business School.
- Shapiro, Helen (1994), *Engines of Growth*, Cambridge: Cambridge University Press.
- Shapiro, Helen (2003), 'Bringing the Firm Back In', in A.K. Dutt and J. Ros (eds), *Development Economics and Structuralist Macroeconomics*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 247–62.
- Shapiro, Helen and Lance Taylor (1990), 'The State and Industrial Strategy', *World Development*, 18 (6): 861–78.
- Stallings, Barbara and Wilson Peres (2000), *Growth, Employment and Equity: The Impact of the Economic Reforms in Latin America and the Caribbean*, Washington, DC: UN and Brookings.
- Streeten, Paul (1963), 'Balance versus Unbalanced Growth', *Economic Weekly*, 20 April.

- Stumpo, G. (1995), 'El Sector de Celulosa y Papel en Chile: Un Case Exitoso de Reestrkturacion Hacia los *Commodities* Industriales', LC/R.1521, Santiago: ECLAC.
- Traca, Daniel A. (2002), 'Imports as competitive discipline: the role of the productivity gap', *Journal of Development Economics*, 69 (1): 1–21.
- UNCTAD (2003), *Trade and Development Report, 2003*, New York and Geneva: United Nations.
- UNDP (2003), *Human Development Report 2003*, New York: United Nations.
- UNIDO (2003), *Industrial Development Report 2002/2003*, Vienna: United Nations.
- Viner, Jacob (1953), *International Trade and Economic Development*, London.
- Wade, R. (1990), *Governing the Market: Economic Theory and the Role of Government in East Asian Industrialization*, Princeton, NJ: Princeton University Press.
- Westphal, L.E. (1990), 'Industrial Policy in an Export-Propelled Economy: Lessons from South Korea's Experience', *Journal of Economic Perspectives*, 4 (3): 41–59.
- Westphal, L.E. (2002), 'Technology Strategies for Economic Development in a Fast Changing Global Economy', *Economics of Innovation and New Technology*, 11: 275–320.
- World Bank (1993), *The East Asian Miracle*, New York: Oxford University Press.
- World Bank (2002), *World Development Report*, Oxford and New York: Oxford University Press.

33 The informal sector

Victor E. Tokman

The questions addressed in this chapter are: Why is the informal sector important? What is its meaning? And what policies have been proposed to support it?

Why is the informal sector important?

The introduction of the concept of the informal sector, in more than three and a half decades ago in 1972, and particularly its increasing acceptance in the development literature, constitutes a proof of its relevance. Few new concepts in the development field that have been introduced during the second half of the twentieth century lasted for such a long period.

The development literature contains diverse readings on the existence of the informal sector. For a large number of development economists, the level of informality is a function of economic development and thus, it will tend to fall over time. Informality is only a symptom of low overall productivity and it constitutes a temporary feature of developing countries. In other interpretations, like De Soto (1987), while it is recognized that the sector is growing rapidly, it constitutes a good sign: that of emerging entrepreneurial talent among the poor. For others like Fields (1990) and Maloney (2004), the informal sector is voluntary. They sustain that it is an unregulated micro entrepreneurial sector that voluntarily moves out of the formal sector due to the laxity of enforcement of labor laws, and those working there do it willingly because wages or working conditions are better than those they could obtain in the formal sector.

A more widely accepted interpretation among development social scientists takes the existence of a high level of informality as a reflection of low growth and insufficient capacity of employment creation in countries where labor supply is rapidly expanding. In this structural context, the informal sector provides low-income subsistence jobs for the poor (Souza and Tokman, 1976). The inadequacy of institutions, particularly in the labor and enterprise fields, can also lead to non-compliance and low incentives to participate in formal arrangements in countries where there is also a low capacity of enforcement (Tokman, 2002).

The relevance of the informal sector is directly linked to the inclusive nature of the concept, which concentrates a major part of the employment problems that affect developing countries worldwide. Close to 60 percent of

the non-agricultural labor force¹ finds employment in the informal sector. In addition, the sector has been expanding during the last two and a half decades (1980/2005). Seven out of each ten new jobs created during that period were informal. Previously, during the post-war period of expansion, contrary to what was expected, it did not disappear, nor even decreased. Informal employment represented the most accessible option for migrants and urban workers in countries without protection against unemployment.

It is also important because most of the employment in the sector is linked to poverty and vulnerability. Six out of ten poor and almost seven out of ten indigent people are employed in the informal sector. The reproduction of poverty is also almost exclusively concentrated in the sector. Ninety-four percent of the increase in the number of poor and all the increase of indigents are there. In addition, 60 percent of the youth in poverty and 58 percent of women are employed in the sector.

Finally, the failure of prosecution as a policy, and the unacceptability of tolerance of such a large group of people to live in conditions of poverty and vulnerability, open the space for policies to promote better jobs and incomes in the informal sector. This has been reinforced by increasing evidence that such policies are less expensive than those required to support the more organized sectors, that they have received increased approval and resources from governments and international organizations and, finally, that they have proven to be effective.

It is this combination of the need to increase employment and to diminish vulnerability in a creative and feasible manner that is behind the popularity of the concept. This is reinforced by the failure of the traditional strategy to transfer people from the rural and informal sectors to modern activities. This strategy, even when it worked, required a long maturity period.

Alternative conceptualizations of the informal sector

The ILO–Kenya pioneer conceptualization

The concept of the informal sector was first introduced by an ILO (International Labor Office) mission to Kenya back in 1972.² The Mission constituted one of several country missions organized by the World Employment Programme (WEP) with the objective to study and identify solutions of the employment problems in the developing world. The Mission's first observation was that while they expected to find significant unemployment, they were surprised by the fact that most people were working. Unemployment was not the main issue, nor even a relevant one. The employment problem was mainly one of people working at low levels of productivity and hence, obtaining low incomes. They were 'the working

poor'. They were fully employed, even for long hours, producing, repairing and selling things.

The characteristics of the 'way of doing things' were identified and constituted the foundations to define the informal sector. These characteristics were listed as opposed to those of organized production: ease of entry, reliance on indigenous resources, family ownership of resources, small scale of operation, labor-intensiveness and adapted technology, skills acquired outside the formal school system, and unregulated and competitive markets.³ Their finding was that in order to survive these workers have developed skills, abilities and businesses at a small scale. Furthermore, that survival was at the margin or even counter to rules and standards, but without involvement in illegal behavior. Goods and services produced were found to be socially needed, since they were demanded. Hence, the existence of the informal sector was functional to the economy and particularly although not exclusively, to the poor to ensure employment and supply goods and services.

Their activities were performed even if often they were repressed by public authorities. Given the informal sector's endurance in spite of unfavorable conditions, the question was posed as to the potential contribution it could make if, instead of repression, a policy of promotion was introduced. The expected results were more and better jobs providing higher incomes for the poor in an environment where good jobs were almost inexistent. The sector's high visibility, given its concentration in urban areas, allowed for its identification, but its original definition was mainly as the opposite of formal activities. This gave origin to further conceptual developments that, as we will see, are still taking place. They, however, cannot deny the original contribution based in the words of Hans Singer: 'an informal enterprise is like a giraffe; it's hard to describe but you know one when you see one'.

A first stage of the advancement in the conceptual understanding of the informal sector following the Kenya contribution took place in the 1970s and early 1980s. Most of the contributions were made under the umbrella of the WEP but involved an active participation of the academic community in both developed and developing countries. The WEP research, mostly based on the Asian and African experiences, focused on two directions. One was the productive capacity of micro-enterprises and the intermediate technology involved. This was developed in parallel with the introduction of such technology to promote employment creation. It coincided with the times of 'small is beautiful' a strategy that promoted small-size production and more adequate technology for labor-abundant economies. Another direction focused on identifying the theoretical underpinnings of the informal sector from a neo-Marxist perspective. The informal producers were seen as petty

producers and their main function was interpreted as to ensure a reserve army of labor and low-wage goods production. Both allow reduction of wages in the formal sector and ensure high profit margins.⁴

The informal sector from a structuralist perspective

Another major advancement was made in Latin America through the PREALC (Regional Employment Programme for Latin America and the Caribbean), a regional arm of the WEP.⁵ The PREALC benefitted from the conceptual advance made in the Kenya Mission, but it was also heavily influenced by the tradition of structuralism rooted in the region. The discussion of the informal sector was placed in the context of the debate about development in Latin America based on the work of Prebisch (1970), Pinto (1970) and Cardoso and Faletto (1970), among others. The employment question was one of the main focuses of Prebisch's (1981) theory of dependent capitalist development of the 1960s. The problem of employment was concentrated in the low-technology or marginalized sectors, since the rapidly growing labor force could not find jobs in modern sectors using technology imported from countries with different factor proportions. This analytical tradition in Latin America incorporated most of the issues that were involved in the discussion of the informal sector, of course with different labels and in a more 'radical' interpretation. The existence of labor surplus, inadequate technology and dependency were at the core of Prebisch's theory of peripheral capitalism. The heterogeneity caused by the process of adoption and diffusion of foreign technology was also introduced by Pinto to explain the occupational segmentation between traditional and modern sectors.

The PREALC built upon this background to pursue its own research on the informal sector. The main difference was in focusing on the employment question, as had been the case in Kenya. However, to advance from description to prescription a deeper understanding of the origin and functioning of the informal sector was needed. For this, the existing knowledge was important. Furthermore, the informal sector was not a temporary phenomenon, nor a marginalized one. It played a function and could have the potential to contribute to the improvement of the employment situation.

There was coincidence in the dynamics of the origin of the sector. Rapid population growth, migration to the cities and insufficient employment creation in modern sectors forced people to create their own employment by producing or selling goods or services that could provide an income to ensure their personal and family survival. Given their insufficient endowment of physical and human capital, the activities they could perform had to be in sectors of easy entry and low capital or skill requirements. Absence of entry barriers is found in competitive markets or at the bottom of

concentrated ones, and both result in low profits because of heavy competition, or of being price-takers determined by the residual firms. The organization of production, in turn, is determined by the initial scarcities of factors of production and lack of access to product and factor markets. The result is that an informal sector unit will be small in size, use mostly family labor and will only hire labor marginally. Technology and division of labor will be rudimentary and the owner will be an entrepreneur and a worker at the same time. Capital will be restricted to family savings or informal loans. Indeed, these characteristics coincide with those identified in Kenya and they emerge as a consequence of a more comprehensive process of development.

The informal sector provides a response to survival needs but its potential role in an employment policy will depend on the relationships with the rest of the economy. The PREALC introduced a distinction to advance in the analysis since the informal sector is both competitive and complementary to the activities of the formal economy. It competes with lower costs and lower-quality goods, as happens with street sellers and retail stores, or in transport where informal taxis compete with licensed transportation. It is also complementary in subcontracting relationships, with the formal sector supplying inputs and final products to informal producers and sellers. Some activities will be able to gain market access, at least temporarily; others will accompany expansion of formal activities; while a residual sector of low-remunerated services will survive until better-paid jobs are available.⁶

To measure the size of the informal sector the PREALC focused on the form of production as the unit of analysis and, given the data available, categories of occupations were combined with size of establishments. The latter were included up to a certain size limit (less than five or ten employees), including both entrepreneurs and workers in those establishments,⁷ and the self-employed, non-remunerated family members and domestic services were also included. The self-employed can be considered as a one-person unit, while unpaid family members work in family units. Domestic servants were more questionable but their inclusion was made following the criteria that these services are performed for a family under several working arrangements, mostly beyond regulations.

The informal sector: its relationship to the state and regulations

In parallel to the above conceptualization, a different definition of informality was applied to labor market segmentation.⁸ Labor analysts adopted the terminology to differentiate unprotected labor from those working at formal units covered by protection legislation and by trade union organization. Some of the authors used this analysis to differentiate casual from

permanent labor, and in the USA it was used to distinguish between primary and secondary labor markets. The former included fully protected workers, while the latter referred to discriminated labor, mostly migrants and inhabitants of the slums. The unit of analysis adopted in these studies moved from the form of production to the labor relation and, as will be argued below, the divide captures differences in protection, rather than the capacity of the economic unit to afford protection. Indeed, for the informal units, both largely coincide; but this is not the case in the formal sector where lack of protection also exists, albeit to a lesser extent.

A new approach was developed in the late 1980s under the logic of decentralization of the labor process (Portes et al., 1989). This was associated to globalization and the changes in the international division of labor. Modern enterprises have to adjust to compete in a new context characterized by unstable and volatile demand. This required the reform of the system of production in order to increase flexibility and efficiency. Modern enterprises had to resort to the decentralization of the production and labor processes since this allowed them to reduce production costs, particularly labor costs, and to transfer demand fluctuations outside the firm. Decentralization is associated to subcontracting the production of intermediate goods and the provision of labor. Both are less regulated than activities performed directly by modern enterprises and are beyond trade union power. In addition, recent reforms in labor legislation further promoted the use of these practices, opening new possibilities of bypassing legal obligations.

This approach claims universal validity, as opposed to previous conceptualizations restricted to understanding employment in developing countries. Framed within a neo-Marxian analysis of power and class struggle, the informal sector is seen as the last echelon of the chain to absorb the costs of adjustment to the new global situation and to diminish union power. The unloading of labor costs and the erosion of power takes place between enterprises, not only in a given country, but also between countries of different levels of development. International subcontracting and mobile transnational companies are taken as evidence to support the argument.

Case studies of different sectors were undertaken in the USA and Western Europe. Underground activities in the former Soviet bloc were also examined to support the claim of universality of the informal sector. The emphasis of the interpretations has changed through time. The logic of survival has been, and still continues to be, a determinant of the existence of the informal sector in developing countries. Increasingly, emphasis has also been given to a mix of activities introduced as a consequence of the logic of decentralization, particularly in a context of economic opening.

However, a distinction should be made between factors that determine the creation of informal occupations, and the level of integration with the rest of economic activity. An activity generated as a result of the need for survival is not necessarily marginal or isolated from the rest of the economy. Obviously, those that are created as the result of decentralization are, by definition, functional to modern enterprises. Levels of development and different economic systems also contribute to explain differences. A micro-enterprise in Bolivia could apparently look similar to one located in the USA. But that is only how it looks. The former responds to the need for survival and lack of better job opportunities, while the latter will cater for a market opened by the needs of larger companies or to reduce labor costs or bypass regulations. The results will also be different: those occupied in the former will be poor, while those in the latter will be able to appropriate the rent of specific markets. Underground activities in socialist countries with price controls and excess demand respond to supply rationing and open spaces for 'illegal' activities with speculative profits.

The common use of 'illegality' as a synonym for 'informality' is also misleading since it includes black market operations to marketing of smuggled products or drugs and prostitution. This tends to confuse concepts. 'Illegal' is an action executed against criminal law and, although it involves an economic activity, its existence is subject to severe sanctions. Informal activities, although they may involve some degree of operation beyond regulations, do not constitute criminal activities.

Another approach which became increasingly important was also introduced during the second half of the 1980s. It defined the informal sector as an operation beyond the prevailing legal-institutional framework (De Soto, 1987). There is, however, an ongoing debate about whether the operation beyond regulations is a cause or a consequence of the informal activity. De Soto accepts previous interpretations of the origin of the informal sector as the result of rural-urban migration and the barriers the migrants faced to find jobs or develop economic activities. The barriers identified were prohibitions and government regulations. The need to survive forces the newcomers to perform multiple activities that usually violate laws and regulations. The cost involved in complying with them exceeds the potential benefits. In this approach, the state, the bureaucracy and the excessive rules and laws of mercantilism choke the potential development of the informal sector.

De Soto's main proposals were to reduce state intervention, simplify procedures, decentralize to the local authorities and deregulate markets. In spite of the merits of some of the proposals, his popularity responded more to the coincidence of the prescription with the conventional package of reforms known as the 'Washington Consensus' that promoted a transformation of

the economies based on opening markets, deregulation and privatization. All this involved a reduction in the role of the state while allocating a greater role to markets in the allocation of resources. Another attraction of the proposal was the automatic process that would surge as a response to the suggested actions. Those operating in the informal sector are expected to become dynamic entrepreneurs. Withdrawal of the state, deregulation, liberalization of markets and private initiative were keywords that reinforced a broader ideological change that was taking place after the fall of the Berlin Wall.

Unprotected labor and the informal sector: a renewed conceptual discussion

The most recent conceptual development in relation to the informal sector was introduced by the ILO, the organization that, as mentioned, was the pioneer in the introduction of the concept in 1972. Three decades later (2002), the ILO introduced the concept of the 'informal economy'. This new concept includes the informal sector, but adds all those workers without protection, even if they are employed in formal activities.

The International Conference of Labour Statisticians in 1991 adopted a resolution to define the informal sector in order to identify its contribution to gross national product in the national accounts. It was agreed that the sector should be defined by the characteristics of the productive units in which these activities are developed (entrepreneurial approach), rather than by the characteristics of the persons involved or by their occupational positions (labor approach). The Report of the Director General of the ILO submitted to the International Labour Conference in 2002 suggests that as the above definition does not capture all dimensions of informal employment, all workers should be classified as formal or informal depending on their labor status. It did not, however, propose to eliminate the concept of the informal sector, in view of its recognized merits, but expanded it to the 'informal economy'. The new concept includes the previous one but expanded it to incorporate labor relations, in addition to production relations.

Jobs, and not people occupying the posts, were adopted as the unit of analysis. Units of production were added as in the previous definition. Employment status and its characteristic of formal or informal is used to define the type of job, without explicitly identifying the prevailing unit of analysis to decide the condition of informal. The main innovation is then to include as informal all workers, independently of where they work, whose labor relation is not subject to labor legislation, taxes, social protection or rights to labor benefits (vacations, sick leave, fire indemnity, and so on). To sum up, the informal economy includes all unprotected workers working in enterprises of more than five employees.⁹

Two comments should be made in relation to the above change of definition. First, conceptually both units of analysis (production and

labor) are combined and related to coverage of social protection. However, the latter relation is only partially made, since the concept of the informal sector is not changed and still includes workers in informal units that have access to protection. Hence, the informal economy is not equivalent to unprotected workers since a percentage of those in the informal sector are protected. This percentage, on average for Latin America, was estimated at 35.6 percent in the year 2002.

Secondly, from a labor policy perspective concepts are mixed and can lead to confusion. Unprotected workers in formal enterprises have the right to be protected as part of their labor relation, formalized by labor legislation and contracts. As a consequence, those enterprises are forced to provide it. The failure to do so constitutes evasion and, as such, is subject to inspection and penalties. Its existence can be attributed to weakness of labor inspection, to looseness of the labor authority or to grey zones incorporated in atypical labor contracts. This is an entirely different situation to what happens in the informal sector. The lack of protection in this case is the result of the informal enterprise inability to finance protection. As studies have shown (Tokman, 1992), the informal sector units in general comply with labor obligations, although not with all of them and, particularly, not with contributions to old age pensions. For those employed in the informal sector lack of protection is a result rather than a characteristic of the post. Access to protection in those productive units can only be afforded if they are able to evolve out of the low productivity–scarce surplus trap.

What are the policy options for the informal sector?

Most of the conceptual frameworks analyzed, based on the historical evidence, assume that the existence of the informal sector is permanent. The expectation that the informal sector will be progressively absorbed in the modern sectors as the economy grows has proven to be unfounded.¹⁰ The diversity of conceptual approaches leads to different policy proposals on what to do with the informal sector. As Emmerij (1974) put it at the time of the introduction of the concept in Kenya, the issue of an increasing informal sector was clear; the only remaining strategic question was whether this expansion would be evolutionary. The size of the sector, measured by the number of people employed, is expected to grow but it is unknown whether that will be accompanied by increasing incomes per employee. This would constitute evolutionary growth while the opposite would result in an involutory situation. An increasing share of urban employment would be in the informal sector at stagnant or even decreasing incomes.

Two approaches have been followed in relation to strategies and policies. The first is to focus on the poverty of those working in the informal sector

and, hence, the policies are conceived as anti-poverty with a bias towards social assistance. A second approach, consistent with the original definition and some of the later conceptual developments, focuses on the incorporation of the informal sector in the modernization process. Although it is recognized that some survival activities without the expectation of future progress are included, the bulk of activities can be promoted in such a way as to integrate them to the main economic circuits. In fact, their possibilities of evolutionary growth depend on their capacity for integration.

This approach pursues the achievement of a self-sustained path of growth where governmental support will only be necessary during a transition stage. The integration of the informal sector to modernity can be promoted by diverse policies, not necessarily exclusive. These policies feature, in general, a triple dimension. The first dimension is the productive support aiming at the development of micro-enterprises by easing access to markets and productive resources. The usual programs are directed to ensure credit, training and access to more dynamic markets. There is ample experience worldwide on such instruments as well as evaluations of their effectiveness. They constitute the 'conventional wisdom' set of policies for the informal sector, albeit in very different modalities. In general, they are effective but unable to promote a systemic change.

The second dimension is related to social welfare support. Policies prescribed under this dimension are closer to anti-poverty ones. They incorporate the interrelationships between the small productive units and the families involved. Both overlap and social support can have an expanded economic effect. Private houses play a dual role for the family, constituting a home and a business location. Housing policies targeted at the poor are incorporating this dimension to ensure both family privacy and adequate business facilities. Health services are also more important because of the dual role of the persons involved. The family needs protection, as poor people do, but at the same time, their business performance is affected by their health problems, and more so if this demands allocating scarce resources to sickness. Social policies in this context can have an important productive effect.

The last dimension, the more recent one in the literature, is the focus on the regulatory system. Although the informal sector is not the result of the inadequacy of the regulatory system (De Soto, 1987), its correction can contribute to the integration of the informal sector to modernity (Tokman, 2002). The debate has evolved greatly during recent years. There has been a substantive rapprochement between an originally simplistic proposition expecting that legal changes would automatically solve all the problems of the informal sector, and the denial that legislation and institutions matter.

The informal sector operates beyond regulation. It is not underground nor is it entirely legal. It operates in a grey zone that allows for avoiding or, at least, reducing costs,¹¹ but at the price of missing opportunities of access to the process of modernization. Those working in the sector, as entrepreneurs or workers, lack the credentials needed to be able to perform in the formal economy. In fact, they constitute second-class citizens in the social and economic sphere. Several strategic options to move in the above direction have been suggested. A first one is to recognize a duality in relation to regulations, tolerating without sanctioning the violations because of structural inability. This would lead to a deepening of duality when the objective should be integration.

A second approach could be to adapt the requirements of entry to formality bringing them near to informal possibilities of compliance. This can be achieved by reducing the costs of legality, diminishing the procedures and redesigning the mechanisms of entry to formality. Entry costs could be reduced when required for all, without affecting the uniqueness of the system. This can be applied, for instance, to labor and fiscal matters. Non-wage labor costs and the tax burden and structure should be changed if justified by potential benefits for those already integrated, but also for the easing of entry costs to the informal sector. Procedures can be simplified, reduced and consolidated, and the time involved can be shortened by introducing information technology. The benefits expected are for all, but particularly for those in the informal sector who cannot afford to spend the time and resources needed to comply with obligations.¹²

Finally, another policy possibility relates to mechanisms to formalize the informal sector, both enterprises and workers. Several measures have been proposed, from simplified methods of entitlement to property, to the introduction of new forms of societies adapted to the need of micro and family enterprises. It could also include the recognition of labor contracts which are verbal, the recognition of existence for tax purposes, and the allocation of sites for street sellers.

The way to do it has been explored in different studies,¹³ but it is important to note the potential effect of these changes on facilitating entry into formality. The legal recognition of titles is required as collateral for credits. The recognition of labor contracts constitutes a necessary step to become a subject of protection. The same happens with the registration as contributor for tax purposes. A similar situation emerges when new legal arrangements are introduced to allow for business operations of micro and family enterprises.

The reformed instruments can contribute to the transition from the informal to the formal sector. Access to capital will be easier when collateral is available. Labor protection can progressively improve when a contract

exists. The legal separation of personal and family assets from the business operation will allow those in the informal sector to undertake risks without involving non-business assets in case of failure. Registration as a tax contributor induces the introduction of accountancy in small ventures that usually lack this fundamental instrument. This is a necessary step towards a more efficient business administration.

The reforms suggested need to be framed within a major strategic change in relation to the informal sector. The usually accepted sequence is to place the priority in the compliance of obligations while the new proposal is to change the order. Ensure the rights to those operating in the informal sector to enable them to expand their businesses and incomes, and only then they will be able to comply with obligations. As a result they will become full economic and social citizens. This would provide incentives for the outsiders to pursue integration, a necessary condition to advance in this direction. The present options are designed to respond to the interest of the insiders without ensuring that it will be in the interest of the outsiders to integrate to modernity.

Notes

1. This percentage is the average for Latin America. Other percentages in what follows also refer to that region at the beginning of this century (2000–2005). The percentages for other developing regions, if anything, are larger than the one mentioned for illustrative purposes.
2. The Mission was organized under the auspices of the World Employment Programme and was headed by Hans Singer and Richard Jolly, both from the Institute of Development Studies (IDS) of the University of Sussex. A member of the mission, K. Hart (1970), wrote a pioneering contribution introducing the concept that was later developed during the mission.
3. As Lubell (1991) correctly noted, the characteristics should be qualified with the adverb 'relatively'. As subsequent research has shown entry was not especially easy, resource inputs are often of foreign origin and markets are regulated but enterprises ignore or evade government attempts to impose regulations.
4. See among others, McGee (1973), Gerry (1974), Hugon et al. (1977), Bromley (1979), Nihan et al. (1979), Sethuraman (1981) and Lubell (1991).
5. I had the privilege of managing the PREALC during those years and became directly involved in research and policy development of the informal sector. A long list of publications of the PREALC about the informal sector in Latin America can be found in Tokman (2004). One of the first published articles was Souza and Tokman (1976).
6. See Tokman (1978b, 1978a, 1989).
7. Before surveys especially designed to measure the informal sector were available most of the estimates did not include micro-enterprises. However, their inclusion could be generalized afterwards when the household surveys incorporated a specific question with that purpose as a standard practice.
8. See, for instance, Mazumdar (1976) and Gordon et al. (1983).
9. An estimate made for Mexico in 2000 to measure the impact of the change in definition shows that 25 percent of the workers in enterprises of more than five employees are not covered by social protection and that the informal sector, measured with the usual definition, includes 57 percent of total employment in the country. The informal economy, with the new definition, amounts to 65.4 percent of total employment. A

- similar exercise for Chile in 2003 results in an expansion of the informal sector from 31.1 percent to an informal economy of 36 percent.
10. The approaches developed under the neo-Marxian perspective argue that the informal sector is integrated but exploited. On the other hand, those that propose neoliberal policies also implicitly assume that the informal sector will disappear with development.
 11. The main costs are fines for operating without permits that could involve, in some cases, the confiscation of stocks and jail.
 12. De Soto (1987) identified the 'paper wall' as a barrier to entry. This was related to the number of requirements and the inefficiency of the bureaucracy in the processing of those requirements. Present policies have gone beyond, unifying procedures by establishing a 'unique window' and modifying administrative behavior by, for instance, adopting the rule that the response to the request is assumed to be affirmative if after a defined period of time there is no resolution. This will shorten the processing. Another example is the diffusion of e-government for processing requests and consultation with government agencies.
 13. De Soto (2000) has proposed a simplified recognition of property rights. Tokman (2002) also contains suggestions in the other areas mentioned.

References

- Bromley, Ray (ed.) (1979), *The Urban Informal Sector: Critical Perspectives on Employment and Housing Policies*, Oxford: Pergamon.
- Cardoso, F.H. and E. Faletto (1970), *Dependencia y desarrollo en América Latina*, México: Siglo XXI Editores.
- De Soto, H. (1987), *El otro sendero*, Lima: Instituto Libertad y Democracia.
- De Soto, H. (2000), *El misterio del capital*, Empresa Editora, Lima: El Comercio SA.
- Emmerij, L. (1974), 'A new look at some strategies for increasing productive employment in Africa', *International Labor Review*, 110 (3): 199–234.
- Fields, G. (1990), 'Labour market modeling and the urban informal sector: theory and evidence' in D. Tunham, B. Salomé and A. Schwarz (eds), *The Informal Sector Revisited*, Paris: Development Centre, OECD, pp. 44–69.
- Gerry, C. (1974), 'Petty Producers and the Urban Economy: A Case Study of Dakar', ILO WEP, Urbanisation and Employment Research Programme No. 8, International Labour Office (ILO), Geneva.
- Gordon, D., R. Edwards and R. Reich (1983), *Segmented Work, Divided Workers: The Historical Transformation of Labor in the United States*, Cambridge: Cambridge University Press.
- Hart, K. (1970), 'Small Scale Entrepreneurs in Ghana and Development Planning', *Journal of Development Studies*, 6 (4): 104–20.
- Hugon, P., N.L. Abadie and A. Morice (1977), 'La petite production marchande et l'emploi dans le secteur informel –le cas africain', Paris: IEDES.
- Lubell, H. (1991), *The Informal Sector in the 1980s and 1990s*, Development Centre Studies, Paris: OECD.
- Maloney, W.F. (2004), 'Informality Revisited', *World Development*, 32 (7): 1159–78.
- Mazumdar, D. (1976), 'The Urban Informal Sector', *World Development*, 4 (8): 655–79.
- McGee, T.G. (1973), 'Hawkers in Hong Kong: A Study of Planning and Policy in a Third World City', Hong Kong: University of Hong Kong, Centre of Asian Studies.
- Nihan, G., E. Demol and J. Comlavi (1979), 'The modern Informal Sector in Lome', *International Labor Review*, 118 (5): 631–78.
- Pinto, A. (1970), 'Naturaleza e implicaciones de la heterogeneidad estructural de la América Latina', *El Trimestre Económico*, enero-marzo, Fondo de Cultura Económica, México, 37 (1).
- Portes, A., M. Castells, and L. Benton (1989), *The Informal Economy*, Baltimore, MD: Johns Hopkins University Press.
- Prebisch, R. (1970), *Transformación y desarrollo. La gran tarea de América Latina*, Santiago de Chile: Comisión Económica para América Latina y el Caribe (CEPAL).

- Prebisch, R. (1981), *Capitalismo periférico. Crisis y transformación*, México: Fondo de Cultura Económica.
- Sethuraman S.V. (ed.) (1981), *The Urban Informal Sector in Developing Countries: Employment, Poverty and Environment*, ILO: Geneva.
- Souza, P.R. and V.E. Tokman (1976), 'The Urban Informal Sector in Latin America', *International Labor Review*, 114 (3): 355–65.
- Tokman, V.E. (1978a), 'Competition Between the Informal and Formal Sectors in Retailing: The Case of Santiago', in *The Urban Informal Sector: Critical Perspectives*, special issue of *World Development*, 6 (9–10): 1187–98.
- Tokman, V.E. (1978b), 'An Exploration into the Nature of Informal–Formal Sector Relationships', *World Development*, 6 (9–10): 1065–75.
- Tokman, V.E. (1989), 'Policies for a Heterogeneous Informal Sector in Latin America', *World Development*, 17 (7): 1067–76.
- Tokman, V.E. (1992), *Beyond Regulations: The Informal Sector in Latin America*, New York: Lynne Rienner Publishers.
- Tokman, V.E. (2002), *De la informalidad a la modernidad*, Santiago, Chile: OIT.
- Tokman, V.E. (2004), *Una voz en el camino. Empleo y equidad en América Latina: 40 años de búsqueda*, Santiago, Chile: Fondo de Cultura Económica.

34 Services and development

Dilip Dutta

Understanding services in modern societies

Many less-developed countries (LDCs) are dualistic in the sense that the informal or unorganized traditional segment coexists side by side with the formal or organized modern segment in the same geographical boundary. In the traditional segment, many non-market service activities are performed within the extended family and/or community.¹ Industrially developed countries are predominantly 'modern' with either a non-existent or a very insignificant 'traditional' component. In more complex modern economies, strangers in the marketplaces generate a wide range of service activities, bringing together economic agents such as consumers, producers and suppliers, borrowers, lenders and investors, and so forth. In this process, human intelligence and knowledge have been systematically applied to generate more and more marketable service activities; science and technology are dynamic supply factors, while education and training have a similar impact on demand.² Although human capital provided by skilled labour has been the main constituent in the generation of service activities, the physical services provided by the unskilled labour are not considered part of it.

All functioning economies require a productive interplay among the three economic sectors – agricultural, industrial and services. However, nations in a globalized world can survive using primarily imported agricultural products.³ For example, the agricultural sector's contribution in gross domestic product (GDP) in 2003 is 0 per cent in Hong Kong and Singapore, while it is 1 per cent in Belgium, Germany, Japan and the United Kingdom.⁴ Similarly, countries can also survive and thrive using imported industrial products, with more rapid growth in services than in industrial products. Australia, Brazil, Jordan, Hong Kong, the Netherlands, the UK and the USA are all cases in point, which have a very high percentage of their GDP originated in services (Table 34.1). But it is unlikely that a modern economy can survive without a service sector. For the functioning of even the fully agricultural economies in modern societies, services such as public administration and utilities, transportation, communication and environmental infrastructures, financial and educational facilities, medical and social supports are all essential.

Services typically involve the provision of human value added in the

Table 34.1 *Percentage contribution in GDP in 2003*

	Industrial sector	Service sector
Australia	26	71
Brazil	21	73
Hong Kong	12	88
Jordan	26	72
Netherlands	26	71
UK	26	73
USA	23	75

Source: World Bank (2005b).

form of labour, advice, managerial skill, entertainment, training, intermediation and the like.⁵ Services are activities or processes that also produce changes in unique service inputs, which are persons or their possessions (that is, the goods they possess). Although service outputs seem to be similar to the goods-producing activities that bring about changes in raw materials, the relationship between the producer and customer in the service sector is very complex. This is because of the involvement of the customer or recipient⁶ as a potential input in different ways. Contrary to popular belief, not all service operations involve direct, sustained interaction between producer and customer. In fact, one can identify three types of producer–customer relationships as inputs for the production of services. Thus, services could be produced by:

1. the producer acting for the recipient (for example, in many repair and professional services);
2. the recipient providing a part of the labour (for example, in the usage of equipment and/or procedures arranged and maintained by the producers, as in rental and leasing services); and/or
3. the recipient and the producer creating services in interaction (for example, in the delivery of many entertainment services).

The *System of National Accounts, 1993* (SNA, 1993) has provided the international guidelines adopted by a majority of countries for data comparability. After noting that the term ‘services’ is difficult to encapsulate within a simple definition, the 1993 SNA first provides a simple definition:

Services are not separate entities over which ownership rights can be established. They cannot be traded separately from their production. Services are

heterogeneous outputs produced to order and typically consist of changes in the condition of the consuming units realised by the activities of the producers at the demand of the customers. By the time their production is completed they must have been provided to the consumers. (MSITS, 2002, p. 7)

Then, after acknowledging the tangible nature of some services, it qualifies this relatively simple definition in the following way:

There is a group of industries, generally classified as services industries, that produce outputs that have many of the characteristics of goods, i.e., those concerned with the provision, storage, communication and dissemination of information, and entertainment in the broadest sense of those terms – the production of general or specialized information, news, consultancy reports, consumer programs, movies, music, etc. The outputs of these industries, over which ownership rights may be established, are often stored on physical objects – paper, tapes, disks, etc. – that can be traded like ordinary goods. Whether characterized as goods or services, these products possess the essential characteristic that they can be produced by one unit and supplied to another, thus making possible division of labour and the emergence of markets. (MSITS, 2002, pp. 7–8)

Role of services in economic growth and development

The conventional view of services suggests that as economies evolve from ‘pre-industrial’ to ‘post-industrial’ or service economies,⁷ growth of the service sector is a natural consequence of development.⁸ The implicit reasoning behind the conventional view is that as countries become rich, the service sector would tend to expand since the income elasticity of demand for services would exceed unity, or, more appropriately, be regarded as high.⁹ Services which contribute to the production of goods or to the production of other services as independent activities are called primary services – the elements of global supply. Services which are not linked to any particular process of production are known as final services – components of final demand such as entertainment services or sports. Intermediate service activities play an indispensable complementary role in the production of goods and other services and hence are integrated or embodied into their production process. It may, however, sometimes be very difficult to separate out the complementary contributions of the intermediate service activities because of the lack of a clear mutual exclusiveness between the two categories. Also, the relationships of complementarity and substitution either between various services or between certain services and certain goods evolve over time due to either structural change in the economy and/or technological development.

In the early 1980s, economists had noted that the kaleidoscope of services and goods constantly changed as technical change, in a broad sense embracing organizational innovations as well, occurred. Technical change

leads to what Bhagwati (1984) called the ‘splintering process’ where goods splinter from services and services, in turn, from goods. Services splinter off from goods mainly through being shifted from the ‘inhousehold’ to the ‘outhousehold’ mode of production.¹⁰ In the process of development, service splintering often follows increased economies of scale as enough demand generated makes it attractive to set up specialized firms supplying these services. It needs, however, to be pointed out that ‘service destruction’ and/or ‘reverse splintering’, as against ‘service splintering’, do also take place with economic growth and development. While ‘service destruction’ arises as stagnant services get increasingly expensive and therefore retreat into the household as unrecorded do-it-yourself amateur activities, the ‘reverse splintering’ process occurs as a result of rapid technical change that allows goods to spring from services (as from musical services to records, magnetic tapes or compact discs). In the latter case, the service is disembodied from the physical presence of the provider and embodies those services in goods that can be bought in market places. This embodiment phenomenon in the form of services splintering into goods could alternatively be extended to services becoming, not goods, but ‘long-distance tradable services’ so that the physical presence of the provider of the services is no longer necessary with the user.¹¹

In light of the above-mentioned various characteristics of the services activities, the Organisation for Economic Co-operation and Development (OECD, 2000, pp. 10–12) highlights the following implications of various trends which have been emerging in the contemporary service industries:

- The convergence of services and manufacturing in many areas is making it increasingly difficult to classify firms uniquely under either category, particularly as manufacturers expand their businesses into service-related areas.
- Advanced business, professional and technical services are likely to have stronger communication capabilities in terms of connectivity and receptivity.
- Advanced business services can also improve the interaction between tacit and codified knowledge¹² and, as a result, could lead to higher innovative capacity.
- Although relatively high emphasis placed on intellectual capital, or ‘intangibles’, in many service activities may hold the key to value creation, their contribution to companies and their intrinsic worth could go unrecognized – a major drawback for obtaining investment finance.
- Most of the fast-growing knowledge-based services are (relatively) environmentally benign, especially when these services help to

increase efficiency in the production and distribution of other goods and services, thereby having a beneficial effect on resource use.

The trend towards globalization since the early 1980s, reinforced by technological advances as well as policies of economic liberalization and institutional reforms for removal of regulatory obstacles to economic activities, has enabled steady growth of international investment as well as trade in goods and services. According to UNCTAD (2000), services are the largest recipients of international investment flows, accounting for about one-fifth of worldwide trade in balance-of-payments terms.¹³

Trend of services contribution to GDP and employment

In spite of some difficulty in disentangling the exact contribution of the service sector to the gross domestic product (GDP) in the sectoral view of service activities, the United Nations Statistics Division has made significant refinements in the production based classification of various economic activities. Its sectoral classification called International Standard Industrial Classification (ISIC) is widely used to determine the sectoral origin of value added. Currently, the World Bank uses the ISIC Revision 3 approach: agricultural value added (ISIC divisions 1–5 including forestry and fishing), industrial value added (ISIC divisions 10–45 comprising mining, manufacturing, construction, electricity, water and gas) and services value added (ISIC divisions 50–99).¹⁴

Table 34.2 based on ISIC Revision 3 shows that since 1980 services have been playing more and more of an important role in their sustained value-added contribution to GDP across different types of countries and also in different regions of the world. The share of services value added in world GDP increased from 55.35 per cent in 1980 to 66.48 per cent in 2000, and in 2007, services represent more than two-thirds of world GDP. During the same period, while high-income economies' services contribution to GDP increased from 58.28 per cent to 69.74 per cent, that for middle-income and low-income economies increased respectively from 41.84 per cent to 54.98 per cent and from 38.19 per cent to 43.70 per cent (Table 34.2). Wide differences however in the share of the service sector's contribution to GDP in, say, year 2003 can be found among LDCs within the same group: as for example between 38.61 per cent (Tanzania) and 51.20 per cent (India) in low-income countries, between 65.20 per cent (South Africa) and 75.12 per cent (Brazil) in lower-middle-income countries, and between 49.5 per cent (Malaysia) and 63.3 per cent (Mauritius) in upper-middle-income countries.¹⁵ These differences are due partly to structural variation among the LDCs and partly to unavailability of data.¹⁶

A similar trend is also mirrored in services employment statistics (as a

Table 34.2 *Share of services as a percentage of GDP*

	1971	1980	1990	1995	2000
Low-income economies	37.36	38.19	41.38	42.44	43.70
Middle-income economies	42.93	41.84	46.77	51.12	54.98
Lower-middle-income economies	41.51	39.35	44.46	47.95	52.78
Upper-middle-income economies	46.60	48.13	52.66	59.28	60.67
High-income economies	54.82	58.28	63.73	67.32	69.74
High-income non-OECD	<i>n.a.</i>	49.67	58.24	65.81	69.82
High-income OECD	54.93	58.49	63.87	67.36	69.74
East-Asia & Pacific	31.36	28.84	36.79	36.64	37.42
South Asia	35.68	38.68	42.73	44.99	48.58
Europe & Central Asia	<i>n.a.</i>	<i>n.a.</i>	39.40	52.32	55.93
Latin America & Caribbean	49.69	49.15	55.09	58.42	64.41
Middle East & North Africa	38.49	39.27	47.35	47.74	47.02
Sub-Saharan Africa	50.16	44.21	48.07	50.75	53.37
World	52.69	55.35	60.30	63.92	66.48

Note: *n.a.* denotes data 'not available'.

Source: World Bank (2005a).

percentage of labour force) at least in high-income and upper-middle-income economies, although the employment trend in various regions (except in the region of East Asia and the Pacific) apparently is not clear due to unavailability of data. As can be seen from Table 34.3, the number of people employed in services increased in high-income economies from 55.78 per cent in 1980 to 68.92 per cent in 2000. Although a similar trend is present in upper-middle-income economies (from 44.78 per cent in 1990 to 58.30 per cent in 2000), but the trend is not encouraging in the case of lower-middle-income economies (from 15.16 per cent in 1980 to 20.13 per cent in 2000) and for low-income economies (20.01 per cent in 1990 – the only data available to the author's knowledge). Unlike among the developed countries, significant differences in the number of people employed in services in, say, year 2000 exist among LDCs: as for example between 14.6 per cent (Mongolia) and 37.9 per cent (Nicaragua) in low-income countries, between 12.9 per cent (China) and 73.3 per cent (Colombia) in lower-middle-income countries, and between 54.14 per cent (Argentina) and 66.5 per cent (Venezuela) in upper-middle-income countries.¹⁷ As has been mentioned in the previous paragraph, the data of the LDCs do have serious deficiencies. For example, labour employed in their informal or unorganized services sector, such as in small or self-

Table 34.3 *Employment (as a percentage of labour force) in services*

	1980	1990	1995	2000
Low-income economies	<i>n.a.</i>	20.01	<i>n.a.</i>	<i>n.a.</i>
Middle-income economies	16.64	23.34	25.11	24.47
Lower-middle-income economies	15.16	21.54	21.70	20.13
Upper-middle-income economies	<i>n.a.</i>	44.78	56.16	58.30
High-income economies	55.78	64.21	67.09	68.92
High-income non-OECD	<i>n.a.</i>	<i>n.a.</i>	<i>n.a.</i>	<i>n.a.</i>
High-income OECD	55.73	64.17	67.02	68.88
East-Asia & Pacific	14.71	13.48	16.76	17.33
South Asia	<i>n.a.</i>	18.26	<i>n.a.</i>	<i>n.a.</i>
Europe & Central Asia	<i>n.a.</i>	34.89	<i>n.a.</i>	<i>n.a.</i>
Latin America & Caribbean	<i>n.a.</i>	54.36	59.41	<i>n.a.</i>
Middle East & North Africa	<i>n.a.</i>	47.56	<i>n.a.</i>	<i>n.a.</i>
Sub-Saharan Africa	<i>n.a.</i>	<i>n.a.</i>	<i>n.a.</i>	<i>n.a.</i>
World	<i>n.a.</i>	29.23	<i>n.a.</i>	<i>n.a.</i>

Note: *n.a.* denotes data 'not available'.

Source: World Bank (2005a).

employed trade and transport activities as well as in personal and domestic services, may very likely be excluded from the official employment statistics.

International trends in services

Due to the intangible nature of the majority of services, distance trade in services inherently incurs more constraints than that in goods. For example, the need for proximity between suppliers and customers in delivering many commercial services (that is, excluding government services)¹⁸ has led many providers to establish foreign affiliates. This type of international trade in services has become as important as 'conventional' international trade in services between residents and non-residents.¹⁹ The shares of services traded in the international market are very lopsided because the developed countries' dominant share in world service exports ranged from 80 per cent to 74 per cent during 1980–2004, while their share in world service imports ranged from 67 per cent to 77 per cent during the same period. The trade balances in services (that is, value of service exports minus value of service imports) during the same period indicate that the developed countries had been running a net surplus and therefore the LDCs had been running net deficit, but the gap is gradually narrowing down.

Table 34.4 *Growth of foreign direct investment in less-developed countries*

	Annual growth rate (%)						
	1986-90	1991-95	1996-2000	2000	2001	2002	2003
FDI inflows	22.9	21.5	39.7	27.7	-41.1	-17.0	-17.6

Source: UNCTAD (2004, Table 1, p. 2).

Of the different forms of capital inflows²⁰ in LDCs, foreign direct investment (FDI) has been significantly growing since the 1980s²¹ (Table 34.4) especially due to the impetus for liberalization of FDI coming from developing and transition economies. But since the 1990s, the structure of FDI has shifted more towards services, compared to FDI in manufacturing activities. While the services sector accounted for only one-quarter of the world FDI stock in the early 1970s and this share was less than one-half in 1990, it had risen to about 60 per cent or an estimated \$4 trillion by 2002.

As it is evident from UNCTAD (2004, p. 17), outward FDI in services continues to be dominated by the firms from developed countries. Unlike the dominance of the firms from the United States on the entire outward stock of services FDI a few decades ago, this had, by 2002, become more evenly distributed among the firms from the United States, Japan and the European Union. Also, LDCs' outward FDI in services began to grow visibly during the 1990s. Their overall share in the global outward FDI services stock rose from 1 per cent in 1990 to 10 per cent in 2002, with even faster increase in construction (20 per cent), followed by business activities (16 per cent). On the inward side, the distribution of services FDI stock has been relatively more balanced, though developed countries still account for the largest share. There has been a recent tendency of many services transnational corporations (TNCs) to enter new markets through cross-border mergers and acquisitions (M&As), rather than through greenfield FDI. In fact, most M&As (inclusive of privatization) that took place during the second half of the 1990s were in services.

Service sector expansion and its effect on economic growth

Empirical evidence suggests that economic growth accompanies expansion of the service sector because services possess high income elasticity of demand in aggregate compared to that in goods output. It is often claimed that the extension of an economy's service sector relative to that of its goods-producing sector has a negative effect on the rate of overall economic growth. The underlying basis of such claims lies in the perishable or non-storable nature of services right after their production, unlike goods

that are storable or non-perishable. Dutt and Lee (1993) provide three different explanations of why the service sector has been singled out as a retarding or stagnant one.

Firstly, it was Smith (1776) who, after identifying the perishability of services upon production, concluded that increased spending on services necessarily implied a reduction in the accumulation of capital. According to Smith's analysis, the rate of economic growth is positively related to the rate of capital accumulation and, therefore, the service sector expansion has a retarding effect on economic growth.

Secondly, Baumol (1967), in his simple two-sector model with only labour input and same wage rate across the sectors, has argued that labour itself is the end product in the service sector, whereas it is primarily an instrument in the goods-producing sector. As a result, the possibility for increasing productivity in the former sector is far more limited than that in the latter. In Baumol's above unbalanced growth model which has been further developed in Baumol et al. (1989), he assumes that labour productivity grows at a constant rate in the goods-producing sector, but it remains constant in the services producing sector.²² He then shows that if the magnitude of relative output in the two sectors is maintained (due to sufficient price inelasticity or income elasticity of services, or government intervention) in an attempt to maintain balanced growth in a world of unbalanced productivity, the service sector will absorb a larger part of the total labour force and manufacturing employment will decline, and as a result, the rate of growth of per capita output will fall, and eventually approach zero.

Thirdly, the manufacturing sector is claimed to be an engine of economic growth because its activities generate technological progress, whose spin-off effects, in turn, favourably affect the whole economy. Therefore, a relative expansion of the service sector hampers the effectiveness of the engine of economic growth due to reduction of the share of the manufacturing sector in an economy's overall output growth. The two steps of this third mechanism have been investigated in more detail in both theoretical and empirical literature.²³

Although the above mechanisms, as suggested for explaining the adverse effects of service sector expansion, have received most attention in the context of developed economies, the recent growth of services in many LDCs has also attracted some attention. Gemmell (1982, p. 60), for example, has provided some empirical evidence on patterns of structural change during the development process across developed and developing countries in the 1970s: the service share increases sizeably in both groups of countries, but only in the developed countries is this associated with a declining manufacturing share; and services increase much faster in the LDCs because of their positive correlation with the manufacturing sector.

This empirical evidence seems to be in line with the theoretical argument that as per capita income increases, industrial outputs initially rise faster than service outputs (in the early stage of development), but later rise slower (due to the coincidence of higher income elasticities of demand with higher income levels) and eventually decline.

The claim against the service sector's retarding effect on economic growth has, however, been losing ground. Apart from having very little support for the claim in their own empirical analysis, Dutt and Lee (1993, p. 313) also provide the following theoretical arguments against the claim.

First, contrary to Smith's claim, although services may perish, some of them (such as education and health care) have long-lasting positive effects on economic growth. Because of inherent difficulty in measuring productivity in the service sector, it may not be easy to substantiate Baumol's claim. Service activities such as research and development may have high spin-off effects on technological change, thereby weakening the engine-of-growth claim.

Secondly, Dutt and Lee (1993, p. 313) questioned the validity of the macroeconomic structure and general state of the economy implicitly assumed in the claim. While capital accumulation is determined by savings in Smith's analysis, Malthus (1820) emphasized the role of demand. Although Malthus defined services in the same way as Smith did, he pointed out however that increased expenditure on services (or unproductive employment) could solve the glut problem. In an economy constrained by demand rather than supply, an expansion of the service sector could, by generating greater demand, increase the rate of economic growth.²⁴

Other responses to the gloomy prognostication of the service sector have been in terms of arguments that not all services have low productivity growth²⁵ (Baumol, 1985), and that errors in measuring service sector output are likely to be quite large (see for example Griliches, 1992, 1994). Recently, Oulton (2001) notes that the stagnationist argument of the unbalanced growth model is logically correct only if all industries produce final goods. But, because the reality is otherwise even if applied to the advanced countries, he refutes the validity of the argument. More specifically, the reason he puts forward is the following:

Quite a different conclusion results if some of the industries produce intermediate goods. And this could be the relevant case in practice since the service industries which have been expanding particularly rapidly are the ones such as financial and business services which are large producers of intermediate inputs (Oulton, 2001, p. 606).

Using data over the period 1973–95 on market services in the UK, Oulton shows that shifting of resources towards the so-called stagnant

intermediate industries like financial and business services whose productivity is growing slowly has raised, not lowered, aggregate growth rate of productivity.

Transforming services with ICT advances

Due to rapid technological advances in the past few decades in transport, computing and telecommunications, many enterprises worldwide have been availing themselves of more distant resources for production on the one hand, and serving ever wider markets on the other. In particular, the information and communication technology (ICT) revolution, including the development of the Internet and electronic commerce, has been making the differences between services and other economic activities narrower and narrower:

(I)nformation and communication technology (ICT) now enables people to participate in a growing number of service-related activities in real, or deferred, time, without having to be physically present. Copies of movies and most other performances can be recorded and mass-produced for future consumption, like manufactured products. Software is developed and boxed like any other manufactured product, and is considered, for all intents and purposes, a good – albeit with a high service-related content. In these instances services have, in a sense, taken on the characteristics of commodities – one provider is mass-producing a common product for many people. Service providers are thus increasingly able to benefit from economies of scale. (OECD, 2000, pp. 7–8)

The service sectors that have benefited the most from economies of scale include banks, telephone and telecommunications networks, distribution and retailing firms, and health systems. We are now living in a world where we have started to see the presence of global-scale service companies, compared to the presence of global manufacturing companies for the last 80 years, since Henry Ford introduced the technique of mass production in the 1920s.

The relationship between service providers and consumers is also changing in other ways with significant economic implications. Technology now allows providers to produce a single product which, though not mass-produced, is capable of being mass-consumed, either on a standardized or customized basis. An example of such a product is online Internet access to dictionaries, encyclopaedias, newspapers, museum collections, and so on. Technology is also affecting the relationship between providers and consumers in previously unthinkable areas, such as health care, where the personal contact to diagnose and treat ailments is becoming less essential. 'Internet' banking, real estate, retail and financial services provide other examples where personal, or on-site, contact with service providers is no longer essential for the services to be performed. In many instances such

services can, in fact, be provided far more efficiently via the Internet or through other remote communication modes.

Since the 1990s the increasing information intensity of economic activity, coupled with rapid technological change and worldwide demand growth, means ICT has been critical to economic development and competitiveness.²⁶ There is a growing consensus in the development community that ICT allows firms from LDCs greater opportunities to get connected to foreign markets and, therefore, to boost their participation in international trade. In their study on links between access to the Internet and trade growth in 56 developed and developing countries for the period 1997–99, Freund and Weinhold (2004) found that LDCs with the fewest established trade links benefited the most from using the Internet. It seems that ICT reduces the historical advantages of long-established firms. Similarly, in the ‘Overview’ based on data from its Investment Climate Surveys 2000–2003, the World Bank (2006, p. 2) notes that firms using ICT grow faster, invest more, and are more productive and profitable than those that do not; their sales growth is 3.4 percentage points higher and value added per employee \$3400 more among LDC firms that use email to communicate with clients and suppliers. Based on the same data as above, Qiang et al. (2006, p. 59), in a separate study on enterprise performance in LDCs, observe that the services sector firms appear to be the heaviest users of both websites and email. While about 90 per cent of firms in the IT and telecommunications industries themselves use these ICT applications to interact with clients and suppliers, the corresponding figure in both the real estate and the hotel and restaurant sectors is 70 per cent. Although data on email and website use are not available for the accounting and finance sector, this sub-sector has the highest percentage (67 per cent) of employees that use computers at work.²⁷

Political economy of international trade in services

Even before the enormous speed with which ICT started progressing in the 1990s, the potential tradeability of long-distance services had been anticipated by the US multinational banks such as the American Express Company in the 1970s, when they were seeking for greater access, and the right to establish, in other countries. The United States started playing a catalytic role in the ongoing exploration for an international framework to regulate and facilitate trade in services in the early 1980s, given an acute sense in the United States that national interest, and not just the narrow interest of the lobbies or the general interest of the world at large, dictates that services be brought into the trading order. This was a result of the increasing perception that the comparative advantage of the USA has shifted to service transactions (Bhagwati, 1987, p. 26). When the GATT evolved into the

World Trade Organization (WTO) in 1995, one of the most important WTO agreements was the General Agreement on Trade in Services (GATS).

The GATS is a set of rules and discipline which governs the use by WTO member countries of trade measures in services. As stressed in the GATS, international trade in services can take place through cross-border supply (where only the service crosses the border through telecommunications), consumption abroad (where consumers consume the service abroad when travelling, such as ship repair abroad), commercial presence (in a market abroad, of not only juridical persons in the strict legal sense, but also legal entities such as representative offices, branches or subsidiaries), and the presence of natural persons (when an individual has moved temporarily into the territory of the consumer in the context of the service supply, whether self-employed or as an employee). Since 2000, WTO members have begun entering into successive rounds of trade liberalizing negotiations regarding these services.

Although the GATS seems to have diffused the earlier worries of LDCs to some extent by providing a 'built-in agenda' that requires WTO members to enter into successive rounds of negotiations aiming at progressive services liberalization, the North–South division has been visibly resurfaced in the Doha Round of WTO negotiations that began in November 2001.

The most controversial development in the Doha Round services negotiations has been the strong push by some developed countries to establish mandatory minimum market access commitments (benchmarks), to improve market access for their services supplied cross-border and through commercial presence. The overwhelming majority of LDCs are fiercely opposed to any kind of benchmarks, arguing that mandatory market opening commitments go against the very nature of the GATS commitment structure, which explicitly recognizes countries' right to liberalize in accordance with their level of development.

For many LDCs, the 'movement of natural persons' is one of the few areas that is expected to offer concrete benefits from services liberalization. Efforts by some of them to improve commitments in this mode have not met with much success due to opposition from developed countries.

Since the mid-1990s, cross-border trade in information technology services²⁸ (ITS), and IT-enabled services (ITeS), including business process outsourcing²⁹ (BPO), have been the fastest-growing areas of international trade in the globalizing world. Advances in information and communication technology (ICT), substantial investment in education in a number of LDCs and the absence of commensurate employment opportunities in these countries are usually identified as the main factors responsible for this phenomenon.

Most offshored IT services are concentrated in a few countries. For

instance, in 2001, Ireland, India, Canada and Israel, in that order, accounted for over 70 per cent of the total market for offshored IT services, mostly in software development and other IT-enabled services. There is, however, scope for more countries to benefit from the offshoring trend, taking into account specific needs in terms of language skills, time zones and geographical or cultural affinity. Although growth rates of 'BPO and other services'³⁰ exports have recently been faster in a handful of developing and emerging economies, most exports of business services still originate in Organisation for Economic Co-operation and Development (OECD) countries: their share in exports of business services was roughly 76 per cent in 2002, while it was approximately 17 per cent in East Asia and the Pacific region, and only about 7 per cent in the other three regions together.³¹

Offshoring is closely related to technological progress: both follow a process of creative destruction. Being driven by competitive pressures to reduce costs, both lead to displacement of existing jobs. The faster growth of services offshoring has given rise to concerns for job losses in some developed countries. There has even been a push in the United States and Australia to introduce legislation that would limit the outsourcing activities of firms with government contracts. What has received less attention are the net benefits. Amiti and Wei (2004) of the Trade Unit of the International Monetary Fund's (IMF) Research Department undertook a study on the effects of foreign outsourcing of services on employment and labour productivity in US industries during 1992–2001 and UK industries during 1995–2001. This study suggests that: 'service outsourcing not only would not induce a fall in aggregate employment but also has the potential to make firms and sectors sufficiently more efficient, leading to enough job creation in the same broadly defined sectors to offset the lost jobs due to outsourcing' (Amity and Wei, 2004, p. 39).

Conclusion

A number of characteristics are usually associated with the concept of services – they are intangible, non-storable, labour-intensive and follow simultaneous interaction between service producers and consumers; although some service activities could also be identified with the opposite characteristics. There have been a number of refinements of the definition of services, but economists and other social scientists have not yet come to any universal agreement. For general acceptance, service activities are now defined in such a way that the definition not only distinguishes them from agricultural and industrial activities, but also identifies their essential characteristics in a parallel fashion.

Since the 1980s, services have been playing more and more of an import-

ant role in their sustained value-added contributions to GDP across many countries – developed and developing – in different regions of the world. Growth of service activities supported by globalizing forces and technological progress (particularly in ICT) has created many challenges and prospects worldwide. There have been increasing trends of international trade in services, global FDI's structural shift towards services and, more recently, global mergers and acquisitions in service industries. The developed countries and the transnational corporations (TNCs) are reaping relatively more benefits from the growth of service activities worldwide. There is, however, a growing consensus in the development community that firms from LDCs find more opportunities to get connected to foreign markets and, therefore, to boost their participation in international trade in services. Because of their comparative advantage in certain areas, a small number of LDCs have also been successful in utilizing the opportunities to some extent.

In many industrially developed countries, the contribution of the services sector is rising to more than 70 per cent of GDP while that of manufacturing is slipping to less than 20 per cent. With an increasing bundling of services with products, the two sectors are becoming more and more interrelated. In order to focus on core competencies and performance improvement in key areas, many companies in the developed world are increasingly relying on sourcing more service-related functions from specialized firms – domestic and/or foreign. In the process of offshore outsourcing, participating developed countries have recently experienced some low-skilled services job loss, although very small as a percentage of their labour force.³² However, this has prompted political opposition to this process and pressure for trade barriers in a number of developed countries. Recent creation of new high-skilled services jobs³³ in these countries is expected to counter protectionism in services trade. Similarly, one can hope that the growing volume and scope of tradeable services through the mechanisms of FDI and M&As cannot escape but call for determined and innovative GATS negotiation strategies for liberalization of the movement of individual service providers³⁴ in the foreseeable future.

Notes

1. 'Hundreds of millions of poor people in developing countries make their living as microentrepreneurs – as farmers, street vendors, and homeworkers, and in a range of other occupations, a large share of them women . . . They are a big part of the informal economy, which is substantial in many developing countries' (World Bank, 2005b, p. 33).
2. Nusbaumer (1987, p. 1).
3. Riddle (1986, p. 2).
4. World Bank (2005b), Table 3, pp. 260–61.
5. OECD (2000, p. 7).

6. The term 'recipient' is sometimes more appropriate than the term 'customer'. While parents are the customers in purchasing educational services for their children, the latter are the recipients of the service.
7. Bell (1973) used the term 'post-industrial sector' to refer to the service sector in the context of his description of a post-industrial society in which the service sector is dominant.
8. Antecedents to this conventional view can be found in Fisher (1935, 1939), Clark (1940), Fuchs (1968) and Kuznets (1971).
9. Bhagwati (1987, p. 18).
10. These terminologies of inhousehold and outhousehold are used in Bhagwati (1987, p. 20).
11. Examples of these long-distance tradeable services include banking transactions from computer terminals at home or abroad, professional services communicated via satellite, and medical diagnosis by video transmission.
12. 'Tacit' knowledge is specific, experimental and heavily influenced by user needs, and hence less amenable to replicate and transmit. But 'codified' knowledge is generic and easy to transfer among the firms in an industry or among the industries within a sector.
13. MSITS (2002, p. 9).
14. Note that construction services and utilities have been excluded from the service sector and included in the industrial sector using the argument that these items reflect capital-intensive production methods.
15. All data are from World Bank (2005a).
16. As long as the informal or unorganized traditional segment of many LDCs remains disproportionately intensive in services and escapes systematic documentation by the statistical organizations, the service share in their national income would be understated (Bhagwati, 1987, pp. 20–21).
17. World Bank (2005a).
18. The public sector both consumes and provides various services. Governments are the major providers of health, education and social services in most of the countries. They also provide various trade-related services in the context of international transactions of goods and services through specialized government institutions such as export credit agencies and export–import banks.
19. The framework of balance of payments (BOP) on services transactions between residents and non-residents provides a sound basis for the measurement of trade in services in the conventional sense. Services international transactions are in practice referred to as trade in services. Each service item under these transactions display a credit and a debit value, representing respectively export and import of that service.
20. Different forms of capital inflow include foreign aid, foreign direct investment (FDI) and commercial borrowings.
21. High annual growth rates of global inflows of foreign direct investment (FDI) during 1986–2000 had, however, dramatically reversed during 2000–03 (as shown in Table 34.4), although prospects for a steady recovery of FDI flows have been expected to be promising since 2004.
22. Baumol's (1967, pp. 415–16) assertion is based on the basic premise that economic activities can be grouped into two types: technologically progressive activities in which innovations, capital accumulation, and economies of large scale all make for a cumulative rise in output per labour hour, and technologically non-progressive activities which, by their very nature, permit only sporadic increases in labour productivity.
23. Dutt and Lee (1993, pp. 312–13) have provided some arguments of these investigations. Regarding the link between faster rate of output growth and manufacturing sector's productivity growth, theoretical arguments have been made in terms of the economies of longer production runs, demand effects, and learning by doing. A large empirical literature on this link has also been developed around the so-called Verdoorn's law. Regarding spin-off effects of the third mechanism, production of various intermediates (for example, chemical as well as electrical and non-electrical machineries) as a result of technological progress in the manufacturing sector, and thereby reducing costs and prices of inputs and improving their quality, are seen as productivity effects on other sectors.

24. 'Of course, if a relative expansion of the service sector slows down the rate of technological change and therefore dampens investment incentives, it is possible for service sector expansion to reduce the rate of growth even in a demand-constraint economy' (Dutt and Lee, 1993, p. 313).
25. Baumol (1985, pp. 301–2) argues that services are too heterogeneous and all services to some degree, and some services to an extraordinary degree, do permit innovation and productivity growth. He classifies the services into three broad categories: stagnant personal services (such as haircutting, teaching, and live artistic performance), the progressive impersonal services (such as telecommunications), and the asymptotically stagnant impersonal services (such as broadcasting, computation, and research and development). He however emphasizes that the three groups shade off into one another, and a particular service may well move from one category to another as circumstances change.
26. The author (Dutta, 2006) has recently reviewed how ICT as a production sector or as an enabler of socio-economic development has potential for economic development and competitiveness in the context of the South Asian region.
27. Compared to the services sector firms, the manufacturing firms' usage rates range from 30 to 50 per cent for websites and from 50 to 70 per cent for e-mail, while the traditional sectors that have driven many LDCs forward, such as the agro-industry and automotive industry sectors, seem to be lagging in their ICT use (Qiang et al., 2006, p. 60).
28. Information technology services (ITS) are computer and related services such as software development and implementation services, IT support services, application development and maintenance, business intelligence and data warehousing, content management, e-procurement and business-to-business (B2B) marketplaces, enterprise security, package implementation, system integration, enterprise application integration, total infrastructure outsourcing, web services (Internet content preparation, and so on), web-hosting and application service providers (ASPs) (Mattoo and Wunsch-Vincent, 2004, p. 767).
29. Mattoo and Wunsch-Vincent (2004) have compiled three categories of business process outsourcing (BPO) services, although they consider neither that the list of activities is exhaustive, nor that the categories are mutually exclusive. They are: (a) customer interaction services (sales support, membership management, claims, reservations for airlines and hotels, subscription renewal, customer services helpline, handling credit and billing problems, telemarketing and marketing research services, and so on); (b) back-office operations (data entry and handling, data processing and database services, medical transcription, payment services, financial information and data processing and handling, human resource processing services, payroll services, warehousing, logistics, inventory, supply chain services, ticketing, insurance claims adjudication, mortgage processing); and (c) more independent professional or business services (human resource services including hiring, benefit planning and payroll, and so on; finance and accounting services including auditing, book-keeping, taxation services, and so on; marketing services, product design and development).
30. The 'BPO and other services' category prepared by Mattoo and Wunsch-Vincent (2004) includes 'total services' in the IMF statistics minus transportation, travel and government services. The chosen category of services thus includes: communication, construction, insurance, computer and information, 'other business' and personal, cultural and recreational services, as well as royalties and licence fees.
31. Mattoo and Wunsch-Vincent (2004, p. 768).
32. For example, job loss due to offshoring in the USA is estimated to be 0.25 million per year. It is small relative to total US employment of 137 million, and accounts for less than 2 per cent of the roughly 15 million Americans who involuntarily lose their jobs each year (Brainard and Litan, 2004, p. 2).
33. For example, an analysis of US Bureau of Labour statistics shows that the overwhelming majority of new jobs created in recent years in the USA have been in occupations in which tacit interactions – such as complex negotiations – are the main component. This type of job now makes up 41 per cent of the US labour force (Manyika, 2006).
34. A recent study by Winters et al. (2002) shows that if OECD countries allow temporary

access to foreign service providers (from developing and least-developed countries) equal to just 3 per cent of their labour force, the global gains would be over \$150 billion – more than three times their total overseas development assistance (ODA) flows (quoted from Chaudhuri et al., 2004, p. 363).

Bibliography

- Amiti, M. and S.-J. Wei (2004), 'Demystifying Outsourcing', *Finance and Development*, December: 36–9.
- Baumol, W.J. (1967), 'Macroeconomics of Unbalanced Growth: The Anatomy of Urban Crisis', *American Economic Review*, 57: 415–26.
- Baumol, W.J. (1985), 'Productivity Policy and the Service Sector', in R.P. Inman (ed.), *Managing the Service Economy: Prospects and Problems*, London, New York: Cambridge University Press.
- Baumol, W.J., S.A. Batey Blackman and E.N. Wolf (1989), *Productivity and American Leadership*, Cambridge, MA: MIT Press.
- Bell, D. (1973), *The Coming of the Post-Industrial Society: A Venture in Social Forecasting*, New York: Basic Books.
- Bhagwati, J. (1984), 'Splintering and Disembodiment of Services in Developing Nations', *World Economy*, (June): 133–44.
- Bhagwati, J. (1987), 'International Trade in Services and its Relevance for Economic Development', in O. Giarini (ed.), *The Emerging Service Economy*, Oxford: Pergamon Press.
- Brainard, L. and R. Litan (2004), "'Offshoring" Service Jobs: Bane or Boon – and What to Do?' Policy Brief, No. 132, Brookings Institution, April.
- Chaudhuri, S., A. Mattoo and R. Self (2004), 'Moving People to Deliver Services: How can the WTO help?' *Journal of World Trade*, 38 (3): 363–93.
- Clark, Colin (1940), *The Conditions of Economic Progress*, London: Macmillan.
- Dutt, A.K. and K.Y. Lee (1993), 'The Service Sector and Economic Growth: Some Cross-Section Evidence', *International Review of Applied Economics*, 7 (3): 311–29.
- Dutta, D. (2006), 'Role of ICT in Development Process: A Review of Issues and Prospects in South Asia', in H. Rahman (ed.), *Information and Communication Technology for Economic and Regional Developments*, Hershey, PA: Idea Group.
- Fisher, Allan (1935), *The Clash of Progress and Security*, London: Macmillan.
- Fisher, Allan (1939), 'Production, primary, secondary, and tertiary', *Economic Record*, 15 (June): 24–38.
- Fuchs, Victor (1968), *The Service Economy*, New York: National Bureau of Economic Research.
- Gemmell, Norman (1982), 'Economic Development and Structural Change: The Role of the Service Sector', *Journal of Development Studies*, 19 (1): 37–66.
- Giarini, Orio (ed.) (1987), *The Emerging Service Economy*, Oxford: Pergamon Press.
- Griliches, Z. (ed.) (1992), *Output Measurement in the Service Sectors*, Studies in Income and Wealth, Vol. 56, Chicago, IL and London: University of Chicago Press.
- Griliches, Z. (1994), 'Productivity, R&D, and the Data Constraint', *American Economic Review*, 84: 1–23.
- ICTSD and IISD (2005), *Doha Round Briefing Series: Hong Kong Update*, International Centre for Trade and Sustainable Development & International Institute for Sustainable Development, Geneva, Vol. 4 (November).
- Inman, R.P. (ed.) (1985), *Managing the Service Economy: Prospects and Problems*, New York: Cambridge University Press.
- Kuznets, Simon (1971), *Economic Growth of Nations: Total Output and Production Structure*, Cambridge, MA: Harvard University Press.
- Malthus, Thomas (1820), *Principles of Political Economy*, New York: Augustus M. Kelley.
- Manyika, James (2006), 'The Coming Imperative for the World's Knowledge Economy', *Financial Times*, Asian edition, 16 May.
- Mattoo, A. and S. Wunsch-Vincent (2004), 'Pre-empting protectionism in services: The GATS and outsourcing', *Journal of International Economic Law*, 7 (4): 765–800.

- MSITS (2002), *Manual on Statistics of International Trade in Services 2002*, Geneva Luxembourg, New York, Paris, Washington, DC: UN, EC, IMF, OECD, UNCTAD and WTO.
- Nusbaumer, Jacques (1987), *The Services Economy: Lever to Growth*, Boston, MA: Kluwer Academic Publishers.
- OECD (2000), *The Service Economy*, Paris: Organisation for Economic Co-operation and Development.
- Oulton, Nicholas (2001), 'Must the Growth Rate Decline? Baumol's Unbalanced Growth Revisited', *Oxford Economic Papers*, 53: 605–27.
- Qiang, C., G. Clarke and N. Halewood (2006), 'The Role of ICT in Doing Business', in *Information and Communications for Development 2006: Global Trends and Policies*, Washington, DC: World Bank pp. 57–85.
- Riddle, D. (1986), *Service-Led Growth: The Role of the Service Sector in World Development*, New York and London: Praeger.
- Riddle, D. (1987), 'The Role of the Service Sector in Economic Development: Similarities and Differences by Development Category', in O. Giardini (ed.), *The Emerging Service Economy*, Oxford: Pergamon Press, pp. 83–104.
- Smith, Adam (1776), *An Inquiry into the Nature and Causes of the Wealth of Nations*, Oxford: Oxford University Press.
- System of National Accounts, 1993* (1993), a joint publication by Commission of the European Communities, International Monetary Fund, Organisation for Economic Co-operation and Development, United Nations and World Bank, Brussels/Luxembourg, New York and Washington, DC.
- UNCTAD (2000), *World Investment Report 2000: Foreign Direct Investment and the Challenges of Development*, New York and Geneva: United Nations Conference on Trade and Development, United Nations.
- UNCTAD (2004), *World Investment Report 2004: The Shift Towards Services—Overview*, New York and Geneva: United Nations Conference on Trade and Development, United Nations.
- Winters, A., T. Walmsley, Z. Wang and R. Grynberg (2002), 'Negotiating the Liberalisation of the Temporary Movement of Natural Persons', Economics Discussion Paper 87, University of Sussex, Brighton.
- World Bank (2005a), *World Development Indicators 2005* (CD-ROM), Washington, DC: IBRD, World Bank.
- World Bank (2005b), *World Development Report 2005, A Better Investment Climate for Everyone*, Washington, DC: World Bank.
- World Bank (2006), *Information and Communications for Development 2006: Global Trends and Policies*, Washington, DC: World Bank.
- WTO (2006), *Measuring Trade in Services*, World Trade Organization, March.

35 Urbanization and rural–urban migration

Charles M. Becker

Introduction: patterns of urbanization

This chapter surveys the pace, causes and problems associated with urbanization and rural–urban migration in developing and middle-income countries.¹ The timing is appropriate: the world’s population, for the first time in human history, is on the verge of containing more city dwellers than rural inhabitants. Moreover, the world is experiencing a great population deceleration. Thus, while urbanization (defined here as the proportion of a population living in cities and towns) continues to rise, urban population growth rates have fallen in most developing countries. Hence, less is being written about uncontrolled city growth or megacity explosions. Yet, severe urban problems remain, and in many cases are unaltered by the declining pace of growth and in-migration.

For developing countries as a whole, excluding China, annual urban population growth averaged about 4 percent between 1955 and 1985, but after the mid-1960s began an erratic but sustained decline (UN, 2002, pp. 19–21). This decline became precipitous in the 1990s, with urban population growth rates falling to 3.2 percent in the first half of the decade, 3 percent in the second half, and a projected 2.8 percent during the years 2001–05. Rural population deceleration was even greater and the decline began earlier. Between 1950 and 1975, rural population growth rates ranged from 1.6 percent to 2.1 percent for developing countries, but fell to 1.3 percent during the next decade, and thereafter decelerated at about 0.2 percentage points every five years. By 2000–05, the UN forecast for rural population growth in developing countries was only 0.5 percent per annum.

These patterns have several implications. First, rapid deceleration in rural population growth implies that, even with annual urban population growth dropping from 4.15 percent in 1960–65 to 2.81 percent four decades later, the urbanization growth rate decline is more modest (from 1.80 percent to 1.33 percent per annum). Second, as natural population growth is higher in rural than urban areas, rural–urban migration must account for 100 percent or more of the urban–rural population growth differential.² These trends are crude, glossing over regional and national differences, differing definitions of the ‘urban’ population and accuracy of demographic estimates, and the

appropriate set of ‘developing’ countries. Nonetheless, the trends are sufficiently sharp that there can be little doubt that the patterns are broadly accurate.

The importance of urban in-migration varies markedly by both region and city size (NRC, 2003). Chen et al. (1998) estimate that migration and population reclassification account for about 40 percent of all developing-country urban population growth from the 1950s to the 1980s. Migration accounted for most urban growth in Latin America, though its share is declining, since the region is now predominately urban (meaning a smaller pool of potential migrants). Migration’s share is smaller but rising in Asia, which is less urbanized, and has very rapid urban economic growth. Rural–urban migration declined from the 1960s to the 1980s in Africa, which has suffered largely stagnant urban economies.

More recently, observed patterns imply that migration will account for less than 40 percent of urban growth, save for in countries and regions (notably, Ethiopia, Ghana and Central Asia) where rural–urban fertility differentials are vast. The National Research Council (NRC) (2003, p. 126) presents information on the birthplace of women aged 15–49 in USAID-sponsored Demographic and Health (DHS, www.measuredhs.com) surveys. In cities with fewer than 100 000 people, the proportion of fecund women born in rural areas ranges from 36 percent in North Africa to 61 percent in South-East Asia. In cities of 0.5 to 1 million, the rural birth share ranges from 21 percent (Latin America) to 35 percent (South, Central and West Asia). In megacities over 5 million, the rural birth share ranges from 12.5 percent (Latin America) to 38 percent (South-East Asia). China is very different from other regions: urban growth has accelerated following the modernization reforms, and reached 4.7 percent in 2001 (Chan and Hu, 2003). With accelerating economic growth and a fairly strict one-child policy in urban areas, net rural in-migration accounted for 74 percent to 80 percent of urban population growth in China between 1978 and 2000, and that percentage appears to be increasing.

Models of structural transformation

The key features of urbanization have long been known, as has the understanding that sustained economic growth without urbanization is nearly inconceivable. Urban areas by definition are the locations for activities with substantial economies of scale internal to the firm, along with urbanization and localization agglomeration economies that are external. Higher densities also mean reduced cost of infrastructure and public service provision. These forces result in higher productivities in cities, and transformation from low to high productivities is the essence of economic growth. Moreover, urban activities can better use skills whose application is most

effective in the presence of sophisticated division of labor. Urbanization raises returns to differentiation of labor, and hence the acquisition of refined sets of skills, rather than broad but shallow knowledge. Specialization enables the application of skills learned in many years of apprenticeship or education, to which there may be significant positive externalities to the community at large.

While not using the jargon above (which emerged in the urban economics literature in later decades), W. Arthur Lewis (1954, 1955) was acutely aware of the advantages of urbanization, and emphasized it in a revolutionary approach to economic growth – through the lens of the dualistic economy, with a traditional agricultural world coexisting with a more vibrant modern urban society. Lewis divided poor economies into two sectors, a ‘modern’, capitalist, industrial sector, and a backward, traditional, agricultural (and traditional services) sector. Labor migrates across sectors to equate expected utility from each activity. In modern industry, labor is paid the value of its marginal product. The difference between the net value of output and the wage bill equals profits, some fraction of which is reinvested. In the traditional sector, farmers, artisans and traders effectively behave as tiny monopolistic competitors, each receiving an average product. For other than a few fortunate individuals, the amount earned is equal to a subsistence-plus level that is roughly constant, while marginal product is zero. Economic growth occurs by reinvestment of modern sector profits and drawing workers from the traditional to the modern sector. Real wages begin to rise only when the ‘surplus labor’ era ends and a standard neoclassical economy emerges in both sectors.

As Lewis did not seek to explain formally how earnings were determined in the traditional sector, neoclassical economists wondered why earnings would not rise if labor were withdrawn, or why they would not fall if population growth exceeded the rate of outmigration to the modern sector. Lewis himself answered that institutions adjusted, while Sen (1966) provided conditions under which ‘unlimited labor’ could be generated in a neoclassical setting. More critically, Hansen (1979) showed that in an African context with unlimited supply of low-quality land, the Lewis model would also emerge in a world of neoclassical labor markets.

The standard dual-economy model seemed to accord with the erstwhile stylized fact of limited increases in living standards for the poor in developing countries. Yet, evidence has mounted that in most countries neither rural nor urban earnings are stable in the face of either positive or negative shocks (Becker and Morrison, 1999; Jamal and Weeks, 1988). There is now irrefutable evidence that real wages are rising in rapidly growing areas of China and South-East Asia today, even though a vast number of low-income workers remain in rural areas (*The Economist*, 2007).

Even before detailed labor market studies became commonplace, economists realized that the simple dualistic framework had no place for the many urban dwellers who earned little or were openly unemployed. The concept of the ‘informal sector’, a term popularized by the International Labour Office, took hold, as did recognition that demand for plum high-income ‘formal’ sector jobs exceeded potential supply.³ In response, Todaro (1969) and Harris and Todaro (1970; synthesized in Blomqvist, 1978) put forth a model with an institutionally determined urban wage above rural incomes. Equilibrium in this setting requires possibly unattractive outcomes to urban migration as well. Migration from rural areas then will occur until the *ex ante* expected utility gain is zero, with the anticipated rewards from landing a high-paying formal sector job offset against possible losses due to extended periods of unemployment or employment in unrewarding informal sector jobs. Wages will be highest in the formal sector, while living standards will be lowest among those unemployed or in the informal sector.

The expectations-driven model proved highly attractive, and the framework has remained. However, the model is not supported by evidence on labor markets in developing-country cities. Mazumdar (1983) was perhaps the first to note that unemployment was a luxury status that few poor could afford, and that it was less prevalent among recent migrants than among long-standing city dwellers. Considerable evidence from many countries also has pointed to a pattern of prearranged jobs for new migrants, who tend to be assisted by relatives and networks based on common origin regions (Becker and Morrison, 1999).

Alternative models

The institutionally determined wage also came in for severe criticism. Stiglitz (1969) began by examining different incentives for migration, and ultimately developed a labor turnover model (1974). In this paradigm, training is costly, and quit rates depend in part on wages offered. A rational firm with some wage-setting power will therefore pay a premium to reduce turnover and compensate for firm-specific skills. Wage gaps and unemployment still emerge, but urban wages now move procyclically if quit rates are suppressed by visible unemployment.

Still others, led first by Stiglitz (1976) and Bliss and Stern (1978), argued that urban formal sector wages above rural and informal levels could be explained by efficiency wages. In particular, firms have an incentive to pay wages to ensure adequate health and nutrition on the part of their workers when productivity gains exceed the higher wage bill.

A different story of wage gaps comes from Sabot’s (1979) analysis of East Africa, emphasizing the importance of education credentialism for

formal sector employment not tied directly to productivity differentials. Small, segmented labor sub-markets form, leading to a 'queuing' pattern of migration. During periods of rapid formal sector employment growth and restricted supply of fresh urban graduates, credentials requirements decline, providing incentives for migration, especially when rural secondary schooling is widespread. Under the reverse circumstances, migration will be limited, even if average income disparities between urban and rural areas are increasing. Sabot's model is intuitively reasonable, but has received little examination outside of East Africa, partly because of poor data.

Education also figures prominently in demographic shift models of urbanization (Becker and Morrison, 1993; Becker and Grewe, 1996). These models recognize that migration is highly age-specific, with young adults more likely to move than older adults or children. Furthermore, shifts in age structure are correlated with economic growth, so that econometric studies of migration using aggregate population flow data often suffer from problems of observational equivalence. With disaggregation, it turns out that standard attraction forces of cities (higher expected wages, educational and employment opportunities, better services) and repellent forces from rural areas (deteriorating agricultural conditions) still matter. But the importance of labor market attractions is strongest for young, prime-age people, especially men, and many other aspects of cities matter as well. Furthermore, growing rural prosperity does not have an invariably negative impact, since rising rural incomes are associated with rising education, and secondary school graduates have much greater migration rates than their less-educated peers. Moreover, some rural prosperity is needed to bankroll the cost of urban migration and job search.

By the 1980s, economists had branched out from the simple story of household location decisions being driven by the choices of prime-aged (and presumably male) workers. Short-term, seasonal migration and migration of women began to draw attention. So, too, did the notion of an individual's migration as part of a household's optimizing decision. The 'new migration economics' literature is most strongly associated with the work of Stark (1991), though others (notably, Taylor, 1987) were major contributors as well. This literature has added greatly to the sophistication of economic analysis of migration, and by focusing on the returns to migration by risk-averse households, qualifies the straightforward results of simpler models. The most valuable lesson from this literature is that sending young adults to the city may be beneficial beyond the direct monetary rewards if, for instance, overall household income variance is reduced, or if the establishment of one family member in the city leads to subsequent migration and higher returns for others.

Micro empirics

The ‘modern era’ of development economics – namely, that informed by a proliferation of micro household and labor market surveys – thus commenced with increasingly sophisticated models giving somewhat contradictory predictions. There was little dispute that migrants reacted to opportunity differentials, and that markets rather than institutional rules determined earnings and prices. Yet, few common points beyond this bland statement emerged. Indeed, the key lesson of urban growth since the mid-1980s lies in the heterogeneity of developing-country migration and city growth. In China and South-East Asia, booming urban manufacturing has led to very high in-migration rates for traditional ‘pull’ reasons. Much of Africa has been beset by stagnant urban economies, but cities continue to grow, reflecting high birth rates as well as a high share of young adults in the population. And, in some countries visited by economic decay or collapse, urbanization rates actually have declined.

With deceleration in urbanization leading to a de-emphasis on ‘runaway cities’, empirical work today asks questions about inter-sectoral and social class mobility, about the extent of urban inequality and poverty, about the nature of informal economies, and about the nature of remittances back to home areas. Regarding poverty, perhaps the most useful cross-country, urban–rural assessments are based on comparisons of the numerous demographic and health surveys. Sahn and Stifel (2003) use 41 such surveys in their analysis of conditions in Africa alone. They relate measures of quality of life, from infant mortality rates to asset ownership to education indicators to urban residence, finding huge urban–rural gaps with very few exceptions. Their rich set of results yields no evidence of pan-continental urban–rural convergence, save for infant mortality. Individual countries that appeared to experience convergence (mainly in the 1990s) across a wide range of measures include Burkina Faso, Cameroon, Ghana, Mali, Niger, Senegal and Tanzania. Widening gaps appear in Madagascar, Nigeria, Zambia and Zimbabwe: convergence seems more likely where initial gaps are greatest, and where governance is relatively good.

Household surveys point to the rich diversity of migration flows from rural areas. Confirming strong selective outmigration, Mberu (2006) finds that migrants in Ethiopia actually have higher living standards than non-movers, with no effective difference once one controls for education and non-agricultural income differences, a commonplace finding (Becker and Morrison, 1999). Agesa and Kim (2001) focus on rural–urban migration of individuals versus whole families in Kenya. Older households with children are more likely to migrate as a unit, rather than to split up. As migrants account for about 70 percent of Kenya’s urban labor force and most

migrants are from split households, Agesa and Kim argue that decreasing split migration may also reduce urbanization.

Their paper reflects renewed attention to the heterogeneity of migrant populations, also a theme of Beauchemin and Bocquier's (2004) review of West African migration. West African data also indicate much lower unemployment rates among migrants than permanent residents. Moreover, much as in transition nations, economic decay in West Africa is associated with outmigration from secondary towns to capital cities and abroad. There is also substantial migration back to rural areas, and not simply to retire. Most urban families maintain rural ties, and will send children or adult family members back to the countryside when urban conditions deteriorate in terms of both income and security.

The explosion of micro surveys also has made it much easier to track urban inequality, which appears to rise during early stages of development but not indefinitely. Part of this increase is due to growing intercity or inter-regional inequality, especially in the absence of correction for living cost differentials. But inequality rises within a given city as well, partly because of the increased importance of skilled labor in the production of goods and services. Inequality may then decline as skills generally increase across the labor force, as skill premia decline, and as unskilled labor shortages emerge (Knight and Sabot, 1990). The well-known rise of urban inequality, and less well-known decline, in China are documented in Démurger et al. (2006). An apparently universal phenomenon is that migrants are not locked out of the formal sector: indeed, West African studies generally find that migrants are more likely than non-migrants to find formal sector employment (Beauchemin and Bocquier, 2004). However, it is also clear that formal sector employment does not guarantee high living standards, though it does seem likely to reduce income variance.

Liang et al. (2002) find from a 1 percent sample of the 1990 Chinese census that rural industrialization has no significant effect on the likelihood of either inter- or intra-provincial migration. A reasonable interpretation is that the variable is capturing multiple effects that run in opposite directions. These multiple effects and endogeneity issues are discussed at length for specific types of infrastructure, public services and employment in an exceptionally careful and elegant study of migration from rural areas and secondary towns to Burkina Faso's main cities (Beauchemin and Schoumaker, 2005). They find that secondary schooling and paved roads have large impacts on migration to large cities. Non-agricultural employment opportunities in towns and markets in rural areas both deter outmigration. There are also very strong age (peaking at 20–29) and distance effects. The presence of health centers increases migration likelihood, though there may be an endogeneity problem. Other infrastructure

(electricity, piped water, telephone service) appears to have no impact. Thus, rural and small-town development efforts are unlikely to deter migration to large cities. However, small towns can be made more viable to existing residents (and attractive to those in nearby rural areas) if they experience increasing non-agricultural employment.

There is also a growing literature on the role of remittances. Often, the figures are huge. For example, Kyrgyzstan's central bank has estimated that remittances in 2005 equaled 50 percent of exports and 14 percent of GDP (Ukueva, 2007). The purpose of remittances varies greatly from one setting to another. In a country such as Tajikistan, which is poor and whose migrant workers are mainly unskilled married men employed abroad, remittances are used mainly for food and basic necessities (Ukueva, 2007). Elsewhere, remittances may be used to finance education, housing construction, purchase of a vehicle, or migration by other family members. Page and Plaza's (2006) survey finds that remittances are associated with both higher education and improved health of origin families. McCormick and Wahba (2003) in a study of international return migration in Egypt find that while 45 percent of those who return came from public sector jobs, only 9 percent return to the public sector. Returning migrants from urban areas are far more skilled than their rural counterparts. But even controlling for human capital characteristics, those from urban areas are much more likely to start a non-farm enterprise (and, relative to existing urban small businesses, will employ more people).

Skill selectivity in both internal and international migration has been well documented. Moreover, as Kanbur and Rapoport (2005) show theoretically, the impact of a skill brain drain on the source region is ambiguous. There are costs (relatively productive workers are lost), but the possibility of migration also raises returns to education, thereby increasing supply. The presence of information networks can also lead to increasing returns through externalities. Indeed, the dynamics are complex: if skilled emigration raises returns to additional skilled labor emigration and lowers returns (fewer agglomeration economies) to staying, the origin economy may be permanently stunted. But this outcome is not inevitable, as returns at home to increasingly scarce skills might make return migration attractive. Remittances plus further skill acquisition in destination regions, and their application when migrants return, further reduce brain drain costs to origin regions.

Taylor et al.'s (2003) empirical analysis of the rural China finds that out-migration increases origin-household self-employed income and, of course, remittance income – but at the expense of even greater losses in cropping, wage and other income. This study is careful to correct for migration endogeneity and selectivity, which is essential for the results to be credible. Ideally, one would also correct for remittance endogeneity, but this can be

difficult, since few surveys contain information on potential demands both in origin and destination areas for household members. Taking remittances as exogenous, Taylor et al. (2003) find that the impact of migration on total origin family income is not significantly different from zero, while per capita income increases since the number of family members has been reduced. These findings appear consistent with empirical work elsewhere (Özden and Schiff, 2006).

The nature of migration flows, which depend on relative rate-of-return differentials, is more varied. Mora and Taylor (2006) examine outmigration from rural Mexico to competing Mexican and US destinations, and find that unskilled labor flows mainly to the United States, while skilled labor flows to Mexican cities. The skills' nature, and the extent and distance of migration also affect the impact on origin regions' economies. It is difficult to generalize: some areas will be emptied of young workers; others will experience construction booms as the result of remittance flows.

Finally, the urban informal sector is now well documented (Guha-Khasnobis and Kanbar, 2006). To crudely characterize recent findings, the extent of segmentation is far less than was anticipated in earlier empirical studies, though returns to human capital attributes do vary by sector, even controlling for selection effects. The informal sector has a high elasticity of employment with respect to output, and it thrives, both in terms of output and earnings, when the economy as a whole is doing well.

Urbanization and growth: new macro empirics

While recent emphasis has been on the micro side, there also have been substantial contributions from researchers using aggregate analysis. Again, the role of remittances has been central, reflecting the huge increase in the numbers of both temporary and permanent international migrants, with studies quantifying their impact on recipient areas. Labor-exporting countries now commonly receive remittance flows equal to several percent of GDP. Page and Plaza (2006) summarize the literature on macro determinants, and in their own empirical work find that exchange rate distortions and the presence of black markets for currency strongly deter remittances. They also infer unofficial (or unrecorded) remittances as a share of the total: their estimates vary from virtually zero in South Asia to 27 percent in the Middle East and North Africa, to 73 percent in sub-Saharan Africa, and about 50 percent elsewhere.

More negative is the brain drain effect, with large losses relative to the stock of skilled workers occurring in the poorest and least stable countries (Lucas, 2006; Amin and Mattoo, 2007). For poor and unstable countries, brain drain estimates can be large. The costs are perhaps greatest in countries such as Nigeria, with a tradition of high-quality tertiary education

and a substantial middle class, from which it is estimated that more than one-third of those with university education are now in OECD countries.

There is also some potential for transmitting macroeconomic shocks from advanced host countries to poorer sending countries: in their time-series analysis of aggregate remittances from the USA to Mexico, Vargas-Silva and Huang (2006) find that US macroeconomic conditions are more important than home-country conditions in determining remittances. However, Quartey and Blankson (2004) use micro data from Ghana to find that remittances also run countercyclically with respect to the origin macroeconomy, increasing during periods of economic crisis; Lucas (2006) draws a similar conclusion. However, Adams (2006) does not find a link between aggregate remittances and the origin country's poverty. Taken together, these findings suggest short-term stabilization effects of remittances, but no systematic contribution to reduced global GDP inequality.

Macro data sets also have improved, enabling new sorts of analysis that were impossible a few years back. For example, Becker et al. (2005) use data on monthly migration from Kazakhstan to Russia to examine the time-series impact of economic crisis on migration flows of different age groups. They find that some economic news is perceived more rapidly than others, and that there may be threshold effects: small shocks get lost in the noise, while large shocks are noticed and have rapid impacts.

On the modeling side, Brueckner and Zenou (1999) and Brueckner and Kim (2001) fill an important, neglected niche by adding an endogenous land market to the Harris–Todaro model. Much of the benefit of a higher urban formal sector wage will be dissipated through higher land rents. Furthermore, to the extent that formal sector workers live apart from – and closer to the urban center than – those in the urban informal sector, rental gradients will reflect labor market distortions. To my knowledge, no empirical work on this link has been conducted, though Malpezzi's (1998) analysis of rents in Cairo finds virtually no pattern of rent transfers, at least in that setting. Other notable work on urban land markets in developing countries includes Dowall and Leaf's (1991) study of Jakarta, and Lanjouw and Levy's (2002) analysis of land prices in Ecuador. There are few surprises: weak registration and property rights reduce land values, lower densities, increase sprawl, and therefore increase commuting costs and reduce labor market opportunities for the urban poor.

Since the mid-1980s, growth economists have become increasingly aware of the importance of cities as a source as well as a consequence of vibrant economic growth. This comes as no surprise to urban economists, who emphasize scale and agglomeration economies. However, measuring them is not easy, though Henderson (1988) has provided many creative ways of doing so. Nor has it been easy to disentangle causality and demonstrate a

clear link from urbanization to growth, though consensus empirical evidence is now that there is a strong, positive effect (Soludo and Kim, 2003).

The importance of urbanization and its endogeneity in the growth process is central to the literature known as the 'new economic geography' (Venables, 2000). Scale and agglomeration economies are at the heart of this analysis, and are hypothesized to be critical contributors to overall economic growth. These forces also give rise to uneven rather than smooth patterns of urban development: some cities grow rapidly, while others progress slowly. Early development is likely to be characterized by increasing population and production concentration within systems of cities, with deconcentration following at later stages of development. This literature also emphasizes the importance of transportation costs and barriers to trade, and, indirectly, the social cost of protectionism on behalf of rural areas and smaller cities.

Coping with city growth

Large urban areas are among the greatest social inventions of all time. (Mills, 2000, p. 73)

It is indisputable that overall urbanization is strongly associated with a level of economic development (for regressions, see NRC, 2003). However, changes in urbanization and urban population growth rates are not exclusively linked to development level, nor to its rate of change (economic growth). In countries such as China, rapid growth means rapid urbanization, as spatial economic models of growth predict.⁴ Those who come to the cities are poor and mainly work in low-paid jobs. But, most live far better than in their origin regions. Migration is large, and the gains are great even when growth is widespread, since rising prosperity and improved rural schooling and roads all make it easier to migrate. The reverse holds elsewhere: when economic chaos ensued with the collapse of the USSR, urban populations stopped growing, and many secondary cities virtually died.

Cities also grow when economic conditions in origin regions are bad or deteriorating. Sub-Saharan Africa experienced the most rapid population growth, and the greatest urbanization increases after East Asia, yet few of its countries enjoyed rapid economic growth. Urban economic stagnation will not prevent rapid growth in the face of even greater rural decay, a pattern also predicted by economic models (Becker and Morrison, 1988).

Migration and attendant urbanization are best thought of as equilibrating responses to positive opportunities and negative shocks. Migrants are not a particularly disadvantaged group and few urban problems are unique to migrants. Nor are cities bad. It is true that crime and a range of negative

externalities, such as pollution and congestion, generally rise with city size.⁵ But scale and agglomeration economies work the other way, and most economists would argue that these latter effects are more important in most settings.

It is also true that governments distort urban structure. These distortions include efforts to restrict the growth of large cities, to restrict housing space and to underinvest in urban infrastructure and public health while overinvesting in showcase projects and favored industries (Mills, 2000). These distortions are costly, in terms of both static and dynamic efficiency. Elsewhere, and often concurrently, governments engage in activities that comprise an ‘urban bias’ (Lipton, 1976). If large cities’ services and infrastructure are neglected, it is usually much worse in small cities and rural areas. Protectionist trade policies favor industrial cities, and showcase projects (from airports to hospitals to stadiums to skyscrapers) almost invariably go to a few large cities, or to a handful of small, favored cities (new capitals near the President’s home). Yet, none of this implies that urbanization, or large city growth, is bad. Rather, these issues remind us of the inefficiencies caused by government intervention aimed at favoring particular groups, classes or industries.

Three features of urbanization seem reasonable to anticipate in the coming decades. First, the world will continue to become increasingly urban. In nations with large rural populations and large rural–urban fertility differentials, this growth will be fueled mainly by migration, and urbanization will increase rapidly. In countries with slow economic growth and initially high levels of urbanization, further urbanization will be modest, and urban growth will be driven mainly by natural increases. Second, in much of the world, intra-urban inequality is likely to increase further. Greater mobility and globalization will ensure that those with skills earn high and likely rising incomes, even if they are in poor cities. Globalization is likely beneficial to the poor, but far more modestly. The same forces will increase differences in growth and prosperity across cities within developing countries. Those that are well connected and have industries and services that benefit quickly from technological advances elsewhere will boom; others will lag.

Finally, future urban growth seems likely to be dominated by service sectors, especially in large cities. Total manufacturing and extractive industry employment is likely to continue to rise, but the greatest gains will be in services, both sophisticated and unskilled labor-intensive. In nations with weak public administration or repressive registration practices, with low levels of education, or that restrict the growth of large commercial, financial and transportation enterprises, much if not most of this growth will be in the informal sector. To the extent that informal activities are

poorly measured relative to formal sector production, it may appear that little economic activity is taking place, even if in reality substantial growth is occurring. Indeed, an important, unaddressed empirical issue involves reassessment of urban production and incomes in sub-Saharan Africa, where the informal sector is most prevalent.

Notes

1. Unless otherwise noted, data are taken from United Nations (2002) or the World Bank (2006).
2. NRC (2003, p. 127) reports a mean total fertility rate (TFR) of 5.55 for rural areas across 56 Demographic and Health Surveys (DHS) in developing countries, as opposed to 4.16 percent in urban areas. Crude birth rate differentials should be smaller, since cities have younger adult populations, but the difference should not be enough to overturn the much higher TFR, and will also work to increase the urban crude death rate advantage.
3. While definitions vary, and must be modified to suit data, the formal sector is generally taken to comprise registered, tax-compliant activities – large-scale manufacturing and mining, highly-skilled services and the public sector. The informal sector includes unregistered, labor-intensive activities in such areas as commerce, personal services, private transportation and small-scale manufacturing.
4. Shifting definitions make analysis of urbanization trends in China a complicated subject (Becker and Morrison, 1999); one gets an appreciation of the problem in light of the emergence of peri-urban regions from Lin (2001). An excellent discussion of data problems and a reconciliation of different series appears in Chan and Hu (2003).
5. For evidence of the association of crime incidence and city size (at least for very large cities) in the Latin American context, see Heinemann and Verner (2006). This positive association may reflect greater anonymity, greater opportunity, and lower 'social capital' of large cities. However, this association is not universal: Fafchamps and Moser (2003) find that the incidence of crime, controlling for population composition and risk factors, declines with population density in Madagascar.

References

- Adams, Richard H., Jr. (2006), 'International Remittances and the Household', *Journal of African Economies*, 15 (Supplement 2): 396–425.
- Agesa, Richard U. and Sunwoong Kim (2001), 'Rural to Urban Migration as a Household Decision: Evidence from Kenya', *Review of Development Economics*, 5 (1): 60–75.
- Amin, Mohammad and Aaditya Mattoo (2007), 'Migration from Zambia: Ensuring Temporarity through Cooperation', Washington, DC: World Bank Policy Research Working Paper 4145.
- Beauchemin, Cris and Philippe Bocquier (2004), 'Migration and Urbanization in Francophone West Africa: An Overview of Recent Empirical Evidence', *Urban Studies*, 41 (11): 2245–72.
- Beauchemin, Cris and Bruno Schoumaker (2005), 'Migration to Cities in Burkina Faso: Does the Level of Development in Sending Areas Matter?' *World Development*, 33 (7): 1129–52.
- Becker, Charles and Christopher Grewe (1996), 'Cohort-Specific Rural–Urban Migration in Africa', *Journal of African Economies*, 5 (2): 228–70.
- Becker, Charles M. and Andrew R. Morrison (1988), 'The Determinants of Urban Population Growth in Sub-Saharan Africa', *Economic Development and Cultural Change*, 36 (2): 259–78.
- Becker, Charles M. and Andrew R. Morrison (1993), 'Observational Equivalence in the Modeling of African Labor Markets and Rural–Urban Migration', *World Development*, 21 (4): 535–54.

- Becker, Charles M. and Andrew R. Morrison (1999), 'Urbanization in Transforming Economies', in Paul Cheshire and Edwin S. Mills (eds), *Handbook of Regional and Urban Economics, Vol. 3: Applied Urban Economics*, Amsterdam: North-Holland, pp. 1673–1790.
- Becker, Charles M., Erbolat Musabek, Ai-Gul Seitonova and Dina Urzhumova (2005), 'The Migration Response to Economic Shock: Lessons from Kazakhstan', *Journal of Comparative Economics*, 33 (1): 107–32.
- Bliss, Christopher and Nicholas Stern (1978), 'Productivity, Wages and Nutrition, Part I: The Theory', *Journal of Development Economics*, 5: 331–62.
- Blomqvist, Åke, G. (1978), 'Urban Job Creation and Unemployment in LDCs: Todaro vs. Harris and Todaro', *Journal of Development Economics*, 5 (1): 3–18.
- Brueckner, Jan K. and Hyun-A Kim (2001), 'Land Markets in the Harris–Todaro Model: A New Factor Equilibrating Rural–Urban Migration', *Journal of Regional Science*, 41 (3): 507–20.
- Brueckner, Jan K. and Yves Zenou (1999), 'Harris–Todaro Models with a Land Market', *Regional Science and Urban Economics*, 29 (3): 317–39.
- Chan, Kam Wing and Ying Hu (2003), 'Urbanization in China in the 1990s: New Definition, Different Series, and Revised Trends', *China Review*, 3 (2): 49–71.
- Chen, N., P. Valente and H. Zlotnick (1998), 'What Do We Know about Recent Trends in Urbanization?' in Richard E. Bilsborrow (ed.), *Migration, Urbanization, and Development: New Directions and Issues*, New York: United Nations Population Fund, pp. 59–88.
- Démurger, Sylvie, Martin Fournier and Li Shi (2006), 'Urban Income Inequality in China Revisited (1988–2002)', GATE: unpublished manuscript, Ecully, France: Université Lyon 2.
- Dowall, David and Michael Leaf (1991), 'The Price of Land for Housing in Jakarta', *Urban Studies*, 28 (5): 707–22.
- The Economist* (2007), 'The problem with Made in China', 13 January, 382 (8511): 68–70.
- Fafchamps, Marcel and Christine Moser (2003), 'Crime, isolation and law enforcement', *Journal of African Economics*, 12 (4): 625–71.
- Guha-Khasnobis, Basudeb and Ravi Kanbar (eds) (2006), *Informal Labour Markets and Development*, Houndmills: Palgrave Macmillan.
- Hansen, Bent (1979), 'Colonial Economic Development with an Unlimited Supply of Land: a Ricardian Case', *Economic Development and Cultural Change*, 27 (4): 611–27.
- Harris, John R. and Michael P. Todaro (1970), 'Migration, Unemployment and Development: A Two-Sector Analysis', *American Economic Review*, 60 (1): 126–42.
- Heinemann, Alessandra and Dorte Verner (2006), 'Crime and Violence in Development: A Literature Review of Latin America and the Caribbean', Washington, DC: World Bank Policy Research Working Paper No. 4041.
- Henderson, J. Vernon (1988), *Urban Development: Theory, Fact, and Illusion*, New York: Oxford University Press.
- Jamal, Vijay and John Weeks (1988), 'The Vanishing Rural–Urban Gap in Sub-Saharan Africa', *International Labour Review*, 127 (3): 271–92.
- Kanbur, Ravi and Hillel Rapoport (2005), 'Migration Selectivity and the Evolution of Spatial Inequality', *Journal of Economic Geography*, 5 (1): 43–57.
- Knight, John and Richard Sabot (1990), *Education, Productivity and Inequality: The East African Natural Experiment*, New York: Oxford University Press.
- Lanjouw, Jean and Philip Levy (2002), 'Untitled: A Study of Formal and Informal Property Rights in Urban Ecuador', *Economic Journal*, 112 (482): 986–1019.
- Lewis, W. Arthur (1954), 'Economic Development with Unlimited Supplies of Labour', *Manchester School*, 22 (2): 139–91.
- Lewis, W. Arthur (1955), *The Theory of Economic Growth*, London: George Allen & Unwin.
- Liang, Zai, Yiu Por Chen and Yanmin Gu (2002), 'Rural Industrialization and Internal Migration in China', *Urban Studies*, 39 (12): 2175–87.
- Lin, George C.S. (2001), 'Evolving Spatial Form or Urban–Rural Interaction in the Pearl River Delta, China', *Professional Geographer*, 53 (1): 56–70.
- Lipton, Michael (1976), *Why Poor People Stay Poor: Urban Bias in World Development*, Cambridge, MA: Harvard University Press.

- Lucas, Robert E.B. (2006), 'Migration and Economic Development in Africa', *Journal of African Economies*, 15 (Supplement 2): 337–95.
- McCormick, Barry and Jackline Wahba (2003), 'Return International Migration and Geographical Inequality: The Case of Egypt', *Journal of African Economies*, 12 (4): 500–532.
- Malpezzi, Stephen (1998), 'Welfare Analysis of Rent Control with Side Payments: A Natural Experiment in Cairo, Egypt', *Regional Science and Urban Economics*, 28 (6): 773–95.
- Mazumdar, Dipak (1983), 'The Rural–Urban Wage Gap, Migration, and the Working of Urban Labor Markets: An Interpretation Based on a Study of the Workers of Bombay City', *Indian Economic Review*, 18 (2): 169–98.
- Mberu, Blessing Uchenna (2006), 'Internal Migration and Household Living Conditions in Ethiopia', *Demographic Research*, 14 (21): 509–40.
- Mills, Edwin S. (2000), 'The Importance of Large Urban Areas – and Governments' Roles in Fostering Them', in Shahid Yusuf, Weiping Wu and Simon Evenett (eds), *Local Dynamics in an Era of Globalization*, New York: Oxford University Press, pp. 69–73.
- Mora, Jorge and J. Edward Taylor (2006), 'Determinants of Migration, Destination, and Sector Choice: Disentangling Individual, Household, and Community Effects', in Çağlar Özden and Maurice Schiff (eds), *International Migration, Remittances and the Brain Drain*, Houndmills: Palgrave Macmillan, pp. 21–51.
- National Research Council, Panel on Urban Population Dynamics (Mark R. Montgomery, Richard Stren, Barney Cohen and Holly Reeds (eds) for the Committee on Population, Division of Behavioral and Social Sciences and Education) (2003), *Cities Transformed: the Dynamics of Demographic Change in Developing Countries*, Washington, DC: National Academies Press.
- Özden, Çağlar and Maurice Schiff (eds) (2006), *International Migration, Remittances and the Brain Drain*, Houndmills: Palgrave Macmillan.
- Page, John and Sonia Plaza (2006), 'Migration, Remittances and Economic Development', *Journal of African Economies*, 15 (Supplement 2): 245–336.
- Quartey, Peter and Theresa Blankson (2004), 'Do Migrant Remittances Minimize the Impact of Macro-Volatility on the Poor in Ghana?' Legon, Ghana, final report submitted to the Global Development Network.
- Sabot, Richard H. (1979), *Economic Development and Urban Migration: Tanzania, 1900–1971*, Oxford: Clarendon Press.
- Sahn, David E. and David C. Stifel (2003), 'Urban–Rural Inequality and Living Standards in Africa', *Journal of African Economies*, 12 (4): 564–97.
- Sen, Amartya K. (1966), 'Peasants and Dualism With or Without Surplus Labor', *Journal of Political Economy*, 74 (5): 425–50.
- Soludo, Charles C. and Jongil Kim (2003), 'Sources of aggregate growth in developing regions: still more questions and answers?' in Gary McMahon and Lyn Squire (eds), *Explaining Growth*, Houndmills: Palgrave Macmillan, pp. 32–76.
- Stark, Oded (1991), *The Migration of Labor*, Cambridge, MA and Oxford: Basil Blackwell.
- Stiglitz, Joseph E. (1969), 'Rural–Urban Migration, Surplus Labour, and the Relationship between Urban and Rural Wages', *Eastern Africa Economic Review*, 1 (1): 1–27.
- Stiglitz, Joseph E. (1974), 'Alternative Theories of Wage Determination and Unemployment in LDCs: The Labor Turnover Model', *Quarterly Journal of Economics*, 88 (2): 194–227.
- Stiglitz, Joseph E. (1976), 'The Efficiency Wage Hypothesis, Surplus Labor, and the Distribution of Income in LDCs', *Oxford Economic Papers*, 28 (7): 185–207.
- Taylor, J. Edward (1987), 'Undocumented Mexico–US Migration and Returns to Households in Rural Mexico', *American Journal of Agricultural Economics*, 69 (3): 626–38.
- Taylor, J. Edward, Scott Rozelle and Alan de Brauw (2003), 'Migration and Incomes in Source Communities: A New Economics of Migration Perspective from China', *Economic Development and Cultural Change*, 52 (1): 75–101.
- Todaro, Michael P. (1969), 'A model of Labor, Migration, and Urban Unemployment in Less Developed Countries', *American Economic Review*, 59 (1): 138–48.
- Ukueva, Nurgul (2007), 'Migration, Remittances, and Growth', unpublished manuscript, Durham, NC: Duke University Department of Economics.
- United Nations (2002), *World Urbanization Prospects: the 2001 Revision*, New York: UN.

- Vargas-Silva, Carlos and Peng Huang (2006), 'Macroeconomic Determinants of Workers' Remittances: Host versus Home Country's Economic Conditions', *Journal of International Trade & Economic Development*, 15 (1): 81–99.
- Venables, Anthony J. (2000), 'Cities and trade: external trade and internal geography in developing economies', in Shahid Yusuf, Weiping Wu and Simon Evenett (eds), *Local Dynamics in an Era of Globalization*, New York: Oxford University Press, pp. 58–64.
- World Bank (2006), *06 World Development Indicators*, Washington, DC: World Bank.

Index

- Acemoglu, D. 68, 72
- Africa
- labor markets 333
- Africa, sub-Saharan 239
- growth of food production 460–62
- agglomeration economies 426–7, 526
- aggregate demand
- and economic growth 196–9, 201–2
 - relevance in less-developed countries 196
 - and underdevelopment 117–19
- Aghion, P. 214
- agricultural surplus 100
- agriculture
- education and technological change
 - in 348
 - effect on growth 240–41, 497
 - and the environment 408–9
 - in India 20–21
 - interaction with industry 239, 241–50
 - technological change in 453–68
- Ahlburg, D. 316–27
- AIDS 77, 318
- Alderman, J. 356–70
- America, Central 73–6
- America, Latin *see* Latin America
- Amiti, M. 510
- Anand, S. 9
- Annales School 27
- apprenticeship 378
- Arestis, P. 290–300
- Asia, growth of food production in 461–2
- Asia, East 21–2, 177–8
- development path of 474–8
 - labor markets in 335–6
- Asia, South
- labor markets in 335–6
- Atkinson, G. 406
- Auty, R. 388–403
- Ayres, C. 164–5, 171
- Bacon, F. 19
- Bagchi, A. K. 16–31
- balance-of-payments, data 91
- balanced growth 240, 243, 247–8
- banks 290–98, 372
- Baran, P. 27, 127–8, 151–3, 191
- Bardhan, P. K. 68, 172
- Barrett, C. 408–9
- basic needs 8
- Basu, S. 290–300
- Bauer, P. T. 182
- Baumol, W. J. 505, 512, 513
- Becker, C. 516–31
- Becker, G. 358–9
- Behrman, J. 356–70
- Berry, A. 328–40
- Bhagwati, J. 499
- big push 114, 120, 165–6, 470
- birth rate 456, 464
- Botswana 76–8
- brain drain 524
- Bukharin, N. 239, 240, 246
- bureaucracy 76, 182–3
- business groups 385
- business process outsourcing (BPO) *see* outsourcing
- capabilities 3
- capacity utilization 218–20
- capital accumulation 104–6
- see also* investment
- capital flows *see* international capital flows
- capitalism 144–5, 147, 153
- Cardoso, F. H. 126–7
- Carnegie, A. 380
- Cassen, R. 316–27
- caste 381
- centre-periphery model 136–7, 141, 252–3
- Chang, H.-J. 171
- Chenery, H. 183
- China 16–22
- labor markets in 336
- civil conflict 465
- class struggle 23

- classical development theory 111–22
- classical economics 97–109, 207–8
- classical growth theory 190–91
- clusters
 - and growth 426–7
- Coase, R., 182, 371
- Colombia 73
- common property, and the environment 411–12
- Commons, J. R. 164, 171
- comparative advantage 178, 244
 - models of 51, 105–6, 473
- computable general equilibrium (CGE)
 - models 86–8, 93, 172
- contract enforcement 216
- convergence 122
 - clubs 42–3
 - conditional 34
 - time series approaches to 36–7
 - unconditional 34
 - β - 34–5
 - σ - 35–6
- corruption 201
- Costa Rica 73–4
- cost-benefit analysis 351, 363–4, 462–3
- ‘creative destruction’ 59–62, 224
- credit 290–300
- credit markets 376–7
- culture 164, 184, 216
- cumulative causation 167

- Daly, H. 406
- Dasgupta, P. 11, 405
- data problems 88–92, 501–2
- De Soto, H. 377, 489, 492
- debt, corporate 296–7
- deindustrialization 20, 55–6
- ‘demographic bonus’ 319–20
- dependency theory 108, 125–34, 152–4, 167
- devaluation
 - contractionary effects of 259–60
- development
 - historical perspectives on 16–31
 - interdisciplinary approaches to 68–81
 - meaning of 3–11
 - measurement of 11–14
- developmental state 394
- division of labor 102–4, 145
 - international 137, 140, 142
 - social 100
- disease 68
- disguised unemployment *see* unemployment, disguised
- distribution, income 6, 12
 - and economic growth 198, 203–4
 - effect of education on 343–5
 - and health 361, 363
- Dobb, M. 25
- dual economy 241
- dual economy model 50, 191, 518 *see also* Lewis, W.A.
- Durlauf, S. N. 32–47
- Dutch disease 260, 262, 389–91
- Dutt, A. 189–206, 504–6, 512
- Dutta, D. 497–515

- East Asia *see* Asia, East
- ECLAC 126–7, 130–31, 136–42
- econometric models 84–5
- economies of scale
 - in services 507
 - see also* increasing returns
- education 8–9, 210–11, 341–55
 - effect on growth 343–4
 - effect of health on 367
 - and entrepreneurship 377–8
 - primary 344–5
 - quality of 349
 - return to 341–3
 - secondary 344–5
 - university 343
 - and urbanization 520
- education policy 350–51
- effective demand *see* aggregate demand
- efficiency
 - dynamic 328
- El Salvador 73–4
- elasticity of factor substitution 418, 423
- employment
 - in services 501–3
- endogenous growth theory
 - see* growth models, endogenous; new growth theory
- Engel’s law 54, 57
- Engels, F. 144–6
- England 18–19

- entrepreneurs 202–3, 371–3
 entrepreneurship 371–87
 gender and 381–2
 policy for promoting 385–6
 socio-economic factors and 379–82
 environment 404–16
 and agriculture 409–8
 and common property 411–12
 and the Green Revolution 467
 and growth 404–8
 and population growth 321–4
 and poverty 407–8
 and services 500
 urban 409–10
 valuation of 410–11
 environmental Kuznets curve 406–7
 ethnic fractionalization 393–4
 ethnicity 216
 and entrepreneurship 380
 Europe, Central and Eastern 165–6
 Europe, Western 16–20
 Evans, P. 68–81
 Evenson, R. E. 453–69
 evolutionary economics 220–21
 exchange rate 258–61
 policy 229–31
 regime 230, 263
 export base theory *see* staple theory
 export promotion 421 *see also* outward
 orientation
 and total factor productivity growth
 426
 exports 374–5
 external disequilibrium 138
 externalities 210, 359, 361
 from education 346–9
 and the environment 413
 from health and nutrition 368
 information 474
 pecuniary 112
 technological 112, 114
 extreme bounds analysis 40

 factor accumulation 116–17
 contribution to growth 421
 factor market
 imperfections 419, 435–50
 family, and entrepreneurship 379
 farm size
 and agricultural yield 443

 Fay, M. 301–15
 Feldstein-Horioka puzzle 282
 fertility 316–17, 319, 345
 fertilizers 463, 467
 finance 290–300
 financial institutions 290–91 *see also*
 banks
 financial liberalization 21, 232, 290–91,
 295
 financial markets 295
 financial repression 290
 fiscal policy 199, 225–6
 foreign direct investment (FDI) 374–6,
 512
 foreign direct investment
 in services 503–4
 foreign exchange constraint 255–7 *see*
 also gap models
 free trade 23
 freedom 4, 13–14, 21–2
 freedom index 13
 functionings 3

 Galton's fallacy 36
 game theory 179
 gap models. 87, 255–7
 Gemmell, N. 505
 gender 8
 and entrepreneurship 381–2
 General Agreement of Trade in
 Services (GATS) 508–9, 511
 General Agreement on Trade and
 Tariffs (GATT) 508
 geography 68–9, 216
 Gerschenkron, A. 26, 384
 Ghosh, P. 435–52
 Gibson, B. 83–94
 global climate change 412
 global production networks 374–6, 385
 globalization 52, 282–4, 428, 500, 527
 Goodwin, R. 24
 governance. 377
 government, and services 512
 government failure 471–2
 government intervention, in financial
 sector 291–8
 Grameen Bank 382
 Green Revolution 408–9, 453–69
 Green Revolution Modern Varieties
 (GRMV) 453–60

- Gresham's law 98
 Griliches, Z. 562
 Gross Domestic Product (GDP) 64–5
 data on 90–91
 per capita 11
 growth 5–6, 189–206, 207–21
 accounting 38–9, 421–4
 collapses 390–91, 396–400
 contribution of education to 343–4
 contribution of services to 501–2, 504–6
 in East Asia 177
 effect of health and nutrition on 364–5
 empirics of 32–43
 and the environment 404–8
 of food production 460–62
 fundamental and proximate determinants of 39–41, 204
 stylized facts of 32–7
 growth models
 endogenous 34, 115–16, 120–21 *see also* new growth theory
 environment in 405
 neoclassical 34, 112, 115–16, 121, 191–4 *see also* Solow, R.
 Von Neumann model of 50
 see also Keynesian growth theory, neoclassical growth theory
 Guatemala 73–4, 367
 Gunder Frank, A. 108, 126, 128–9, 141, 152
 Gutierrez, M. 269–89

 Hagen, E. 379
 Hamilton, A. 26
 Hamilton, B. 378
 happiness 4
 Harris-Todaro model 519, 525
 Harrod, R. 219
 Harrod-Domar model 191, 212
 Heal, G. 405
 health 8, 201, 345–6, 356–70
 health policy 361–4, 369
 Hilferding, R. 148–9
 Hirschman, A. O. 162–3, 166–8, 170, 173
 Hirschmeier, J. 378
 HIV 77, 318
 Hoddinott, J. 356–70

 Hodgson, G. 171–2
 Howitt, P. 214
 human capital 210–12, 341–55, 358–9, 497
 human development 5–11, 14
 Human Development Index 9, 11–14, 65
 Hume, D. 100

 ILO *see* International Labour Organization
 imperialism 20, 25, 27–8, 125, 132–3, 146, 148–51, 153
 import substitution 182
 as export promotion 254
 industrialization 131, 140, 257, 417–18, 420, 471–2, 477
 'impossible trinity' 227, 231, 261–2
 income, national 6
 increasing returns 112, 115–18, 195, 201, 238, 241, 245, 500
 India 16–21, 376–7, 379, 381
 India
 financial sector in 292–9
 industrial policy 470–78
 rationale for 470–71
 Industrial Revolution 17–19
 industrialization 53, 117, 138–42, 257
 in resource poor economies 394–6
 rural 522
 industry *see* industrialization
 interaction with agriculture 241–50
 inequality
 and entrepreneurship 381
 and the Green Revolution 467
 and infrastructure 305–6
 international 37
 see also distribution
 infant industry protection 18, 107
 inflation
 effects on growth 225
 targeting 227–8, 263
 informal sector 89, 330–31, 334, 483–94, 497, 501, 519, 524, 527
 alternative conceptualizations of 484–91
 policies towards 491–4
 and state regulations 487–90
 information and communications technology (ICT) 375, 507–8, 513
 information externalities 474

- information technology services (ITS)
509, 513
- information technology *see*
Information and communications
technology
- infrastructure 301–13
barriers to progress 308–12
and income inequality 305–6
and poverty 305
private investment in 310–12
progress in delivery of 306–8
regulation 303, 312–13
and total factor productivity 306
- Ingram, G. K. 301–15
- innovation 213–14, 372
- institutional approach 68–80
- institutional change 80
- institutional economics 162–75
- institutional economics, new 169–70
- institutions 68–80, 215–16, 377
- intermediate products 213
- intermediate technologies 485
- international capital flows 254, 261,
282–4, 500
controls on 231, 235, 262–4
liberalization of 231, 262
'sudden stops' of 234
- International Labour Organization
(ILO) 484, 490
- international migration 523
- international production networks 427
- international trade 54, 374–5
liberalization of *see* trade
liberalization
in services 503, 508–10
theory of 252 *see also under* specific
theories
- internet 507–8
- investment 219–20, 270–72
determinants of 276–80
and growth 280–82
- Jameson, K. P. 162–75
- Japan 26, 374, 378–9, 384–5
- Johnson, S. 72
- Kaldor, N. 24, 219
- Kalecki, M. 24, 25, 116–17, 196, 219,
270–71, 275
- Kautsky, K. 149
- Kenya 367
- Keynes, J. M. 118–19, 196, 276
- Keynesian growth theory 196–9,
218–20
- Knight, F. 371
- knowledge 210, 213–15
- knowledge
codified 500, 512
tacit 500, 512
- Korea, South 384–5
- Korea, South, credit and financial
sector in 291–3, 295–9
- Kotwal, A. 435–52
- Kremer, M. 367–8
- Krugman, P. 245
- Kurz, H. 207–22
- Kuznets, S. 48, 50–51, 58–9, 64–5, 183
- Kourtellos, A. 32–47
- Lal, D. 182–3
- labor
productive and unproductive 103
- labor market 328–38
imperfections 330–31, 333, 337
relations 335
- land reform 181, 443, 445, 450
- late industrialization 26, 384, 385
- Latin America
development path of 130–31, 142,
474–8
growth of food production in 461–2
labor market in 333–5
- learning by doing 244, 246, 248
- Lee, D. R. 409
- Lee, K.-Y. 504–6, 512
- Leibenstein, H. 371–2
- leisure 6
- Lenin, V. 130, 147–9, 151, 153
- Lewis model 191
- Lewis, W.A. 3, 111, 113, 116, 118, 191,
241, 329–30, 518
- life expectancy 11, 357
- limited liability 440–41, 447
- linkages
backward 137, 391–2
diffuse 393
final demand 391, 393
fiscal 391–2
forward 137, 391–2
point source 393

- List, F. 107, 147
 literacy 11, 377
 low-level equilibrium trap *see* poverty trap
 Lucas critique 84–5
 Lucas, R. E. B. 195, 210–11
 Luxemburg, R. 149
- Mahoney, J. 74–6
 malnutrition, child 357, 364
 Malthus, T. R. 105, 316
 Maluccio, J. 367
 market imperfections 195–6, 201, 359, 361–2
 markets 176–8, 182–3
 Marx, K. 23, 107–8, 144–56, 196, 204, 217
 Marxian development economics 191
 Marxian growth theory 190–91
 Marxism 131, 136, 142, 144–56
 Matsuyama, K. 239, 241, 245–6, 249–50
 Mattoo, A. 513
 McClelland, D. 379–80
 McPeak, J. 404–16
 mercantilism 98–9
 mergers and acquisitions 504
 Mexico 366
 micro-enterprises 485, 492–3
 micro-lending 382, 448
 Middle East and North Africa, growth of food production in 461–2
 migration
 international *see* brain drain
 rural–urban *see* rural–urban migration
 Miguel, E. 367–8
 military 74–5
 Mill, J. S. 106
 Millennium Development Goals 365
 Ministry of International Trade and Industry (MITI) in Japan 384
 Mitchell, W. C. 165
 modernization theory 178
 modes of production 107–8
 Modigliani, F. 272
 monetary policy 199, 226–9
 Montesquieu, C. L. 100
 Montiel, P. 223–37
 moral hazard 232, 436–7, 445, 447
 mortality 22, 317–19
 child 456, 464
 and HIV/AIDS 318
 infant 345
 multinational firms *see* transnational corporations
 multiple equilibria 35–7, 41–3, 200–201
 Mun, T. 99
 Myint, H. 182
 Myrdal, G. 167, 171
- Nafziger, E. W. 371–87
 Narodniks 130, 147
 nationalism 146–7
 nationalization, of banks 291–2, 294
 natural disasters 412–13
 natural monopolies 302
 natural resources 69, 72, 76–7, 388–401, 404
 in growth models 388–90
 need-for-achievement 379–80
 Nelson, R. 221
 neoclassical development economics 176–86
 neoclassical growth theory 23–4, 191–6, 209–14
 Netherlands, The 18
 new growth theory 24, 194–5, 202, 203, 210–22, 341
 Nigeria 379
 North, D., 69–70, 169, 171–2
 North–South interaction 24
 Nugent, J. 73
 Nurkse, R. 111, 113–14, 118
 nutrition 201, 356–70
- Ocampo, J. A. 252–65
 optimization 179
 Ostrom, E. 413
 Oulton, N. 506
 outsourcing 509–11, 513
 outward orientation 472
- Pack, H. 417–32
 Palma, J. G. 125–43
 parameter heterogeneity 41–3
 Patrinos, H. 341–55
 Pearce, D. 406
 Petty, W. 63, 99–100, 212
 Pezzy, J. 405

- Philippines, The 178–9
 Physiocracy 99–102
 planning 180–82
 policy, for entrepreneurial development 385–6
 population 22, 216–17, 316–24
 and environment 321–4
 growth of 190, 200, 453
 macroeconomic role of 318–20
 and poverty 316, 320–21
 poverty 366
 and agricultural production 450
 and the environment 407–8
 and the informal sector 484
 and infrastructure 305
 and population growth 316, 320–21
 poverty trap 114, 119, 199–201, 241
 PREALC 486–7
 Prebisch, R. 136, 252–3, 486
 Prebisch-Singer thesis 23, 182
 Preobrazhensky, E. 239, 240, 243
 privatization 365
 of banks 297
 production function
 econometric estimation of 424
 productivity growth 244–50, 505–6 *see also* technological change
 property rights 69–72, 201, 215–16, 377
 Protestantism, and entrepreneurship 380
 ‘proto-industrialization’ 18
 Psacharopoulos, G. 341–55
 public goods 305, 302
 public sector
 and agriculture 458–9
 and services 512
 see also government, state
 quality of life 5
 Quesnay, F. 100–102 *see also* physiocracy
 randomized experiments 360, 368–9
 Rao, R.S. 376–7
 Rashid, S. 176–86
 Rattso, J. 238–51
 religion, and entrepreneurship 380
 remittances, migrant 521, 523
 rent contract, fixed 438
 rent seeking 362, 376–7, 474
 research and development 210, 213
 in agriculture 457–8
 resource allocation
 patterns of 56–7
 and productivity growth 58–9
 resource curse 57, 388 *see also* Dutch disease
 Ricardo, D. 22–3, 105–6
 risk, political 71
 Robinson, J. 72–3, 197
 Rodrik, D. 68
 Romer, P. 195, 210, 212–14
 Ros, J. 111–24
 Rosenstein-Rodan, P. 23, 111–12, 114, 116, 165–6, 180
 Rostow, W. 25–6, 181
 rural industrialization 522
 rural–urban migration 516–28
 Sachs, J. 68–9
 Salvadori, N. 207–22
 savings 270–72
 determinants of 272–6
 Say’s law
 in underdeveloped countries 118
 Schultz, T. P. 341
 Schumpeter, J. 202, 214, 220, 371–3, 386
 Schumpeterian growth model 214, 244
 sectoral interaction 238–51
 Self employment, and entrepreneurship 378
 Sen, A. K. 3, 9, 168
 services 497–515
 services
 definition of 498–9
 types of 499
 Shapiro, H. 470–82
 sharecropping 438
 simulation models 85–8
 Smith, A. 22, 102–5, 162, 176–7, 207, 214, 341, 505
 social accounting matrix (SAM) 87, 91–2, 172
 social capital 396–7
 social factors, entrepreneurship and 379–82
 socialism
 transition to 125–6, 128, 150
 Solimano, A. 269–89

- Solow growth model 191–4, 209–10, 217–18
- Solow, R. M. 271, 389, 405
- South Asia *see* Asia, South
- Soviet Union 238–40
- specialization, pattern of
 and coordination failures 120
 and external disequilibrium 138
 and growth 473
- stage theory of economic growth 25–6, 181
- staples theory 391–4
- state 75–6, 78–80, 165–6
 as entrepreneur 384–5
 role in agriculture 457–8, 65
- Steindl, J. 219
- Stiglitz, J. 405, 519
- stock markets 290
- Streeten, P. 3–15
- structural adjustment programs 21
- structural change 48–65, 505–6
 stylized facts of 52–6
- structural heterogeneity 137–42
- structuralism 126, 136–42, 167, 252–3, 255
- Subsistence sector 113, 115, 118
- Sunkel, O. 167
- surplus labor 58, 112–13, 118, 191, 196, 249, 329–30, 396, 444
- sustainability, environmental 405–6
- sustainable development 404–5
- Sutcliffe, B. 144–61
- Swidler, N. 77–8
- Syrquin, M. 48–67, 183
- Tan, C. M. 32–47
- Taylor, L. 196, 204
- Taylor rule 228–9
- technical choice 418–19
- technological change 116–17, 139, 199, 202–3, 210–15, 244–9, 258, 348, 373–6, 421–9
 see also innovation, learning by doing, total factor productivity growth
 in agriculture 453–68
 in ICT 507–8, 510
 and services 499–500, 505
- technology 164
- technology transfer 203, 374–6, 385
- terms of trade, agriculture–industry 239, 243
- terms of trade, international 23
 deterioration of 137–42
 shocks 233–4
- Toman, M. 405
- Tornqvist index 422–4
- total factor productivity 38–9
 growth 55, 58, 61, 280–81, 373, 417–18, 421–8
 growth, in agriculture 455, 463, 465
 and infrastructure 306
- trade, international *see* international trade
- trade liberalization 249, 477–8, 509
- tragedy of the commons 449
- transactions costs 445
- transnational corporations 375, 476, 504
- Trotsky, L. 147–8
- unbalanced growth 240
- underdevelopment 111, 127–9
- unemployment 137–8, 202, 329–30, 336
 data on 89–90
 unemployment, disguised 166, 191
 uneven development 23
- unorganized sector *see* informal sector
- urban bias 527
- urbanization 516–28
 and the environment 409–10
- United States of America 380–81
- van Wijnbergen, S. 245
- Veblen, T. 164–5
- Verdoon's law 512
- Viner, J. 181–2
- volatility
 and long run growth 224–5
- wage differentials
 by education 332
 formal–informal sector 331, 334, 336
 by gender 331
 public–private 331
 rural–urban 333
- wage flexibility 332
- war 27–8 *see also* civil conflict

- Warren, B. 153–4
- Washington Consensus 169, 176
- water, access to safe 365
- Weber, M. 182, 380
- Wei, S.-J., 514
- Weitzman, M. 214–15
- well-being 4
- Williamson, O. 169, 171
- Winter, S. 221
- women, education of 345–6
- world-system theory 152
- World Trade Organization (WTO)
508–9
- Wunsch-Vincent, S. 513
- Yamamoto, K. 378–9

INTERNATIONAL HANDBOOK
OF DEVELOPMENT ECONOMICS
VOLUME TWO

International Handbook of Development Economics

Volume Two

Edited by

Amitava Krishna Dutt

University of Notre Dame, USA

and

Jaime Ros

University of Notre Dame, USA

Edward Elgar

Cheltenham, UK • Northampton, MA, USA

© Amitava Krishna Dutt and Jaime Ros 2008

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical or photocopying, recording, or otherwise without the prior permission of the publisher.

Published by
Edward Elgar Publishing Limited
The Lypiatts
15 Lansdown Road
Cheltenham
Glos GL50 2JA
UK

Edward Elgar Publishing, Inc.
William Pratt House
9 Dewey Court
Northampton
Massachusetts 01060
USA

A catalogue record for this book
is available from the British Library

Library of Congress Control Number: 2008927966

ISBN 978 1 84542 327 8 (2 volume set)

Printed and bound in Great Britain by MPG Books Ltd, Bodmin, Cornwall

Contents

PART VI INTERNATIONAL ISSUES

36	International trade and development	3
	<i>Francisco Rodríguez</i>	
37	Terms of trade and economic development	16
	<i>David Sapsford</i>	
38	Trade policy and development	30
	<i>Henry J. Bruton</i>	
39	Foreign direct investment	48
	<i>V.N. Balasubramanyam</i>	
40	Private capital flows and development	59
	<i>Stephany Griffith-Jones</i>	
41	International capital flows to emerging economies: short- and long-run effects	71
	<i>Graciela L. Kaminsky</i>	
42	Foreign aid	98
	<i>Steven Radelet</i>	
43	International migration and the brain drain	119
	<i>Francisco L. Rivera-Batiz</i>	
44	International technology transfer: the role of foreign direct investment	137
	<i>Amy Jocelyn Glass and Kamal Saggi</i>	
45	International institutions and development	150
	<i>Kunibert Raffer</i>	
46	North–South issues	163
	<i>Amitava Krishna Dutt</i>	

PART VII DISTRIBUTION AND POVERTY

47	Measures of the distribution of income and their interpretation in developing countries	179
	<i>Samuel A. Morley</i>	
48	Income distribution: effects on growth and development	193
	<i>Nancy Birdsall</i>	
49	Measuring progress in reducing poverty	213
	<i>Lyn Squire</i>	

50	Gender issues in development <i>Diane Elson</i>	227
51	Children and development <i>Paul Glewwe and Amy Damon</i>	250
52	Ethnicity and economic development <i>William A. Darity Jr and Russell E. Triplett</i>	262

PART VIII THE STATE, INSTITUTIONS AND DEVELOPMENT

53	The role of the state and markets in development <i>Louis Putterman</i>	281
54	Monetary policy <i>Paul R. Masson</i>	293
55	Fiscal policy <i>Raghbendra Jha</i>	305
56	Stabilization policy and structural adjustment <i>Valpy FitzGerald</i>	319
57	Economic planning in developing economies <i>Bill Gibson</i>	341
58	State-owned enterprises and privatization <i>Anindya Sen</i>	356
59	Corruption and development <i>Cheryl W. Gray</i>	369
60	Law and development <i>Pranab Bardhan</i>	381
61	Institutions, property rights and development <i>Jean-Philippe Platteau and Jean-Marie Baland</i>	394
62	Culture and development <i>Roland Hoksbergen and Charles K. Wilber</i>	407
63	The economics of war: causes and consequences <i>Frances Stewart and Graham Brown</i>	423

PART IX ECONOMIC DEVELOPMENT EXPERIENCE

64	Economic development in the Third World: an international perspective <i>Richard Kozul-Wright and Paul Rayment</i>	445
65	Latin America and the Caribbean <i>Albert Fishlow</i>	469
66	Sub-Saharan Africa <i>Benno J. Ndulu and Stephen A. O'Connell</i>	490

67	The Middle East and North Africa <i>Omar S. Dahi and Firat Demir</i>	522
68	China <i>Yasheng Huang</i>	536
69	South Asia <i>Prabhat Patnaik and Jayati Ghosh</i>	554
70	The East Asian newly industrializing countries <i>Jomo K.S.</i>	569
71	Transition economies: lessons for development <i>Ruud Knaack and Henk Jager</i>	579
	<i>Index</i>	601

PART VI

INTERNATIONAL ISSUES

36 International trade and development

Francisco Rodríguez

In July of 2007 more than 1000 economists – including four Nobel prize winners – signed a letter to Congress petitioning it not to impose tariffs on China. The phrasing of the petition is revealing of the extent to which many economists believe in the beneficial effects of free trade:

As economists, we understand the vital and beneficial role that free trade plays in the world economy. Conversely, we believe that barriers to free trade destroy wealth and benefit no one in the long run . . . There is no foundation in economics that supports punitive tariffs.¹

This view of trade is widely shared by the overwhelming majority of economists with a neoclassical training. In a recent survey of graduate students at the six top economics departments in the USA, only 7 percent of respondents disagreed with the statement ‘tariffs and quotas reduce general economic welfare’ (Colander, 2005, p. 189) Of all the policy statements presented in the survey, trade policy was the one that commanded the broadest consensus among students. As Alan Blinder recently commented, in the economics profession ‘anyone who says anything even obliquely that sounds hostile to free trade is treated as an apostate’.²

This consensus is not due to the absence of theoretical results showing the existence of situations in which greater trade can decrease welfare. Students of international trade are commonly exposed to examples of optimal tariffs, infant industries and strategic trade policy, all of which can serve to justify intervention in international trade. Rather, the consensus in the profession emerges from the vision that these examples are not relevant in most real-world circumstances and that the dangers from attempting to carry out activist trade policy far outweigh its potential benefits. As one of the architects of strategic trade theory has himself written:

The gains from intervention are limited by uncertainty about appropriate policies, by entry that dissipates the gains, and by the general equilibrium effects that insure that promoting one sector diverts resources from others . . . It is possible, then, both to believe that comparative advantage is an incomplete model of trade and to believe that free trade is nevertheless the right policy. In fact, this is the position taken by most of the new trade theorists themselves. (Krugman, 1987, p. 143)

Deciding whether to accept this conventional wisdom or to question it is a key step in the framing of a country's development strategy. The purpose of this chapter is to analyze critically the basis for the pro-trade consensus, and to lay out clearly the arguments of supporters and detractors of activist trade policy. Such an exercise requires a dual discussion of the theoretical literature and the empirical evidence, which I present in the next two sections. As I will argue, a careful examination of these two literatures provides grounds for the raising of serious questions about the desirability of outward-oriented trade strategies for development.

What the theory says

In its simplest version, the gains-from-trade theorem (Kemp, 1962; Samuelson, 1962) states that in the absence of distortions and when lump-sum transfers are feasible, all individuals in an economy can be made better off from an appropriate combination of full liberalization and compensating transfers. This result follows from the basic Ricardian insight that it will be more efficient for countries to shift production from goods for which their opportunity cost of production is high to goods for which it is low.

The static intuition can be readily extended to an intertemporal model of growth, as shown by Smith (1979). In such models, the efficiency effects of trade will show up in higher steady-state levels of consumption and welfare. This extension can generate considerable confusion as openness will not be associated with higher long-run growth rates. This is more than anything a consequence of the fact that the long-run growth rate in a Ramsey-style model of economic growth is determined by the rate of technological progress and unaffected by any other variables. However, openness does generate higher steady-state levels of income and higher growth rates on the transition to these new steady states.

The intuition readily carries forward to endogenous growth models. In an AK model of growth, static efficiency losses generated by distortions turn up as declines in the level of efficiency captured by the productivity term in the production function. Trade policy thus has an unambiguously negative effect on the growth rate. In more complex endogenous growth models that seek to endogenize productivity as a consequence of decisions to innovate, the public goods nature of knowledge introduces a distortion. The existence of this distortion opens up the possibility of second-best effects in which trade policy can potentially enhance welfare, a possibility that we discuss in more detail below. A full discussion of alternative models of this type is given by Grossman and Helpman (1991).

In all of these renderings, the gains-from-trade theorem is little more than an extension of the first and second fundamental theorems of welfare economics. The possibility of trading shifts the consumption possibilities

frontier outwards, and the competitive equilibrium allows us to attain any point on that consumption possibilities frontier. The possibility of lump-sum transfers allows us to redistribute the gains from trade in such a way that all individuals can be made better off.

The theorem is thus open to the same caveats as the first and second welfare theorems. In the first place, it is vulnerable to the existence of distortions. One of the earliest recognized distortions in the trade literature is the possibility of the home country being able to affect international prices. In this case the world economy becomes a monopolist in the world market and perfectly competitive levels of production are no longer optimal. In the second place, the result requires the existence of non-distortionary taxes and subsidies that allow the redistribution of the gains from trade. In the absence of these instruments, some individuals may – and generally will – be made worse off by greater openness.

Obviously, in the real world distortions do exist and lump-sum taxes do not. Most of the academic literature's defense of the gains-from-trade principle as a useful guide to policy action comes from the interpretation of results that appear to indicate that it will generally be suboptimal to deal with distortions through trade policy, and that reasonable approximations to lump-sum transfers exist.

Let us first discuss the issue of distortions. In a set of classic articles, Bhagwati (1971) and Bhagwati and Srinivasan (1969) showed that in the presence of distortions, trade tariffs or taxes will generally not be the optimal way to address the distortions or objectives that are commonly taken as rationales for activist trade policy. The exception is the case in which the distortion is actually generated by the existence of monopoly power in trade, in which the first-best policy is a tariff. But in the case in which there is an externality that leads to underproduction of a given good, the optimal policy is to subsidize the production of that good. A tariff is suboptimal because it is actually a combination of a production subsidy and a consumption tax, and there is generally no reason to expect that one would simultaneously want to increase production and reduce consumption of any particular good. These seminal contributions are thus generally taken as the demonstration that trade policy is a misguided way to address most of the problems generated by the existence of distortions.

The Bhagwati–Srinivasan contributions reflect a vision of policy-making in which the first-best equilibrium can be attained and thus serves as a useful policy benchmark. The ideal policy will be the one that accurately identifies all existing distortions and introduces an optimal intervention to eliminate the effects of each distortion. The existence of a government with the capabilities to carry out this complex exercise is presumed. One way of understanding this presumption is as a reflection of the view that the

majority of distortions are policy-induced. In this case, the policy prescription is clear: a laissez-faire elimination of all government-induced distortions could lead the economy sufficiently close to its first-best optimum to take advantage of the full gains from trade.³

While such a vision of the world is certainly coherent, it is also quite reasonable to hold to an alternative vision of the world as completely ridden with distortions. In this world, interacting complex processes such as innovation, knowledge networks, geographical clusters, public goods and international market power combine to create an economy in which distortions are a fact of everyday life. The task of identifying all such distortions and crafting interventions to address each one of them is beyond the realm of rationally effective policy-making. Even if one could identify all existing distortions, the design of interventions to eliminate some of them may be out of the sphere of possible policy actions by the government because of institutional or political constraints.

In such a world, policy-making should not try to replicate the first-best equilibrium. Rather, it makes sense to think of policy reforms as taking place in a setting of radical Knightian uncertainty, where the expected effects of removing a policy distortion can only be deduced from local experimentation. Hausmann et al. (2004) have recently proposed such an approach to policy reforms. Rather than attempting to eliminate all distortions at once, they suggest that reformers should concentrate on the reforms that have the greatest expected pay-off, given that other distortions are in place. They suggest an empirical method to infer whether certain distortions are in effect binding constraints on growth, and thus to identify whether altering them will lead to greater growth. The resulting method of policy-making is akin to the use of non-linear programming algorithms to search for local maxima, which do so through the search for incremental improvements rather than by the explicit calculation of a global solution.

One of the consequences of taking seriously such a world is that it turns a common free trade argument on its head. Advocates of trade often argue that even though some level of intervention in trade may be optimal theoretically to address trade-induced externalities, identifying such interventions in a context of considerable uncertainty would be beyond the capacities of most governments (Krugman, 1987). However, if we view the real world as ridden by trade and non-trade induced distortions, the same dose of realism that leads us to conclude that the government cannot address all of these distortions also leads us to recognize that we are unavoidably in a second-best world, in which the incremental effects of trade policy on welfare could well be positive.

Let us now turn to the issue of lump-sum transfers. Non-distortionary taxes require conditioning on characteristics that individuals cannot

change. It is doubtful whether such taxes actually exist or not in real life, and even if they do it would be extremely difficult to design them to redistribute the gains from trade – as the observable characteristics that are out of the control of individuals are unlikely to correlate perfectly or even reasonably well with the benefit or loss from trade that these individuals experience.

The existence of compensation to losers is particularly relevant precisely because trade theory predicts stark effects on income distribution from trade openness. In the absence of compensation, the Stolper–Samuelson theorem predicts that the real return to a country's scarce factor will decrease with greater openness. In this case developing-country unskilled workers would actually benefit from greater openness, so that trade may be distributively beneficial for poor countries with an abundance of unskilled labor. However, the factor endowments model of trade appears to have little empirical support (see Feenstra, 2004, Chapter 2), so that this may not be the most appropriate theory with which to think about the distributive effects of trade. Alternative theoretical frameworks can produce different predictions concerning income distribution. For example, a set of recent theoretical and empirical contributions (Rodrik, 1997; Reddy and Dube, 2000; Ortega and Rodríguez, 2006) have argued that trade can diminish the bargaining power of unions and thus lead to a decline in labor shares.

In the case of compensation, the literature's optimism comes from the belief that while lump-sum taxes do not exist, reasonable approximations can be constructed in practice to carry out the necessary compensations to losers. Indeed, it has been shown that lump-sum taxes are not even theoretically necessary in this respect: factor and income taxes will suffice to enact the desired redistribution (Dixit and Norman, 1980, pp. 79–80). This result appears intuitive even if the real-world setting is much more complex than that of our models: it is generally possible to identify – at least *ex post* – the key groups that gain and lose from trade openness and to design transfer programs to redistribute the gains.

The key question regarding compensation is not whether it is feasible in a technical and operational sense to design and implement such compensation, as it almost certainly is, but whether this compensation is likely to take place in practice. There are a number of political economy reasons why one may expect that such compensation is unlikely to take place. One is that while it may be optimal for the gainers to promise to compensate the losers before the reforms are carried out, such promises are likely to be time-inconsistent, particularly if their gains are protected by some degree of irreversibility in trade reforms. While the manipulation of activist trade policies by interest groups has received considerable attention in the literature (for example, Grossman and Helpman, 1994) and has often been adduced as a reason for

the superiority of simple rules such as free trade, the political economy of compensation arrangements has received much less attention. An early exception can be found in the work of political scientists such as Ronald Rogowski (1987), who argued that increased trade alters the political landscape, making owners of abundant resources much more powerful and assertive, and thus much less likely to accept demands for compensation.⁴

In sum, whether one considers the free trade case a reasonable one or not on theoretical grounds depends on whether one considers two basic assumptions of the gains from trade theorem reasonable. The first one is that the extent of distortions is sufficiently restricted so that the government can identify them and deal with them through policies designed on the basis of the theory of optimal intervention. The second one is that redistributive policies can and will be implemented to compensate the losers from trade liberalization, particularly when these are the most disadvantaged groups in society.

The decision of whether these two assumptions are reasonable or not on purely theoretical grounds is far from clear-cut. What should be clear is that a critical vision, which is based on skepticism about the appropriateness of the first-best model as a guide for policy and the likelihood of implementation of appropriate redistributive strategies, cannot be deemed insensible on a purely theoretical basis. The belief in the optimality of free trade should thus be based on the belief that the empirical evidence decisively points in favor of a beneficial effect of trade on growth. We turn now to that issue.

What the data say

Broadly speaking, the empirical literature that has studied the effect of openness on growth has taken one of two vantage points. The first one is to analyze the correlation between openness and growth in data sets that cover a large section of developing and developed countries, in the tradition of cross-country growth empirics initiated by Robert Barro (1991). The second one is to concentrate on country or region-level analytical case studies of economic growth. Both literatures have been appealed to by proponents and detractors of trade-oriented development strategies. In what follows, I will attempt to shed some light on the reasons behind these differences in interpretation.

The empirical literature on openness and growth is voluminous indeed. I will not attempt to provide a full survey of the main contributions (the interested reader may consult Rodríguez, 2008). Broadly speaking, however, a number of findings appear to emerge from this literature.

First, there is no strong unconditional or conditional correlation between economic growth and a number of direct measures of trade policy,

such as weighted or unweighted tariffs, import quotas or other non-tariff barriers. This point was first made by Rodríguez and Rodrik (2001) and generated some surprise in the literature. It has since been confirmed by, among others, DeJong and Ripoll (2006), who argue that there may be a non-linear relationship where the effect of tariffs on growth depends on the initial level of a country's income and may be positive or negative. In Rodríguez (2007), I have also shown evidence in favor of a non-linear effect, although I have argued that the precise form of that effect may be difficult to discern.

Second, there appears to be a reasonably strong correlation between growth or productivity and the ratio of trade to gross domestic product (GDP), especially when the latter is measured in prices of a constant base year (Dollar and Kraay, 2002; Alcalá and Ciccone, 2003). Some attempts have been made to discern whether this correlation actually embodies a causal relationship. The most well-known attempt, formulated by Frankel and Romer (1999), consists in using instrumental variables estimates of the effect of trade volumes on growth where the latter is instrumented with its geographic determinants as derived from the estimation of gravity equations. These results are controversial – as Rodríguez and Rodrik (2001) and Irwin and Tervio (2000) have shown, they are not robust to controlling for the direct effect of geographical variables on income or productivity. Other attempts to discern causality using alternative methods to instrumental variables do not confirm the existence of causal effect (Rigobón and Rodrik, 2005).

A drawback of using the trade to GDP ratio as an indicator of openness is that it may capture many non-policy-induced changes in trade openness which are largely irrelevant if one is preoccupied with designing a developing country's trade strategy. Natural resource booms, the emergence of new export sectors, changes in other countries' trade policies, and changes in foreign aid can all have an effect on the trade to GDP ratio without necessarily having an obvious link to trade policy. In sum, the key problem of the trade–GDP ratio is that it is an indicator of results and not of policy actions. To take just one example, if the infant-industry argument for protection were correct, initial levels of trade protection would lead to the development of productive, competitive domestic industries that would later on be capable of competing internationally. Tariffs would be associated with higher growth, but so would exports. A correlation between trade volumes and growth may thus not be very informative about the desirability of activist trade policies.

Some authors have tried to produce compound measures of trade policy that capture the different ways in which an economy can be closed to international trade. According to these authors, one would not expect to observe

a simple correlation between simple measures of trade policy such as tariffs and economic growth because countries can use many policy devices to impose trade protection, of which only one is import tariffs. The most famous of these measures was provided by Sachs and Warner (1995) and recently updated by Wacziarg and Welch (2003). What these indices actually measure is very controversial. Rodríguez and Rodrik (2001) argue that the Sachs and Warner variable's effect on growth was purely driven by two subcomponents of the index – black market premia and export marketing boards – which are not obviously linked to trade policy. For example, they argue that the effect of export marketing boards on growth in the Sachs–Warner study comes from the fact that the variable was taken from a 1994 World Bank study called *Adjustment in Africa* that covered only 29 African economies undergoing adjustment programs during the 1980s, leading to the exclusion of non-African or African non-adjusting economies from the sample and strongly biasing the results in favor of a trade–growth correlation. Rodríguez (2008) levies similar criticisms at the Wacziarg and Welch (2003) exercise.

In recent years, there has been growing skepticism of the possibility of establishing strong conclusions regarding causal growth effects using the cross-country regression framework. A growing consensus appears to have emerged around the belief that the problems of causality, robustness and specification are simply too pervasive and difficult to solve in the context of highly aggregated cross-national empirical data. This skepticism has led authors such as Bhagwati and Srinivasan (2001) to discount the aggregate growth evidence altogether, and to call for concentrating exclusively on the evidence from case studies. While these criticisms should be taken seriously, it is important to note that even if one takes the cross-country evidence at face value, accepting the framework without questioning, it does not appear to lend the strongest of supports to the pro-trade view. As in the case of the theoretical literature, it appears to be open to multiple interpretations, some of which are consistent with the view that protection is not unequivocally harmful for growth.

Country-level studies of openness and growth are also open to multiple interpretations. Bhagwati and Srinivasan (2001) cite the Organisation for Economic Co-operation and Development (OECD) and National Bureau of Economic Research (NBER) studies of more than a dozen major developing countries carried out in the 1960s and 1970s, which uncovered key differences between the constraints on economic performance in countries that pursued import substitution strategies and those that pursued export promotion. A revised interpretation of this view was given by the World Bank's 1993 study *The East Asian Miracle*. Broadly speaking, the key argument of this study was that the openness to trade and reliance on market

forces of East Asian economies played a fundamental role in making possible their sustained growth acceleration.

The World Bank's characterization of the high-growing East Asian tigers as economies that followed a strategy of free trade has, however, been strongly questioned by several authors. Some of these criticisms were collected in a 1994 volume published by the Overseas Development Council (Fishlow et al., 1994) in which Dani Rodrik, Robert Wade and Stephen Haggard disputed the key findings of the World Bank study. In Robert Wade's words, 'the [World Bank's] report uses standards of inference so elastic that practically anything could be confirmed' (2003, p. xix)

One of the key points of dispute concerns whether East Asia can adequately be characterized as a region that followed a non-activist trade policy. The World Bank study had concluded that East Asia's relative prices were closer to international averages than those of other regions, supporting the contention that its international trade was relatively undistorted. Wade pointed out that this is only true when one uses an unweighted average that includes the island economies of Hong Kong and Singapore, where price distortions were necessarily negligible. In contrast, during the 1976–85 period, relative prices in Japan, South Korea and Taiwan deviated more from international prices than those of countries which are generally perceived to have had strong records of intervention, such as India, Pakistan, Brazil, Mexico and Venezuela in the period 1976–85. Similarly, Alice Amsden's (1992) in-depth study of South Korea's industrialization contends that the success of its industrial policies was largely due to an active intervention in the determination of relative prices, a strategy that she labels 'getting relative prices wrong'.

During the 1990s, the set of liberalization experiences that could be the subject of in-depth studies expanded dramatically. Between 1990 and 2002, the average tariff rate in the world went down from 10.5 percent to 6 percent, and the ratio of imports plus exports to GDP rose from 75.2 percent to 86.8 percent (World Bank, 2005a). In 1990, the General Agreement on Tariffs and Trade (GATT) had been signed by 96 countries: between 1990 and 2005, 65 countries joined it either as the GATT or in its most recent incarnation as the WTO.

While the result of these liberalization experiences has not yet been fully analyzed, what is clear is that aggressive trade liberalization proved to be very far from a necessary condition for a growth take-off. Some of the most aggressive liberalizers of this period were former communist economies such as Mongolia, Ukraine and Moldova, which suffered some of the deepest growth collapses in post-World War II history. But openness did not only fail to pay off in the former Soviet Bloc. With the exception of Cuba, the evidence suggests that virtually all Latin American

economies moved in a direction of greater trade liberalization during the 1990s. Yet the region's growth performance during the post-reform period has been disappointing to say the least, with per-worker GDP and total factor productivity growing respectively at annual rates of only 0.1 percent and 0.2 percent between 1990 and 2002 (Ocampo, 2004). The region is said to have entered an era of 'reform fatigue' (Lora et al., 2004) in which voters are increasingly willing to vote for political platforms to roll back reforms.

In sum, neither cross-national empirical studies nor country-level case studies seem to give strong support to the idea that openness is unequivocally good for growth. A reading of the evidence in support of activist trade strategies is certainly possible and indeed has been carried out by reputable mainstream economists. These conclusions mirror our interpretation of the theoretical literature, which can also be interpreted as supporting a case for intervention in trade policy.

Concluding comments

One way to explain the apparent divorce between the favorable view that the majority of economists have about free trade and the lessons given by the empirical and theoretical literature is by thinking about free trade as one of the components of our discipline's 'hard core' (in the sense of Lakatos, 1976), a set of beliefs and methodological assumptions that are not considered the appropriate subject of empirical tests. Since these core beliefs are never tested without auxiliary assumptions, any failure to explain the evidence can be handled by altering the assumptions but not the core belief. As a senior faculty member once quipped after seeing a presentation of my work, 'if the data does not say that trade is good for growth, then the data must be wrong'.

It is not easy for a discipline to abandon or even begin to question a hard-core belief, but neither is it impossible. To take one example, during the 1990s the assumption of rationality has made the transition from a hard-core belief to an auxiliary hypothesis that is not even taken very seriously most of the time. This change has opened up a burgeoning new area of research in behavioral economics which has transformed our understanding of individual economic behavior.

Signs that this may be starting to happen in the study of the relationship between trade and development are beginning to appear. In 2005, the World Bank published a comprehensive assessment of the experience of the 1990s with economic reforms (World Bank, 2005b). The sobering assessment of this disappointing period recognizes that the results of economic reforms were far below what its proponents had expected and rejects the one-size-fits-all approach to reform that the institution espoused during

the greater part of the period in question. On the concrete matter of trade policy, the report concludes that: '[w]hile trade reforms can help accelerate integration in the world economy and strengthen an effective growth strategy, they cannot ensure its success', and 'the distributive effects of trade liberalization are diverse, and not always pro-poor' (pp. 131–2). On the fairness of the world trading system, it states that 'global markets are the most hostile to the products produced by the world's poor'. As Dani Rodrik wrote in his review of this volume: 'occasionally, the reader has to remind himself that the book he is holding in his hands is not some radical manifesto, but a report prepared by the seat of orthodoxy in the universe of development policy' (Rodrik, 2006, pp. 974–5).

A reconsideration of the role of openness in countries' development strategies would fundamentally alter the nature of the debate on generating and sustaining growth. Whether this occurs will probably depend not only on the internal dynamics of academia, but also on the extent to which outside reality exerts pressure for such a change. Political discontent with the experience of the 1990s is undoubtedly a key reason for the World Bank's reappraisal of the reform experience. In the same way, the results of the current reassertion of state involvement in much of the developing world are likely to influence deeply the direction that development research will take in the future. Perhaps, to turn Keynes on his head, economists are nothing more than the slaves of long-defunct practical men.

Notes

1. At the moment of writing (August 2007), the process of signature collection is not yet finished and thus the letter has not been published. The text of the letter was obtained through personal communication with Andy Roth of the Club for Growth.
2. See Cohen (2007).
3. Obviously, even if only a few distortions remain there is no theoretical presumption that the resulting equilibrium will be better than the pre-laissez-faire equilibrium. It can be argued that at this point it becomes feasible to target the remaining distortions through optimal interventions. Alternatively, it can be argued that theoretical models are always approximations of the real world, and that a distortion-free model should be a reasonable approximation to a reality in which there is a reduced number of distortions.
4. For a recent exploration of this issue, see Davidson et al. (2006).

References

- Alcalá, Francisco and Antonio Ciccone (2003), 'Trade, Extent of the Market, and Economic Growth 1960–1996', Barcelona: Universitat Pompeu Fabra.
- Amsden, Alice (1992), *Asia's next Giant: South Korea and Late Industrialization*, Oxford: Oxford University Press.
- Barro, Robert (1991), 'Economic Growth in a Cross-Section of Countries', *Quarterly Journal of Economics*, 106 (2): 407–43
- Bhagwati, Jagdish (1971), 'The Generalized Theory of Distortions and Welfare', in J. Bhagwati, R.W. Jones, R.A. Mundell and J. Vanek (eds), *Trade, Balance of Payments and Growth: Papers in International Economics in Honor of Charles P. Kindleberger*, Amsterdam: North-Holland, pp. 69–90.

- Bhagwati, Jagdish and T.N. Srinivasan (1969), 'Optimal Intervention to Achieve Non-Economic Objectives', *Review of Economic Studies*, 36: 27–38.
- Bhagwati, Jagdish and T.N. Srinivasan (2001), 'Outward-Oriented and Development: Are the Revisionists Right?' in Deepak Lal and Richard H. Snape (eds), *Trade, Development, and Political Economy: Essays in Honour of Anne O. Krueger*, London: Palgrave Macmillan, pp. 3–26.
- Cohen, Patricia (2007), 'In Economics Departments, a Growing Will to Debate Fundamental Assumptions', *New York Times*, 11 July.
- Colander, David (2005), 'The Making of an Economist Redux', *Journal of Economic Perspectives* 19 (1): 175–98.
- Davidson, Carl, Steven J. Matusz and Douglas R. Nelson (2006), 'Can Compensation Save Free Trade', *Journal of International Economics*, 71 (1): 167–86.
- DeJong, D.N. and M. Ripoll (2006), 'Tariffs and Growth: An Empirical Exploration of Contingent Relationships', *Review of Economics and Statistics*, 88 (4): 625–40.
- Dixit, A. and V. Norman (1980), *Theory of International Trade*, Cambridge: Cambridge University Press.
- Dollar, David, and Aart Kraay (2002), 'Trade, Growth, and Poverty', *Economic Journal*, 114 (493): F22–F49.
- Feenstra, Richard (2004), *Advanced International Trade: Theory and Evidence*, Princeton, NJ: Princeton University Press.
- Fishlow, Albert, Catherine Gwin, Stephan Haggard, Dani Rodrik and Robert Wade (1994), *Miracle or Design: Lessons from the East Asian Experience*, Washington, DC: Overseas Development Council.
- Frankel, Jeffrey and David Romer (1999), 'Does Trade Cause Growth?' *American Economic Review*, 89 (3): 379–99.
- Grossman, Gene M. and Elhanan Helpman (1991), *Innovation and Growth in the Global Economy*, Cambridge, MA: MIT Press.
- Grossman, Gene M. and Elhanan Helpman (1994), 'Protection for Sale', *American Economic Review*, 84 (4): 833–50.
- Hausmann, Ricardo, Dani Rodrik and Andrés Velasco (2004), 'Growth Diagnostics', Cambridge, MA: Harvard University.
- Irwin, Douglas and Marko Tervio (2000), 'Does Trade Raise Income: Evidence from the Twentieth Century', NBER Working Paper No 7745.
- Kemp, M. (1962), 'The Gain from International Trade', *Review of Economic Studies*, 72: 803–19.
- Krugman, Paul (1987), 'Is Free Trade Passé?' *Journal of Economic Perspectives* 1 (2): 131–44.
- Lakatos, Imre (1976), 'Proofs and Refutations', in John Worrall and Elie Zahar (eds), Cambridge and New York: Cambridge University Press.
- Lora, Eduardo, Ugo Panizza and Myriam Quispe-Agnoli (2004), 'Reform Fatigue: Symptoms, Reasons, and Implications', Atlanta, GA: Federal Reserve Bank of Atlanta, *Economic Review*, 89 (2): 1–28.
- Ocampo, José Antonio (2004), 'Latin America's Growth and Equity Frustrations During Structural Reforms', *Journal of Economic Perspectives*, 18 (2): 67–88.
- Ortega, Daniel and Francisco Rodríguez (2006), 'Trade Policy and Factor Prices: An Empirical Strategy', Middletown, NJ: Wesleyan University.
- Reddy, Sanjay and Arindrajit Dube (2000), 'Liberalization, Income Distribution and Political Economy: The Bargaining Channel and Its Implications', New York: Columbia University.
- Rigobón, Roberto and Dani Rodrik (2005), 'Rule of Law, Democracy, Openness and Income: Estimating the Interrelationships', *Economics of Transition*, 13 (3): 533–64.
- Rodríguez, Francisco (2007) 'Cleaning Up the Kitchen Sink: On the Consequences of the Linearity Assumption for Cross-Country Growth Empirics', Wesleyan Economics Working Paper No. 2006-004.
- Rodríguez, Francisco (2008), 'Openness and Growth: What Have We Learned?' in J.A. Ocampo, Jomo K.S. and Rob Vos (eds), *Growth Divergences: Explaining Differences in Economic Performance*, New York: Zed Books.
- Rodríguez, Francisco and Dani Rodrik (2001), 'Trade Policy and Economic Growth: A Skeptic's Guide to the Cross-National Evidence', in Ben S. Bernanke and Kenneth Rogoff

- (eds), *NBER Macroeconomics Annual 2000*, Cambridge, MA: National Bureau of Economic Research.
- Rodrik, Dani (1997), *Has Globalization Gone Too Far?*, Washington, DC: Institute for International Economics.
- Rodrik, Dani (2006), 'Goodbye Washington Consensus, Hello Washington Confusion? A Review of the World Bank's *Economic Growth in the 1990s: Learning from a Decade of Reform*', *Journal of Economic Literature*, 44 (4): 973–87.
- Rogowski, Ronald (1987), 'Political Cleavages and Changing Exposure to Trade', *American Political Science Review*, 81: 1121–37.
- Sachs, Jeffrey D. and Andrew Warner (1995), 'Economic Reform and the Process of Global Integration', *Brookings Papers on Economic Activity*, 1995 (1): 1–118.
- Samuelson, P.A. (1962), 'The Gains from International Trade Once Again', *Economic Journal*, 37: 47–61.
- Smith, M.A.M (1979), 'Intertemporal Gains from Trade', *Journal of International Economics*, 9: 239–48.
- Wacziarg, Romain and Karen H. Welch (2003), 'Trade Liberalization and Growth: New Evidence', NBER Working Paper 10152, National Bureau of Economic Research, Cambridge, MA.
- Wade, Robert (2003), *Governing the Market: Economic Theory and the Role of Government in East Asian Industrialization*, 2nd edn, Princeton, NJ: Princeton University Press.
- World Bank (1993), *The East Asian Miracle: Economic Growth and Public Policy*, Washington, DC: World Bank.
- World Bank (1994), *Adjustment in Africa: Reforms, Results, and the Road Ahead*, Washington, DC: World Bank.
- World Bank (2005a), *Economic Growth in the 1990s: Learning from a Decade of Reform*, Washington, DC: World Bank.
- World Bank (2005b), *World Development Indicators*, Washington, DC: World Bank.

37 Terms of trade and economic development

David Sapsford¹

Introduction

Why do countries trade with one another? What determines the terms on which trade between countries is conducted in the world marketplace? These two questions are perhaps the most fundamental to be considered in any discussion of international economic relations, be it trade between developed and developing countries or trade amongst countries in either the developing or the developed world. These questions are of especial importance in the context of economic development since if there are ‘gains from trade’ to be earned, the distribution of such gains between trading partners (especially when we are thinking of trade between developed or industrialized countries on the one hand, and developing nations on the other) carries important implications not only for living standards and economic welfare within the participating countries, but also for the continued willingness of developing and newly emerging economies to engage in processes, such as those initiated under the auspices of the World Trade Organization, designed to bring about further reductions in barriers which impede the process of international trade.

The classical economists, most notably Adam Smith (1723–90) and David Ricardo (1772–1823) (see in particular, Smith, 1776 and Ricardo, 1817), initially addressed the first of these questions and their, respective, analyses of absolute and comparative advantage as the basis for trade are widely recognized. Indeed, to this day the Ricardian model, despite its various (over)simplifying assumptions (including that of a two-country, two-good, one-factor world) still occupies centre stage in the economic theory of international trade. Despite its undoubted logical elegance, a major limitation of the Ricardian analysis is encountered when we move to the second of these two questions, because the model leaves us analytically with a range of indeterminacy, somewhere within which the terms of trade (defined in the usual way as the ratio of the price of a country’s exports to the price of its imports) must lie if trade is to offer benefits (in terms of increased output or consumption with unchanged resource endowments) to at least one of the trading partners. This indeterminacy is, in practical terms, more than a mere theoretical loose end, especially when we move

outside of the static world of Ricardo's model and begin to seek answers to important real-world questions such as who gains most from trade and to what, if any, extent the pattern of winners and losers has changed over time with the evolution that has occurred in both the structure and performance of the world economy. Such questions are of particular importance in the context of the changes that have occurred over the long run in the terms on which trade between developed and developing nations has been conducted, especially since this carries important implications for the pace and nature of economic development in the Third World.

Who gains from trade?

While the elegance of Ricardo's analysis and its correctness within the confines of its own assumptions can not be faulted it does, as noted, beg a question that is of central importance in the context of the trade that takes place between countries of the developed or industrialized world, on the one hand, and those of the developing or Third World on the other. While the analysis demonstrates quite clearly the potential benefits to trading partners from engaging in international trade in the world marketplace, it has nothing whatsoever to say about the division of these potential gains between them. Suppose that the two countries comprising the Ricardian world are, say, the USA and China, then it follows from the model that if relative prices in the world marketplace (the so called net barter terms of trade) were equal to US internal relative prices then China would effectively appropriate all of the gains from trade for herself, whereas at the opposite end of the spectrum, the USA would scoop all of the gains if Chinese relative prices prevailed.

In order to focus ideas let us consider trade between the countries of the developed or industrialized world and those of the developing world and, for simplicity, assume that the former produce manufactured goods while the latter produce primary commodities. The fact that Ricardo's analysis did not shed any light on the issue of how the potential gains from trade are shared out in practice did not seem to constitute a problem in the minds of classical economists since in a related context Ricardo, like Smith before him, had argued that as an inevitable consequence of the twin forces of diminishing returns in the production of primary commodities from a fixed stock of land (including mineral resources) as population increased, and the downward pressures on production costs in manufacturing generated by the moderating influences of surplus population and urbanization upon wages, the price of primary products would rise over the long run in relation to the price of manufactured goods, thereby giving rise to an upward drift in the net barter terms of trade between primary commodities and manufactured goods.² On the above assumptions this movement

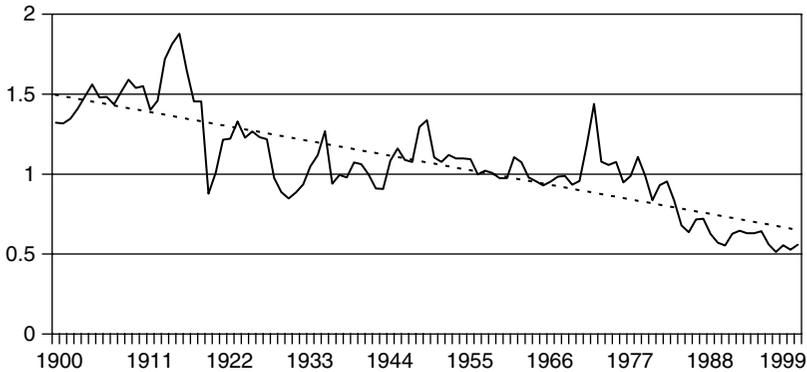
will translate into an improvement in the terms of trade of developing countries *vis-à-vis* the developed countries. On the basis of this argument there was little, if any, reason to be concerned about the plight of developing countries in the context of their trading relations with the industrialized world since it predicted that over the long run the terms of trade would shift, as a consequence of the workings of the invisible hand of the market, steadily in their favour, with the result that they would in due course enjoy an increasing share of the potential gains from trade. Indeed, casual analysis of the views expressed by many politicians and international policy-makers over at least the last century would seem to suggest an underlying acceptance of this particular prediction of classical economics (Sapsford and Singer, 1998).

The evidence

In the early 1950s, the classical prediction of a secular improvement in the terms of trade of primary commodity-dependent developing countries was challenged by both Prebisch (1950 [1962]) and Singer (1950).³ Both argued forcefully that in direct contravention of the then still prevailing classical orthodoxy, the terms of trade faced by primary producers had actually as a matter of statistical fact been historically subject to, and could be expected to continue to be subject to, a declining as distinct from an improving trend. Both analyses therefore implied that contrary to the classical view, developing countries were actually obtaining a falling proportion of the potential gains from their trade with the countries of the developed world (Sapsford et al., 1992). The statistical techniques available for revealing and estimating trends in the middle of the twentieth century were by today's standards extremely rudimentary. However, it is relevant to notice that some 50-plus years later the conclusion of a downward secular movement remains strongly intact, despite a rapid evolution in the sophistication of the time-series methods to which terms-of-trade data have been subjected.⁴ Figure 37.1 provides a time plot covering the period 1900 to 1999 of the log of most commonly analysed terms-of-trade series, together with the estimated least-squares trend line which provides clear evidence of the presence of a steady downward trend over the last century.⁵

From Singer I to Singer II

When the Prebisch–Singer thesis was launched in 1950 the world was, perhaps, a simpler place economically than it is today. In particular, the fact that the industrialized countries of the North specialized heavily in the production and subsequent export of manufactured goods, while those in the South concentrated intensively in the production or extraction and export of primary commodities, meant that the Prebisch–Singer focus upon the



Source: Updated version of the Grilli-Yang index, based on data compiled and kindly made available by Dr Stephan Pfaffenzeller, University of Liverpool.

Figure 37.1 Terms-of-trade index

terms of trade between manufactured and primary goods carried direct implications for the terms of trade of less-developed countries (LDCs) *vis-à-vis* developed countries. As the post-1950 decades unfolded, and many of the LDCs of the South moved increasingly down a path of industrialization, Singer (1975) argued that terms-of-trade issues continued to be of vital importance to LDCs despite their greater reliance on manufactured exports. In this context he distinguished between what he called 'Singer I' and 'Singer II'; with Singer I referring to terms of trade between different commodities (manufactures versus primaries) and Singer II referring to terms of trade between different countries (less-developed economies, the 'South', versus developed economies, the 'North').

As we shall see in the following section there has been a range of alternative explanations put forward in the literature for the observed secular deterioration in the terms of trade. However, we will also see that these explanations relate at least as much to the characteristics and structures of the countries themselves as to the commodities they trade. The distinction between Singer I and Singer II is an important one because it alerts us to the possibility that potentially serious terms-of-trade problems for LDCs can continue to persist despite their increasing tendency to switch towards manufacturing exports. Various pieces of empirical evidence support this concern. For example, econometric evidence reported by Sarkar and Singer (1991) suggested that for a large sample of less-developed countries over the period 1979 to 1987 the terms of trade between the manufactures exported by them and those imported by them from the developed economies was subject to a declining trend. A similar finding has also been

reported by Lutz (1999). Some new evidence relating to the experiences of a sample of the world's very poorest countries is reported later in this chapter. In an overview of the performance of the Prebisch–Singer thesis over its first half-century, Singer (1987) presented some statistical evidence to suggest that the terms-of-trade deterioration experienced by LDCs since the mid-1950s can be attributed, in approximately equal proportions, to three causes: first, the deterioration in the price of primary commodities in relation to manufactures (as emphasized in Singer I); second, a more rapid deterioration in the prices of the primary commodities produced by developing countries than in the prices of those produced by developed countries; and finally, a fall in the price of the manufactures exported by developing countries relative to those exported to them by developed countries (as emphasized by Singer II).

Explaining the downward trend

A number of theoretical explanations have been put forward in the literature to account for the observed downward trend in the terms of trade of developing relative to developed countries, and these can be conveniently summarized under the following four headings:

- Differing price elasticities of demand for primary commodities and manufactured goods (with the inelastic nature of the former resulting in a tendency for increases in the conditions of commodity supply to be felt more strongly in price decreases than in quantity increases).
- Differing rates of growth in the demands for primary commodities and manufactured goods (with the demand for primaries expanding less rapidly than the demand for manufactures due to their lower income elasticity of demand – especially so in the case of agricultural commodities due to the operation of Engel's Law – plus the development of synthetic substitutes and the occurrence in manufacturing of technical progress of the raw material saving sort).
- Technological superiority (the argument being that the price of manufactured goods rise relative to those of primaries because they embody both a so-called Schumpeterian rent element for innovation, plus an element of monopolistic profit arising from the monopoly power of multinational producers).
- Asymmetries in market structure (the argument here is that differences in market structure – with primary commodities typically being produced and sold under competitive conditions, while manufacturing in industrialized countries is often characterized by a high degree of monopoly by organized labour and monopoly producers – mean that while technical progress in the production of primary

commodities results in lower prices, technical progress in manufacturing leads to increased factor incomes as opposed to lower prices).

In his 1950 paper Singer placed particular emphasis on points listed under the first two bullet points, while Prebisch (1950 [1962]) specially emphasized those covered under the third and final bullets. A number of recent authors have developed what have become known as North–South models, which further develop the points grouped under bullets three and four. See, for example, Sarkar (1994) and Dutt (1998).

Some policy implications

Although space constraints do not allow us to discuss in any detail the policy implications of the observed worsening trend in the terms on which trade is conducted in the world marketplace between primary commodity-producing countries and manufacturing countries, it is nevertheless important to note that the Prebisch–Singer hypothesis is sometimes advanced as one argument in favour of development and trade policies of the import-substituting industrialization (IS) as opposed to export promotion (EP) variety.⁶ A number of early enthusiasts of the Prebisch–Singer thesis recommended the adoption of such a policy stance. However, the policy issues here are not clear-cut and the fact, already mentioned, that all four of the above explanations relate as much, if not more, to the characteristics of different types of countries as to the characteristics of different types of traded goods highlights the need to devise and implement policies that address differences and imbalances of the former as well as the latter sort (Singer, 1987). Although a number of recent analysts (for example Krueger, 1997) have drawn an association between below-average growth performance and the adoption of IS as distinct from EP policies, the real-world situation, as Singer frequently pointed out, was rather more complex when one looks beneath the surface, in that countries such as Korea (held up in some circles – including Washington, DC – as a beacon to demonstrate the superiority of outwardly oriented trade policies) appear in reality to have achieved their above-average rate of growth by adopting a subtle dynamically evolving policy mix involving a combination of IS policies in certain sectors of the economy simultaneously accompanied by EP policies in other sectors. As Singer was also quick to point out (for example Sapsford and Singer, 1998), Ricardian comparative advantage is seldom, if indeed ever, exogenously given in the manner assumed in the simple Ricardian model, still so central in trade analysis. Thus while it is logically correct that in the static Ricardian world primary commodity-exporting countries can still potentially gain from trade by specializing in those sectors in which they possess (static) comparative advantage, what is relevant in reality is the

fact that many comparative advantages are actually either consciously engineered or acquired through learning by doing and increasing returns to scale.⁷ Accordingly, it has been argued that the Prebisch–Singer thesis provides one possible case for industrialization based on (a limited period of) infant industry protection. In development economics, as in prospecting, all that glitters is not gold.

Although outside of the scope of the current chapter, it is important to recognize that the terms-of-trade issue carries a range of important implications for issues other than the distribution of the gains from trade. These include questions relating to the level, and pattern, of Southern or LDC growth and the unevenness (or otherwise) of international development patterns. Some brief comments are, however, in order. As far as the implications of terms of trade for Southern growth are concerned, there now seems to be something of a consensus in the so-called ‘New’ (or Endogenous) Growth Theory literature that the terms of trade do matter when it comes to explaining inter-country differences in growth performance. In particular, in a study of the determinants of growth across countries Barro (1997) estimates the (partial) elasticity of real output with respect to the country-specific net barter terms of trade to be 0.137, with a standard error of 0.3 – implying an effect which is significantly different from zero. The long-run trends in the terms of trade, and the forces that drive these, also carry some important implications for our understanding of the observed patterns of international development. The literature here is extensive and is well covered by Dutt (1990, 1998). For the purpose of the current discussion it is relevant to notice that different causes of terms-of-trade deterioration carry with them different implications regarding the evenness of development. In cases where the predominant factor driving terms-of-trade deterioration is the income elasticity of demand, we would expect to observe uneven development. However, in cases where its roots lie in the pace of technological progress in the South then, especially in situations where there are inflows of foreign direct investment into the South accompanied by the possibility of significant spillovers of various sorts to the domestic economy, there is the potential for deteriorating terms of trade to go hand in hand with more even development (Balasubramanyam et al., 1996). In a nutshell: the terms of trade matter, in terms of both the speed of LDC growth and its evenness.

Some recent issues: country-specific evidence and terms-of-trade volatility

Terms-of-trade issues continue to attract the attentions of researchers with interests in a variety of trade-related fields and in this section we consider two of these. The first relates to the strength and periodicity of terms-of-trade volatility, while the second relates, in the spirit of Singer II, to the experiences of particular economies.

Trend versus volatility in the terms of trade

Inspection of the time plot in Figure 37.1 suggests that while the terms of trade declined secularly over the course of the last century they were, at the same time, subject to appreciable variability about that trend. In a statistical analysis of the time-series properties of the Grilli–Yang data (as plotted in Figure 37.1) Sapsford and Balasubramanyam (1999) produced evidence that suggested that the extent of volatility exhibited by the terms of trade about its trend increased over the twentieth century. This increased volatility was particularly marked after 1972 and appeared to have been accompanied by an acceleration in the trend rate of decline of the terms of trade. Taken together these results yielded a depressing picture for the primary commodity-dependent countries of the developing world for they indicated that since 1973 such countries have experienced a marked increase in the trend rate of deterioration in their terms of trade (from about 1.6 per cent per annum to around 4 per cent), accompanied by a marked increase in their volatility estimated to be to the order of 50 per cent. This clearly represents a doubly unpleasant state of affairs.

In a related study, Singer and Lutz (1994) report cross-country panel data estimates which seem to indicate that the magnitude of both the downward trend in the terms of trade and the extent of volatility about this trend exerted significant downward effects on gross national product (GNP) growth in a sample of some 79 non-oil-producing countries. In this study the authors argue that terms-of-trade volatility exerts a detrimental influence on growth performance via a number of channels, including: the increased risks which it imposes on investment (which they see as reducing the incentives to invest and thereby the realized volume of investment); its potential to disrupt development (including structural adjustment) plans, to increase price instability, to destabilize domestic incomes and to distort the structure of domestic prices, including the price of traded relative to non-traded goods.

The arguments developed and empirical results reported in these two studies would seem to imply that the trend and volatility in the terms of trade should not be treated as separable issues in the manner of classical time-series analysis. Instead these two issues are best seen from the analytical standpoint as twin pillars of the same underlying problem faced by LDCs: their heavy dependence on primary commodities (or more recently upon components and parts of manufactured goods) as a source of export revenue. The empirical work reported in these papers focused attention on both the long-run and cross-country evidence regarding the trend and volatility in the terms of trade seen as related dimensions of the same underlying problem. Amongst other things the results reported by Sapsford and Balasubramanyam (1999), as noted above, indicated the

occurrence after 1972 of a major increase in the trend rate of deterioration of the terms of trade and, moreover, that this was accompanied by a marked increase in volatility. This is the true 'double whammy' in which both of the twin pillars of the commodity problem seem to have turned simultaneously against these countries. However, the evidence reported by Singer and Lutz (1994) suggests that the bad news does not end there, in the sense that both the trend rate of terms-of-trade decline and the extent of terms-of-trade volatility would appear to exert a subsequent significant downward influence upon the rate of economic growth; a 'triple whammy'.⁸

Country-specific evidence

A number of studies in the recent literature have moved away from the continued analysis of the aggregate data,⁹ preferring instead to focus attention on country-specific evidence, especially that relating to low-income primary commodity-exporting countries. It is a regrettable fact of current economic life that it is the very poorest economies in the world who are the most dependent on either a single or a small number of primary commodities for the vast majority of their export earnings. In an UNCTAD study, Sapsford (2001) noted that according to World Bank 1996 data, all but two of the world's poorest 20 economies were located in Africa and of these, some 13 were dependent on either a single primary commodity or a small number of commodities for in excess of 90 per cent of their export earnings. Two examples are as follows: Mali (with an estimated annual per capita income of US\$260 and life expectancy at birth of 50 years) earned 99.8 per cent of its export revenue in 1996 from cotton, while Ethiopia (the then poorest economy in the world, with an estimated per capita income of only US\$110 per annum, and life expectancy of only 43 years)¹⁰ depended on coffee for 99.8 per cent of its export earnings. Clearly, for these seriously poor economies both trends and volatilities in the terms on which they undertake their trade in these particular primary commodities with the outside world are of vital importance. One might say that they are quite literally a matter of life and death.

Sapsford (2001) provided a detailed statistical analysis of the terms-of-trade experiences of this sample of the world's poorest economies over the period 1960 to 1998. Using a simplified version of the structural econometric model proposed in Bloch and Sapsford (2000), this study undertook a structural stability analysis of country-specific terms of trade. Notice that unlike cruder previous approaches to trend estimation, this approach controlled for the influence of fluctuations in the level of production in the industrialized world on country-specific terms of trade. Although a detailed discussion of the nature and implication of the results revealed by

this study is beyond the scope of the present short chapter¹¹ it is relevant to notice that the main results can be summarized as follows:

- Of the world's 15 poorest commodity-exporting countries, all but two experienced a significant change in the trend rate of growth of their terms of trade during the period 1960–93.
- In nine of these cases, the change in the trend occurred between 1972 and 1982.
- Nineteen out of 28 reported trend estimates are negative; only three of the reported trend estimates are positive.
- In nine out of the 13 countries where there is a trend shift, the pattern shows a worsening of the situation in respect of terms of trade.
- In six out of the 13 countries where there is a trend shift, the pattern shows an increase in the volatility of the terms of trade.

Taken together these results seem to indicate that over the post-1960 period many of the very poorest commodity-exporting LDCs in the world have indeed been subject to Prebisch–Singer effects on their terms of trade, effects which have exerted a continuous downward pressure on economic and export growth of a magnitude sufficient to more than offset the upward effects which they might have experienced as a result of the positive influence of expanding output in the industrialized countries.

Some concluding remarks

The terms-of-trade is a topic that continues to attract the attentions of researchers in the fields of international trade and development economics: scarcely a week passes without yet another addition to the literature. Within the confines of this short chapter it has only been possible to scratch the surface of the many and varied issues involved. Particular emphasis was placed in the first half of the chapter upon the famous Prebisch–Singer hypothesis, the empirical evidence relating to the hypothesis and its policy implications, especially as these relate to countries situated at the lower end of the per capita income distribution. In the second half of the chapter we focused attention on some recent analyses that have extended the earlier approaches to encompass the effects, from a country-specific perspective, of terms-of-trade volatility upon economic growth performance and the relationship between terms-of-trade trend and volatility.

As we have seen, the Prebisch–Singer hypothesis that there is a long-term decline in the price of primary commodities relative to the price of manufactures has historically been an object of controversy, although given the overwhelming weight of empirical evidence in its favour, it seems to have become widely accepted in the majority of circles since the 1990s.¹² Most

tests of the hypothesis use time-series models to estimate trend growth rates in selected relative prices. The focus of concern has typically been the net barter terms of trade between producers of primary products (equated with developing countries) and producers of manufactures (equated with industrialized countries). A new approach which has been developed recently is to construct a structural model which seeks to identify the various different factors which impinge on the prices of manufactured goods and primary commodities (Bloch and Sapsford, 1997, 2000). Applying this approach, it has been found that the overall trend identified in the time-series models is the net effect of a number of separate divergent influences. On the one hand, there are Prebisch and Singer effects that exert a downward pressure on the commodity terms of trade. These effects arise because of differences in market structure (markets for primary products are more perfectly competitive) and differences in the factor bias of technical change (technical change in manufactures is assumed to save raw material inputs and labour). On the other hand, rising output in the industrialized countries can have an offsetting effect, as primary products used in manufacturing activity experience rising prices when the level of manufacturing activity increases.

The final section of this chapter discussed some country-specific results that were obtained by applying this basic approach. These results seemed to suggest that many of the very poorest economies in the world – which are also, as it happens, heavily (if not totally) dependent on either a single primary commodity or a very small number of them for their export revenue – have been subject over at least the last four decades of the twentieth century not only to the downward pressures of the sort emphasized by both Prebisch and Singer but also to the additional pressures generated by increasing terms-of-trade volatility.

Some two and a half centuries have elapsed since the classical economists first proffered their prediction that the tide would inevitably turn, over the long run, in favour of the LDCs. To offer what is perhaps logically the only admissible defence – namely, that some two and a half centuries is too short a time period for the classical mechanisms to fully work themselves out – seems implausible not only to trained economists but more importantly to those citizens of the world's very poorest economies who seem condemned to remain forever at the very margins of survival.

Notes

1. This chapter is dedicated to the memory of Hans Singer (1910–2006), friend, collaborator and inspiration.
2. For brevity I refer hereafter to the net barter terms of trade between primary commodities and manufactured goods (that is, to the ratio of the price of primary commodities to the price of manufactured goods) as simply their terms of trade.

3. Although conventionally referred to in the literature as the Prebisch–Singer hypothesis recent evidence reported by Toye and Toye (2003) suggests that it should correctly be termed the Singer hypothesis since their detailed archival research indicates that Prebisch's interest in the topic stemmed directly from his receipt of Singer's then unpublished UN manuscript on the subject.
4. Being in contradiction with the then prevailing orthodoxy, these papers attracted criticism on a number of (primarily statistical) grounds. However, detailed scrutiny of the literature seems to indicate that almost six decades after its initial launch the empirical validity of the declining-trend hypothesis first put forward by Prebisch and Singer has become pretty much universally accepted. See Sapsford and Chen (1998) for a review of the second wave of statistical studies that appeared during the late 1980s and the 1990s, and Spraos (1983) and Sapsford (1985) for reviews of the earlier criticisms. This longevity is all the more remarkable when one recognizes the wide array of statistical techniques to which the hypothesis has been subjected. There are few, if any, hypotheses in economics that can claim to have stood the test of time so well as this one. See Spraos (1980, 1983), Sapsford and Singer (1998), Sapsford and Chen (1999) and Bloch and Sapsford (2000) for detailed reviews.
5. The data series employed here were compiled by Grilli and Yang (1988). Dr Stephan Pfaffenzeller of the University of Liverpool compiled the post-1986 data according to the same definitions. This is the terms-of-trade series most commonly analysed in the literature, being compiled as the World Bank's index of the prices of some 24 internationally traded non-fuel primary products deflated by the UN's index of the unit values of manufactured exports from industrialized countries.
6. See Sapsford and Balasubramanyam (1994) for detailed discussion.
7. The early twenty-first-century experience of China may well turn out to offer an important lesson here regarding the distinction between what is logically correct and what is relevant.
8. Although outside of the scope of the present chapter it is important to notice that the existence and indeed persistence of volatility around the trend carries with it a potential policy implication regarding the possible merits of instituting some form of stabilization scheme (for example buffer stocks, compensatory financing mechanisms and the like) designed to at least smooth out, if not remove, fluctuations about the trend. The literature here is extensive and goes back in time beyond Keynes, although the latter's views as expressed at the Bretton Woods Conference are of particular importance. For surveys of the major issues involved see Winters and Sapsford (1990), Maizels (1992) and Sapsford and Morgan (1994).
9. Or, perhaps more correctly, have chosen to move away from the continued reanalysis of the Grilli–Yang (1988) data set.
10. To put these life expectancy figures in perspective it should be noted that the corresponding figures for the USA, the UK and Japan are 76, 77 and 80 years respectively.
11. See UNCTAD (2002) for a summary. Notice that in this study (following Singer and Lutz, 1994, Sapsford and Balasubramanyam, 1999, and others) the standard error of estimate about the estimated regression plane was used as the sub-period measure of terms-of-trade volatility.
12. It is now the case that at least some of the international agencies involved in the world trading system have come to accept that primary commodity producers in developing countries do face real and significant uncertainties and risks regarding the prices that they will actually receive for their products when they come to the world market. At the time of writing in 2006, a task force set up under the auspices of the World Bank is investigating a range of possible 'market-based' approaches (including the formation of futures markets) for dealing with the price risks faced by primary commodity producers in developing countries. As pointed out by Morgan (2001) these approaches appear to represent an attempt to confront price risk by modifying the financial environment within which primary producers in less-developed countries operate. However, it remains to be seen whether such approaches will prove any more, or less, successful than the various policies which have preceded them.

References

- Balasubramanyam, V., M. Salisu and D. Sapsford (1996), 'Foreign Direct Investment and Growth in EP and IS Countries', *Economic Journal*, 106 (434): 92–105.
- Barro, R. (1997), *Determinants of Economic Growth*, Cambridge, MA: MIT Press.
- Bloch, H. and D. Sapsford (1997), 'Some Estimates of Prebisch and Singer Effects on the Terms of Trade Between Primary Producers and Manufacturers', *World Development*, 25 (11): 1873–84.
- Bloch, H. and D. Sapsford (2000), 'Whither the Terms of Trade? An Elaboration of the Prebisch–Singer Hypothesis', *Cambridge Journal of Economics*, 24 (4): 461–81.
- Dutt, A. (1990), *Growth, Distribution and Uneven Development*, London: Cambridge University Press.
- Dutt, A. (1998), 'Direct Foreign Investment and North–South Trade: Uneven Development or Convergent Growth?' in D. Sapsford and J. Chen (eds), *Development Economics and Policy*, London: Macmillan, pp. 261–86.
- Grilli, E. and M. Yang (1988), 'Primary Commodity Prices, Manufactured Goods Prices and the Terms of Trade of Developing Countries: What the Long-Run Shows', *World Bank Economic Review*, 2 (1): 1–47.
- Krueger, A. (1997), 'Trade Policy and Development: How We Learn', *American Economic Review*, 87 (1): 1–22.
- Lutz, M. (1999), 'Commodity Terms of Trade and Individual Countries' Net Barter Terms of Trade: Is There an Empirical Relationship?' *Journal of International Development*, 11 (6): 859–70.
- Maizels, A. (1992), *Commodities in Crisis*, Oxford: Clarendon Press.
- Morgan, W. (2001), 'Commodity Futures Markets in LDCs: A Review and Prospects', *Progress in Development Studies*, 1 (2): 139–50.
- Prebisch, R. (1950), 'The Economic Development of Latin America and its Principal Problem', UN ECLA; also published (1962) in *Economic Bulletin for Latin America*, 7 (1): 1–22.
- Ricardo, D. (1817), *On the Principles of Political Economy and Taxation*, London: Penguin; reprinted (1971).
- Sapsford, D. (1985), 'The Statistical Debate on the Net Barter Terms of Trade Between Primary Commodities and Manufactures: A Comment and Some Additional Evidence', *Economic Journal*, 95 (379): 781–8.
- Sapsford, D. (2001), 'Terms of Trade: Trend, Volatility and the Role of Financial Crises in Late Industrialising Countries', Background paper to *Least Developed Countries Report 2002, Escaping the Poverty Trap*, Geneva: UNCTAD.
- Sapsford, D. and V.N. Balasubramanyam (1994), 'The Long-Run Behaviour of the Relative Price of Primary Commodities: Statistical Evidence and Policy Implications', *World Development*, 22 (11): 1737–45.
- Sapsford, D. and V. Balasubramanyam (1999), 'Trend and Volatility in the Net Barter Terms of Trade, 1900–1992: New Results from the Application of a (Not So) New Method', *Journal of International Development*, 11 (6): 851–7.
- Sapsford, D. and J. Chen (1998), 'The Prebisch–Singer Terms of Trade Hypothesis: Some (Very) New Evidence', in D. Sapsford and J. Chen (eds), *Development Economics and Policy*, London: Macmillan, pp. 25–34.
- Sapsford, D. and J. Chen (1999), 'The Prebisch–Singer Thesis: A Thesis for the New Millennium', *Journal of International Development*, 11 (6): 843–9.
- Sapsford, D. and W. Morgan (eds) (1994), *The Economics of Primary Commodities: Models, Analysis and Policy*, Aldershot, UK and Brookfield, USA: Edward Elgar.
- Sapsford, D., P. Sarkar and H. Singer (1992), 'The Prebisch–Singer Terms of Trade Controversy Revisited', *Journal of International Development*, 4 (3): 315–32.
- Sapsford, D. and H. Singer (1998), 'The IMF, the World Bank and Commodity Prices: A Case of Shifting Sands?' *World Development*, 26 (9): 1737–45.
- Sarkar, P. (1994), 'Terms of Trade of the South vis-à-vis the North: A Macroeconomic Framework', in D. Sapsford and W. Morgan (eds), *The Economics of Primary Commodities: Models, Analysis and Policy*, Aldershot, UK and Brookfield, USA: Edward Elgar, 79–90.

- Sarkar, P. and H. Singer (1991), 'Manufactured Exports of Developing Countries and Their Terms of Trade Since 1965', *World Development*, 19 (4): 333–40.
- Singer, H. (1950), 'The Distribution of Gains Between Investing and Borrowing Countries', *American Economic Review, Papers and Proceedings*, 40: 473–85.
- Singer, H. (1975), 'The Distribution of Gains Revisited', in A. Cairncross and M. Puri (eds), *The Strategy of International Development: Essays in the Economics of Backwardness*, London: Macmillan, pp. 1–38.
- Singer, H. (1987), 'Terms of Trade and Economic Development, in J. Eatwell, M. Milgate and P. Newman (eds), *The New Palgrave: A Dictionary of Economics*, London: Macmillan, pp. 626–8.
- Singer, H. and M. Lutz (1994), 'Trend and Volatility in the Terms of Trade: Consequences for Growth', in D. Sapsford and W. Morgan (eds), *The Economics of Primary Commodities: Models, Analysis and Policy*, Aldershot, UK and Brookfield, USA: Edward Elgar, pp. 91–121.
- Smith, A. (1776), *The Wealth of Nations*, London: Penguin; reprinted (1961).
- Spraos, J. (1980), 'The Statistical Debate on the Net Barter Terms of Trade Between Primary Products and Manufactures', *Economic Journal*, 90 (357): 107–28.
- Spraos, J. (1983), *Inequalising Trade?* Oxford: Oxford University Press.
- Toye, J. and R. Toye (2003), 'The Origins and Interpretation of the Prebisch–Singer Thesis', *History of Political Economy*, 35 (3): 437–67.
- UNCTAD (2002), *Least Developed Countries Report 2002: Escaping the Poverty Trap*, New York and Geneva: United Nations Publications.
- Winters, L.A. and D. Sapsford (1990), *Primary Commodity Prices, Economic Models and Policies*, London: Cambridge University Press.

38 Trade policy and development

Henry J. Bruton

Introduction

The countries of Western Europe, northern North America, and Australia and New Zealand (the North) began to achieve increasing per capita gross domestic product (GDP) in the first part of the nineteenth century. Growth, so measured, while not uninterrupted, became sustained enough that one may say that the routine functioning of these economies produced increasing per capita GDP. Growth became, in effect, built in. As a consequence of 150–200 years of this fairly routine growth, the countries of the North are now very rich, at least in terms of GDP per capita. The Great Question of Development Economics is: Why have not all countries been able to establish economies in which growth is built in to their routine operations? The broad policy question is therefore how to bring about modifications in these non-growing economies (the South) in such a way that growth becomes routine for them too. The theoretical question is to explain why and how one group of countries did grow while, another, much larger group, has not.¹ In the best of all worlds, this explanatory story would be well grounded in fundamental economic principles, supported by convincing empirical evidence, and lead to policies that can be implemented and, when implemented, bring about the achievement of the objective stated above.

The purpose of this chapter is to study the way that foreign trade enters the development quest just defined. The approach is to tell a story of history and learning, of learning through time as evidence and theoretical insight accumulated. Such a historical, learning approach seems appropriate because of the changing views and continuing controversy, and because no widely held model or theory or policy position has become conventional wisdom.

The beginning

In the late 1940s and early 1950s when ‘development economics’ emerged as a distinct field of inquiry within the broad discipline of academic economics, empirical evidence and explanatory hypotheses about the process of development were relatively primitive. Yet it was widely accepted that the huge differences in per capita GDP between the North and South demanded action, demanded that the world ‘do something’.² Trade policy

was, from the beginning, a major element both in terms of explanation and in terms of policy.

Import substitution: argument and practice

The earliest role of trade as an explanation of the failure of the South to grow rested on arguments and assumptions about the way that the 'structure' of production in a country affected its capacity to grow. Structure of production refers mainly to the composition of output of an economy, especially to the distinction between manufacturing on the one hand and agriculture and mining on the other. The economies of the South had, it was widely argued, become locked into agriculture and mining, largely due to foreign trade based (at least to some extent) on assumptions about comparative advantage. While this locked-in state may have resulted in the maximization of current production, the argument continued, it did prevent these economies from growing over time. There were several detailed arguments and assumptions on which these conclusions depended.

The earliest specific argument was that the terms of trade had consistently moved against the South countries. Several reasons were offered as to why this had been and was happening. Productivity growth in manufacturing was thought to be greater than in agriculture and mining, but this higher productivity growth was matched or more than matched by rising wages in the North. The rising wages, combined with the widespread monopoly power of producers in the North, prevented manufacturing prices from falling as productivity grew. While in the South, productivity grew more slowly and an abundant supply of labor prevented wage rates (and prices) from rising. Added to these considerations was the assumption that demand for food (and agricultural products in general) and minerals was inelastic with respect to income in world markets. As output of these commodities increased, their prices relative to those of manufactures tended to fall over time as per capita incomes rose in the North.³ The South could not shift out of agriculture and minerals into manufactures because its technological capacity was such that it could not compete with the producers of the North as long as free trade prevailed.⁴ This disadvantage with respect to technological capacity prevailed despite the fact that the South had access to the same array of equipment and ideas that most of the countries of the North had – mainly imported from Britain.⁵

These arguments disputed any role that conventional trade theory – comparative advantage as formulated by Ricardo and by Heckscher–Ohlin – could play in a development context. Comparative advantage, it was often noted, is a static notion that assumes technical knowledge is the same in all countries, constant returns to scale everywhere, and that equilibrium states always prevail, that is, no second-best issues arise. It thus had nothing to

communicate with respect to formulating a policy aimed at producing marked changes in the structure of an economy. Added to these considerations were references to colonialism and to the manner and extent to which it had drained resources from the poor to the rich countries.

These explanations of the failure of the South to grow had obvious policy implications, especially for trade policy: abandon free trade, and behind protection change the structure of the economy, that is, industrialize. The idea was simple. Products – especially consumer durables – now imported would be subjected to prohibitive tariffs or banned completely and local firms would emerge to take advantage of the already existing demand. These arguments and their policy implications became identified as the import-substitution (IS) approach to development.

Growth theory available at the time was the Harrod–Domar model in which physical capital formation was the basic source of growth. The main (often only) requirement to achieve growth (and changed structure) was a rate of capital formation sufficiently high, given the productivity of capital, to achieve a target rate of growth of output.⁶ Since virtually all capital goods were imported from the North, many countries sought to subsidize capital formation by maintaining an overvalued exchange rate. This practice, along with low, often negative, real interest rates, did induce relatively satisfactory rates of investment. At the same time such policies made it particularly difficult for new firms to enter export markets, they penalized employment and resulted in a great deal of underutilized capacity in the newly created firms.⁷ Physical capital, deemed the major lack in the South, was frequently found to be unused and poorly maintained. The importation of capital goods along with the inability to export new products created balance-of-payments problems which were often met by further exchange controls or more widespread use of tariffs and quotas.⁸

Distortions were added to by the common practice of many countries of establishing individual tariffs to whatever level was deemed necessary to enable a given activity to come into existence. The result was a wide range of tariff rates and an even wider range of effective rates of protection.

In these highly distorted economies, market prices were quite misleading as signals of social costs and social returns. Considerable attention was given to benefit–cost analysis of large-scale projects using shadow prices, rather than market prices. This was difficult to do with any confidence, and even where it was done, faced the problem that costs and prices paid and received were those prevailing in the market, so that profitability of firms in shadow prices did not assure profitability in market prices which somehow had to be achieved.⁹ The use of shadow prices did not, in any way, offset the distortions that the policies to implement the shift in structure imposed on the economies.¹⁰

It also became clear that the distortions, especially those that affected access to foreign exchange and domestic investment permits, resulted in a substantial allocation of resources to rent-seeking, that is, to seeking permits and licenses to access underpriced resources, rather than seeking investment projects, new technical knowledge and focusing on their own business. There is little doubt that such activities were costly in many ways and dampened the search for ways to make the new firms increasingly productive.¹¹

The general approach to development just described, despite violating many of the more established principles of textbook economics, was widely supported. The World Bank, for example, in its 1979 *World Development Report*, noted (pp. 67–8) that the policy had had important positive effects on entrepreneurial and technological capacity in many developing countries, and had induced the growth of a manufacturing sector behind high levels of protection. This bank report noted the inefficiencies described above, but with much more restraint than was the case a decade or so later. There were other favorable reports and very few explicit criticisms expressed in the literature of the day.

Some consequences

Tables 38.1 and 38.2 show the broad outlines of the post-World War II years when IS was widely followed. The 1950s were dominated by the recoveries from the dislocations of the war, but by the 1960s the effects of IS in many countries became paramount. During the 1960–73 period both labor productivity and total factor productivity growth were impressive over a wide range of countries. Most countries also experienced export growth that dispelled doubts about the ability of developing countries to export. An index from Little et al. (1970, p. 245) shows that manufactured exports from all developing countries increased from a base of 100 in 1953 to 283 in 1965. Capital formation also took place at impressive rates in most countries as the protection of domestic activities created many new opportunities. Even agriculture, despite being heavily penalized, grew at respectable rates in most – not all – countries pursuing IS.

In addition life expectancy increased almost everywhere, infrastructure was improved, and literacy increased. Clearly things got better in the 1950s and 1960s as a consequence of the IS strategy.¹² By the end of the 1970s it appeared that IS was a great success and a sure guide to continued growth.

As the tables show however, difficulties appeared as the 1970s wore on. Total factor productivity growth (TFPG) slowed markedly, and many cases of negative TFPG appeared, a sure sign of increasing distortions and coordination failures. Falling growth rates also contributed to negative productivity growth. Export growth slowed and turned negative in many cases.

Table 38.1 *Rates of growth of productivity*

Country	1960/73		1973/84		1984/94	
	GDP/L	TFP	GDP/L	TFP	GDP/L	TFP
<i>East Asia</i>	4.2	1.3	4.0	0.5	4.4	1.6
China	2.2	1.4	4.3	2.2	0.0	4.6
Indonesia	2.5	1.1	4.3	0.5	3.7	0.9
S. Korea	5.6	1.4	5.3	1.1	6.2	2.1
Malaysia	4.0	1.0	3.6	0.4	3.8	1.4
Philippines	2.5	0.7	1.2	-1.3	-0.3	-0.9
Thailand	4.8	1.4	3.6	1.1	6.9	3.3
Taiwan	6.8	2.2	4.9	0.9	5.6	2.8
<i>Latin America</i>	3.4	1.8	0.4	-1.1	0.1	-0.4
Argentina	2.6	0.2	0.4	-1.0	1.1	1.0
Brazil	4.4	2.9	1.0	-0.8	0.5	-0.2
Chile	1.6	0.7	-0.6	0.7	4.7	3.7
Colombia	2.9	1.9	1.2	0.0	1.8	1.0
Ecuador	4.4	3.3	1.7	-0.5	0.0	-0.1
Mexico	3.8	1.6	0.7	-0.8	-1.1	-1.8
Venezuela	1.2	0.9	-3.1	4.3	-0.6	-0.4
<i>Middle East</i>	4.7	2.3	0.5	-2.2	-1.1	-1.5
Egypt	3.0	1.8	6.2	2.3	0.0	-1.5
Iran	6.1	2.4	-2.9	-5.7	-2.2	-2.2
Jordan	2.1	-0.9	6.7	2.3	1.2	-2.9
<i>South Asia</i>	1.8	0.1	2.5	1.2	2.7	1.5
Bangladesh	0.0	-0.6	2.5	1.8	1.1	0.7
India	1.8	0.1	2.4	1.0	3.1	1.6
Sri Lanka	2.1	1.0	3.2	0.7	2.7	1.0
Pakistan	3.9	0.2	2.8	2.0	2.7	1.5
<i>Africa</i>	1.9	0.3	-0.6	-2.0	-0.6	-0.4
Ethiopia	2.2	0.2	0.0	-0.9	-0.2	-1.6
Ghana	0.9	-1.0	-3.2	-3.2	1.8	1.1
Kenya	3.4	3.4	0.4	-0.1	0.1	0.4
Nigeria	1.2	-0.9	-2.3	-4.6	1.3	2.0
Uganda	0.7	-0.3	-2.9	-3.0	1.3	1.1
Tanzania	3.0	2.2	-1.1	-1.7	1.0	0.6
S.Africa	2.3	0.9	1.0	-0.3	-2.0	-1.8
Zambia	1.0	0.2	-2.3	-1.9	-2.5	-1.1
Zimbabwe	2.9	2.7	-0.8	-1.3	0.2	0.4

Source: Rodrik (1999, pp. 71–72). The data in Rodrik are from a more complete study by Collins et al. (1996).

Table 38.2 Growth rates: exports and investment

Countries	Exports		Investment	
	1960/70	1970/77	1960/70	1970/77
<i>Low-income countries</i>	5.0	-1.7	5.7	2.3
Bangladesh	6.6	-7.3	11.1	-7.8
Ethiopia	3.7	-3.8	5.7	-0.9
Malawi	11.6	3.0	13.3	4.5
India	3.1	6.4	5.6	2.1
Pakistan	8.2	-3.0	6.9	-0.7
Tanzania	3.5	-7.2	9.8	2.7
Sri Lanka	4.6	-5.3	6.6	0.9
Kenya	7.2	1.2	7.0	-2.7
Uganda	5.0	-9.6	9.8	-11.5
Indonesia	3.5	7.5	4.8	16.6
<i>Middle-income countries</i>	5.4	5.1	7.6	8.1
Egypt	3.2	-3.3	3.1	23.6
Ghana	0.1	-1.9	-3.2	-8.6
Nigeria	6.1	1.3	6.5	22.9
Thailand	5.2	12.1	15.4	6.3
Philippines	2.2	5.0	8.2	11.7
Zambia	2.2	-2.3	10.6	-5.9
Jordan	10.1	20.8	9.9	...
Colombia	2.2	1.2	4.5	3.6
Ecuador	3.7	9.0	...	12.1
South Korea	35.2	30.7	23.1	12.4
Peru	1.9	-4.4	2.4	7.1
Malaysia	6.1	5.2	7.2	10.0
Turkey	1.6	0.8	8.8	12.7
Mexico	3.3	1.9	9.5	8.1
Chile	0.6	7.7	4.2	-8.9
China	23.7	6.7	16.2	9.1
South Africa	5.5	6.7	9.0	...
Brazil	5.0	6.5	5.3	12.6
Argentina	3.3	5.5	4.1	1.6
Iran	12.7	-0.2	12.2	22.6
Venezuela	2.0	-10.5	7.3	9.8

Source: World Bank, *World Development Report 1979*.

Capital formation in the poorer countries declined and, while remaining strong in a number of middle-income countries, declined in others as investment opportunities dried up. The slowing down of capital formation and productivity growth also dampened an already weak demand for labor,

and unemployment became an increasingly evident issue. Similarly almost all countries had trouble maintaining both internal and external stability.¹³

Two things were clear to most observers by the early 1970s: the first was that for well over a decade IS had been a genuine success, and the second was that, as practiced, it was not the strategy that could create an economy in which growth was built in to the routine operation of the economy.

Several aspects of the strategy became widely appreciated. That the assumption of fixed production coefficients of Harrod–Domar was the source of many difficulties especially became recognized.¹⁴ The assumption of fixed production coefficients was a principal reason for the failure of the high rates of investment to create the demand for labor that would match available supply as well as for the widespread underutilization. The assumption rested largely on the belief, widely held, that the technology of the North was so dominant that it made any other technology, any other factor combination, technologically inefficient, that is, inefficient at any set of factor prices. So in effect firms in the South had no choice of technique or machine if they were to compete in the world economy.

This last assumption rested on another, larger, more encompassing notion. The prevailing idea of development at this time was simply to duplicate the North. This idea is given full exposition by Arthur Lewis in a very influential article published in the early 1950s.¹⁵ Lewis divides the economy of a South country into a large, very poor, traditional sector and a small capitalist sector. Investment was to take place in the latter sector and was to be accomplished by importing physical capital and technology from the North. As investment continued, the capitalist sector would grow relative to the traditional sector until it encompassed the whole economy and the South would be as the North. Development was in effect imported, not indigenous, and capacity to import became a crucial constraint.¹⁶

The dual economy model was thus a story of displacement, not an explanation of how a non-growing traditional economy changed itself into a growing economy. This pervasive idea, development as replication of the North, had several consequences. It detracted attention away from the development notion itself, that is, to achieve the metamorphosis of a traditional, non-changing economy into a growing one. A theory of development is concerned with explaining how this traditional economy can be turned into a growing one in such a way that those characteristics that, in effect, define the society – its history, its values, its ethos, its very meaning – are not violated. Fixed production coefficients, the idea that all technical and organizational knowledge from the North could be codified and immediately utilized as in the North, the achievement of the North's product mix, all contributed to the dampening, even preventing, of any domestic efforts to adapt, to seek and to learn, and to recognize the importance of

building onto and from and within existing institutions and within the boundaries defined by the ideas of the good life of the population. It is evident, from this statement, that the role of trade (and other foreign activities) is vastly important, but subtle and complex.¹⁷

Finally, the strategy as pursued not only neglected, but in general penalized agriculture. Agriculture was often taxed and more often price controls on foodstuffs were imposed to keep their prices (and therefore wages) low in urban areas as yet another means of subsidizing new manufactures. The neglect of agriculture also encouraged a more rapid exodus from rural to urban areas, the result of which was increased unemployment in the cities and the emergence of 'informal sectors' alongside the new manufactures. Agriculture in most countries of the South was the largest sector in terms of both production and employment. To penalize this sector meant that the major sector of the economy was being penalized. In a country where new exports were very slow in emerging, a weak agricultural sector meant that the growth of domestic demand for new, non-agricultural products was severely dampened, and this in turn impeded the learning-by-doing process in the new firms. That agriculture could be safely penalized was a by-product of the 'structural change' argument for import substitution: the other side of creating a new manufacturing sector was the killing off of agriculture. Where exports of the new activities were not possible and agriculture was penalized, then importing food grains often became necessary and was costly. The late 1950s in India were a prime example of this sort of phenomena.

That this set of policies did not result in the creation of a built-in growth process does not now appear surprising. Despite the burst of early growth in the 1950s and 1960s noted above, and the improvement in a number of welfare measures, problems began to emerge that convinced most observers that the import-substitution process was in no sense sustainable. The details of the story just summarized vary markedly among the countries of the South, but the broad picture seems generally applicable. For Africa less so, perhaps, than in most other places, and more so in India, Pakistan and most Latin American countries.¹⁸

The fall of IS and the rise of openness

While recognition that the IS policies had created an unsustainable situation became widespread, alternative strategies did not convince many policy-makers in the South. That distortions were ubiquitous and were penalizing the economies was appreciated by most observers. Economists still had only rough ideas of how to make an economy grow, but of course distortions and optimal allocation of resources were textbook stuff. It is not surprising therefore that the new strategy proposed concentrated

attention on eliminating the sources of these distortions: the ad hoc tariffs, internal controls of many kinds, overvalued exchange rate, soft budget constraint, public ownership of many firms, price controls and inflationary fiscal policies. The catchwords became ‘market friendly’, ‘privatizing’, ‘macro stability’ and ‘openness’. ‘Outward orientation’ and ‘Washington Consensus’ replaced IS as the summary term. This new approach originated largely in the North and the pressure to liberalize came from the North, the World Bank, the United States Agency for International Development (USAID) and Northern academic economists, hence the appellation Washington Consensus. Reluctance of South countries to move sharply toward liberalization was partly a matter of the immediate economic costs (especially the almost inevitable increased unemployment) to withdrawing protection from recently established activities that could not survive without protection, and partly a matter of genuine doubt that the newly pushed theory would in fact have the desired effect. In addition, there were numerous people who profited from IS, and who, therefore, opposed its abandonment.

There was no new ‘theory’ of development that led to this different strategy. The liberalization package as looked upon by the South as simply a return to pre-World War I international arrangements. In particular the idea that if property rights are clear and in place, if the price mechanism follows the textbook stories, if macrostability is maintained, and the investment rate is at least 15 percent, the economy will grow with full employment, was not a widely accepted view outside the Washington Consensus group. It was too easy to see that some countries – for example India, Korea, Vietnam, China and Botswana – were growing well by violating many of the conditions laid down as essential by the Washington Consensus. Similarly other countries – especially in Latin America – had followed the rule rather closely and performed much less well.¹⁹

The period from 1870 to 1914 is especially illuminating. During this period capital and trade moved easily to almost all parts of the world, government interventions in the economy were modest, migrations were large and widespread, foreign exchange markets were fairly stable due to the widespread commitment to the gold standard and to the British pound, the international capital markets were effective, price stability was common, and there were significant social and political changes throughout the world. In addition, there were marked improvements in transportation and communication around the world.²⁰ These circumstances, however, did not produce industrialization in the South, and did not produce productivity growth, nor the growth of factory employment in the South.²¹ While there was some growth in the South during this interval, it is correct to say that the North grew faster and routine growth became even

more firmly established. This experience was rarely confronted by the advocates of openness.²²

Doubts about the Washington Consensus

Distortions of course mean (by definition) that the economy operates within its production possibilities frontier with the given technology that is available. So eliminating the distortions would be expected to result in a once-over increase in output; but the real question was what produced growth after this one-shot effect has been absorbed.

The main evidence leading to the outward-oriented position was the success of Korea and Taiwan. Taiwan had begun to grow rapidly from the early 1950s and Korea from the early 1960s. The most obvious feature of these impressive performances was the high growth rate of exports, often non-traditional products and services that had come into existence behind the protection and the subsidies that IS had provided. Similarly it was widely believed initially that the two economies were largely market driven with very little government intervention. Both countries made marked policy changes in this direction in the 1960s. These changes did reduce distortions, but they also included other government policies that subsidized exports and capital formation. In particular it became clear that both countries maintained an undervalued exchange rate over much of the time after 1960. There were other policies and arrangements that impeded imports. Liberalization as practiced by these two countries did not mean 'free' trade. It also became understood that both countries had accumulated a great deal of collective learning during the Japanese occupancy of the first part of the twentieth century.²³ Korea also learned greatly from the presence of United States armed forces engineers in their country. This collective learning had produced a labor force much more experienced and skilled than in most of the other countries trying the IS strategy. This last item is especially relevant in understanding how and why production in the two countries responded so well and so quickly to the incentives and subsidies offered by their governments.

There was one other characteristic of government policy-making that is highly relevant to this story. Both governments recognized that good policy-making required trial and error and hence willingness to change policies and try different approaches. The idea was not minimal government, but rather a searching, learning government.²⁴ This searching for the right policies was necessary simply because it was recognized that there can be no right policy that can be arrived at in any other way.

The Washington Consensus became widely advocated and strongly pushed by international agencies and key academic figures. To repeat the point made above: that IS, as practiced, had run into a dead end was, in

general, appreciated, but the Washington Consensus seemed simply to go back to an approach to development that had also failed as evidenced in particular by the experience of the 1870–1914 period. Research in the 1970s and 1980s on Korea and Taiwan convinced almost all observers that their success was not due to a minimal government or a market-friendly approach, and that these economies were much less open than was thought to be the case in the 1960s and 1970s.²⁵

There have been, from early on, efforts to examine empirically the effects of IS and outward-orientation policies on growth of GDP, on employment and productivity, and on macrostability. A survey of this literature is beyond the scope of this chapter, but a few remarks may give the flavor of the current status of this work.²⁶ There are two general approaches: case studies of individual countries and cross-country regression analyses involving numerous countries. There are difficulties at all steps along the way in both approaches: definitions and measurement of inward- and outward-looking, quality of available data, the appropriate theoretical formulations, distinguishing the role of trade policy from that of macro stability, education and other possible factors, and many more. Cross-country regressions, once so widely used, are increasingly recognized as subject to so many difficulties that their results are essentially meaningless.²⁷

Three issues of direct relevance to trade policy may be noted briefly. The first is the role of exports. Do countries grow well because they export or do they export because they grow well? At the aggregate level the evidence might well support the former position, but at the firm or sector level, the evidence generally supports the opposite view. The micro data appear somewhat more convincing, that is, firms must find their basic inducements to search and learn that lead to productivity growth within their own indigenous environment, and thereby become able to export. It does seem clear that once they begin to export, then they can gain additional knowledge from that activity. Secondly, evidence supports the view that exporting does not have much effect on productivity growth of traditional activities of the developing countries.²⁸

Thirdly the empirical (and other) evidence supports the view that the exchange rate is a crucial policy variable. The overvalued exchange rate was a major, perhaps the major, reason for the failure of the IS approach. Recent empirical work and case studies show convincingly the strength of an undervalued exchange rate as an instrument that encourages import replacements and the search for foreign markets as well as the search for increased productivity and higher-quality output.²⁹

While it is important that empirical and historical research continue, it now seems likely that for the foreseeable future dispute between the two approaches cannot be resolved by evidence of the conventional sort applied

in the usual way. The most illuminating approach is simply case studies of individual countries to ascertain what has gone on in these countries. Generalizations across countries and through time, supported by rigorous theoretical formulations and by equally rigorous econometric results, do not now appear possible, and certainly not possible to the extent that specific policy formulations emerge that are applicable to all countries. The last section addresses briefly the consequences of this position.

The upshot of it all

An important reason why the role of trade in development is difficult to isolate is that the development process itself is not yet fully understood. There are two main aspects of development that both IS and the Washington Consensus seemed not to have appreciated in their early formulation. The first was the assumption, implicit in most of the literature of both strategies, that technical and other knowledge are public goods, available to everyone and equally productive in all places. Research of the 1980s and 1990s has shown this assumption to be terribly misinformed.³⁰ Productivity levels among firms in the same activity vary widely within a country and across countries, as do rates growth of productivity, profit rates, capital–labor ratios, product quality, and so on.³¹ These great and abiding differences are explained by the fact that much of knowledge cannot be codified and put into manuals, but is tacit and can be accumulated only by producing and is, therefore, unique to the firm where it is accumulated. That this is the case is due to the great differences in entrepreneurship, in the ingenuity of the shop floor people, and in the extent of commitment to searching and learning.³² This finding has had a fundamental impact on development theory and practice: since sustained growth requires sustained increases in productivity, which, in turn, requires continuing increases in knowledge, a growing economy must be one in which firms are, as a matter of routine, actively engaged in searching for new knowledge.³³ Policy incentives must be geared to this objective, different from the incentives aimed at achieving an ‘optimal’ allocation of given resources and given technology.³⁴ Trade policy, in particular, is greatly affected by the notion of tacit knowledge and its source.

The second important research finding has to do with institutions: institutions as norms and rules of the game. As such they have direct and significant effects on the response of economic agents to market incentives. Economic actors are rule-followers as well as profit-seekers, and the rules emerge from and in turn create institutions. So whether a policy will have the desired effect depends on the institutional environment within which it is applied, and any change is sure to be resisted to some extent. In particular it means that the usual market signals can be expected to

have diverse results in different countries and at different times in the same country.

Given the two preceding points, it is evident that the textbook notions of 'optimal' allocation of resources or some other ideal or maximum state are basically meaningless: technology is always changing, that is, its change is part of the firm's activity, and institutions mean that the textbook ideal – for example perfect competition – cannot be defined independently of these institutions. To understand a given economy means to gain some insight into how it works and then to show that its workings can be improved upon according to some practical criteria. This is quite different from the more usual objective of trying to create an economy that follows a general strategy, for example perfect competition, Pareto optimality, and so on.

Incentives are to be aimed at creating inducements to search and learn. Import-substitution strategies created investment opportunities, but also dampened any inducements to search and learn, to create tacit knowledge and thereby continue to grow. The Washington Consensus assumed that simply eliminating distortions by virtue of an unfettered market and a 'realistic' exchange rate would produce growth. Both were and are inadequate, because they misunderstand tacit knowledge and the role of institutions.

The final question is what kind of a trade policy will induce firms to engage in the searching and learning process that does produce growth and improved quality of output. Foreign trade policy should aim at making it very profitable to export and to replace imports. Such an approach is different from that usually practiced of subsidizing inputs – capital, loanable funds, land, fertilizer, and so on – as noted earlier.

There are several trade policies available that reflect this point of view. The most evident is an undervalued exchange rate, that is, one that results in the accumulation of foreign exchange. Such an exchange rate policy has been pursued by China, Taiwan, South Korea, Brazil and Japan at times to considerable advantage. Foreign aid takes the form of guaranteeing favorable prices for a developing country's exports, or prices that rise as the exporting firm increases its productivity or tax relief based on increased export earnings. A policy that rewards firms that increase employment over time with a given capital stock has had favorable effects on both employment and productivity growth.

The general conclusion as to the present state of play of the role of trade may be summed up in this way: we have learned over recent decades that simple notions such as IS and outward orientation lead nowhere. Trade policies – including exchange rate policy – must emerge from a clear vision of how development takes place. That vision must explicitly include knowledge accumulation and application and the recognition that much knowledge is necessarily tacit. It must also recognize that institutions are so

fundamental and so history- and culture-dependent that generalizations across countries are risky indeed. Thus a country, in determining its foreign trade policy, must study how its economy in fact works.³⁵

Notes

1. There is some evidence that inequality among nations increased between 1500 and 1800. See, for example, Lindert and Williamson (2003).
2. President Harry Truman, in his inaugural address in early 1949, committed the United States to supply technical and financial aid to the countries of Africa, Asia, and Latin America.
3. This argument was first developed in Raul Prebisch (1950) and by Hans Singer (1950). A later and slightly different version is Lewis (1978).
4. A formal model that demonstrates that free trade can be harmful to long-run growth in those countries that are locked into sectors where productivity is low and whose indigenous technological change is slow or completely absent, is Rodriguez and Rodrik (2001).
5. See the papers by Gregory Clark and Robert Feenstra (2003) and by Clark (1987).
6. The notion of an optimal rate of saving was often introduced. If this optimal rate of domestic saving was not high enough to produce the target rate of growth of output, then foreign aid or borrowing could supplement domestic saving to the extent necessary to reach the target.
7. Gordon C. Winston (1974) is one of the earliest analyses of the underutilization of capital issues.
8. Countries that had long relied heavily on one or two primary exports often experienced the 'Dutch disease' even before the beginning of import-substitution policies. The overvaluation pursued as a policy to encourage capital formation often exacerbated an already misleading exchange rate regime.
9. See Hollis B. Chenery (1961) for an elaboration of the role of shadow prices in development.
10. The study by Little and Mirrlees (1974) is a fine review of the state and role of cost-benefit analysis at the time.
11. Anne O. Krueger (1974) first drew the attention of the profession to this important cost of the distortions.
12. African countries shared less in this success than in countries in other parts of the world. This failure had little to do with IS as few African countries made a substantial commitment to the strategy.
13. Although the term 'import substitution' seems to imply that imports should be reduced relative to GDP, this was rarely the case. The strategy proved quite import-intensive. See Carlos Diaz-Alejandro (1965).
14. Robert Solow's neoclassical growth model (Solow, 1956) was motivated largely by the recognition that the fixed-coefficient assumption severely penalized employment growth as well as impeding adjustment of production techniques to the factor supply situation in general.
15. See W. Arthur Lewis (1954) and, somewhat later, Fei and Ranis (1964) for elaborations of the labor surplus models of growth.
16. Chenery and Strout (1966) developed a model in which capacity to import was a constraint on growth along with the capacity to save.
17. The role of 'tacit knowledge', in contrast to codified knowledge, is a crucial part of the story and will be discussed later.
18. There were many studies that explored the import substitution experience in specific countries and across several countries. Little et al. (1970) and the six country studies that accompanied it were among the first that attracted attention. Two studies by Balassa (1971) and Balassa et al. (1982) were also important in spreading understanding about the problems of IS, as was Krueger (1978). There are many others.
19. See, for example, Rodriguez and Rodrik (2001), Krueger (2000) and Baldwin (2004).

20. Some calculations show that the correlation between domestic saving and domestic investment was lower in this period than in recent years when capital markets are assumed to work extra well. With 'perfect' international capital markets there should be no or very little relationship between domestic saving and investment.
21. Elaboration of this point is in Lewis (1978) and Dowrick and DeLong (2003). See also Maddison (1970).
22. The story is similar if one begins in 1820 except that from 1820 to 1870 there were more tariffs and other interferences with the flow of trade and other forms of market interventions.
23. See Kohli (1994) for a good discussion of the impact of the Japanese occupancy on the creation of human capital in the two Koreas and Japan.
24. Gustav Ranis (in Winrock International Institute for Agricultural Development and USAID, 1991, pp. 128–9) points out that: . . . 'what happened in Taiwan was not Mandarins sitting around saying is this what we have to do now. There was a lot of bumbling and stumbling and going back and forth.' See also Biggs et al. (1995).
25. The World Bank study (1993) acknowledged this picture in general, but argued strongly that Korea and Taiwan were distinctive in several ways and their experiences and policy packages could not be replicated in other countries. See also Stiglitz (2000).
26. There are numerous surveys of this material. See especially Edwards (1993), Harrison and Revenga (1995), Easterly and Levine (2001), Rodriguez and Rodrik (2001) and Baldwin (2004).
27. Jesus Felipe (1999), Durlauf (2000) and Brock and Durlauf (2001) all have especially helpful discussions of cross-country regressions and growth accounting in general.
28. Sectoral and firm studies have become plentiful in recent years, and results can always be questioned. Michael Hobday (1995) is a thorough study of the relationship between exporting and learning. Roberts and Tybout (1996), Aw et al. (1998), Temple (1999), Tybout (2000) and Bernard and Jensen (2001) all have helpful discussions and much data. Westphal (1990) is a particularly good discussion of the role of exports.
29. See Bruton (1997, Chapter 8) and Bruton (1998) for a discussion of the undervalued exchange rate as an instrument of development policy.
30. See Evenson and Westphal (1995) for further discussion and evidence.
31. There are many studies available that confirm this state of things. See the sources cited in notes 28 and 30.
32. Tacit knowledge has been discussed by many people. Nelson and Winter (1982) were perhaps the first to employ it in a strategic way in economic analysis. More recently good discussions are found in papers by Chandler et al. (1998) and Chandler (1992). See also Helleiner (1992).
33. John Williamson (2003, p. 324) lists ten 'points' that are included in the Washington Consensus and that will (presumably) produce growth. The list does not include any item that acts directly on growth, but essentially defines the conditions for a perfectly competitive economy with no inflation and a government role limited to conventional post office, defense, infrastructure, and so on. See also Krueger (2000).
34. The differences among firms means that it is a dubious business to speak of a country's comparative advantage. Some firms in an industry export, some do not. For a country to seek to produce those goods and services in which it has a comparative advantage is to seek that which is not there. See Hausmann and Rodrik (2002) for an interesting empirical study of the determination of activities in a country.
35. This position is increasingly recognized even by economists who differ sharply in other ways. See almost any of the recent writings of Dani Rodrik, especially Rodrik (1999, 2000) and Stiglitz (2000) on the one hand and Srinivasan and Bhagwati (2001) on the other. This point is also discussed in Bruton (1998).

References

- Aw, Bee Yan, Xiaomin Chen and Mark J. Roberts (1998), 'Firm-level Evidence on Productivity Differentials, Turnover, and Exports in Taiwan', NBER Working Paper No 6235.

- Balassa, Bela (ed.) (1971), *The Structure of Production in Developing Countries*, Baltimore, MD: Johns Hopkins University Press.
- Balassa, Bela, Julio Berlinski, Daniel Schydowski et al. (1982), *Development Strategies in Semi-Industrialized Countries*, Baltimore, MD: Johns Hopkins University Press for World Bank.
- Baldwin, Robert E. (2004), 'Openness and Growth: What's the Empirical Relationship?', in Robert E. Baldwin and L. Alan Winters (eds), *Challenges to Globalization*, Chicago, IL: University of Chicago Press, pp. 499–521.
- Bernard, Andrew B. and J. Bradford Jensen (2001), 'Why Some Firms Export', NBER Working Paper No 8349.
- Biggs, Tyler, Manju Shah and Pradeep Srivastava (1995), *Technological Capabilities, and Learning in African Enterprises*, Washington, DC: World Bank.
- Brock, William A. and Steven N. Durlauf (2001), 'Growth Empirics and Reality', *The World Bank Economic Review*, 15 (2): 229–72.
- Bruton, Henry J. (1997), *On the Search for Well-Being*, Ann Arbor, MI: University of Michigan Press.
- Bruton, Henry J. (1998), 'A Reconsideration of Import Substitution', *Journal of Economic Literature*, 36 (June): 903–6.
- Chandler, Alfred D., Jr. (1992), 'Organizational Capabilities and the Economic History of the Industrial Enterprise', *Journal of Economic Perspectives*, 6 (3): 78–100.
- Chandler, Alfred D., Jr., Peter Hagstrom and Orjan Solvell (1998), *The Dynamic Firm*, Oxford: Oxford University Press.
- Chenery, Hollis B. (1961), 'Comparative Advantage and Development Theory', *American Economic Review*, 50 (2): 18–51.
- Chenery, Hollis B. and Alan M. Strout (1966), 'Foreign Assistance and Economic Development', *American Economic Review*, 56 (4): 674–733.
- Clark, Gregory (1987), 'Why Isn't the Whole World Developed? Lessons from the Cotton Mills', *Journal of Economic History*, 41 (1): 141–73.
- Clark, Gregory and Robert C. Feenstra (2003), 'Technology and the Great Divergence', in Michael D. Bordo, Alan M. Taylor and Jeffrey G. Williamson (eds), *Globalization in Historical Perspective*, Chicago, IL and London: University of Chicago Press, pp. 277–322.
- Collins, Susan, Barry Bosworth and Dani Rodrik (1996), 'Economic Growth in East Asia: Accumulation Versus Assimilation', *Brookings Papers on Economic Activity*, 1996 (2): 135–91.
- Diaz-Alejandro, Carlos F. (1965), 'On the Import Intensity of Import Substitution', *Kyklos* 18 (3): 495–511.
- Dowrick, Steve and J. Bradford DeLong (2003), 'Globalization and Convergence', in Michael D. Bordo, Alan M. Taylor and Jeffrey G. Williamson (eds), *Globalization in Historical Perspective*, Chicago, IL and London: University of Chicago Press, pp. 191–226.
- Durlauf, Steven N. (2000), 'Econometric Analysis and the Study of Economic Growth: A Skeptical Perspective', in Roger E. Backhouse and Andrea Salanti (eds), *Macroeconomics and the Real World, Volume 1*, Oxford: Oxford University Press, pp. 249–62.
- Easterly, William and Ross Levine (2001), 'What Have We Learned from a Decade of Research on Growth? It's Not Factor Accumulation: Stylized Facts and Growth Models', *World Bank Economic Review*, 15: 127–219.
- Edwards, Sebastian (1993), 'Openness, Trade Liberalization and Growth in Developing Countries', *Journal of Economic Literature*, 31 (4): 1358–93.
- Evenson, Robert E. and Larry E. Westphal (1995), 'Technological Change and Technology Strategy', in Jere Behrman and T.N. Srinivasan (eds), *Handbook of Development Economics*, Vol. IIIA, Amsterdam: Elsevier, pp. 2209–99.
- Fei, John C.H. and Gustav Ranis (1964), *Development of the Labor Surplus Economy*, Homewood, IL: Richard Irwin.
- Felipe, Jesus (1999), 'Total Factor Productivity Growth in East Asia: A Critical Survey', *Journal of Development Studies*, 35 (4): 1–42.
- Harrison, Ann and Ana Revenga (1995), 'The Effects of Trade Policy Reform: What Do We Really Know', NBER Working Paper No 5225.

- Hausman, Ricardo and Dani Rodrick (2002), 'Economic Development as Self Discovery', NBER Working Paper 8952.
- Helleiner, Gerald K. (ed.) (1992), *Trade Policy, Industrialization, and Development*, Oxford: Clarendon Press.
- Hobday, Michael (1995), *Innovation in East Asia*, Aldershot, UK and Brookfield, USA: Edward Elgar.
- Kohli, Atul (1994), 'Where Do High Growth Political Economies Come From? The Japanese Lineage of Korea's Development State', *World Development*, 22 (September): 1269–93.
- Krueger, Anne O. (1974), 'On the Political Economy of the Rent Seeking Society', *American Economic Review*, 64 (1): 291–303.
- Krueger, Anne O. (1978), *Foreign Trade Regimes and Economic Development*, Cambridge, MA: Ballinger.
- Krueger, Anne O. (ed.) (2000), *Economic Policy Reform*, Chicago, IL: University of Chicago Press.
- Lewis, W. Arthur (1954), 'Economic Development with Unlimited Supplies of Labor', *Manchester School*, 22 (2): 139–91.
- Lewis, W. Arthur (1978), *Growth and Fluctuations*, London: George Allen & Unwin.
- Lindert, Peter H. and Jeffrey G. Williamson (2003), 'Does Globalization Make the World More Equal', in Michael D. Bordo, Alan M. Taylor and Jeffrey G. Williamson (eds), *Globalization in Historical Perspective*, Chicago, IL and London: University of Chicago Press, pp. 227–76.
- Little, I.M.D. and James Mirrlees (1974), *Project Appraisal and Planning for Developing Countries*, New York: Basic Books.
- Little, I.M.D., Tibor Scitovsky and Maurice Scott (1970), *Industry and Trade in Some Developing Countries*, London: Oxford Press for the OECD.
- Maddison, Angus (1970), *Economic Progress and Policy in Developing Countries*, New York: W.W. Norton & Co.
- Nelson, Richard R. and Sidney G. Winter (1982), *An Evolutionary Theory of Economic Change*, Cambridge, MA and London: Belknap Press of Harvard University.
- Prebisch, Raul (1950), *The Economic Development of Latin America and Its Principal Problems*, Lake Success, NY: United Nations Department of Social Affairs.
- Roberts, Mark J. and James R. Tybout (eds) (1996), *Industrial Evolution in Developing Countries*, New York: Oxford University Press for the World Bank.
- Rodriguez, Francisco and Dani Rodrik (2001), 'Trade Policy and Economic Growth: A Skeptic's Guide to Cross-National Evidence', in Ben Bernanke and Kenneth S. Rogoff (eds), *NBER Macroeconomic Annual 2001*, Cambridge, MA: MIT Press, pp. 261–325.
- Rodrik, Dani (1999), *The New Global Economy and Developing Countries: Making Openness Work*, Washington, DC: Overseas Development Council.
- Rodrik, Dani (2000), 'Development Strategies for the 21st Century', in Boris Pleskovic and Nicholas Stern (eds), *Annual World Bank Conference on Development Economics*, Washington, DC: World Bank, pp. 85–108.
- Rodrik, Dani (2003), 'Growth Strategies', NBER Working Paper 10050.
- Singer, Hans W. (1950), 'The Distribution of Gains Between Investing and Borrowing Countries', *American Economic Review*, 40 (2): 473–85.
- Solow, Robert M. (1956), 'A Contribution to the Theory of Economic Growth', *Quarterly Journal of Economics*, 70 (February): 65–94.
- Srinivasan, T.N. and Jagdish Bhagwati (2001), 'Outward-Oriented and Development: Are Revisionists Right?' in Deepak Lal and Richard Snape (eds), *Trade, Development and Political Economy*, Houndmills and New York: Palgrave, pp. 3–26.
- Stiglitz, Joseph E. (2000), 'Development Thinking at the Millennium', in Boris Pleskovic and Nicholas Stern (eds), *Annual World Bank Conference on Development Economics*, Washington, DC: World Bank, pp. 13–38.
- Temple, Jonathan (1999), 'The New Growth Evidence', *Journal of Economic Literature*, 38 (1): 112–56.
- Tybout, James R. (2000), 'Manufacturing Firms in Developing Countries: How Will They Do and Why', *Journal of Economic Literature*, 38 (1): 11–44.

- Westphal, Larry E. (1990), 'Industrial Policy in an Export-Propelled Economy: Lessons from South Korea's Experience', *The Journal of Economic Perspectives*, 4 (Summer): 41–60.
- Williamson, John (2003), 'Our Agenda and the Washington Consensus', in Pedro-Pablo Kuczynski and John Williamson (eds), *After the Washington Consensus*, Washington, DC: Institute for International Economics, pp. 323–32.
- Winrock International Institute for Agricultural Development and USAID (1991), *African Development: Lessons from Asia: Proceedings of a Seminar on strategies for the future of Africa*, Baltimore, MD: Winrock Publications Sales.
- Winston, Gordon C. (1974), 'The Theory of Capital Utilization and Idleness', *Journal of Economic Literature*, 12 (4): 1301–20.
- World Bank (1979), *World Development Report, 1979*, Washington, DC: World Bank.
- World Bank (1993), *The East Asian Miracle: Economic Growth and Public Policy*, New York: Oxford University Press for the World Bank.

39 Foreign direct investment

*V.N. Balasubramanyam*¹

Introduction

The vast literature on foreign direct investment (FDI) in developing countries may seem disproportionate to the volume of FDI they harbour. Developing countries as a group accounted for 25 per cent of the total stock of FDI of \$8.9 trillion in the world economy at the end of the year 2004, most of which, around 70 per cent, was accounted for by 11 developing countries (Tables 39.1 and 39.2). One reason for the intense interest in FDI is the nature of the beast, which appears to evoke both admiration and opposition in equal measure; admiration because of the unrivalled ability of the multinational enterprises (MNEs), the main purveyors of FDI, to transfer technology and know-how across borders; opposition because the MNE is first and foremost a profit-maximizing entity. In the recent past attitudes towards FDI have turned from a mixture of suspicion and admiration towards one of unqualified admiration, shown by the eagerness of most developing countries to attract FDI. This change in attitude towards FDI on the part of host developing countries is influenced by a number of factors: a steep reduction in alternative sources of finance in the wake of the debt crisis, the collapse of the Soviet Union and with it a waning of ideological opposition to capitalism and its institutions, the demonstrable success of the East Asian countries based in part on FDI, and growth in knowledge and understanding of the nature and operations of FDI on the part of the host countries.

The issues that have surfaced in the recent literature on FDI reflect these changes. These have to do with the specific factors which figure prominently in the choice of locales for investment by foreign firms, and the factors which promote effective transfer of technology to the host countries and maximize the benefits they can expect from FDI. In addition, there is the suggestion mooted principally by the EU and Japan that FDI should be on the agenda of the World Trade Organization (WTO) on a par with trade in goods and services.

This chapter reviews these and other issues in the literature on FDI. The next section discusses the determinants of FDI, drawing on the theoretical and empirical literature. The subsequent section discusses the impact of FDI on growth and development in host developing countries. The final section draws some conclusions.

Table 39.1 Stock of foreign direct investment: 1990–2004

	1990–91	1995–96	2000–01	2004–05
<i>World</i>				
\$ US trillion	1.77	2.76	5.79	8.9
Percentage	100	100	100	100
<i>Developed countries</i>				
\$ US trillion	1.4	2.05	3.98	6.47
Percentage	79.4	74.4	68.7	72.7
<i>Developing countries</i>				
\$ US trillion	0.36	0.7	1.74	2.23
Percentage	20.6	25.2	30.1	25.1

Source: UNCTAD (various issues).

Table 39.2 Stock of inward foreign direct investment in developing countries

	2000–01		2004–05	
	US \$ million	percentage	US \$ million	percentage
<i>All developing economies</i>	1 739 726	100	2 232 868	100
Argentina	67 601	3.9	53 697	2.4
Brazil	103 015	5.9	150 965	6.8
Mexico	97 170	5.6	182 536	8.2
Africa	151 246	8.7	219 277	9.8
China	193 348	11.1	245 467	11.0
China, Hong Kong SAR	455 469	26.2	456 833	20.5
India	17 517	1.0	38 676	1.7
Korea, People's Republic of	37 189	2.1	55 327	2.5
Malaysia	52 747	3.0	46 291	2.1
Singapore	112 571	6.5	160 422	7.2
Thailand	29 915	1.7	48 598	2.2
Total of 11 countries	1 317 788	75.7	1 658 089	74.3

Source: UNCTAD (various issues).

Determinants

Stephen Hymer (1976) initiated the discussion on the determinants of FDI with the thesis that firms go abroad to maximize the rents inherent in the advantages they own. These advantages range from the possession of a

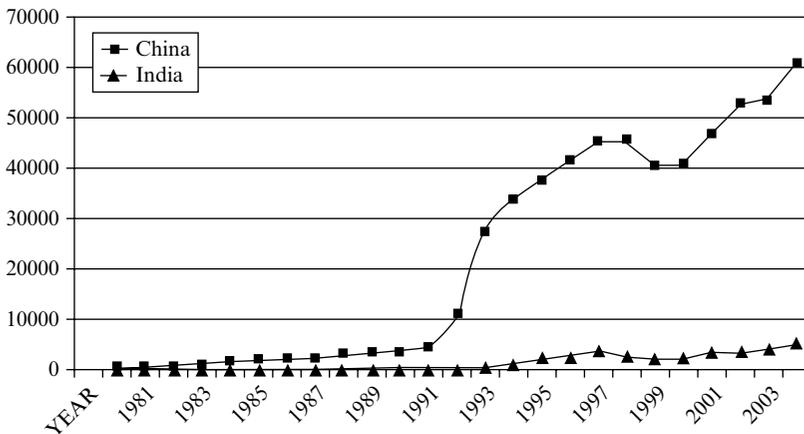
brand name to complex processes and product technologies. Hymer's work set in train a number of studies on the foreign investment decision process of foreign firms: John Dunning (1973, 1981), Buckley and Casson (1991), and Markusen (2004) who encapsulated the key factors in the foreign investment decision process of firms as the 'OLI paradigm'. Simply put, firms must possess ownership advantages (O), location advantages (L) and must be capable of internalizing operations (I) if they are to invest abroad. Internalization here refers to the ability of the firm to set up its own suppliers of materials and equipment (backward linkages) and marketing operations (forward linkages). Such internalization is necessary to overcome imperfections in the market which may result in a number of problems including the loss of ownership advantages through imitation by others, hold-up of operations by suppliers of materials and equipment, and the problems associated with decision-making with imperfect information (Williamson, 1981).

The concern of the host countries is with location advantages (L): how best to lure firms, which possess ownership advantages and are able to internalize, to their locale and benefit from the technology and know-how they bring along and the jobs they create. There is now much received wisdom on the sort of L factors which will attract FDI, the result of years of research and the experience of host countries with FDI. It is now well known that artificial incentives such as tax concessions and tax holidays offered to foreign firms do not always attract foreign investors; such incentives are not a substitute for the sort of economic environment foreign firms seek. Countries endowed with the sort of raw materials the foreign firms seek, those blessed with cheap but productive labour (low efficiency wages), and economies which enjoy macroeconomic stability and pursue stable economic policies tend to attract relatively large volumes of FDI. In the absence of these fundamentals, tax concessions and various sorts of subsidies are of little attraction to foreign firms.

In the same vein, it is suggested that policy-induced incentives such as tariffs on imports and subsidies to exports do not lure foreign firms. In fact a neutral trade policy which favours neither production for exports nor the production of import substitutes for the home market tends not only to attract large volumes of FDI but also promotes efficient utilization of FDI. The import-substituting domestic market-oriented strategy (IS strategy) is characterized by tariffs and quotas on imports, which on average outweigh the average level of subsidies given to exports. In the case of the export-oriented strategy (EP strategy) the opposite is the case. These sorts of incentives, which bias production in favour of either exports or the domestic markets, distort allocation of resources in the economy. They are also artificial and uncertain incentives. They are artificial in the sense that they do

not reflect the true market-dictated comparative advantage of various sectors in the economy; they are uncertain in the sense that they are subject to the vagaries of government policy. Foreign firms in general are not attracted by such artificial incentives and those that do respond and operate in the distorted environment may reap benefits but they tend to be transient. A neutral strategy favours neither the export markets nor the domestic markets; resource allocation in such an environment is dictated by market forces. Foreign firms seek such an environment, which allows for the full play of the comparative advantage they possess. There is robust statistical support for the proposition enunciated by Bhagwati (1978) that a neutral strategy attracts relatively large volumes of FDI and also promotes its efficacy (Balasubramanyam and Salisu, 1991; Balasubramanyam et al., 1996).

The recent experience of India and China with FDI, though, does not seem to conform to received wisdom on the determinants of FDI. They both possess most if not all of the location advantages foreign firms seek. Both countries are endowed with relatively cheap labour, both have liberalized their trade and FDI regimes to a considerable extent, they both possess large domestic markets and they have posted impressive growth rates in recent years. But they differ markedly in the volume of FDI they have attracted. China attracts ten times more FDI than India does: in recent years the annual average inflows into China have averaged around \$50 billion compared with the \$4 billion that India attracts (Figure 39.1 and



Source: UNCTAD

Figure 39.1 Inflows of foreign direct investment in China and India: 1980–2004 (\$ US million)

Table 39.3 Inflows of Foreign Direct Investment in China and India (\$ US million)

	China	India
1980	57	79
1981	265	92
1982	430	72
1983	916	6
1984	1 419	19
1985	1 956	106
1986	2 244	118
1987	2 314	212
1988	3 194	91
1989	3 393	252
1990	3 487	237
1991	4 366	75
1992	11 008	252
1993	27 515	532
1994	33 767	974
1995	37 521	2 151
1996	41 726	2 525
1997	45 257	3 619
1998	45 463	2 633
1999	40 319	2 168
2000	40 715	2 319
2001	46 878	3 403
2002	52 743	3 449
2003	53 505	4 269
2004	60 630	5 335

Source: UNCTAD (various issues).

Table 39.3). Indeed India's FDI regime is reported to be much more liberal than that of China (Nagaraj, 2003). Even so, the volume of FDI in India is lower than that in China by a factor of ten or more, a fact frequently debated in the media.

A number of reasons are offered for the observed differences in the volume of FDI in the two countries including differences in the accounting procedures between the two countries, so-called round-tripping FDI in China, and the horrendous Indian bureaucracy which stifles any sort of enterprise, be it foreign or domestic. There may be other explanations for the observed differences in the volume of FDI the two countries harbour. Yashang Huang (2003) is of the view that a considerable volume of FDI in

China is a substitute for domestic investment. Despite the relatively high savings rate, domestic Chinese investors find it hard to obtain credit lines and borrow from the banks, and they turn to foreign investors for finance. Many of the state-owned enterprises on the verge of bankruptcy also seek FDI, and foreign firms invest in regions of China which local firms hesitate to enter. It may also be the case that the requirements of the Indian manufacturing and services sectors are relatively low compared with those of China, simply because India is better endowed than China with the sort of human skills which FDI provides. In sum the determinants of FDI are rooted in the endowments of the host countries including human skills, the infrastructure facilities they provide and, most importantly, a policy framework which provides distortion-free product and labour markets.

Impact

FDI is defined as ownership of facilities abroad with control over operations. The distinguishing feature of FDI which sets it apart from other forms of capital flows is the control over operations which the parent company exercises over its subsidiaries abroad. As stated earlier, such control over operations is essential to preserve the ownership over advantages the firm enjoys. Such control over operations is exercised by virtue of ownership of equity and possession and control of technology and know-how. The higher the degree of equity participation by the investor firm in an entity, the greater is its ability to exercise control over operations. It is thought possible for a firm to exercise control over operations if it owns say only 4 per cent of the total equity of an entity and the other 60 per cent is disbursed amongst a number of investors. In fact, the International Monetary Fund (IMF) defines any investment by a firm with an equity share of 10 per cent as FDI. The second attribute of FDI – ownership of technology and know-how – is a much more powerful tool for exercising control over operations than majority equity ownership.

The three attributes of FDI, equity ownership, control over operations and transfer of technology, are intertwined. Ownership of equity and technology enables the firm to exercise control over operations and preserve its monopoly over technology and know-how, which in turn enables it to transfer technology and know-how across frontiers. The essential point to note here is that ownership of equity is a means to an end, the end being control over operations and transfer of technology. If the market for technology were perfect and if technology and knowledge were not public goods, the multinational company – the purveyor of FDI – would prefer to enter international markets by contracting technology-licensing agreements. Such agreements, by definition, are bereft of equity participation: the firm transfers technology to the licenses in return for fixed technical fees

and royalties tied to the profits of the licensee. But because technology can be easily imitated and it is difficult to arrive at a price for most technologies, multinationals engage in FDI, that is, they own equity in entities abroad.

This somewhat long-drawn-out discussion of FDI is to emphasize the fact that it is technology and know-how transfer which is the main attribute of FDI, and it is this attribute which is sought after by countries host to FDI. Admittedly the capital that accompanies technology and know-how is also a benefit to the host countries, but for reasons specified above, transfer of capital is not a major feature of FDI. In any case, the larger the equity share of the foreign firm in an entity, the higher would be the control over operations it exercises, something which most host countries would wish to minimize. Also, relatively poor countries, which have very few of the attributes discussed in the section on determinants, may not be able to attract FDI in large enough volumes to meet their demands for capital.

FDI is also sought for the foreign exchange it provides developing countries, in the form of both the capital that accompanies it and the exports it promotes. The contribution of FDI to China's exports is well known (Wei, 2004), and also to those of other East Asian countries such as Malaysia and Singapore (Driffield et al., 2004). FDI can also save foreign exchange for the host countries with productive investments in import-substituting industries. All of this, of course, makes for a healthy balance of payments of the country. But here again it should be noted that any investment, be it export-oriented or domestic market-oriented, contributes to the balance of payments as long as it is socially productive. This is simply because the balance of payments, as Kindleberger (1969), commenting on the balance of payments effects of FDI reminded us, is a general equilibrium phenomenon. In other words, the balance of payments is an integral part of the total economy. An FDI project which is socially productive, in the sense that the private rate of return to the investment does not exceed the social rate of return, will contribute to the balance of payments. Host countries which offer various fiscal incentives such as tax holidays and the institution of export-processing zones to lure foreign firms may be giving away income to the foreigners if the private rates of returns exceed the social rates, a conceivable outcome in the presence of market distortions which the incentives are supposed to rectify.

In sum the most significant benefit from FDI to the host countries is the technology and know-how it transfers. It is such transfers which augment the skill endowments of host countries and promote employment for local labour. How does the technology which is transferred get transmitted to the local economy? There are several channels, including imitation, acquisition of skills, competition and various sorts of tie-ups between the

foreign-owned and locally owned firms. Imitation of the products produced by foreign affiliates through reverse engineering, an activity that enables local firms to copy the processes and design of new products, is a recognized channel for spillovers. The acquisition of skills occurs mainly through the movement of skilled labour employed by foreign affiliates to locally owned firms. Such internal migration of labour, trained by foreign affiliates, is a significant channel for spillovers. Labour employed in foreign affiliates may wish to set up their own establishments with the experience and skills gained from their sojourn in the foreign affiliates. Also, foreign affiliates may, either in response to performance requirements imposed by the host country or because of distinct cost advantages, train or establish local suppliers of components and parts. This too would be a channel for spillovers.

Another potent channel for spillovers – or, more to the point, growth of productive efficiency – is competition. The theory here is that the entry of foreign affiliates increases competition in the marketplace and locally owned firms are compelled to increase their productive efficiency. This is the sort of efficiency recognized in the literature as ‘X-efficiency’ rather than allocative efficiency. Finally, locally owned firms may learn marketing techniques and methods of penetrating export markets from export-oriented foreign affiliates. This would count as a specific sort of technology transfer.

These propositions have been extensively tested in the context of FDI in developed and developing countries (Haddad and Harrison, 1993; Blomstrom and Kokko, 1998; Borensztein et al., 1998; Gorg and Greenaway, 2001). These econometric studies have produced a mixed bag of results: some identify positive spillovers from the presence of foreign affiliates in manufacturing industries, and others find them to be either negligible or negative.

These studies identify a number of factors that are likely to promote spillovers of technology and know-how from foreign affiliates to locally owned firms. First, the magnitude of spillovers tends to be high in industry segments in which the gap in technological capabilities between foreign affiliates and locally owned firms tend to be narrow. Second, spillovers are likely to be high when the competition in the marketplace between locally owned firms and foreign affiliates tends to be intense. Third, the extent and magnitude of spillovers differ between industries and host countries. Fourth, several studies show that spillovers are proportional to the magnitude of foreign presence, measured by shares of foreign affiliates in total equity or sales of the relevant industry groups. Fifth, local capabilities (including research and development – R&D – and human skills) sustain high levels of spillovers. Finally, analogous to the last proposition, the liberalization of foreign trade, increased competition and development of local infrastructure all promote spillovers.

The message of all this is clear. Increased volumes of FDI alone are unlikely to generate widespread spillovers. In the absence of competition and cooperant factors such as local R&D and human skills, spillovers from FDI may be limited. Put another way, FDI is a catalyst of technical change and growth; it cannot be expected to be the prime mover. Indeed empirical research suggests that FDI is most effective as an agent of change in economies that possess a threshold level of human capital and skills and in those economies that have attained a threshold level of growth (Blomstrom et al., 1994; Balasubramanyam et al., 1999).

In sum, in the absence of the necessary ingredients and cooperant factors, large volumes of FDI alone may not be effective in promoting growth and may even be counterproductive. For these reasons the exuberance relating to the role of FDI in the growth process and exhortations that developing countries should adopt a wide open door policy towards FDI should be tempered by a recognition of the conditions necessary for the effective utilization of FDI.

Conclusions

Recent literature on FDI reflects the substantial change in attitudes towards FDI by the developing countries. In the past, attitudes towards FDI and its role in the development process ranged to extremes – from hostility to ardent advocacy. In the recent past, there is a growing appreciation of its role in the development process and most developing countries have sought to attract increasing volumes of FDI. Research on FDI reflects this change in attitude. Much of the literature now is centred on econometric testing of the determinants and impact of FDI rather than the polemical debates on the role of FDI. Detailed case studies which provide analytical insights into the nature of FDI and its impact on development, of the sort done by Sanjaya Lall (1983), are unfortunately few and far between.

The message of the recent literature, though, is clear. First, FDI is attracted to countries that can provide the sort of environment which allows foreign firms to establish a foothold and successfully exploit the rents in the advantages they possess. The sort of environment which is conducive to the operations of foreign firms is characterized by a stable macro-economic environment with stable FDI regimes and an assured supply of cooperant factors including human capital. Second, FDI is a superb catalyst of development but not an initiator. The new-found enthusiasm for FDI on the part of developing countries is based on the success stories of the East Asian countries with FDI and the drying-up of alternative sources of finance such as bank credit. It is, though, worth noting that the successful utilization of FDI is contingent upon a number of factors discussed in the foregoing. In this context it is worth recalling Paul Streeten's observa-

tion written during the 1970s (Streeten, 1971) when controversy on the role of the multinational enterprise as the purveyor of FDI was at its height: 'it is not sensible to transfer income by attempting to transform the MPE [multinational production enterprise] from what it is – a profit seeking animal – into something it is not – a public service'.

Note

1. The author is grateful to MS Jasleen Sindhu for research assistance with this chapter.

References

- Balasubramanyam, V.N. and M. Salisu (1991), 'Export Promotion, Import Substitution and Direct Foreign Investment in Less Developed Countries', in A. Koekkoek and L. Mennes (eds), *International Trade and Global Development*, London: Routledge, pp. 191–207.
- Balasubramanyam, V.N., D. Sapsford and M.A. Salisu (1996), 'Foreign Direct Investment and Growth in EP and IS Countries', *Economic Journal*, 106: 92–105.
- Balasubramanyam, V.N., D. Sapsford and M.A. Salisu (1999), 'Foreign Direct Investment as an Engine of Growth', *Journal of International Trade and Economic Development*, 8: 27–40.
- Bhagwati, J.N. (1978). *Anatomy and Consequences of Exchange Control Regimes, Vol. 1, Studies in International Economic Relations*, New York: National Bureau of Economic Research.
- Blomstrom, M. and A. Kokko (1998), 'Multinational corporations and spillovers', *Journal of Economic Surveys*, 12 (2): 247–77.
- Blomstrom, M., R.E. Lipsey and M. Zejan (1994), 'What Explains Growth in Developing Countries', NBER Discussion Paper, No. 1924.
- Borensztein, E., J. de Gregorio and J.W. Lee (1998), 'How Does Foreign Investment Affect Economic Growth?' *Journal of International Economics*, 45: 115–35.
- Buckley, P.J. and M. Casson (1991), *The Future of the Multinational Enterprise*, London: Macmillan.
- Driffield, N., R. Clarke and Abdul Halim Mohd Noor (2004), 'Inward Foreign Direct Investment and Industrial Development of Malaysia', in Y. Wei and V.N. Balasubramanyam (eds), *Foreign Direct Investment: Six Country Case Studies*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 74–97.
- Dunning, J.H. (1973), 'The Determinants of International Production', *Oxford Economic Papers*, 25: 289–336.
- Dunning, J.H. (1981), 'Explaining the International Direct Investment Position of Countries: Towards a Dynamic or Developmental Approach', *Weltwirtschaftliches Archiv*, 117 (25): 30–64.
- Gorg, H. and D. Greenaway (2001), 'Foreign Direct Investment and Intra-Industry Spillovers: A Review of the Literature', Research Paper Series No. 137, Centre for Research on Globalisation and Labour Markets Programme, School of Economics, Nottingham University, Nottingham, United Kingdom, mimeo.
- Haddad, M. and A. Harrison (1993), 'Are there Positive Spillovers from Foreign Direct Investment? Evidence from Panel Data for Morocco', *Journal of Development Economics*, 42: 51–74.
- Huang, Y. (2003), *Selling China: Foreign Direct Investment During the Reform Era*, New York: Cambridge University Press.
- Hymer, S. (1976), *The International Operations of National Firms: A Study of Foreign Direct Investment*, Cambridge, MA: MIT Press.
- Kindleberger, C.P. (1969), *American Business Abroad: Six Lectures on Direct Investment*, New Haven, CT: Yale University Press.
- Lall, S. (1983), *Technological Change, Employment Generation and Multinationals: A Case Study of a Foreign Firm and a Local Multinational in India*. Geneva: International Labour Office.

- Markusen, J. (2004), *Multinational Firms and the Theory of International Trade*, Cambridge, MA: MIT Press.
- Nagaraj, R. (2003), 'Foreign Direct Investment in India in the 1990s: Trends and Issues', *Economic and Political Weekly*, 26 April.
- Streeten, P. (1971), 'Costs and Benefits of Multinational Enterprises in Less Developed Countries', in J.H. Dunning (ed.), *The Multinational Enterprise*, London: Allen & Unwin, pp. 240–59.
- UNCTAD (various issues), *World Investment Report*, Geneva.
- Wei, Y. (2004), 'Foreign Direct Investment in China', in Y. Wei and V.N. Balasubramanyam (ed.), *Foreign Direct Investment: Six Country Case Studies*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.
- Williamson, O.E. (1981), 'The Modern Corporation: Origins, Evolution, Attitudes', *Journal of Economic Literature*, 19: 1537–68.

40 Private capital flows and development

Stephany Griffith-Jones

Standard economic theory traditionally argued that international private capital flows will make a major contribution to development to the extent that they will flow from capital-abundant industrialized countries to capital-scarce developing countries, and help to smooth spending throughout the business cycle in capital-recipient countries.

In recent years, reality has contradicted both aspects (WESS, 2005). Between 1997 and 2004, developing countries have transferred a large amount of resources to developed countries. In addition to this, private capital flows to developing countries are highly concentrated in a group of large middle-income countries and are particularly insufficient for low-income and small countries. Secondly, private capital flows to developing countries have been highly volatile and reversible; as a consequence, they have been a major factor in causing developmentally costly currency and financial crises. Rather than smooth domestic expenditure, private capital flows seem to have contributed to making it more volatile.

Boom–bust cycles of capital flows have been particularly damaging for developing countries, when they both directly increase macroeconomic instability and reduce the room for manoeuvre to adopt countercyclical macroeconomic policies, and indeed generate strong biases towards adopting procyclical macroeconomic policies (Kaminsky et al., 2004; Stiglitz and Levy, 2005). Furthermore, there is now overwhelming evidence – accepted by institutions like the International Monetary Fund – that procyclical financial markets and procyclical macroeconomic policies have not encouraged growth and, on the contrary, have increased growth volatility in those developing countries that have integrated to a larger extent into international financial markets (Prasad et al., 2003).

The costs of financial volatility for economic growth are high, as it can generate cumulative effects on capital accumulation (Easterly, 2001). Indeed, major reversals of private flows have led to many developmentally and financially costly crises, which have lowered output and consumption well below what they would have been if those crises had not occurred. Eichengreen (2004) estimated that income of developing countries had been 25 per cent lower since 1980 than it would have been had such crises not occurred, with the average annual cost of the crises being just over \$100 billion. Griffith-Jones and Gottshalk (2006) have estimated a similar

though somewhat higher annual average cost of crises in the period 1995–2002, of \$150 billion in terms of lost gross domestic product (GDP).

These features are by no means inevitable. An appropriate domestic and international environment can improve the capacity of developing countries to benefit from private capital flows. In what follows we examine both characteristics of private capital flows to developing countries and some policy options that would improve their development impact.

Main characteristics of private flows

The volatility and reversibility of capital flows to emerging countries and the marginalization of many of the poorer and smaller developing economies with respect to financial markets are rooted in the combination of financial market failures and basic asymmetries in the world economy (Ocampo, 2001).

Instability is inherent in the functioning of financial markets (Keynes, 1936; Minsky, 1982). Indeed, boom–bust patterns in financial markets have occurred for centuries (Kindleberger, 1978). The basic reason for the existence of these patterns is that finance deals with future information that, by its very nature, is not known in advance; therefore, opinions and expectations about the future rather than factual information dominate financial market decisions. This is compounded by asymmetries of information that characterize financial markets (Stiglitz, 2000). Owing to the non-existence or the large asymmetries of information, financial agents rely to a large extent on the ‘information’ provided by the actions of other market agents, leading to interdependence in their behaviour, that is to say, contagion and herding. At the macroeconomic level, the contagion of opinions and expectations about future macroeconomic conditions tends to generate alternating phases of euphoria and panic. At a microeconomic level, it can result in either permanent or cyclical rationing of lending to market agents that are perceived by the market as risky borrowers. In many cases it is the endogenous behaviour of international financial markets that conditions or strongly influences fundamentals in developing countries. A supply-led, large capital inflow affects the domestic economic situation (for example by generating an asset price bubble or an overvalued exchange rate) in a way that can increase inflows. This can lead to costly macroeconomic crises, which makes regulation and other state intervention in international financial markets essential.

Herding and volatility seem to be accentuated by some features of the functioning of modern markets. An important element in the increased volatility of international bank lending is the use of modern risk management models (such as Value at Risk). As Persaud (2003) points out, the intrinsic problem with market-sensitive risk management systems is that

they incorrectly assume that banks act independently when in fact their decisions are interconnected. When many banks try to sell the same asset at the same time, and there are few or no buyers, prices fall and volatility increases. As prices collapse, for liquidity reasons banks try to sell another asset, which may have been previously uncorrelated with the first. This increases the volatility of the second asset and also correlation. This prompts repeated rounds of selling among agents who use similar models, and generalized herding takes place. The adoption of banks' own risk management models to determine their required levels of capital in the internal ratings approach, as proposed in the new Basel Capital Accord, could seriously increase banks' tendency for procyclicality in lending, exacerbating both booms and crashes.

An additional source of concern is the evidence that the Value at Risk (VaR) models first developed by banks are being extensively adopted by fund managers and pension funds, leading to similar herding patterns and to procyclicality in their investment. Therefore, herding is not restricted to one class of actor, but is spreading among many actors. The increasing use of similar market-sensitive risk management techniques (Persaud, 2000) and the dominance of investment managers aiming for very short-term profits, evaluated and paid at very short-term intervals (Griffith-Jones, 1998; Williamson, 2003), seem to have increased the frequency and depth of boom–bust cycles. The downgrade by a rating agency or any other new information available to investors may lead them to sell bonds and stop banks from lending to specific markets; simultaneously, reduced liquidity – owing, for example, to margin calls associated with derivative contracts in these markets – or contagion of opinions about the behaviour of different market segments that are believed to be correlated with a market facing a sell-off, will lead market agents to sell other assets or to stop lending to other markets. Through these and other mechanisms, contagion spreads both across countries and across different flows.

Different types of capital flows are subject, however, to different volatility patterns. In particular, the higher volatility of short-term capital indicates that reliance on such financing is highly risky (Rodrik and Velasco, 1999), whereas the smaller volatility of FDI *vis-à-vis* all forms of financial flows is considered a source of strength. However, even FDI does have volatile components. A particularly recent concern is that multinational companies, especially those selling in domestic markets, hedge their foreign exchange rate risk. This is particularly problematic when such hedging is done far more when a major devaluation is likely, as this will put additional pressure on the exchange rate and on the reserves. Naturally, such risks tend to become less important as national financial development deepens.

Capital account cycles involve short-term fluctuations, such as the very intense movements of spreads and interruption (rationing) of financing. These phenomena were observed during the Asian and, particularly, during the Russian crisis. However, and perhaps more importantly, they also involve medium-term fluctuations, as the experience since 1980 indicates. During those decades, the developing world experienced two such medium-term cycles that left strong imprints on the growth rates of many countries: a boom of external financing (mostly in the form of syndicated bank loans) in the 1970s, followed by a debt crisis in a large part of the developing world in the 1980s, and a new boom in the 1990s (now mostly portfolio flows), followed by a sharp reduction in net flows since the Asian crisis.

Improved economic conditions in developing countries, as well as the higher global growth, drove a recovery of private capital flows to developing countries in 2003, 2004 and 2005, perhaps signalling the beginning of a new cycle.

More importantly, net transfers of financial resources from developing countries have not experienced a positive turnaround and, on the contrary, continued to deteriorate in 2004 for the seventh year in a row, reaching an estimated \$350 billion in 2004 (see Table 40.1). Periods of negative net transfers of financial resources from developing countries (especially from Latin America) have been frequent throughout history; indeed, Kregel (2004) provides evidence that these negative net transfers have been the rule rather than the exception.

Recently, these large and increasing net transfers of financial resources are explained by the combination of relatively low net financial flows and accumulation of very large foreign exchange reserves. Indeed, the most significant aspect of the net outflows from developing countries in recent years has been the growth in official reserves, particularly in Asia (Table 40.1). Accumulation of reserves initially had a large component of 'self-insurance' against financial instability, a rational decision of individual countries in the face of the limited 'collective insurance', often accompanied by what countries see as undesirable conditionality, provided by the international financial system. However, reserve accumulation in Asia has now clearly exceeded the need in several countries for self-insurance, raising increasing questions about the balance of costs and benefits of additional accumulation, especially if such reserves are invested in low-yielding assets and particularly in a currency, the United States dollar, that may at some point fall quite sharply. At a more fundamental level, the fact that countries like China and India, with very low levels of income per capita and large numbers of poor people – even though they have such dynamic growth – are transferring significant resources to finance developed countries, and especially the US, contradicts theory and is ethically undesirable.

Table 40.1 Net transfer of financial resources to developing countries and economies in transition, 1993–2004 (US\$40 billion)

	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004
Developing countries	69.3	35.8	42.9	19.9	-5.2	-37.9	-127.4	-186.5	-153.7	-205.5	-274.8	-353.8
Africa	1.1	4.0	6.4	-5.8	-4.7	15.6	4.3	-26.2	-14.7	-5.6	-20.2	-32.8
Sub-Saharan (excluding Nigeria and South Africa)	8.6	6.7	7.4	5.3	7.5	12.1	9.1	3.0	7.9	6.4	6.5	3.9
Eastern and Southern Asia	18.7	1.0	22.1	18.5	-31.1	-128.2	-142.7	-121.3	-113.1	-142.1	-147.5	-167.8
Western Asia	33.1	13.2	15.6	5.3	6.2	28.5	-0.9	-39.1	-32.0	-26.7	-47.6	-79.9
Latin America	16.4	17.7	-1.2	1.8	24.5	46.2	11.8	0.1	6.1	-31.1	-59.5	-73.4
Economies in transition	1.8	-3.9	-2.3	-6.2	2.7	3.0	-24.0	-48.8	-30.5	-27.0	-34.4	-57.6
Memorandum item: Heavily indebted poor countries (HIPC)	8.5	7.1	6.3	6.8	7.1	8.6	10.1	8.8	8.8	9.9	10.6	11.3

Sources: UN/DESA, WESS (2005).

As private flows recover, an important question for policy-makers in developing countries is whether they will be sufficient and particularly more stable and less reversible than in the past. In this regard, the dominant role of FDI and the fact that it has been relatively stable in times of crises, are positive. However, not all components of FDI are equally stable. Furthermore, multinational companies, especially those producing for the local market, increasingly hedge their short-term foreign exchange risks, particularly when devaluations seem likely. This can lead to major temporary outflows of capital and significant pressure on exchange rates (Ffrench-Davis and Griffith-Jones, 2003; Persaud, 2003). More generally, the increasing use of financial engineering and of derivatives (as well as the growing scale and complexity of derivatives discussed below) seems to make the hypothesis of a hierarchy of volatility, whereby some categories of flows are more stable than others, less clear-cut.

Another potentially positive effect is the greater interest shown by institutional investors (such as life insurers) in investing in emerging countries (European Central Bank, 2005). However, the large rise in ‘carry trade’ – that is to say, investment in high-yielding emerging-market instruments using debt raised at lower cost in mature markets – makes those flows vulnerable to narrowing of interest rate differentials. Furthermore, the large fall in emerging countries’ bond spreads during 2004–05 (while naturally positive in itself for borrowing countries) has raised concerns that this increases the vulnerability of developing countries to international changes.

Finally, there are two structural trends that may add stability. The first is attested by the greater importance of local currency bond markets in developing countries; the second by the fact that international banks have increasingly ‘crossed the border’, lending from their local branches in local currency, and usually fund themselves via domestic deposits. This makes countries less vulnerable to crises, although it also implies that foreign banks are contributing less – or no – foreign savings.

At the time of writing in 2005 there were thus mixed signs in respect of whether the new inflows will be more stable than in the past. Therefore, policy efforts must be made, both in source and in recipient countries, to encourage more stable flows and discourage large flows that are potentially more reversible.

Measures to counter procyclicality of private flows

To counter the boom–bust pattern that characterizes private capital flows, several options are available. We here consider two: (1) designing mechanisms to encourage more stable private flows (countercyclical guarantees) or that distribute better the risk faced by developing countries throughout the business cycle (indexed bonds and bonds denominated in the currency

of developing countries); and (2) introducing prudential regulations on the capital account. We also consider the likely effect of the New Basel Capital Accord (Basel II) on patterns of capital flows to developing countries. The procyclical pattern of private capital flows gives a compensatory role also to official financing, in relation to official development financing and to emergency (balance-of-payments) financing, respectively (for a discussion of these, see for example Griffith-Jones and Ocampo, 2003 and WESS, 2005, Chapters III and VI).

Countercyclical financing instruments

One way of addressing the problems created by the inherent tendency of private flows to be procyclical is for public institutions to issue guarantees that have countercyclical elements (Griffith-Jones and Fuzzo de Lima, 2004). In this regard, multilateral development banks and export credit agencies could introduce explicit countercyclical elements in the risk evaluations they make for issuing guarantees for lending to developing countries. This would imply that when banks or other private lenders lowered their exposure to a country, multilateral development banks or export credit agencies would increase their level of guarantees, if they considered that the country's long-term fundamentals were basically sound. When private banks' willingness to lend increased, multilateral development banks or export credit agencies could reduce their exposure.

There have also been proposals to introduce GDP-indexed bonds. The coupon payments on these bonds would vary in part with the growth rate of the debtor's economy, being higher in years of rapid growth of GDP (measured in an international currency) and lower in years of below-trend growth. It has been argued that such instruments would improve the cushioning of emerging-market borrowers against adverse shocks by making debt payments more contingent on the borrower's ability to pay. GDP-indexed bonds would therefore restrict the range of variation of the debt-to-GDP ratio and hence reduce the likelihood of debt crises and defaults. At the same time, they would also reduce the likelihood of procyclical fiscal policy responses to adverse shocks (Griffith-Jones and Sharma, 2006).

Another alternative for better managing the risks faced by developing countries throughout the business cycle consists in the introduction of local currency-denominated bonds. These bonds offer, in particular, a cure against the currency mismatches that characterize the debt structure of developing countries. At the domestic level, the development of domestic capital markets, especially bond markets, also creates a more stable source of local funding for both the public and private sectors, thereby mitigating the funding difficulties created by sudden stops in cross-border capital. In addition to proposals for institutional measures to develop local capital

markets, there have also been innovative proposals to make local currency investments more attractive to international investors. Spiegel and Dodd (2004) have suggested raising capital in international markets by forming diversified portfolios of emerging-market local currency debt issued by sovereign governments. These portfolios of local currency government debt securities would employ risk management techniques of diversification to generate a return-to-risk that competed favourably with other major capital market security indices.

Prudential capital account regulations

The accumulation of risks that developing countries face during capital account booms depends not only on the magnitude of private and public sector debts but also on maturity and currency mismatches on the balance sheets of financial and non-financial agents. Thus, capital account regulations potentially have a dual role: as a macroeconomic policy tool with which to provide some room for countercyclical monetary policies that smooth out debt ratios and spending; and as a 'liability policy' designed to improve private sector external debt profiles (Ocampo, 2003).

Overall, the experiences with capital account regulations in the 1990s were useful for improving debt profiles, giving governments more latitude in pursuing stabilizing macroeconomic policies, and insulating countries from some of the vagaries of capital markets. There is much evidence that, if well implemented, the benefits far outweigh the costs (Stiglitz and Levy, 2005; Ocampo and Palma, 2005).

One type of capital account regulations are price-based regulations. The basic advantages of price-based instruments are their simplicity and their focus on averting the build-up of macroeconomic disequilibria and, ultimately, preventing crises. A highly significant innovation in this sphere during the 1990s was the establishment in Chile and Colombia of an unremunerated reserve requirement for capital inflows.

It is noteworthy that institutions such as the International Monetary Fund and the Bank for International Settlements have increasingly concluded that these controls were effective in important aspects. There is broad agreement that they were effective in reducing short-term debt flows and thus in improving or maintaining good external debt profiles. There is greater controversy about their effectiveness as a macroeconomic policy tool. Nonetheless, it can be asserted that reserve requirements helped countries maintain higher domestic interest rates during periods of euphoria in international financial markets.

On the other hand, quantity-based capital account regulations might be preferable when the policy objective is to reduce significantly domestic macroeconomic sensitivity to international capital flows.

The experience of the Asian countries that maintained quantity-based restrictions throughout the 1990s suggests that those restrictions might indeed also be particularly effective in preventing crises. China, India, Taiwan Province of the Republic of China and Vietnam offer successful examples in this regard. Despite the slow and cautious liberalization that has taken place in several of these economies since the early 1990s, the use of such traditional regulations has helped them prevent contagion from the East Asian crisis (see for example, in relation to India, Reddy, 2000).

Malaysia offers an interesting example of the effective use of quantitative regulations during the 1990s. Kaplan and Rodrik (2001) and others provide evidence that Malaysian regulations during the Asian crisis gave the government space within which to enact expansionary monetary and fiscal policies that contributed to the speedy recovery of economic activity.

Although quantity-based restrictions can be effective if authorities wish to limit capital outflows during crises, crisis-driven quantitative controls generate serious credibility issues and may be ineffective in the absence of a strong administrative capacity. A tradition of regulation may be necessary, and the tightening or loosening of permanent regulatory regimes through the cycle may be superior to the alternation of different (even opposite) capital account regimes.

It should be emphasized that capital account regulations should always be seen as an instrument that provides an additional degree of freedom to the authorities with respect to their adopting sensible counter-cyclical macroeconomic policies, but never as a substitute for those policies.

Basel II and developing countries

The right regulatory and supervisory regime is essential for maintaining domestic financial stability. In a globalized economy, some common standards of regulation and supervision may be also essential to guarantee global financial stability. This has been the major motivation behind the principles adopted by the Basel Committee on Banking Supervision in recent decades. The second generation of these standards (Basel II), agreed to in June 2004, takes a further step in aligning regulatory capital with the risks in bank lending, and in adapting regulations to the complexities of risk management.

There are fears that Basel II creates the risk of a sharp reduction in bank lending to developing countries, and of an increase in the cost of a significant part of the remaining lending, particularly in the case of low-rated borrowing countries. An equal cause for concern is the danger that Basel II will accentuate the procyclicality of bank lending, which is damaging for all economies, but particularly so for fragile developing ones, which are more vulnerable to strong cyclical fluctuations of financing.

Indeed, the proposed internal ratings-based (IRB) approach of Basel II overestimates the risk of international bank lending to developing countries, primarily because it does not appropriately reflect the clear benefits of international diversification. However, there is a great deal of evidence that by failing to take account of the benefits of international diversification at the portfolio level, capital requirements for loans to developing countries will be significantly higher than is justified on the basis of the actual risks attached to this lending (see, for example, Griffith-Jones et al., 2003).

Therefore, one clear way in which Basle II could be improved so as to reduce the negative and technically incorrect effects on developing countries would be to introduce the benefits of diversification into the internal ratings-based approach. One of the major benefits of investing in developing and emerging economies is their relatively low correlation with mature markets. This hypothesis was tested empirically using a wide variety of financial, market and macro variables (Griffith-Jones et al., 2004a). Every statistical test performed showed that the correlation between developed markets only was higher, in every case, than that between developed and developing markets.

An additional positive effect of taking account of the benefits of diversification is that this makes capital requirements far less procyclical than they otherwise would be. Indeed, if the benefits of diversification are incorporated, simulations show that the variance over time of capital requirements will be significantly smaller than if these benefits are not incorporated. Therefore, introducing the benefits of geographical diversification significantly decreases, though it does not eliminate, the higher procyclicality that the internal ratings-based approach implies. This difference may well be significant enough to prevent a 'credit crunch'.

However, even if the benefits of diversification are incorporated, the internal ratings-based approach will still be more procyclical than the standardized approach, which is closer to the principles of the first Basel Capital Accord (Basel I). Therefore, as well as introducing the benefits of diversification, it seems desirable to introduce countercyclical measures (for example, countercyclical provisioning against losses) at the same time as Basel II is implemented.

We can conclude that several measures can be taken to reduce boom–bust patterns of private flows, and thus enhance their contribution to development. However, given that the risk of costly crises will remain – even if such measures are introduced – maintaining and improving the supply of countercyclical official liquidity and development finance is essential.

Acknowledgement

The author would like to acknowledge the excellent assistance by Carmen Seekatz.

References

- Easterly, William (2001), *The Elusive Quest for Growth: Economists' Adventures and Misadventures in the Tropics*, Cambridge, MA, and London: MIT Press.
- Eichengreen, Barry (2004), 'Financial instability', in Bjorn Lomborg (ed.), *Global Crises, Global Solutions*, Cambridge and New York: Cambridge University Press.
- European Central Bank (2005), 'Financial Flows to Emerging Market Economies: Changing Patterns and Recent Developments', Monthly Bulletin (January 2005), Frankfurt am Main, Germany, pp. 59–73.
- Ffrench-Davis, Ricardo, and Stephany Griffith-Jones (2003), *From Capital Surges to Drought: Seeking Stability for Emergency Economies*, UNU-WIDER Studies in Development Economics and Policy Series, Basingstoke: Palgrave Macmillan.
- Griffith-Jones, Stephany (1998), *Global Capital Flows: Should They be Regulated?*, Basingstoke and New York: Macmillan and St. Martin's Press.
- Griffith-Jones, Stephany and Ana Fuzzo de Lima (2004), 'Alternative Loan Guarantee Mechanisms and Project Finance for Infrastructure in Developing Countries', in Inge Kaul et al. (eds), *The New Public Finance, Responding to Global Challenges*, New York: Oxford University Press.
- Griffith-Jones, Stephany and Ricardo Gottschalk (2006), 'Costs of Currency Crises and Benefits of International Financial Reform', currently available from www.stephanygj.com.
- Griffith-Jones, Stephany and José Antonio Ocampo (2003), 'What Progress on International Financial Reform? Why so Limited?', Prepared for the Expert Group on Development Issues, Sweden.
- Griffith-Jones, Stephany and Krishnan Sharma (2006), UN DESA Discussion Paper, ST/ESA/2006/DWP/21, New York, April 2006, http://www.un.org/esa/desa/papers/2006/p2006_54GDP-Indexed Bonds: Making It Happen.
- Griffith-Jones, Stephany, Miguel Angel Segoviano and Stephen Spratt (2003), 'Submission to the Basel Committee on Banking Supervision: CP3 and the Developing World', Institute of Development Studies, University of Sussex, Brighton, United Kingdom, available from www.stephanygj.com.
- Kaminsky, Graciela L., Carmen M. Reinhart and Carlos A. Végh (2004), 'When it rains, it pours: procyclical capital flows and macroeconomic policies', NBER Working Paper, No. 10780, Cambridge, MA: National Bureau of Economic Research.
- Kaplan, Ethan, and Dani Rodrik (2001), 'Did the Malaysian Capital Controls Work?', John F. Kennedy School of Government, Harvard University, Cambridge, MA.
- Keynes, John Maynard (1936), 'General Theory of Employment, Interest and Money', in Donald Moggridge (ed.), *Collected Writings of John Maynard Keynes, vol. 7: The General Theory*, London: Macmillan, for the Royal Economic Society, 1973.
- Kindleberger, Charles P. (1978), *Manias, Panics, and Crashes: A History of Financial Crises*, New York: Basic Books.
- Kregel, Jan (2004), 'External Financing for Development and International Financial Instability', G-24 Discussion Paper Series, No. 32 (October 2004), United Nations Conference on Trade and Development and Intergovernmental Group of Twenty-four, New York and Geneva: United Nations.
- Minsky, Hyman P. (1982), *Inflation, Recession and Economic Policy*, Sussex: Wheatsheaf Books.
- Ocampo, José Antonio (2001), *International Asymmetries and the Design of the International Financial System*. Serie Temas de Coyuntura, N° 15, Sales N° E, 01.11.6.70, Santiago: Economic and Social Commission for Latin America and the Caribbean, p. 23.
- Ocampo, José Antonio (2003), *Capital-Account and Counter-Cyclical Prudential Regulations in Developing Countries*, United Nations Publications.
- Ocampo, José Antonio and José Gabriel Palma (2005), 'Dealing with Volatile External Finances at Source: The Role of Preventive Capital Account Regulations', Initiative for Policy Dialogue, Columbia University, New York.
- Persaud, Avinash (2000), 'Sending the Herd off the Cliff Edge: The Disturbing Interaction Between Herding and Market-sensitive Risk Management Practices', Institute of

- International Finance Competition in Honour of Jacques de Larosiere, First Prize Essay on Global Finance for 2000, Washington, DC.
- Persaud, Avinash (2003), 'Liquidity Black Holes: Why Modern Financial Regulation in Developed Countries is Making Short-term Capital Flows to Developing Countries Even More Volatile', in Ricardo Ffrench Davis and Stephany Griffith-Jones (eds), *From Capital Surges to Drought: Seeking Stability for Emergency Economies*, UNU-WIDER Studies in Development Economics and Policy Series. Basingstoke: Palgrave Macmillan.
- Prasad, E., K. Rogoff, S. Wei and M. Kose (2003), *Effects of Financial Globalization on Developing Countries*, International Monetary Fund, Washington, DC.
- Reddy, Y.V. (2000), 'Capital Account Liberalisation: The Developing Country Perspective', Overseas Development Institute, London.
- Rodrik, Dani and Andrés Velasco (1999), 'Short-term Capital Flows', NBER Working Paper, No. 7364, Cambridge, MA: National Bureau of Economic Research.
- Spiegel, Shari, and Randall Dodd (2004), 'Up from Sin: A Portfolio Approach to Salvation. A Study for Presentation to the XVIII G24 Technical Group Meeting', Palais des Nations, Geneva, 8–9 March.
- Stiglitz, Joseph E. (2000), 'The Contributions of the Economics of Information to Twentieth Century Economics', *Quarterly Journal of Economics*, November, pp. 1441–78.
- Stiglitz, Joseph E. and A. Levy (2005), 'Liberalización de los mercados de capitales, Globalización y el FMI Desarrollo Económico' 45 (177) Apr–Jun, 2005, pp. 3–23.
- UN's World Economic and Social Survey*, 2005.
- Williamson, John (2003), 'Proposals for curbing the boom-bust cycle in the supply of capital to Emerging Markets', in Robert Ffrench-Davis and Stephany Griffith-Jones (eds), *From Capital Surges to Drought: Seeking Stability for Emergency Economies*, UNU-WIDER Studies in Development Economics and Policy Series, Basingstoke: Palgrave Macmillan.

41 International capital flows to emerging economies: short- and long-run effects

*Graciela L. Kaminsky*¹

Introduction

The explosion of capital flows to emerging markets in the early and mid-1990s and their reversal following the crises in Asia, Latin America, and the transition economies have reignited a heated debate on the benefits and drawbacks of financial globalization. Many have argued that globalization has gone too far and that international capital markets have become extremely erratic, with ‘excessive’ booms and busts in capital flows triggering bubbles and financial crises and magnifying the business cycle. In contrast, the traditional view asserts that international capital markets enhance growth and productivity by allowing capital to flow to its most attractive destination.

Even if international capital flows do not trigger excess volatility in domestic financial markets, it is still true that large capital inflows can spark off inflation in the presence of a fixed exchange rate regime. Moreover, transitory capital inflows may distort relative prices, with the domestic economy losing competitiveness as a result of the appreciation of the real exchange rate. Therefore, it is no wonder that policy-makers have used a variety of tools to manage these flows, especially flows of the ‘hot money’ type.

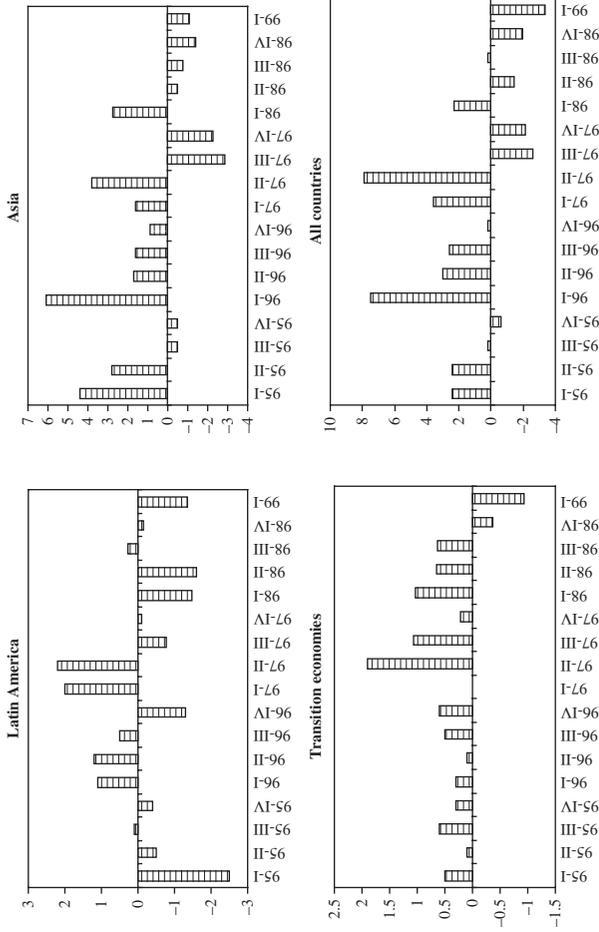
This chapter re-examines the evidence on the characteristics of international capital flows to emerging economies, with particular attention to portfolio equity flows and bank lending around the time of the crises of the 1990s. The results suggest that episodes of surge in capital inflows do, in fact, end abruptly – whether owing to home-grown problems or contagion from abroad. This chapter also reviews the evidence on the short- and long-run effects of financial deregulation on financial and real cycles. Interestingly, the stylized evidence suggests that although financial liberalization may trigger excessive booms and busts in the short run, financial markets tend to stabilize and growth accelerates in the long run, in part because financial globalization seems to trigger institutional reform. The conclusion summarizes what we know about financial globalization and examines policy options.

The behavior of mutual funds

The booms and busts in international capital flows have brought international investors into the limelight. In this section, I examine the behavior of equity mutual funds in emerging markets. I use the information on portfolio allocations provided by Emerging Market Funds Research, Inc., which covers the positions of nearly 1400 international emerging market equity funds, with an average position of about US\$120 billion in 1996. It includes United States registered and offshore funds as well as funds registered in Luxembourg, the United Kingdom of Great Britain and Northern Ireland, Ireland, the Cayman Islands, Canada and Switzerland. Both open- and closed-end funds are also included in this data set, which starts at 1995.

Figure 41.1 shows the average quarterly net flows to these regions from 1995 to 1999. Mutual fund flows to emerging markets peaked in the second quarter of 1997, reaching about US\$8 billion. Overall, booms in mutual fund flows were followed by reversals. Reversals were not persistent after the 'Tequila crisis'. Outflows from Latin America reached about US\$4 billion in 1995, but mutual funds increased their positions in Latin America by about US\$2 billion in the first half of 1996. The Tequila crisis did not have any spillovers in Asia or in transition economies. In fact, flows to Asia ballooned to almost US\$11 billion in 1996, while flows to transition economies remained stable throughout 1995–96. The picture changed after the Asian crisis. This time, mutual funds pulled out not only from Asia but also from Latin America, with net outflows in the latter region reaching about US\$1 billion in the six months following the collapse of the Thai baht. Mutual fund withdrawals took a turn for the worse in 1998, reaching about US\$4 billion in Asia and also in Latin America, with substantial outflows from transition economies after the Russian crisis.

Figure 41.2 assesses the problem of the sudden stops in times of financial turmoil. It reports the average quarterly flows (as a percentage of the mutual funds' initial positions) to countries in Asia and Latin America, as well as to transition economies in the two quarters following three crises. The top panel looks at the aftermath of the Mexican devaluation in December 1994, the middle panel examines the aftermath of the collapse of the Thai baht in July 1997, and the bottom panel studies the aftermath of the Russian devaluation and moratorium in August 1998. To capture the magnitude of the sudden-stop syndrome, this figure reports total flows relative to average flows (also as percentages of their initial positions) during the whole sample (1995–99). Following the Mexican devaluation, for example, mutual funds sold about 5 percent of their Brazilian positions (relative to their average quarterly buying/selling from 1995 to 1999). Thus, as shown in the first panel in Figure 41.2, Brazil experienced unusual withdrawals of about 5 percent in the aftermath of the Mexican devaluation. As shown in the last

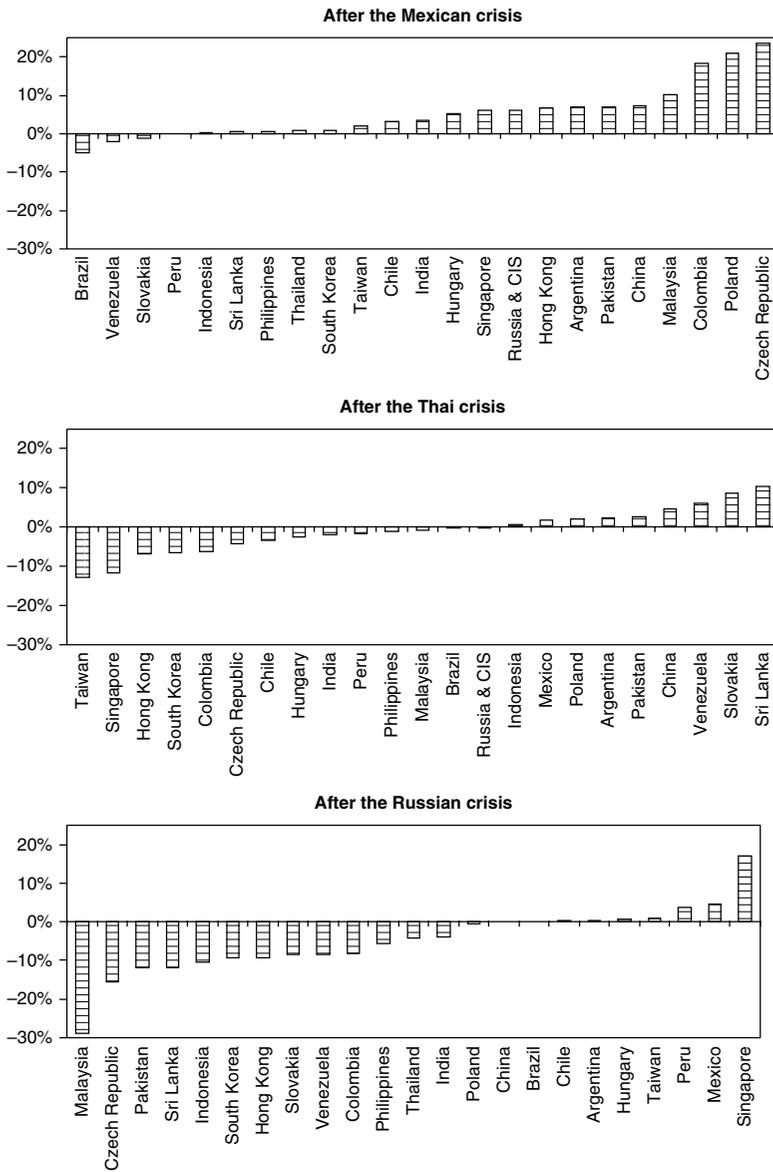


Notes:

Latin America includes Argentina, Brazil, Chile, Colombia, Mexico, Peru and Venezuela. Asia includes China, Hong Kong, India, Indonesia, Malaysia, Pakistan, the Philippines, Singapore, South Korea, Sri Lanka, Taiwan and Thailand. Transition economies include Armenia, Azerbaijan, Belarus, the Czech Republic, Georgia, Hungary, Kazakhstan, Kyrgyzstan, Moldova, Poland, Russia, Slovakia, Tajikistan, Turkmenistan, Ukraine and Uzbekistan.

Source: Kaminsky et al. (2002).

Figure 41.1 Mutual funds: quarterly flows to emerging countries (billions of dollars)



Source: Kaminsky et al. (2002).

Figure 41.2 *Mutual fund flows: global spillovers*

panel, Malaysia was the country most affected in the aftermath of the Russian crisis, with abnormal outflows of approximately 30 percent.

The extent of the mutual fund sudden stop in the aftermath of the three crises was substantially different. The so-called Tequila crisis was circumscribed to Latin America. Moreover, 'abnormal' mutual fund withdrawals in the aftermath of the collapse of the Mexican peso were confined to a handful of Latin American countries, with only Brazil and the Bolivarian Republic of Venezuela – besides the crisis country, Mexico – suffering average withdrawals of 5 and 2 percent, respectively, in the two quarters following the devaluation. In contrast, mutual funds increased their exposure to Asian countries and transition economies, with (above-trend) flows oscillating around 4 percent for Asia and 11 percent for the transition economies.

The aftermath of the collapse of the Thai baht presents a different picture of the international mutual funds industry. It is in this episode that we first observe signs of a more general retrenchment of mutual funds in emerging markets. Mutual fund flows to Asian economies were well below trend in the two quarters following the collapse of the Thai baht. Only flows to China, Pakistan and Sri Lanka were above average. Interestingly, after the collapse of the Thai baht, we observe substantial withdrawals from Hong Kong Special Administrative Region, Singapore and Taiwan Province of the People's Republic of China, with average quarterly withdrawals oscillating at about 12 percent above average in the case of Singapore and Taiwan and about 7 percent for Hong Kong. The retrenchment this time also affected Latin America and the transition economies, with withdrawals reaching about 6 percent for Colombia and 4 percent for the Czech Republic during the two quarters following the outbreak of the Thai crisis. Colombia, the Czech Republic, Chile, Hungary and Peru were the countries most affected in this episode, with sales averaging about 3 percent above average.

The flight away from emerging markets became more pronounced during the Russian crisis, with about half of the countries in the sample experiencing abnormal sales of about 10 percent or even larger. In some cases, withdrawals were massive. For example, average mutual funds sales (relative to trend) in Malaysia reached 30 percent while those in the Czech Republic were in the order of 16 percent. Some Latin American countries were also dramatically affected in the aftermath of the Russian collapse. Colombia and Venezuela, for example, suffered average quarterly outflows of about 8 percent. Mutual funds investments in Mexico and Peru were the only ones that did not suffer following the worldwide turmoil triggered by the Russian default. In fact, inflows to Mexico were 5 percent above the average observed in the 1995–99 period.

Table 41.1 examines in detail why some countries were severely affected by mutual fund withdrawals while others were left unscathed. Three factors

Table 41.1 *The behavior of mutual funds during crises*

The Mexican crisis			
Region	Percentage of countries with		
	Fragility	Liquid financial markets	Risk
<i>Asia</i>			
With MF withdrawals
Without MF withdrawals	0	42	25
<i>Latin America</i>			
With MF withdrawals	67	33	67
Without MF withdrawals	0	67	33
<i>Transition economies</i>			
With MF withdrawals	..	0	0
Without MF withdrawals	33	75	50
The Thai crisis			
Region	Percentage of countries with		
	Fragility	Liquid financial markets	Risk
<i>Asia</i>			
With MF withdrawals	43	86	29
Without MF withdrawals	25	0	25
<i>Latin America</i>			
With MF withdrawals	75	50	25
Without MF withdrawals	0	100	0
<i>Transition economies</i>			
With MF withdrawals	100	100	33
Without MF withdrawals	0	50	0
The Russian crisis			
Region	Percentage of countries with		
	Fragility	Liquid financial markets	Risk
<i>Asia</i>			
With MF withdrawals	40	40	60
Without MF withdrawals	0	100	0
<i>Latin America</i>			
With MF withdrawals	50	100	0
Without MF withdrawals	20	60	0

Table 41.1 (continued)

Region	The Russian crisis		
	Percentage of countries with		
	Fragility	Liquid financial markets	Risk
<i>Transition economies</i>			
With MF withdrawals	50	33	0
Without MF withdrawals	100	0	100

Notes:

This table relates the mutual fund (MF) withdrawals (injections) of funds to the emerging markets shown in Figure 41.4 with indicators of fragility, liquidity of financial markets, and economic and political risk in those economies.

.. Data not available.

Source: Kaminsky et al. (2002).

are examined: economic fragility, liquidity of financial markets² and economic and political risk. Economic fragility is captured using the probabilities of crises in Kaminsky (1998) that measure the likelihood of crises conditional on 18 indicators reflecting macroeconomic vulnerabilities in each country. These indicators provide information about fiscal and monetary imbalances, financial and real vulnerabilities, current account and capital account problems, and world factors. For Table 41.1, I classify an economy as fragile if the probability of a crisis is higher than 50 percent; otherwise it is considered healthy.

Liquidity is captured using four indicators. The first one – the volume traded in the stock market – provides an overall measure of the size and depth of the stock market. The second one – the share of the mutual funds portfolio in each country at the onset of the crisis – is related to mutual funds liquidity in each country, since investors cannot sell in countries in which they have basically no exposure. These first two indicators provide two different pictures of liquidity of financial markets. The third indicator dates the time when firms in emerging markets start to trade in mature and more liquid financial markets. The fourth indicator captures the ability of investors to change their portfolio rapidly in a particular country. In particular, this last indicator evaluates the extent of restrictions to capital mobility in each country. Restrictions could add ‘sand in the wheels’ of capital markets and thus curtail liquidity.³

Finally, the risk indicator captures both political and economic uncertainty. The political risk indicator captures uncertainty due to expected changes of authorities or future policy actions, and it also identifies

widespread social unrest. In particular, it includes major changes in the political arena or events of political instability that took place six months before and after the crisis. The risk indicator also captures economic risk, such as imposition of restrictions to capital mobility in response to crises. A country is classified as risky when there is at least either political or economic risk.

Table 41.1 shows the characteristics of countries that suffer abnormal withdrawals and injections in the aftermath of the three crises.⁴ The table groups the countries into three regions: Asia, Latin America and transition economies. As shown in the first column, countries with fragile economies constitute the bulk of the countries that suffer withdrawals. During the Mexican crisis, for example, Latin America was the only region that suffered withdrawals. Interestingly, 67 percent of the countries that suffered withdrawals in this episode were also countries with deteriorated fundamentals. Again, during the Thai crisis, at least 75 percent of the countries that suffered withdrawals in the transition economies group and Latin America were countries with economic vulnerabilities. Similarly, 43 percent of the Asian countries affected by abnormal withdrawals also had deteriorated economies. The Republic of Korea (South Korea), Colombia, the Czech Republic and Chile, for example, suffered huge withdrawals in the aftermath of the Thai crisis – the Czech Republic and South Korea were the two most vulnerable countries during the Asian crisis (Thailand ranked fourth) in the sample of 25 countries, while Colombia ranked sixth. In contrast, countries that did not experience mutual fund withdrawals were less fragile in general (see Goldstein et al., 2000).

Domestic fragilities were not the only explanation for the sudden-stop syndrome, however. China, for example, did not even suffer a mild hiccup in the midst of the Asian crisis, even when devaluation fears were widespread among investors and the vulnerability of its financial system was widely known. In contrast, Singapore, Taiwan and Hong Kong – countries with the most liquid financial markets in the region – suffered pronounced capital flow reversals even when their economies looked far healthier than that of China. Overall, 86 percent of the countries in the Asia-Pacific region that suffered withdrawals were countries with quite liquid financial markets. In contrast, all the countries in that region unaffected by the Thai crisis had illiquid financial markets.

Finally, risk also had an important role, with 40 percent of the countries most affected by withdrawals also experiencing political and economic risk. In 1994, for example, in the midst of the banking crisis, Venezuela abandoned convertibility. Far from discouraging capital outflows, the implementation of restrictions to capital mobility seems to have also contributed to the fire sales of Venezuelan assets. Similarly, Malaysia suffered

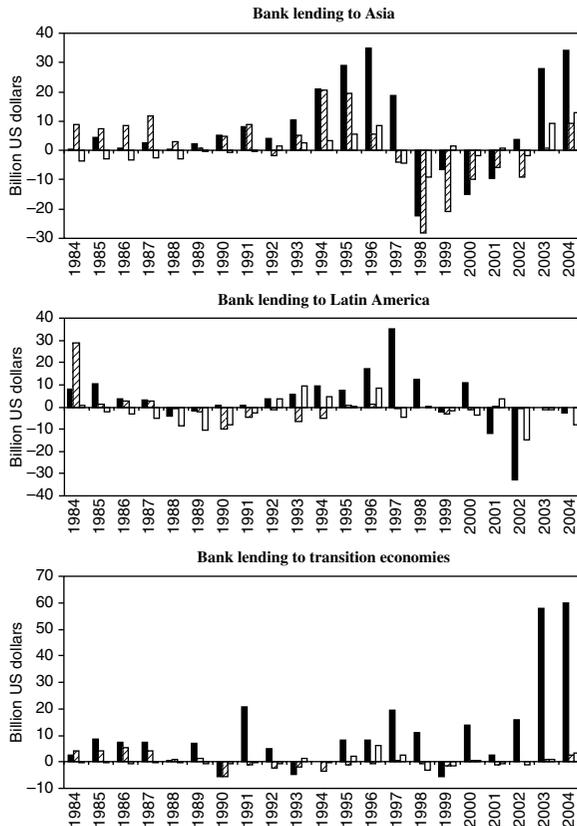
substantial losses in the aftermath of the Russian crisis when it introduced outright controls on capital outflows. Interestingly, the withdrawals may have been triggered by the increased risk – perceived or real – associated with the country.

The behavior of banks

Bank-related lending has also been quite volatile since the late 1970s. This section examines the role of European, Japanese and United States banks in spreading the crises of the 1990s. The Bank for International Settlements (BIS) Consolidated Banking Statistics database is used to examine the role of the three international banking clusters. In particular, international claims of reporting BIS banks in emerging economies, including both total cross-border claims and local claims in foreign currency booked by foreign offices, are studied.

As shown in Figure 41.3, bank flows poured into Asia throughout most of the 1990s and accelerated following the Mexican crisis. Bank loans to emerging Asia expanded by 89 percent from June 1994 to June 1997. Part of the rise in lending was due to the European banks' goal of achieving a higher profile in emerging markets, particularly in South Korea. Much of the lending boom, especially in the case of Thailand, Indonesia and South Korea, was due to a rapid expansion in credit from Japanese banks. Faced with a slumping economy and little domestic loan demand, Japanese banks increasingly looked overseas to the rapidly growing economies of South-East Asia as potential borrowers. United States bank lending to Asia was modest before the crisis. By June 1997, the United States banks' positions in emerging Asia had only reached US\$32 billion and only accounted for 20 percent of all United States bank lending to developing countries. In contrast, by the onset of the Thai crisis, Japanese banks had exposure to Asia four times as much as United States banks (US\$124 billion). European bank lending to emerging Asia was also significant and, by the onset of the Thai crisis, the exposure of European banks to Asia surpassed that of Japanese banks, reaching US\$161 billion. The exposure of European banks to emerging Asia accounted for about a half of all their lending to emerging markets; South Korea alone accounted for 40 percent of their lending to the developing world.

Japanese banks, heavily exposed to Thailand, were the first to pull out of emerging Asia. Between June and December of 1997, lending by Japanese banks fell by 8 percent. European banks, heavily exposed to South Korea, only began to pull out following the start of the crisis in that country in November 1997. In net terms, European bank lending to Asia continued to increase from June to December 1997. By June 1998, however, lending to emerging Asia was reduced across the board. Bank lending to Asia fell by

*Notes:*

Asia includes Afghanistan, Armenia, Azerbaijan, Bangladesh, Bhutan, British Overseas Territories, Brunei, Cambodia, China, Fiji, French Polynesia, Georgia, India, Indonesia, Kazakhstan, Kiribati, Kyrgyzstan, Laos, Macau, Malaysia, the Maldives, Mongolia, Myanmar, Nauru, Nepal, New Caledonia, North Korea, Pakistan, Papua New Guinea, the Philippines, the Solomon Islands, South Korea, Sri Lanka, Taiwan, Tajikistan, Thailand, Tonga, Turkmenistan, Tuvalu, US Pacific Islands, Uzbekistan, Vietnam, Wallis Futuna and Western Samoa.

Latin America includes Argentina, Belize, Bolivia, Brazil, Chile, Colombia, Costa Rica, Cuba, Dominica, the Dominican Republic, Ecuador, El Salvador, the Falkland Islands, Grenada, Guatemala, Guyana, Haiti, Honduras, Jamaica, Mexico, Nicaragua, Paraguay, Peru, St Lucia, St Vincent, Suriname, Trinidad and Tobago, Turks and Caicos, Uruguay and Venezuela.

Transition economies include Albania, Belarus, Bosnia and Herzegovina, Bulgaria, Croatia, Cyprus, the Czech Republic, Czechoslovakia, Estonia, the German Democratic Republic, Hungary, Latvia, Lithuania, Macedonia, Malta, Moldova, Poland, Romania, Russia, Serbia and Montenegro, Slovakia, Slovenia, the Soviet Union, Turkey and Ukraine.

Source: Bank for International Settlements.

Figure 41.3 Bank lending: European banks, Japanese banks, US banks

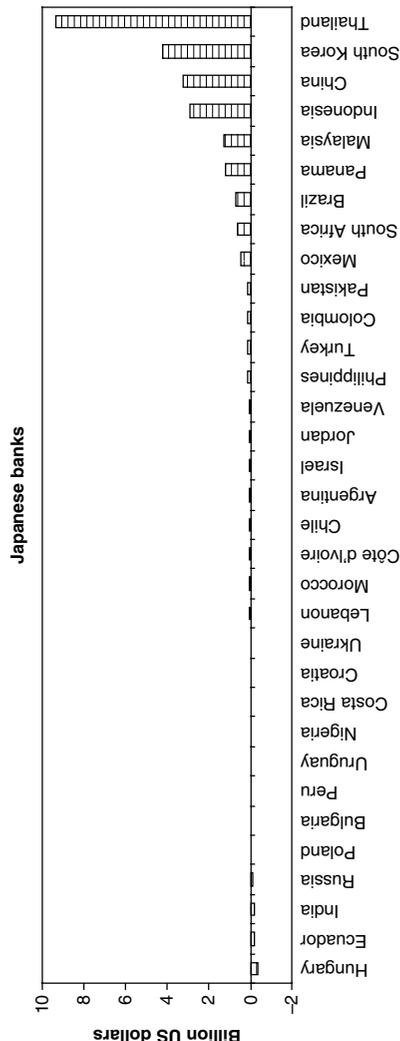
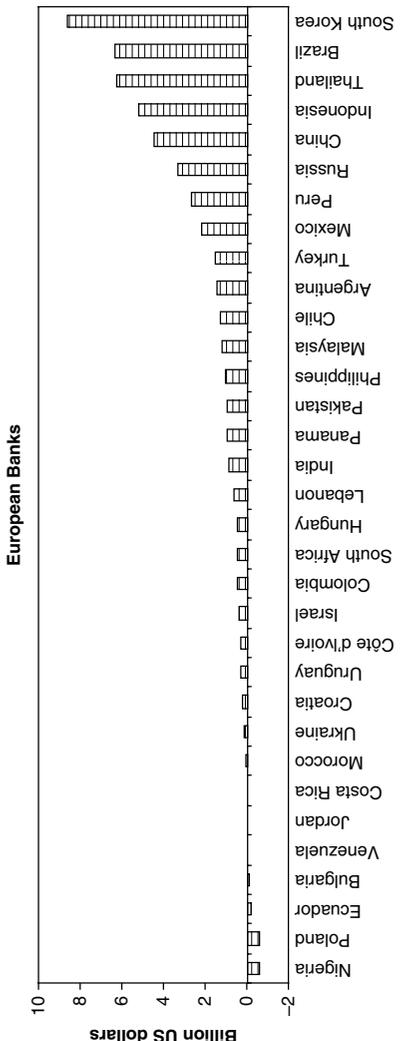
US\$46 billion, with European banks recalling US\$12 billion, Japanese banks US\$25 billion and United States banks US\$9 billion, respectively.

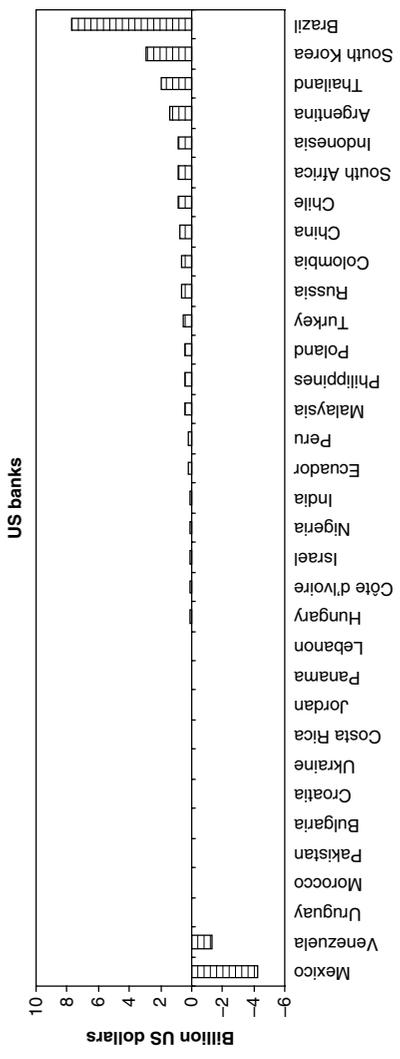
Figure 41.3 also reports bank lending to Latin America and transition economies. Exposure to these regions increased sharply in the mid-1990s (in large part driven by the purchase of domestic banks by European banks), with claims on these regions increasing by about 50 percent from June 1994 to June 1998, immediately before the onset of the Russian crisis. During the 1990s, European banks had the largest exposure to these regions – accounting for 67 percent to Latin America and 84 percent to transition economies. The Russian crisis led to some withdrawals of Japanese and United States lending from both regions, but this was not the case with European banks that had acquired local banks. Total exposure to Latin America by European banks peaked in December 2000.

Figures 41.4 to 41.6 tally country-by-country bank flows originating in European, Japanese and United States banks in the aftermath of the Mexican, Thai and Russian crises. Each figure focuses on the year following the crisis. Figure 41.4 shows that, with the exception of Mexico and Venezuela (which had a banking crisis of its own making), Latin American countries did not suffer major reversals in bank lending following the Mexican crisis. Moreover, within a year of the crisis, lending to Latin America recovered and even surpassed the levels observed before the crisis. Brazil was the prime beneficiary of bank flows during 1995, with lending from European and United States banks reaching US\$15 billion. Even in the case of Mexico and Venezuela, withdrawals were not made across the board. Only United States banks recalled loans from these countries. Figure 41.4 also shows that in Asia, the major recipients of capital flows in 1995 were South Korea, Thailand and Indonesia.

Figure 41.5 shows the behavior of bank lending in the aftermath of the Thai crisis. In contrast to the Tequila crisis, the Thai crisis triggered major reversals in bank flows from banks in Europe, Japan and the United States. Thailand, South Korea, Indonesia and Malaysia were the countries that suffered major withdrawals. Contagion was only regional in nature, with almost all of the Latin American countries, and to a lesser degree transition economies, continuing to have uninterrupted access to bank lending.

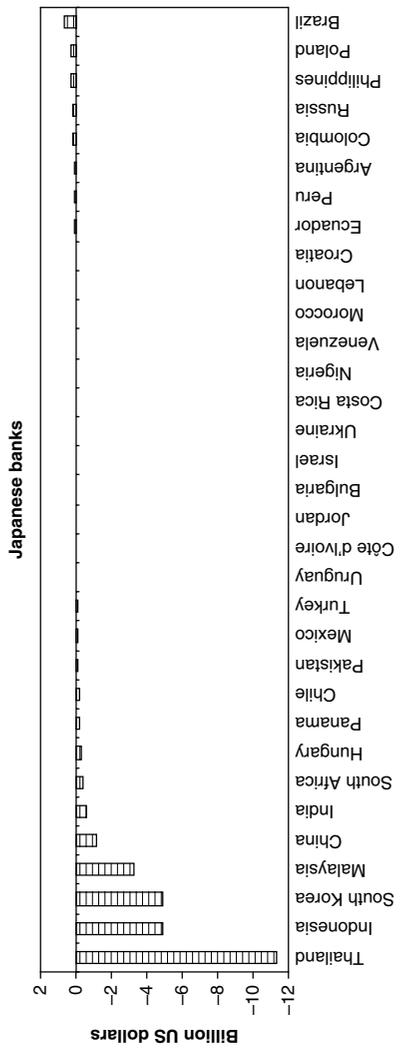
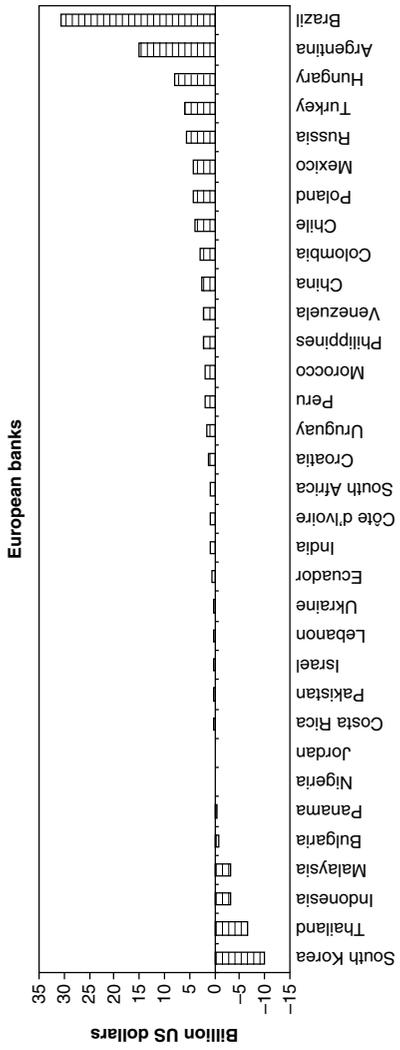
Figure 41.6 shows contagion from the Russian crisis. As was the case with mutual funds, the reversal in bank lending following the Russian default was not restricted to the Russian Federation or neighbouring countries. This time, the reversal was more widespread, and affected countries as far away as Brazil and South Africa. While Japanese banks continued to recall loans from Thailand, Indonesia and South Korea, reversals were not just restricted to these countries. Japanese banks, as well as United States banks, also recalled loans from Brazil, Mexico, India and South Africa.

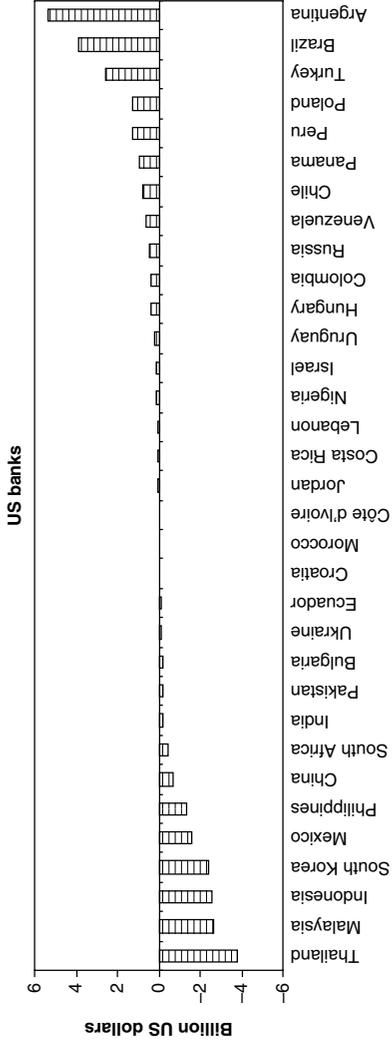




Source: Bank for International Settlements.

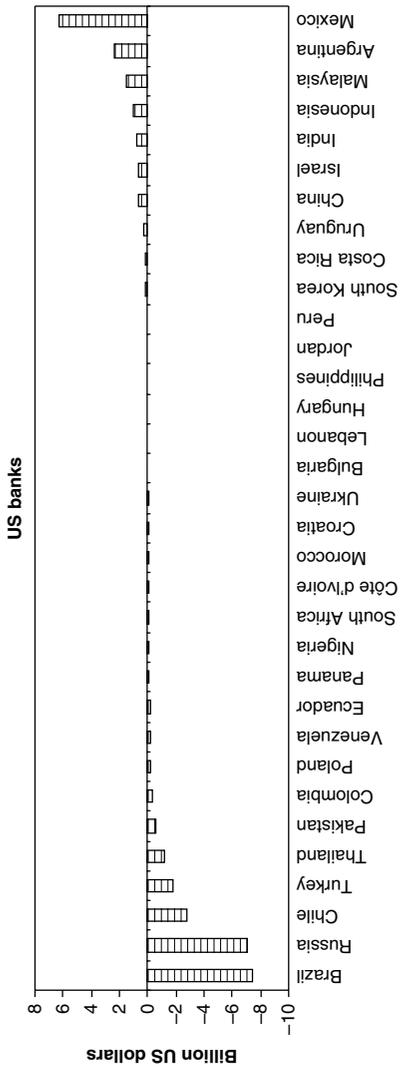
Figure 41.4 Bank flows: global spillovers – after the Mexican crisis: December 1994–December 1995





Source: Bank for International Settlements.

Figure 41.5 Bank flows: global spillovers – after the Thai crisis: June 1997–June 1998



Source: Bank for International Settlements.

Figure 41.6 Bank flows: global spillovers – after the Russian crisis: June 1998–June 1999

More formal evidence suggests that international banks were at the centre of financial contagion in the late 1990s. For example, Kaminsky and Reinhart (2000) examine contagion during the debt crisis in 1982, the Mexican crisis in 1994 and the Asian crisis in 1997, and find that United States banks were at the core of the contagion during the debt crisis, while Japanese banks spread the Thai crisis to Indonesia, South Korea and Malaysia. Van Rijckeghem and Weder (2003) examine the Tequila, Asian and Russian crises and the flows to 31 emerging countries from 11 creditor countries using BIS banks. Their evidence supports the idea that the degree to which countries compete for funds from common bank lenders is a fairly robust predictor of the incidence of contagion. Finally, Caramazza et al. (2000) extend earlier work on indicators of vulnerability to currency crises by examining the role of financial linkages, while controlling for the roles of internal and external macroeconomic imbalances and trade spillovers. Their results indicate that financial links do matter while exchange rate regimes and controls on capital flows do not seem to.

Globalization and volatility

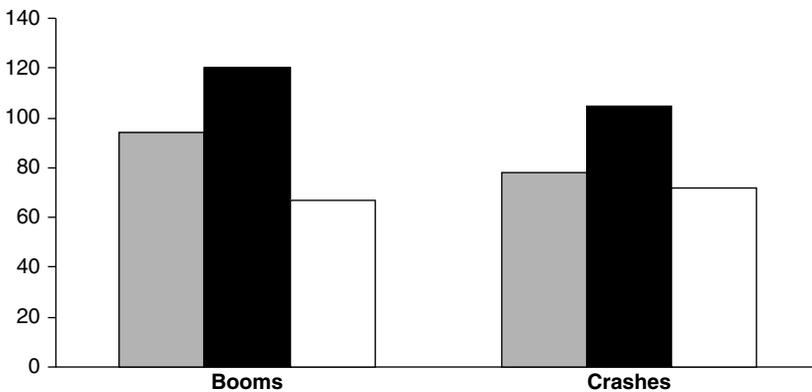
As discussed in the introduction, the views on the effects of financial globalization have been diverse; there are those who defend capital controls (Rodrik, 1998; Stiglitz, 1999) and those who maintain that capital should be allowed to move freely (Dornbusch, 1998). The rationale for restricting international capital flows is grounded in the belief that market failures and distortions pervade capital markets around the world. One of the most frequently cited distortions is that of asymmetric information, which is rampant in international capital markets due to geographical and cultural differences that complicate the task of obtaining information. In addition, imperfections in international markets are magnified by the difficulties in enforcing contracts across borders. With imperfect information, investors may overreact to shocks, withdrawing massively from countries at the first signs of economic problems, or become euphoric and pour in capital in quantities beyond those justified by 'good' fundamentals. On the other hand, those who consider international capital markets to be efficient favor unrestricted capital movements. Financial liberalization is believed to improve the functioning of financial systems, increasing the availability of funds and allowing cross-country risk diversification. Moreover, it is also claimed that financial integration tends to facilitate economic growth.

This section will summarize some of the findings in the literature on the effects of globalization, paying particular attention to the short- and long-run effects of financial integration on real and financial volatility.

Financial markets

The evidence from the crises of the 1990s suggests that crises are preceded by ‘excessive’ capital inflows that, in turn, fuel large expansions in domestic credit and bubbles in financial markets (see, for example, Sachs et al., 1996). There is also evidence that most episodes of banking crises are preceded by financial liberalization (see, for example, Kaminsky and Reinhart, 1999; Demirguc-Kunt and Detragiache, 1999). To reconcile the evidence that globalization is at the heart of financial crises with the hypothesis that international capital markets allow capital to move to its most attractive destination and promote more stable financial markets, I examine the possible time-varying effects of financial liberalization on stock market price cycles.⁵

Figure 41.7 shows the average amplitude of booms and crashes in stock prices for 14 emerging markets⁶ during periods of repression, in the immediate aftermath of liberalization (the four years following liberalization), and in the long run. The evidence in this figure seems to point to excessive cycles, with larger booms followed by larger crashes in the immediate aftermath of financial liberalization. However, liberalization does not permanently bring about more volatile financial markets. If liberalization persists, stock markets in emerging countries become more stable. Kaminsky and Schmukler (2003) argue that these conflicting effects arise because during episodes of financial repression, banks are protected from outside competition and do not have the pressure to run efficiently. Liberalization in this



Note: Gray: repression; Black: short-run liberalization; White: long-run liberalization.

Source: Kaminsky and Schmukler (2003).

Figure 41.7 Average amplitude of booms and crashes in stock prices in 14 emerging markets (in percentage points)

scenario unveils a new problem, as protected domestic banks suddenly get access to new sources of funding, triggering protracted financial booms. But financial liberalization triggers reforms and better-functioning financial markets as domestic investors, now with access to international capital markets, demand better enforcement rules to continue to invest in domestic financial markets. Moreover, as pointed out by Stulz (1999), the liberalization and gradual integration of emerging markets into international financial markets may help strengthen the domestic financial sector, as foreign investors generally have better skills and more information and can thus monitor management in ways that local investors cannot. Liberalization also allows firms to access mature capital markets. Firms listed on foreign stock markets are in the jurisdiction of a superior legal system with higher disclosure standards that will promote more transparency in the management of the firm and can trigger improvements in corporate governance.

Business cycles and growth

The evidence in the previous section is suggestive of excessive booms and busts in financial markets in developing countries following globalization, but of more stable financial markets in the long run if globalization persists. This section will examine the relationship between globalization and business cycle fluctuations and growth.

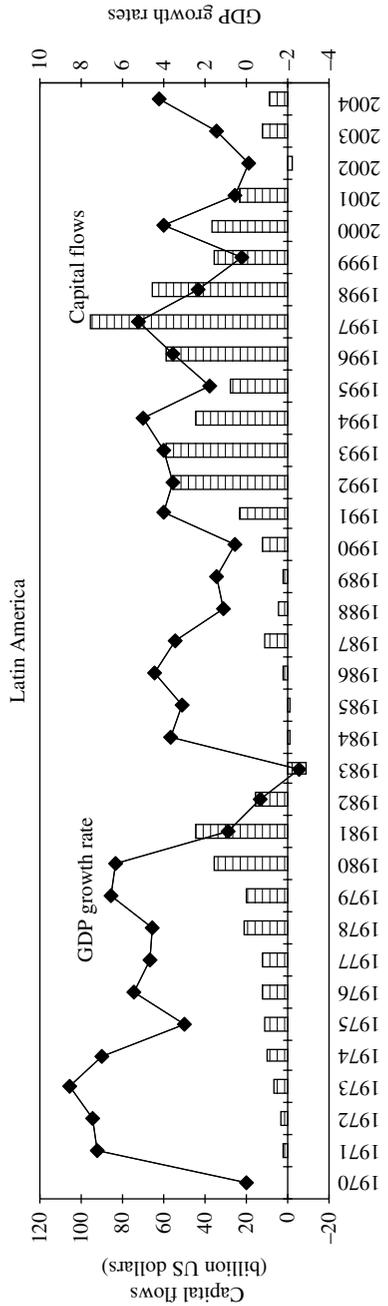
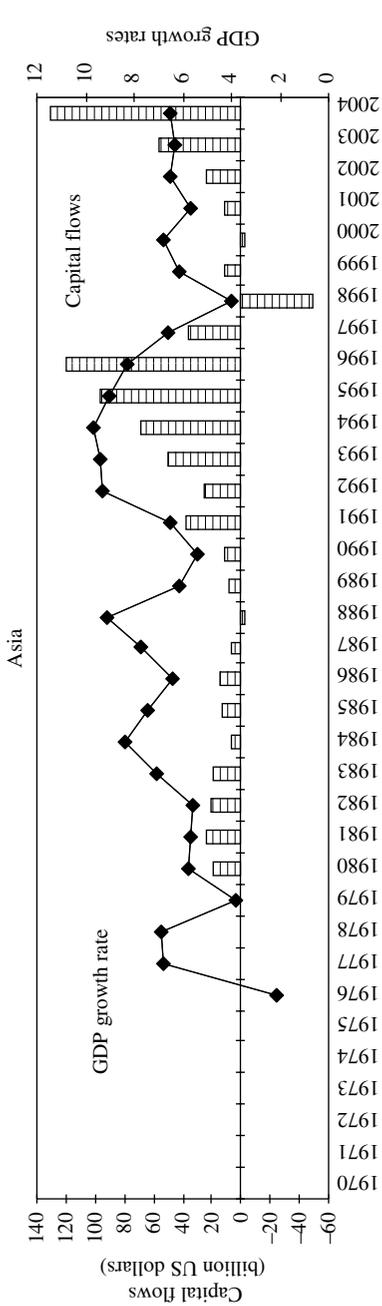
Figure 41.8 shows international capital flows to emerging markets in Asia, Latin America and transition economies, as well as annual output growth rates. The panels suggest that capital flows have been procyclical, with large inflows in good times and outflows during recessions. For example, Latin America's growth rates oscillated at around 4.5 percent in periods of capital inflows, while growth rates were about 1 percent in periods of sudden stops. Similarly, Asia's economic activity collapsed to about 5.5 percent after the sudden stop in capital flows in the late 1990s, after growing at an average annual growth rate of 8.5 percent during the earlier period of large capital inflows. This evidence contrasts sharply with the prescription that international capital markets should allow countries to smooth out the effect of the business cycle. Countries seem to have lost access to international credit markets during recessions on a systematic basis.

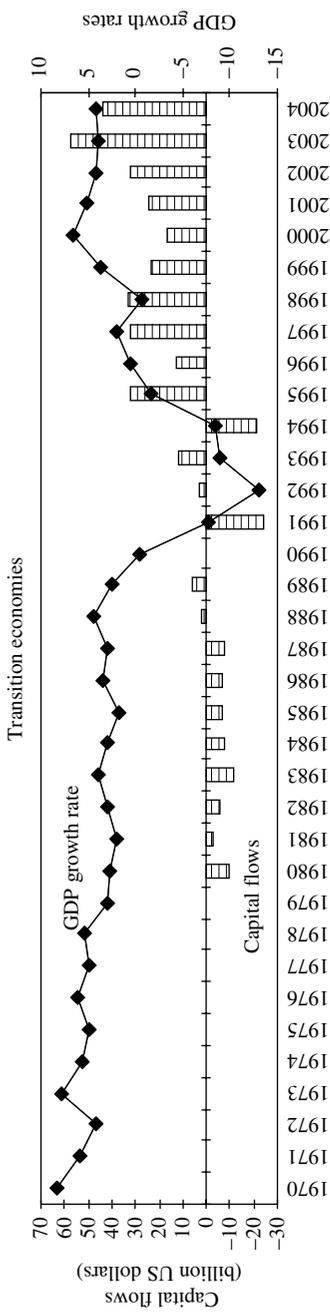
This non-optimal behavior of international capital flows has also been studied by Calvo et al. (2004), who observe that sudden reversals in capital flows to emerging economies lead to large real depreciations and profound downturns. As reported in Kaminsky et al. (2004), macro-policies tend to be procyclical in developing countries while they are countercyclical or acyclical in industrialized countries. That is to say, macro-policies tend to

smooth out the business cycle in industrial countries but magnify it in developing countries, as shown in Table 41.2. The left panel in this table reports the correlation between the cyclical components of fiscal and monetary policy with the business cycle. The right panel shows the correlations of the cyclical components of fiscal and monetary policy with net capital inflows. Interestingly, the evidence suggests that international capital flows to developing countries may trigger procyclical macro-policies. Government expenditure (inflation tax) is positively (negatively) correlated with net capital inflows, indicating that periods of capital inflows are associated with expansionary fiscal policies, and periods of capital outflows with contractionary fiscal policies. While more research is needed, the stylized evidence suggests that international capital flows may trigger more volatile business cycles in emerging economies.

While this evidence points to links between financial integration and output instability over the business cycle, there is also evidence that financial integration promotes growth. A variety of authors have examined the effects of domestic and external deregulation of financial markets in emerging economies and found that they generally trigger sustainable growth in the long run. Bekaert et al. (2005), for example, examine the effects on growth of the opening of the stock market to foreign investors in a sample of about 90 developing countries and find that, overall, liberalization triggers an increase in growth by approximately one percentage point. Similarly, Galindo et al. (2002) study whether financial liberalization promotes economic growth by analyzing its effect on the cost of external financing to firms. They find that the liberalization of domestic and external financial markets reduces the cost of external funds faced by firms. In particular, they find that industries that depend on external finance grow almost 1 percent faster, relative to industries with low external financing dependence, in episodes of globalization compared to episodes of repression. The evidence on the links between financial liberalization and growth is not conclusive, however. Edison et al. (2002), for example, using data from 57 countries from 1980 to 2000, conclude that there is no robustly significant effect of financial integration on economic growth. Similarly, Kraay (1998), using a sample of 117 countries, finds no effect of financial liberalization on growth or, at best, mixed results.

Perhaps the inability of past research to agree on the effects of financial globalization on economic growth lies in the fact that liberalization has time-varying effects on growth. Loayza and Ranciere (2002) present some evidence that suggests this might be the case. These authors estimate transitory and trend effects of financial deepening on growth using a sample of about 80 countries and find that financial deepening, which in general is closely related to financial liberalization, harms growth in the short run but





Notes:

The countries comprising Asia are Bangladesh, China, Hong Kong, India, Indonesia, Malaysia, Pakistan, the Philippines, Singapore, South Korea, Taiwan, Thailand and Vietnam.

The countries comprising the transition economies are Albania, Armenia, Azerbaijan, Belarus, Bosnia and Herzegovina, Bulgaria, Croatia, the Czech Republic, Estonia, Georgia, Hungary, Kazakhstan, Kyrgyzstan, Latvia, Lithuania, Macedonia, Moldova, Mongolia, Poland, Romania, Russia, Serbia and Montenegro, Slovakia, Slovenia, Tajikistan, Turkmenistan, Ukraine and Uzbekistan.

The countries comprising Latin America are Argentina, Brazil, Chile, Colombia, the Dominican Republic, Ecuador, Guatemala, Mexico, Peru, Uruguay and Venezuela.

Source: World Economic Outlook.

Figure 41.8 Private capital flows to emerging markets and GDP annual growth rates

Table 41.2 Correlations between the cyclical components of macropolicies, real GDP and net capital inflows

Countries	Correlations with real GDP			Correlations with net capital inflows		
	Fiscal policy		Monetary policy	Fiscal policy		Monetary policy
	Government expenditure	Inflation tax	Lending interest rate	Government expenditure	Inflation tax	Lending interest rate
	OECD	-0.13	0.16	0.23	0.03	0.04
Non-OECD	0.33	-0.15	-0.05	0.20	-0.16	-0.06

Notes:

A positive (negative) correlation between government expenditure (inflation tax) and real GDP indicates pro-cyclical fiscal policy. A negative correlation between lending interest rates and real GDP indicates pro-cyclical monetary policy. A positive (negative) correlation between government expenditure (inflation tax and lending interest rates) and net capital inflows indicates that contractionary macropolicies are linked to episodes of low net capital inflows.

The cyclical component of the various indicators was obtained using the HP filter.

Source: Kaminsky et al. (2004).

leads to higher growth in the long run. These latest results are closely linked to the evidence from stock market cycles discussed before and suggest that financial liberalization triggers growth in the long run because it fuels institutional reform.

Conclusions

The explosion of capital flows to emerging markets in the early and mid-1990s and the recent reversal following the crises around the globe have reignited a heated debate on how to manage international capital flows. Capital outflows worry policy-makers, but so do capital inflows, as they may trigger bubbles in asset markets and lead to an appreciation of the domestic currency and a loss of competitiveness. Policy-makers also worry that capital inflows are mostly of the ‘hot money’ type, which is why capital controls have mostly targeted short-term capital inflows. While capital controls may work, at least in the very short run, the introduction of restrictions to capital mobility may have undesirable long-run effects. In particular, capital controls protect inefficient domestic financial institutions and thus may trigger financial vulnerabilities.⁷ Capital controls may also delay improve-

ments in corporate governance of non-financial firms because, as countries liberalize their capital accounts, domestic corporations start participating in international capital markets, mainly through cross-listing in major world stock exchanges, with higher disclosure standards and under the jurisdiction of a superior legal system. This certainly promotes more transparency in the management of the firm and can trigger improvements in corporate governance (see, for example, Stulz, 1999). Thus, regulation of capital flows may not only provoke financial vulnerabilities but also lower economic growth. Policy-makers have also resorted to sterilization of capital flows to regain control of monetary policy. While sterilization may provide some relief, it may also be quite costly to central banks. Moreover, the ability of governments to control international capital flows or to sterilize them diminishes with globalization.

In conclusion, there is no optimal policy to deal with the risks of volatile international capital flows, as policies that may work in the short run may have adverse effects in the long run. Since there is evidence that currency and banking crises tend to occur in economies with deteriorated fundamentals, conservative macroeconomic policies should be at the heart of dealing with volatile capital flows. Further research should examine whether countries can deregulate financial systems without becoming vulnerable to crises. Since the costs of crises have been quite large, this last question deserves much attention.

Notes

1. This chapter draws on previous research with Richard Lyons, Carmen Reinhart, Sergio Schmukler and Carlos Végh and is a shorter version of a paper entitled 'International Capital Flows, Financial Stability and Growth' (Kaminsky, 2006).
2. Liquidity may have an important effect on investors' portfolio allocations since investors may want to avoid illiquid markets to minimize the price collapses always present when there is no ready market.
3. To identify liquid markets, countries are ranked by region according to their volume traded and according to their share in the mutual funds portfolio at the onset of the crisis. The dummy variable related to volume traded is given a value of one if the country ranks among the top 30 percent of most-liquid countries in the region in that category, and a value of zero otherwise. Similarly, countries are classified as liquid (that is to say, the dummy variable is given a value of one) if they rank among the 30 percent of the countries with the largest share in mutual fund portfolios for the region. A third dummy is created to capture whether emerging-market firms are trading in mature financial markets: the variable is given a value of one if they do, and zero if they do not. Finally, the variable capturing restrictions to entry and exit of foreigners in the stock markets of emerging economies is given a value of one if there are no restrictions, and zero if there are. All of this information is collapsed into a liquidity variable that is the average of the four univariate liquidity dummy variables. Thus, the general index of liquidity, the average of the four components, can have five values: 0, 1/4, 2/4, 3/4 and 1, with a value of one indicating a highly liquid market. I classify a country as having liquid financial markets when this dummy takes a value of 2/4 or higher.
4. See Kaminsky et al. (2002) for a country-by-country detail on fragility, liquidity, risk and mutual fund withdrawals.

5. The results discussed below are from Kaminsky and Schmukler (2003).
6. The 14 emerging economies are Argentina, Brazil, Chile, Colombia, Hong Kong, Indonesia, South Korea, Malaysia, Mexico, Peru, the Philippines, Taiwan, Thailand and Venezuela.
7. Claessens et al. (1998) present evidence that liberalization of the capital account and foreign bank entry lead to improvements in banking system efficiency.

References

- Bekaert, Geert, Campbell Harvey and Christian Lundblad (2005), 'Does Financial Liberalization Spur Growth?' in Geert Bekaert, Campbell R. Harvey and Christian Lundblad, *Journal of Financial Economics*, 77: 3–55.
- Calvo, Guillermo, Alejandro Izquierdo and Luis Mejía (2004), 'On the Empirics of Sudden Stops the Relevance of Balance-Sheet Effects', NBER Working Papers 10520.
- Caramazza, Francesco, Luca Ricci and Ranil Salgado (2000), 'Trade and Financial Contagion in Currency Crises', IMF Working Paper No. 55, International Monetary Fund, Washington, DC.
- Claessens, Stjin, Asli Demirgüç-Kunt and Harry Huizinga (1998), 'How Does Foreign Entry Affect the Domestic Banking Sector?', World Bank Policy Research Working Paper No. 1918.
- Demirguc-Kunt, Asli and Enrica Detragiache (1999), 'Financial Liberalization and Financial Fragility', in Boris Pleskovic and Joseph E. Stiglitz (eds), *Annual World Bank Conference on Development Economics*, Washington, DC: World Bank, pp. 303–31.
- Dornbusch, Rudiger (1998), 'Capital Controls: An Idea Whose Time is Past', in *Should the IMF Pursue Capital-Account Convertibility?* Essays in International Finance 207, International Finance Section, Department of Economics, Princeton University, 20–27 May.
- Edison, Hali, Michael Klein, Luca Ricci and Torsten Slok (2002), 'Capital Account Liberalization and Economic Performance: A Review of the Literature', Working Paper No. 120, International Monetary Fund, Washington, DC.
- Galindo, Arturo, Alejandro Micco and Guillermo Ordoñez (2002), 'Financial Liberalization and Growth: Empirical Evidence', paper presented at the Conference on Financial Globalization: A Blessing or a Curse?, World Bank, Washington, DC.
- Goldstein, Morris, Graciela Kaminsky and Carmen Reinhart (2000), *Assessing Financial Vulnerability: An Early Warning System for Emerging Markets*, Washington, DC: Institute for International Economics.
- Kaminsky, Graciela (1998), 'Currency and Banking Crises: The Early Warnings of Distress', International Finance Discussion Paper No. 629, October, Board of Governors of the Federal Reserve System, Washington, DC.
- Kaminsky, Graciela (2006), 'International Capital Flows, Financial Stability, and Growth', in José Antonio Ocampo, Jomo K.S. and Sarbuland Khan (eds), *Policy Matters: Economic and Social Policies to Sustain Equitable Development*, New York: United Nations; London: Zed Books; New Delhi: Opus, pp. 123–67.
- Kaminsky, Graciela and Carmen Reinhart (1999), 'The Twin Crises: Causes of Banking and Balance of Payment Problems', *American Economic Review*, 89 (3): 473–500.
- Kaminsky, Graciela and Carmen Reinhart (2000), 'On Crises, Contagion, and Confusion', *Journal of International Economics*, 51 (1): 145–68.
- Kaminsky, Graciela and Sergio Schmukler (2003), 'Short-Run Pain, Long-Run Gain: The Effects of Financial Liberalization', NBER Working Paper No 9787, June.
- Kaminsky, Graciela, Richard Lyons and Sergio Schmukler (2002), 'Fragility, Liquidity, and Risk: The Behavior of Mutual Funds During Crises', mimeo article, World Bank, Washington, DC.
- Kaminsky, Graciela, Carmen Reinhart and Carlos Végh (2004), 'When it Rains, it Pours: Procyclical Capital Flows and Macro Policies', in Mark Gertler and Kenneth S. Rogoff (eds), *NBER Macroeconomics Annual 2004*, Cambridge, MA: National Bureau of Economic Research, pp. 11–53.

- Kraay, Art (1998), 'In Search of the Macroeconomic Effects of Capital Account Liberalization', unpublished, World Bank.
- Loayza, Norman and Romaine Ranciere (2002), 'Financial Development, Financial Fragility, and Growth', Working Paper No. 145, Central Bank of Chile, Santiago.
- Rodrik, Dani (1998), 'Who Needs Capital Account Convertibility?' in S. Fischer R.N. Cooper, R. Dornbusch, P.M. Garber, C. Massad, J.J Polak, D.Rodrik and S.S. Tarapore (eds), *Should the IMF Pursue Capital-Account Convertibility?* Essays in International Finance No. 207, Department of Economics, Princeton University, May.
- Sachs, Jeffrey, Aaron Tornell and Andrés Velasco (1996), 'Financial Crises in Emerging Markets: The Lessons from 1995', *Brookings Papers on Economic Activity*, 1: 147–215.
- Stiglitz, Joseph (1999), 'Bleak Growth for the Developing World', *International Herald Tribune*, 10–11 April, p. 6.
- Stulz, Rene (1999), 'Globalization of Equity Markets and the Cost of Capital', NBER Working Paper No 7021, March.
- Van Rijckeghem, C. and Beatrice Weder (2003), 'Spillovers through Banking Centers: A Panel Data Analysis', *Journal of International Money and Finance*, 22: 483–509.

42 Foreign aid

*Steven Radelet*¹

Introduction

Controversies about aid effectiveness go back decades. Critics such as Milton Friedman (1958), Peter Bauer (1972) and William Easterly (2001) have leveled stinging critiques, charging that aid has enlarged government bureaucracies, perpetuated bad governments, enriched the elite in poor countries, or just been wasted. They cite widespread poverty in Africa and South Asia despite four decades of aid starting in the 1960s, and point to countries that have received substantial aid yet have had disastrous records such as the Democratic Republic of the Congo, Haiti, Papua New Guinea and Somalia. In their eyes, aid programs should be dramatically reformed, substantially curtailed or eliminated altogether.

Supporters counter that these arguments, while partially correct, are overstated. Nicholas Stern (2002), Joseph Stiglitz (2002), Jeffrey Sachs et al. (2004) and others have argued that although aid has sometimes failed, it has supported poverty reduction and growth in some countries and prevented worse performance in others. They believe that many of the weaknesses of aid have more to do with donors than recipients, and point to a range of successful countries that have received significant aid such as Botswana, Indonesia, Korea and, more recently, Tanzania and Mozambique, along with successful initiatives such as the Green Revolution, the campaign against river blindness, and the introduction of oral rehydration therapy.

This chapter explores trends in aid, the motivations for aid, its impacts, and debates about reforming aid. It begins by examining aid magnitudes and who gives and receives aid. It discusses the multiple motivations and objectives of aid, some of which conflict with each other. It then explores the empirical evidence on the relationship between aid and growth, most (but far from all) of which concludes there is a positive relationship (at least under certain circumstances). It examines some of the key challenges in making aid more effective, including the principal-agent problem and the related issue of conditionality, and concludes by examining some of the main proposals for improving aid effectiveness.

Donors and recipients

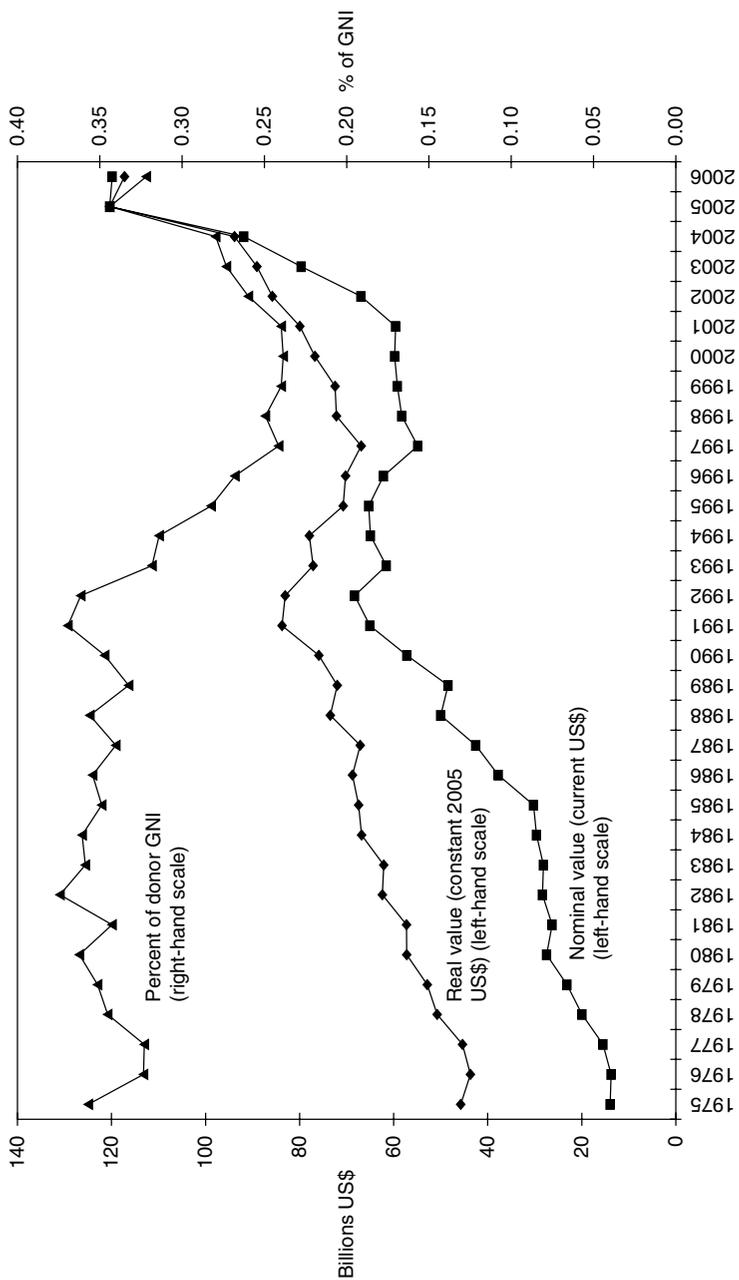
What is foreign aid?

The standard definition of foreign aid comes from the Development Assistance Committee (DAC) of the Organisation for Economic Co-operation and Development (OECD), which defines foreign aid (or the equivalent term, foreign assistance) as financial flows, technical assistance and commodities that are: (1) designed to promote economic development and welfare as their main objective (thus excluding aid for military or other non-development purposes); and (2) provided as either grants or subsidized loans.

Grants and subsidized loans are referred to as concessional financing, whereas loans that carry market or near-market terms (and therefore are not foreign aid) are non-concessional financing.² According to the DAC, a loan counts as aid if it has a 'grant element' of 25 percent or more, meaning that the present value of the loan must be at least 25 percent below the present value of a comparable loan at market interest rates (usually assumed by the DAC – rather arbitrarily – to be 10 percent with no grace period). Thus, the grant element is zero for a loan carrying a 10 percent interest rate, 100 percent for an outright grant, and something in-between for other loans.

The DAC classifies aid flows into three broad categories. Official development assistance (ODA) is the largest, consisting of aid provided by donor governments to low- and middle-income countries. Official assistance (OA) is aid provided by governments to richer countries with per capita incomes higher than approximately \$9000³ (for example, the Bahamas, Cyprus, Israel and Singapore) and to countries that were formerly part of the Soviet Union or its satellites. Private voluntary assistance includes grants from non-governmental organizations, religious groups, charities, foundations and private companies.

When discussing foreign aid, most people have in mind ODA. Global ODA increased steadily from the 1960s until it reached a peak of \$68 billion in 1992, just after the end of the Cold War (Figure 42.1), and then declined sharply to just under \$55 billion in 1997. Aid flows began to rebound in the late 1990s following calls for greater debt relief and increased aid to new democracies, and accelerated very sharply after the attacks of 11 September 2001, reaching \$120 billion in 2006 (all of these figures would be slightly higher if they included OA). In real terms, total ODA in 2002 was about the same as in 1992, and by 2006 was about 15 percent higher. Measured as a share of donor income ODA fell sharply during the 1990s, and has rebounded only slightly. Donors have pledged to continue to increase aid, most recently in July 2005 when the heads of state of the



Source: OECD/DAC 2008.

Figure 42.1 Global ODA 1975–2006

Group of 8 industrialized countries promised to double aid to sub-Saharan Africa by 2010 and triple it by 2015, but growing budget tensions in donor countries may undermine these pledges.

Who gives aid, and who receives it?

Historically most aid has been given as bilateral assistance directly from one country to another. Donors also provide aid indirectly as multilateral assistance, which pools resources together from many donors. The major multilateral institutions include the World Bank, the International Monetary Fund (IMF), the African, Asian and Inter-American Development Banks, and various United Nations agencies such as the United Nations Development Programme.

In terms of total dollars, the United States has consistently been the world's largest donor (except in the mid-1990s when Japan briefly topped the list). In 2006 the USA provided \$22.9 billion in ODA, with Japan, France the United Kingdom, and Germany the next-largest donors. However, when aid is measured as a share of donor income, the most generous donors are Norway, Denmark, Luxembourg, the Netherlands, and Sweden, each of which provided between 0.81–1.02 percent of GNI in 2006. Saudi Arabia provided aid equivalent to about 0.57 percent of its income. The United States is one of the smallest donors by this measure, at about 0.18 percent of US income in 2006, just over half of the 1970 level of 0.32 percent and less than one-third of the US average during the 1960s. Donors have pledged since the 1960s to devote 0.7 percent of their income as aid, most recently at the Financing for Development Conference in Monterrey in March 2002, but only a handful of small donors have achieved this level of aid.

One hundred and forty-five countries and territories around the world received aid in 2006. Table 42.1 shows the ten largest recipients, each of which received more than \$1.8 billion. Nigeria was at the top of the list in 2006, with measured aid of \$11.4 billion. But this figure is misleading because it includes \$9.5 billion for a one-time debt relief deal. Debt relief is accounted for differently than other components of ODA – the value of debt relief is the charge to the creditor country's budget for writing off the debt in the year of the debt relief, and does not represent new funding to the recipient (although it does capture a future reduction in debt service obligations). Nigeria's actual inflow of new finance in 2006 was \$1.9 billion. Iraq and Afghanistan together received nearly \$12 billion dollars, nearly unprecedented amounts, accounting for about 10 percent of the global total. Amounts to the other countries shown in Table 42.1 are more typical (by historical standards) for large recipients. Total dollar amounts are important, but they do not tell the entire story. On a per capita basis, the aid flows to some of these countries are fairly small. Vietnam received \$1.8

Table 42.1 *Major aid recipients, 2006*

		Total ODA (millions US\$)
1.	Nigeria	11434
2.	Iraq	8661
3.	Afghanistan	3000
4.	Pakistan	2147
5.	Sudan	2058
6.	Congo, Dem. Rep.	2056
7.	Ethiopia	1947
8.	Vietnam	1846
9.	Tanzania	1825
10.	Cameroon	1684
		Aid as % of recipient GNI
1.	Solomon Islands	61
2.	Tuvalu	58
3.	Liberia	54
4.	Burundi	53
5.	Micronesia, Fed. States	41
6.	Afghanistan	36
7.	Palestinian Adm. Areas	35
8.	Malawi	30
9.	Marshall Islands	29
10.	Guinea-Bissau	28
		Aid per capita (US\$)
1.	Palau	1866
2.	Mayotte	1777
3.	Nauru	1740
4.	Cook Islands	1614
5.	Tuvalu	1534
6.	Marshall Islands	786
7.	Palestinian Adm. Areas	426
8.	Solomon Islands	409
9.	Iraq	304
10.	Cape Verde	282

Source: OECD 2007 *Development Cooperation Report* and UN Statistics *National Accounts Main Aggregates Database* where GNI information missing.

Table 42.2 Official aid receipts by region, 2006

	Billion US\$	% of GNI	US\$ per person
Sub-Saharan Africa	38.2	5.8	50.2
South Asia	9.2	0.8	6.1
East Asia & Pacific	7.4	0.2	3.9
Europe & Central Asia	2.6	0.4	17.6
Middle East & North Africa	14.6	2.0	44.0
Latin America & Caribbean	6.0	0.2	11.0
Low-income	47.5	3.1	20.2
Lower-middle income	26.7	0.5	11.0
Upper-middle income	3.7	0.2	10.1

Source: Author's calculations based on data from OECD 2007 *Development Cooperation Report* and *World Development Indicators 2007*.

billion in aid in 2004, but this was equivalent to just 3 percent of its gross national income (GNI) or about \$22 per person. By contrast, Cameroon received a similar amount, \$1.7 billion in 2006, but for its 16.6 million people this was equivalent to about \$100 dollars per person. For small countries, a little bit goes a long way. Tiny Solomon Islands received just \$205 million, but this translated into 61 percent of GNI and about \$409 per person. Aid is typically measured as a share of GNI, but this can be misleading as a high ratio can just as easily be indicative of low GNI as of a large amount of aid.

On a regional basis, sub-Saharan African countries received aid flows of 5.8 percent of GNI in 2006, or \$50.2 per person (Table 42.2), although close to one-third of this amount was due to several large one-time debt relief deals that are not new inflows and are not indicative of long-term trends. North Africa and the Middle East received more than \$44 per person (largely on account of Iraq), and Europe and Central Asia received about \$18 per person. For low-income countries around the world, donors provided aid averaging about \$20.2 per recipient in 2006, although once again these figures are inflated by several one-time debt relief deals.

Generally speaking, aid is one of the largest components of foreign capital flows to low-income countries, but not to most middle-income countries, where private capital flows are more important. Aid flows averaged 3.1 percent of GNI in low-income countries in 2004, but just 0.2 percent of GNI in upper-middle-income countries. It is commonly claimed that the decline in aid flows to developing countries in the 1990s was more than offset by a rise in private capital. While this is true for developing countries in aggregate, the rise in private capital flows was heavily

concentrated in a handful of middle-income countries. In low-income countries, private capital rose much more slowly, and remained significantly smaller than aid.

Why do donors give aid?

Donors have a variety of motivations for providing aid, only some of which are directly related to economic development. There is little question that foreign policy and political relationships are the most important determinants of aid flows. During the Cold War, both the United States and the Soviet Union used aid to vie for the support of developing countries with little regard as to whether the aid was actually used to support development. The two largest recipients of US foreign aid (including both OA and ODA) from 1980 until very recently were Israel and Egypt, as the USA provided financial support to back the 1979 Camp David peace agreement. Beginning in 2002 Iraq became the largest aid recipient in the world, and its reconstruction is likely to become the largest single foreign aid program ever recorded. Taiwan and China have used aid (among other policy tools) to try to gain support and recognition for their governments from countries around the world. Many donors provide significant aid to their former colonies as a means of retaining some political influence (Alesina and Dollar, 2000).

Many people see the main rationale for aid as fighting poverty, and although this is less important than political considerations in donor allocation decisions, it still plays an important role. Donors generally provide their most concessional aid to the poorest countries, and some aid programs are designed explicitly with this objective in mind. For example, the World Bank's concessional financing arm – the International Development Association (IDA) – has an income ceiling (\$965 per capita in 2004). Once countries reach that ceiling, in most cases they 'graduate' from IDA to non-concessional International Bank for Reconstruction and Development (IBRD) loans. Other programs have less formal graduation rules, but still tend to provide less aid as incomes grow.

Country size matters as well. Large countries, such as Bangladesh, Indonesia, Nigeria and Pakistan receive relatively small amounts of aid on a per capita basis, even though hundreds of millions of people live in poverty in these countries. By contrast, some small countries receive very large amounts. For political reasons, donors generally want to influence as many countries as possible, which tends to lead to a disproportionate amount of aid going to small countries.

Bilateral aid is often designed at least partially to help support the economic interests of certain firms or sectors in the donor country. Multilateral aid is less prone to these pressures, although by no means immune. Many donors 'tie' portions of their aid by requiring that certain

goods and services be purchased from firms in the donor's home country, or that it be used for specific purposes that support groups in the donor countries (such as universities or business consulting firms). Automobiles, airline tickets and consulting services financed by USA foreign aid in most cases must be purchased from USA firms. Tying aid can give it more political support at home, but it can also make it more costly and less effective. If funds must be spent in the donor country, it reduces competition for services so that donors do not always use the least-cost provider. For example, the USA requires that food aid be purchased in the USA and shipped in US carriers to recipient countries, which can be much more expensive and take much longer than if food was purchased in a neighboring country. This means that recipients receive much less value for each dollar of aid allocated than they otherwise could. One study found that tying aid added 15–20 percent to its cost, thus significantly reducing its impact on recipient countries. Donors have begun to reduce the amount of aid that they tie, but the practice is still widespread among some donors. The USA no longer reports the share of its aid that is tied, but historically it has been around 75 percent. Greece ties about 70 percent of its aid, and Canada and Austria more than 40 percent. By contrast, Ireland, Norway, and the UK do not tie any of their aid.

Aid, growth and development

Most foreign aid is designed to meet one or more of four broad economic and development objectives: (1) to stimulate economic growth through building infrastructure, supporting productive sectors such as agriculture, or bringing new ideas and technologies; (2) to strengthen education, health, environmental or political institutions or systems; (3) to support subsistence consumption of food and other commodities, especially during relief operations or humanitarian crises; or (4) to help stabilize an economy following economic shocks.

Despite these broader objectives for aid, growth has always been the main yardstick used to judge aid's effectiveness. Debate has raged about the relationship between aid and growth for years, but there are some broad parameters of agreement. Even most aid pessimists agree that aid has been successful in some countries (such as in Botswana or Korea, or more recently in Mozambique and Tanzania), that aid has helped improve health by supplying essential medicines, and that aid is an important vehicle in providing emergency relief following natural disasters. Similarly, aid optimists concede that much aid has been wasted or stolen, such as by the Marcos regime in the Philippines and the Duvalier regime in Haiti, and that even under the best circumstances aid can create certain adverse economic incentives. Debate continues on the overall general trends, the conditions

under which aid works or does not work, and on what steps can be taken to make aid more effective. Although the majority of research since the mid-1990s has found a positive relationship between aid and growth, several studies have found no relationship. Three broad views have emerged on the relationship between aid and growth.⁴

Aid has a positive relationship with growth on average across countries (although not in every country), but with diminishing returns as the volume of aid increases

There are three key channels through which aid might spur growth:

- First, the classic view is that aid augments saving, finances investment and adds to the capital stock. In this view, poor countries are unable to generate sufficient amounts of saving on their own to finance the investment necessary to initiate growth, or at best only enough for very slow growth. In the strongest version of this view, the poorest countries may be stuck in a poverty trap in which their income is too low to generate the saving necessary to initiate the process of sustained growth (Sachs et al., 2004). A related argument is that aid might help relax a foreign exchange constraint in countries that earn relatively little foreign exchange, a view that was popularized through the early ‘two-gap’ models of economic growth.
- Second, aid might increase worker productivity through investments in health or education.
- Third, aid could provide a conduit for the transfer of technology or knowledge from rich countries to poor countries by paying for capital goods imports, through technical assistance, or through direct transfer of technologies such as the introduction of new seeds and fertilizers in the Green Revolution.

Several early studies found a positive relationship between aid and growth (for example, Papenek, 1973; Levy, 1988), but this strand of the literature took a significant turn in the mid-1990s when researchers began to investigate whether aid might support growth with diminishing returns. Oddly – given Solow’s response to the Harrod–Domar model in the 1950s – research until the mid-1990s only tested a linear relationship, a specification which persists in some studies today. A large group of studies that allow for diminishing returns have found a positive relationship although the direction of causality is a subject of ongoing debate.⁵ These studies do not conclude that aid has always worked in every country, but rather that on average and controlling for other factors, higher aid flows have been associated with more rapid growth. These studies have received much less public attention than

those that have found a zero or conditional relationship. The robustness of the conclusions of several of these studies has been the subject of on-going debate, as has the robustness of the conclusions of several studies that have reached the opposite conclusion, as discussed below. But since the mid-1990s the majority of published research on the topic has found a positive relationship either by allowing for diminishing returns, or by testing for conditional relationships as explored below.

Aid could also have a positive impact on development outcomes other than growth, such as health, education or the environment. Perhaps the best-documented area is health, where aid-supported programs have contributed to the eradication of smallpox, the near eradication of polio, control of river blindness and other diseases, the spread of oral rehydration tablets to combat diarrhea, and the dramatic increase in immunization rates in developing countries since 1970 (Levine et al., 2004). A recent cross-country study found a positive and significant relationship between health aid and infant mortality (Mishra and Newhouse, 2007). Undoubtedly, much aid aimed at health has also been squandered. But beyond the examples listed here, there is little systematic evidence on the relationship between aid and health, education, income distribution or other outcomes.

Aid has no affect on growth, and may actually undermine growth

Peter Bauer was perhaps the most outspoken proponent of this view (for example, Bauer, 1972), but he never provided systematic empirical evidence to support his argument. Many later empirical studies did reach this conclusion,⁶ although once again the robustness of these results is the subject of on-going debate. These researchers have suggested a variety of reasons why aid might not support growth:

- First, aid simply could be wasted, such as on limousines or presidential palaces, or it could encourage corruption, not just in aid programs but more broadly.
- Second, it can help keep bad governments in power, thus helping to perpetuate poor economic policies and postpone reform. Some argue that aid provided to countries in the midst of war might inadvertently help finance and perpetuate the conflict, and add to instability.
- Third, countries may have limited absorptive capacity to use aid flows effectively if they have relatively few skilled workers, weak infrastructure or constrained delivery systems. (Aid could help redress these weaknesses, but it may not be aimed to do so.)
- Fourth, aid flows can reduce domestic saving, both private saving (through its impact on interest rates) and government saving (through its impact on government revenue).

- Fifth, aid flows could undermine private sector incentives for investment or to improve productivity. Aid can cause the currency to appreciate, undermining the profitability of the production of all tradable goods (known as the Dutch disease). Food aid, if not managed appropriately, can reduce farm prices and hurt farmer income.

The last two points merit further discussion. On aid and saving, while foreign aid adds to total saving (since aid is a form of foreign saving), some studies have shown that a dollar of aid adds less than a dollar to total saving and investment, since domestic savings may fall as aid increases. Some of these studies conclude that aid is ineffective because it 'leaks' to consumption. This approach is not particularly helpful in the aggregate since large portions of aid are in fact designed specifically to directly increase consumption and not investment, including food aid, immunization programs, purchases of textbooks, technical assistance, and the like. Nevertheless, even where aid is aimed at investment, the impact could be partially offset by a reduction in either private saving (through a decline in the rate of return on private investment) or government saving (through a fall in tax revenues). There is a wide range of estimates of the offset effect, but most find that \$1 in aid translates to an increase in investment in the range of 33 to 67 cents. Much depends on the particular country, the type of aid, and other factors.

Aid also could undermine incentives for private sector activity. Aid can spur inflation and cause a real appreciation of the exchange rate, which reduces the profitability of production of all tradable goods, creating 'Dutch disease' effects.⁷ Aid flows can enlarge the size of the government and related services supporting aid projects, drawing workers and investment away from other productive activities such as agro-processing, garments or footwear exports. To the extent that these tradable activities are a key source of productivity gains, long-term growth may suffer. Similarly, food aid can sometimes undermine local food production if an influx of food drives down prices (it has less adverse impact on production when it displaces food imports).

The empirical studies that have found no relationship between aid and growth have been influential. However, very few published studies have reached that conclusion since the mid-1990s (a recent exception is Rajan and Subramanian, 2005a). Most of those that do use restrictive models that impose constraints such as a linear relationship between aid and growth, ruling out by assumption the possibility of diminishing returns. Most also only examine aggregate aid, imposing the restriction that all aid has a similar impact on growth, which is not particularly realistic, since famine

relief, immunization programs and road projects are all likely to have very different impacts on growth.

Aid has a conditional relationship with growth, helping to accelerate growth under certain circumstances

This view holds that aid supports growth in some circumstances but not others, and searches for key characteristics associated with the difference. This ‘conditional’ strand of the literature has three subcategories, with the effectiveness of aid depending on the characteristics of the recipient country, the practices and procedures of the donors, or the type of activity that the aid supports.

Recipient-country characteristics Isham et al. (1995) found that World Bank projects had higher rates of returns in countries with stronger civil liberties. Burnside and Dollar (2000), in a very influential study, concluded that aid stimulated growth in countries with good policies, but not otherwise. Other researchers have proposed different country characteristics that might affect the aid–growth relationship, including export price shocks, climatic shocks, the terms of trade, macroeconomic and trade policies, institutional quality, warfare, type of government and location in the tropics.⁸ All of these studies rely on an interaction term between aid and the variable in question, and (not surprisingly) many of the interaction terms are fragile. Easterly et al. (2004) find that the original Burnside and Dollar results do not hold up to modest robustness checks. Roodman (2007) tests several other ‘conditional’ studies and find most of them to be relatively fragile, although the conclusions of Dalggaard and Tarp (2004) are more robust.

Nevertheless, the view that aid works better (or in a stronger version, aid works only) in countries with good policies and institutions has become the conventional wisdom among donors, partly based on this research and partly due to development practitioners that believe this to be the case based on their own experience. The appeal of this approach is that it can explain why aid seems to have supported growth in some ‘well-behaving’ countries but not others. These findings have had an enormous impact on donors (World Bank, 1998). The concept feeds directly into the World Bank’s Performance-Based Allocation (PBA) system for distributing concessional IDA funds, and was the foundation for the United States’ new Millennium Challenge Account (Radelet, 2003).

Donor practices Many analysts have argued that donor practices strongly influence aid effectiveness. For example, multilateral aid might be more effective than bilateral aid, and ‘untied’ aid is thought to have higher

returns than 'tied' aid, as discussed previously. Many observers argue that donors that have large bureaucracies, do not coordinate with other donors, or have poor monitoring and evaluation systems undermine the effectiveness of their own programs. Two influential and overlapping views argue that aid would be more effective if there were greater 'country ownership' or broader 'participation' among government and community groups in recipient countries in setting priorities and designing programs. There has been substantial debate about these issues, and in some cases these ideas have begun to change donor practices. But to date there has been very little systematic research connecting specific donor practices to aid effectiveness.

Type of aid Different kinds of aid might affect growth in different ways. Clemens et al. (2004) disaggregated aid into types most likely and least likely to affect growth within a few years, if at all. They separated aid into three categories: (1) emergency and humanitarian aid (likely to be negatively associated with growth, since aid tends to increase sharply at the same time growth falls following an economic shock); (2) aid that might only affect growth after a long period of time, if at all, and so the relationship may be difficult to detect (such as aid for health, education, the environment, and to support democracy); and (3) aid that is directly aimed at affecting growth (building roads, ports and electricity generators, or supporting agriculture). It found a strong positive relationship between the third type of aid (about half of all aid) and growth, a result which stood up to a wide variety of robustness checks. As expected, the relationship with the other types was less detectable.

To summarize the aid and growth research, it appears that aid has been successful in some countries but not others. The overall trend is a subject of debate, but most research has found a positive relationship although the direction of causality is not always clear. This research is only beginning to scratch beneath the surface and investigate what types of aid are most effective and the conditions under which aid has the largest impact on growth. Since disputes continue about the determinants of economic growth more broadly, perhaps it is not surprising that the aid-growth relationship continues to be a matter of sharp debate.

Donor relationships with recipient countries

The criticisms about aid have led to debates about how aid programs can be improved to support growth and development more effectively. But the challenge is not easy. Aid programs face some inherent difficulties in trying to achieve a wide range of objectives, provide financial oversight and ensure results.

The principal–agent problem

A key issue facing aid agencies is that there is only an indirect and distant relationship between the people actually providing the financing – taxpayers in donor countries – and the intended ultimate beneficiaries of aid projects – poor people living in low-income countries. In most aid programs, there is a long and complex chain of principal–agent relationships, starting with the taxpayers that delegate authority to elected officials, who in turn become principals that delegate authority to a new set of agents, the heads of aid agencies, which delegate to agency employees, contractors and consultants. In the recipient country, there are similar relationships between citizens, their government and those that actually implement programs. The objectives, incentives and information available to these agents are not always well aligned with the objectives of either the taxpayers or the beneficiaries.

All public sector agencies and many private companies are faced with these principal–agent problems, but the international dimension and physical separation between the original taxpayers and ultimate beneficiaries makes it an even greater challenge for aid.⁹ In domestic public programs (such as rubbish collection or local schools) the taxpayers and ultimate beneficiaries are the same people, so they have clearer information about success or failure and can reward or penalize their agents accordingly by re-electing them or voting them out of office. But this feedback loop is broken for aid agencies. Taxpayers cannot tell if their money is well spent, beneficiaries sometimes do not even know about local programs, and each have limited mechanisms for penalties and rewards.

The principal–agent problem affects nearly all aspects of aid delivery including program design, implementation, compensation, incentives, evaluation and allocation of funding. The problem can never be fully solved – private companies face similar issues between owners, managers and employees, as do private aid foundations and charities. The challenge is to design institutions and incentives that mitigate these problems as much as possible to clarify goals, objectives, incentives and rewards. In this regard, one of the key challenges for donors is if, when, and how to apply conditions to their aid, a subject to which we now turn.

Conditionality

Partly as a result of the principal–agent problem, donors often apply conditions on aid programs to encourage recipients to act more in accord with the donors' (and possibly the ultimate beneficiaries') interests. Donor conditions on recipient actions or policies are among the most controversial aspects of aid. Policy conditionality is most often associated with the IMF and World Bank, but all donors use conditions to some extent.

The rationale for economic policy conditions is straightforward: donors believe that certain policies and actions in different countries are important for growth and development, and that without them providing aid is futile. If government policies have led to high rates of inflation, massive inefficiencies and waste of public spending, and extensive corruption, then providing aid – whatever the specific purpose – without requiring fundamental change would provide no benefits and perhaps could perpetuate damage. Some even argue that the primary purpose of aid is not the money, but for aid to act as a lever for the policy reforms.

There are three key problems with conditionality. First, it is not always clear what policy conditions are the most appropriate to ensure sustained growth and development. Development doctrine has swung from a state-led approach in the 1950s and 1960s, to basic human needs in the 1970s, to a macroeconomic approach focused on open markets in the 1980s and 1990s, to a greater focus on institutions beginning in the mid-1990s. As a result, the list of conditions is constantly evolving. Debate has raged for decades about whether specific IMF and World Bank conditions are justifiable and whether they support or hurt stabilization, growth and development. And who should bear the costs if donor-imposed conditions make things worse?

Second, while donors are often criticized for imposing too many conditions, they are almost as often criticized for not imposing enough conditions. Some advocates that criticize the IMF for imposing too much fiscal austerity also insist that it should require governments to spend a minimum amount on health and education. The World Bank is often asked to add conditions to force governments to take specific actions, for example on projects that have potential adverse environmental consequences.

Third, conditionality does not seem to work. Most analysts agree that governments implement reforms only when it is in their interests to do so, and donor conditions have little if any impact on that decision. Many donors continue to disburse aid even when recipients fail to meet conditions, sometimes repeatedly so. Donors are faced with their own internal incentives to continue to disburse aid to support the contractors and recipients that depend on it. They also face a ‘Samaritan’s dilemma’ that withdrawing aid would create short-term pain for the very people it is aimed to help.¹⁰

The nature of conditionality has changed over time as the most pressing issues have changed and as donors continue to wrestle with the best ways to apply conditions. During the 1980s, most conditions focused on macroeconomic issues, trade reforms and privatization, as reflected in IMF and World Bank-sponsored structural adjustment programs. During the 1990s as macroeconomic imbalances improved and following the end of the Cold

War, attention shifted to governance, corruption and institution-building. Debate has re-emerged as to whether aid should be conditioned on democratic reforms in recipient countries. Whether governance-focused conditionality is a good idea, or whether it will be more successful than structural and policy conditionalities, remains to be seen.

There are no clear-cut rules for conditionality. Striking the right balance between responsible oversight and accountability on the one hand, and ensuring against high bureaucratic obstacles and the imposition of unnecessary controls or unwarranted policy changes on the other, requires flexibility, judgment and the ability to balance multiple objectives – none of which are easy for aid agencies to achieve.

Improving aid effectiveness

The debates about the strengths and weaknesses of aid have led to specific ideas for change, some of which donors have begun to put into practice. Four stand out.

Country selectivity One influential idea is that donors should be more selective about the countries to which they provide aid, based on the view that aid works best in countries with good policies and institutions. In the strongest version, aid should be provided only to countries that meet these criteria. A more moderate view is that more aid should be allocated to countries with stronger policies and institutions, but some aid should be targeted to countries with weaker policies, especially post-conflict countries. This proposal turns the conditionality debate: instead of providing aid to encourage reforms, give it to countries that have already demonstrated a desire to implement key reforms. In the language of the principal-agent problem, donors should spend less time trying to write contracts that force an alignment of incentives and instead give more aid to countries that on their own demonstrate similar motivations and objectives. Some donors have begun to be more 'selective', including the World Bank in the allocation of its concessional IDA funds, some European donors in terms of providing budget support, and the USA with its new Millennium Challenge Account. But since so much aid is allocated for political, security and other foreign policy reasons, there are limits to how far donors are likely to go in this direction.

Recipient participation Many analysts argue that aid has been weakened by donor domination in setting priorities, designing programs and implementing projects, and push for either a more 'country led' approach in which recipient governments take a stronger role, or a 'participatory' approach in which various groups in recipient countries (government,

NGOs, charities, the private sector) play a more active role. The idea is to eliminate some of the problems in the long chain of principal-agent relationships, and more tightly integrate the ultimate beneficiaries in key aspects of the aid delivery process. The World Bank and IMF (by requiring Poverty Reduction Strategy Papers), the Global Fund to Fight AIDS, Tuberculosis and Malaria, and the Millennium Challenge Corporation have all moved towards greater local participation in designing and implementing the programs they finance. This approach is new, so there is no evidence yet on the extent to which (or the circumstances under which) it improves aid effectiveness. There is a clear and inescapable tension between country ownership on the one hand, and donor priorities and conditionality on the other. Donors are more likely to facilitate a participatory approach in countries in which governments show a strong commitment to sound development policies, and less so in countries with corrupt and dictatorial governments.

Harmonization and coordination Managing aid flows from many different donors is a huge challenge for recipient countries, since different donors usually insist on using their own unique processes for initiating, implementing and monitoring projects. Recipients can be overwhelmed by requirements for multiple project audits, environmental assessments, procurement reports, financial statements and project updates. According to the World Bank, developing countries typically work with 30 or more aid agencies across a wide variety of sectors, with each sending an average of five missions a year to oversee their projects.¹¹ The donors all want to meet with the same top government officials, leaving them with much less time to deal with pressing matters. These concerns have led to numerous suggestions for donors to coordinate their activities more closely, harmonize their systems or 'pool' their funds (Kanbur and Sandler, 1999). But while there has been some progress, the pace of change amongst the donors seems glacial.

Results-based management The emphasis on demonstrating the effectiveness of aid has led to calls for improved monitoring and evaluation and results-based management. In this view, aid programs should aim to achieve very specific quantitative targets, and decisions about renewing or reallocating aid going forward should be based on those results. There are three basic objectives: (1) helping donors allocate funds towards programs that are working; (2) detecting problems at an early stage to help modify and strengthen existing programs; and (3) improving the design of future programs. Stronger monitoring and evaluation would help improve principal-agent relationships so that aid agencies have clearer incentives and

taxpayers have better information about the impact of aid on its intended beneficiaries.

Summary and conclusions

Aid flows fell in the 1990s after the end of the Cold War and aid was widely attacked for being ineffective in spurring growth and development. However, aid began to grow again in the late 1990s and indications are that it will continue to grow throughout the first decade of the twenty-first century, although probably less rapidly than donors have pledged.

Most empirical research on aid and growth conducted since the mid-1990s has found a positive relationship, in contrast to popular perceptions, particularly studies that have allowed for diminishing returns and have controlled for other factors that affect growth. Some studies have found that the aid–growth relationship is conditional on the policy or institutional environment, but many of those results have been fragile. Some studies have concluded that there is no relationship or even a negative one, but while influential, these studies are few in number and tend to use restrictive assumptions. Recent research that has explored how different types of aid might have different impacts on growth has suggested one key reason why earlier research has reached mixed conclusions.

Nevertheless, there is little doubt that aid has been less effective in spurring development than is often expected. Aid can keep bad governments in power for too long, and can undermine incentives for saving, tax collection and private sector production. Aid relationships are made much more difficult by a complex chain of principal–agent problems that weaken information flows, introduce myriad motivations for different actors, and make monitoring and accountability more difficult. Attempts to solve the principal–agent problem through conditionality have not been very successful. The newest wave of reform efforts aims to solve some of the weaknesses of aid and the principal–agent problem through greater donor selectivity in choosing aid recipients, increased recipient participation in setting priorities and designing programs, streamlining aid bureaucracies, increasing donor coordination, and establishing clearer goals for aid and stronger monitoring and evaluation of aid-financed activities. These ideas have been very influential in designing aid programs in recent years, but there is no systematic evidence at this point as to whether these changes will lead to greater aid effectiveness.

Notes

1. This chapter draws heavily from Chapter 14 of Perkins et al. *Economics of Development*, 6th edn, 2006 (New York: W.W. Norton & Co.), (used by permission), and from Radelet et al. (2006). I thank Bilal Siddiqi and Sami Bazzi for their research assistance, and Amitava Dutt and Jaime Ros for comments on an earlier draft. I also thank the William and Flora Hewlett Foundation for financial support.

2. Non-concessional loans from donor agencies are counted as part of official development finance, but not as official development assistance.
3. More precisely, assistance to countries with per capita incomes (for three consecutive years) above the World Bank's 'high income' threshold, but the DAC makes some exceptions.
4. This summary draws heavily from the review in Clemens et al. (2004). For another recent review of the literature see Hansen and Tarp (2001).
5. Hadjimichael et al. (1995), Durberry et al. (1998), Dalgaard and Hansen (2000), Hansen and Tarp (2000, 2001), Lensink and White (2001), Dalgaard and Tarp (2004) and Clemens et al. (2004).
6. Griffin and Enos (1970); Mosley (1980); Mosley et al. (1987); Dowling and Hiemenz (1982); Singh (1985); Boone (1994) and Rajan and Subramanian (2005b).
7. See Younger (1992), Bulir and Lane (2002), Rajan and Subramanian (2005b) and Tressel and Prati (2006).
8. Burnside and Dollar (2000), Collier and Dehn (2001), Guillaumont and Chauvet (2001), Chauvet and Guillaumont (2002), Collier and Dollar (2002), Collier and Hoeffler (2002), and Dalgaard and Tarp (2004).
9. For an excellent discussion of the principal-agent problem in aid programs, see Martens (2004).
10. See for example Easterly (2001), Svensson (2003) and Kanbur (2006).
11. 'Cutting the Red Tape', World Bank Development News Media, 21 February 2003, available at <http://go.worldbank.org/BD8VODLW00>.

References

- Alesina, Alberto and David Dollar (2000), 'Who Gives Foreign Aid to Whom and Why?' *Journal of Economic Growth*, March: 33–63.
- Bauer, Peter, (1972), *Dissent on Development*, Cambridge, MA: Harvard University Press.
- Boone, Peter (1994), 'The Impact of Foreign Aid on Savings and Growth', Centre for Economic Performance Working Paper No. 677, London School of Economics.
- Bulir, Ales and Timothy Lane (2002), 'Aid and Fiscal Management', IMF Working Paper 02/112.
- Burnside, Craig and David Dollar (2000), 'Aid, Policies, and Growth', *American Economic Review*, 90 (4): 847–68.
- Chauvet, Lisa and Patrick Guillaumont (2002), 'Aid and Growth Revisited: Policy, Economic Vulnerability and Political Instability', paper presented at the Annual Bank Conference on Development Economics, Oslo, June.
- Clemens, Michael, Steven Radelet and Rikhil Bhavnani (2004), 'Counting Chickens When They Hatch: the Short-Term Effect of Aid on Growth', Center for Global Development Working Paper No. 44, November.
- Collier, Paul and Jan Dehn (2001), 'Aid, Shocks and Growth', Working Paper 2688, World Bank, October.
- Collier, Paul and David Dollar (2002), 'Aid Allocation and Poverty Reduction', *European Economic Review*, 45 (1): 1–26.
- Collier, Paul and Anke Hoeffler (2002), 'Aid, Policy and Growth in Post-Conflict Societies', Policy Research Working Paper 2902, World Bank.
- Dalgaard, Carl-Johan and Henrik Hansen (2000), 'On Aid, Growth, and Good Policies', CREDIT Research Paper No. 00/17, Centre for Research in Economic Development and International Trade, University of Nottingham.
- Dalgaard, Carl-Johan and Finn Tarp (2004), 'On the Empirics of Foreign Aid and Growth', *Economic Journal*, 114 (496): 191–216.
- Dowling, J. Malcolm and Ulrich Hiemenz (1982), 'Aid, Savings, and Growth in the Asian Region', Report No. 3, Economic Office, Asian Development Bank.
- Durberry, Ramesh, Norman Gemmill and David Greenaway (1998), 'New Evidence on the Impact of Foreign Aid on Economic Growth', CREDIT Research Paper No. 98/8, Centre

- for Research in Economic Development and International Trade, University of Nottingham.
- Easterly, William (2001), *The Elusive Quest for Growth: Economists' Adventures and Misadventures in the Tropics*, Cambridge, MA: Massachusetts Institute of Technology Press.
- Easterly, William, Ross Levine and David Roodman (2004), 'Aid, Policies and Growth: Comment', *American Economic Review*, 94 (3): 774–80.
- Friedman, Milton (1958), 'Foreign Economic Aid', *Yale Review*, 47 (4): 501–16.
- Griffin, K.B. and J.L. Enos (1970), 'Foreign Assistance: Objectives and Consequences', *Economic Development and Cultural Change*, 18 (3): 313–27.
- Guillaumont, Patrick and Lisa Chauvet (2001), 'Aid and Performance: A Reassessment', *Journal of Development Studies*, 37 (6): 66–92.
- Hadjimichael, Michael T., Dhaneshwar Ghura, Martin Muhleisen, Roger Nord and E. Murat Ucer (1995), 'Sub-Saharan Africa: Growth, Savings, and Investment 1986–93', IMF Occasional Paper No. 118, Washington: International Monetary Fund.
- Hansen, Henrik and Finn Tarp (2000), 'Aid Effectiveness Disputed', *Journal of International Development*, 12: 375–98.
- Hansen, Henrik and Finn Tarp (2001), 'Aid and Growth Regressions', *Journal of Development Economics*, 64: 547–70.
- Isham, Jonathan, Daniel Kaufmann and Lant Pritchett (1995), 'Governance and Returns on Investment: An Empirical Investigation', World Bank Policy Research Working Paper No. 1550, Washington, DC: World Bank.
- Kanbur, Ravi (2006), 'The Economics of International Aid', in Serge Christophe-Kolm and Jean Mercier-Ythier, (eds), *The Economics of Giving, Reciprocity, and Altruism*, North Holland, pp. 1559–88.
- Kanbur, Ravi and Todd Sandler (1999), 'The Future of Development Assistance: Common Pools and International Public Goods', Overseas Development Council Policy Essay No. 5, Washington, DC: Overseas Development Council.
- Lensink, Robert and Howard White (2001), 'Are There Negative Returns to Aid?' *Journal of Development Studies*, 37 (6): 42–65.
- Levine, Ruth, and the 'What Works' Working Group (with Molly Kinder) (2004), *Millions Saved: Proven Success in Global Health*, Washington, DC: Center for Global Development.
- Levy, Victor (1988), 'Aid and Growth in Sub-Saharan Africa: The Recent Experience', *European Economic Review*, 32 (9): 1777–95.
- Martens, Berten (2004), 'Introduction' in Berten Martens (ed.), *The Institutional Economics of Foreign Aid*, Cambridge: Cambridge University Press.
- Mishra, Prachi and David Newhouse (2007), 'Health Aid and Infant Mortality', International Monetary Fund Working Paper WP/07/10, April.
- Mosley, Paul (1980), 'Aid, Savings, and Growth Revisited', *Oxford Bulletin of Economics and Statistics*, 42 (2): 79–96.
- Mosley, Paul, John Hudson and Sara Horrell (1987), 'Aid, the Public Sector and the Market in Less Developed Countries', *Economic Journal*, 97 (387): 616–41.
- Organisation for Economic Co-operation and Development (2008), *2007 Development Cooperation Report*, Paris: OECD.
- Papanek, Gustav, F. (1973), 'Aid, Foreign Private Investment, Savings, and Growth in Less Developed Countries', *Journal of Political Economy*, 81 (1): 120–30.
- Perkins, Dwight, Steven Radelet and David Lindauer (2006), *Economics of Development*, 6th edn, New York: W.W. Norton and Company.
- Radelet, Steven (2003), *Challenging Foreign Aid: A Policymaker's Guide to the Millennium Challenge Account*, Washington, DC: Center for Global Development.
- Radelet, Steven, Michael Clemens and Rikhil Bhavnani (2006), 'Aid and Growth: The Current Debate and Some New Evidence', in Peter Isard, Leslie Lipschitz, Alexandros Mourmouras and Peter Hellas (eds), *Macroeconomic Management of Foreign Aid: Opportunities and Pitfalls*, Washington, DC: International Monetary Fund.
- Rajan, Raghuram and Arvind Subramanian (2005a), 'Aid and Growth, What Does the Cross-Country Evidence Really Show?' IMF Working Paper No. 05/127, June.

- Rajan, Raghuram and Arvind Subramanian (2005b), 'What Undermines Aid's Impact on Growth', IMF Working Paper 05/126, International Monetary Fund.
- Roodman, David (2007), 'The Anarchy of Numbers: Aid, Development and Cross-Country Empirics', *The World Bank Economic Review*, 21 (2): 255–77.
- Sachs, Jeffrey, John W. McArthur, Guido Schmidt-Traub, Margaret Kruk, Chandrika Bahadur, Michael Faye and Gordon McCord (2004), 'Ending Africa's Poverty Trap', *Brookings Papers on Economic Activity*, 1: 117–240.
- Singh, Ram D. (1985), 'State Intervention, Foreign Economic Aid, Savings and Growth in LDCs: Some Recent Evidence', *Kyklos*, 38 (2): 216–32.
- Stern, Nicholas (2002), 'Making the Case for Aid', in James D. Wolfensohn (ed.), *World Bank, A Case for Aid: Building a Consensus for Development Assistance* (Washington, DC: World Bank, pp. 15–24.
- Stiglitz, Joseph (2002), 'Overseas Aid is Money Well Spent', *Financial Times* (April 14).
- Svensson, Jakob (2003), 'Why Conditional Aid Does Not Work and What Can Be Done About It', *Journal of Development Economics*, 70 (2): 381–402.
- Tresselt, Thierry and Alessandro Prati (2006), 'Aid Volatility and Dutch Disease: Is There a Role for Macroeconomic Policies?', IMF Working Papers 06/145, International Monetary Fund.
- World Bank (1998), *Assessing Aid: What Works, What Doesn't, and Why*, Oxford: Oxford University Press.
- Younger, Stephen (1992), 'Aid and the Dutch Disease: Macroeconomic Management When Everybody Loves You', *World Development*, 20 (11): 1587–97.

43 International migration and the brain drain

Francisco L. Rivera-Batiz

Introduction

International migration flows have expanded by historical proportions in recent years. In 1960, there was a stock of slightly over 75 million people residing in countries other than their country of birth. By 2005, this number had grown to 190 million. For many developing countries, international labor flows have become a natural outcome of the globalization process, as much as trade and investment flows. The money the migrants send back home has become a major source of income for families in developing countries. In 2005, migrant remittances amounted to \$600 billion. For some countries, the income received from the services of workers abroad is now a major item of the balance of payments.

But international migration flows are not without their costs. For many years, the issue of the brain drain has been studied by international economists. From software engineers in India to doctors in the Philippines and nurses in sub-Saharan Africa, the exodus of skilled migrants has been a policy concern for many developing countries. The impact on the skilled labor forces of some countries has been substantial, especially in sub-Saharan Africa. In Ghana, for example, over 40 percent of persons with a college degree or more have migrated to other countries. In Gambia, the corresponding proportion is close to 65 percent, and in Somalia it is 59 percent. What are the consequences of these labor outflows? What benefits or costs do they impose in the sending nations?

This chapter provides an analysis and survey of the key recent trends in international migration, their determinants and major consequences. The next section presents data on the growth of global migration flows since the mid-1950s. The section after that proceeds to examine the causes of these migration flows, focusing on the main economic factors involved. The subsequent section analyzes the consequences of international migration for source countries. The concluding section discusses implications for future research.

Trends in international migration flows

International migration refers to the movement of people across national borders. Data on migration are available directly from national immigration

authorities in recipient countries. However, the data available for many countries are sketchy and difficult to compare with data for other countries due to differences in migration policies, definitions of what constitutes immigrants, the presence of undocumented migration, and so on. Some international organizations gather cross-country data on migration and seek to provide more uniform, comparable statistics. The Organisation for Economic Co-operation and Development (OECD) has for many years collected information on the migration of OECD countries and has an extensive database for these countries (see, for example, OECD, 2006). The United Nations has the most comprehensive worldwide database on the number of migrants residing in different countries (United Nations, 2007). There are also data collection efforts supported by the World Bank (Docquier and Marfouk, 2006), the International Monetary Fund (Carrington and Detragiache, 1998), and other institutions such as the Development Research Centre on Migration, Globalisation and Poverty at the University of Sussex (see Parsons et al., 2007).

Table 43.1 shows the massive growth in the estimated number of international migrants between 1960 and 2005. In 2005, there were 190 million people residing in countries other than their country of birth, up from 75 million in 1960. Most of these migrants originated in developing countries: it is estimated that 123 million migrants, or 65 percent of the total, were born in developing countries. Mexico had the highest number of persons residing abroad, equal to over 10 million in 2000, largely in the United States. This was followed by India, which had close to 9 million persons residing abroad in 2000, with the country's diaspora spread all over the world. Other countries with mass emigration include Bangladesh (6.6 million emigrants), the Philippines (3.4 million), Pakistan (3.4 million), Turkey (3.0 million), Afghanistan (2.7 million), Morocco (2.6 million), Egypt (2.5 million) and Algeria (2.1 million), among others.

The destination of emigrants from developing countries is evenly split between high-income and developing countries. In 2005, close to 62 million people born in developing countries were residing in high-income economies, but 61 million resided in other developing nations. Among high-income countries, those with the largest number of immigrants included the United States (38.4 million), Germany (10.1 million), France (6.4 million) and Canada (6.1 million). Developing countries with the highest immigrant populations included Russia (12.1 million, who moved mostly from elsewhere in the former Soviet Union), Ukraine (6.8 million), Saudi Arabia (6.4 million), India (3.3 million) and the United Arab Emirates (3.2 million).

The emigration of skilled workers is one of the major concerns relating to migration flows in developing countries. To measure the magnitude of

Table 43.1 Estimates of the Stock of International Migrants in the World, 1960–2005

Year	Stock of migrants		Change between years
1960	75 463 352		
1965	78 443 933]	2 980 581
1970	81 335 779]	2 891 846
1975	86 789 304]	5 453 525
1980	99 275 898]	12 486 594
1985	111 013 230]	11 737 332
1990	154 945 333]	43 932 103
1995	165 080 235]	10 134 902
2000	176 735 772]	11 655 537
2005	190 633 564]	13 897 792

Source: United Nations (2007).

skilled emigration or brain drain, a country’s labor force is divided into those who are skilled – generally considered to be those who have some tertiary education, that is, 13 years of schooling or more – and those who are less skilled, who have achieved less than tertiary education. The emigration of the skilled has been rising, just as global migration flows have increased in general. In OECD countries, for example, there were 12.5 million skilled immigrants of working age in 1990, but by 2000 the number had risen to 20.4 million, equal to 34.6 percent of the total number of immigrants in these countries.

The skilled emigration rate is defined as the stock of skilled migrants from a country (all persons with tertiary education living abroad) calculated as a percentage of the total skilled labor force in the source country augmented by the skilled migrants themselves. This shows the percentage of workers with tertiary education who were born in a country but are residing outside its borders. Hence, it is a measure of the relative impact of the emigration on the sending country’s skilled labor market. The highest skilled emigration rates in the world prevail in the Caribbean, where in 2001

as much as 42.8 percent of the region's tertiary labor force resided outside its borders (Docquier and Marfouk, 2006). Skilled emigration has also been enormous for a number of countries in Africa. In East Africa, the skilled emigration rate was close to 20 percent in 2001.

The determinants of international migration flows

What has caused the mass migration documented in the previous section? There is a massive literature examining the determinants of migration flows. This section presents the main forces and conceptual approaches that seek to explain migration flows.

The economic approach to migration: costs and benefits

At the level of theory, the classical economic model of the decision to migrate was formalized by Sjaastad (1962) and has been extended in a number of directions (see, for example, Lucas, 1985; Borjas, 1999). In this approach, the decision to migrate is seen as an investment decision that depends on individual assessments of the net balance of the present and future costs and benefits of migration. For a worker i , the net gain from migration, G_i , is equal to the present discounted value of the benefits minus the costs of migrating:

$$G_i = \sum_{t=1}^T [B_{it}/(1+r)^t - C_{it}/(1+r)^t] \quad (43.1)$$

where T is the lifespan of the worker, r is a discount rate, B_{it} are the benefits at any given time t , generally in the form of higher wages or improved employment opportunities in the destination country, and the costs, C_{it} , include the direct costs of the move (transportation costs), the foregone earnings when the individual migrates, and any utility losses associated with leaving the homeland.

Note that different individuals will face varying costs and benefits, and the probability of a person migrating from an origin to a destination area will vary. Older workers, for example, may suffer the greatest net losses in foregone earnings and may also face higher psychic costs of leaving the source country. As a result, the likelihood of migration declines with age. Location also matters. Indeed, most migration flows occur among geographically close, often neighboring, countries, where the costs of migration are lower. Empirically, the role of geographical and cultural proximity in determining migration movements has been emphasized by the so-called 'gravity model of migration flows', for which there is ample supporting evidence (see Karemera et al., 2000; Hatton and Williamson, 2005).

Labor markets and the returns to international migration

Due to the difficulties of estimating future costs and benefits, most studies examining the decision to migrate focus on the current labor market gains from migrating, I_i , given by:

$$I_i = \log W_{iD} - \log W_{iO} \quad (43.2)$$

where W_{iD} is the wage rate that individual i can obtain in the destination and W_{iO} is the equivalent wage at home. These wage rates are influenced by the various characteristics of workers, including their schooling, Ed_i , their on-the-job experience, Ex_i , their motivation, M_i , health, marital status, number of children, and so on. Suppose, for simplicity, that the following characterizes the wages of individual i in the destination and origin regions:

$$\log W_{iD} = \alpha_D + \alpha_{1D}Ed_i + \alpha_{2D}Ex_i + \alpha_{3D}M_i \quad (43.3)$$

$$\log W_{iO} = \alpha_O + \alpha_{1O}Ed_i + \alpha_{2O}Ex_i + \alpha_{3O}M_i \quad (43.4)$$

where the α 's represent how the various individual characteristics (education, experience, and so on) are rewarded in each country. The net gain from migration is thus:

$$I_i = \log W_{iD} - \log W_{iO} = (\alpha_D - \alpha_O) + (\alpha_{1D} - \alpha_{1O})Ed_i + (\alpha_{2D} - \alpha_{2O})Ex_i + (\alpha_{3D} - \alpha_{3O})M_i \quad (43.5)$$

For any individual, the incentive to migrate will vary on the basis of the skills (education, experience, and so on) that the worker is endowed with, his or her motivation, and how skills as well as motivation are rewarded in relative terms in the source and destination countries.

A large body of research has now accumulated studying the labor market outcomes of developing-country emigrants in their destinations (see, for example, the collection of research in Zimmermann and Constant, 2004, as well as the surveys by Borjas, 1999 and Hanson, 2006). This literature suggests that the economic returns to migrating are substantial for most workers. But the economic progress of migrants varies according to the characteristics of the migrants themselves (age, schooling, immigration status, and so on), the timing of the migration, and the country of destination. In some European countries, for instance, immigrants have substantially lower rates of labor force participation as well as higher unemployment rates than the native-born population. The relatively poor labor market outcomes of some immigrants are partly related to lack of skills, but they also depend on labor market conditions in – and

time spent in – recipient countries (Fertig and Schmidt, 2002; Rivera-Batiz, 2007).

Migrant selectivity

One of the most discussed issues in the international migration literature is whether emigrants are the most qualified, skilled workers in the origin economy or not. If migrants are positively selected, then they will be more likely to succeed abroad, but their exit from the source country will drain the most capable, most skilled population from the nation, with a potentially negative impact on its economy and society.

What determines whether migrants are positively or negatively selected? There are forces that favor a positive selectivity and others that encourage a negative selectivity. The best-known hypothesis is that emigrants tend to be positively selected because in order to compensate for the substantial costs of migration, only those who have the strongest drive and motivation – and the expectations of great rewards – will actually undertake the migration process (see Chiswick, 1978, 1999). Of course, if costs of migration decline, then this aspect of selectivity will tend to become less significant. In addition, as equation (43.5) suggests, the impact of motivation on migration depends on whether motivational skills are more strongly rewarded in the destination region. Indeed, one suspects that holding constant the distribution of motivational skills in a population, if these skills are poorly rewarded at home but richly rewarded abroad, the incentives to migrate from source to destination regions will increase.

A second hypothesis (referred to in the literature as the ‘Roy model’) is that, if those with greater skills or abilities are rewarded more highly compared to the less-skilled in the origin area when compared to the destination region, this will generate less incentives for those at the top of the skills or ability distribution to emigrate compared to those at the bottom of the distribution, causing a negative selectivity of migrants. Therefore, the emigrant contingent will be positively or negatively selected depending on the relative inequality of the distribution of income at home and abroad. For instance, the more unequal the income distribution in the origin area as compared to the destination, the lower the incentives for the highly skilled to emigrate relative to the less skilled. This result is easily obtained from equation (43.5). Since $\partial I_i / \partial Ed_i = (\alpha_{1D} - \alpha_{1O})$, if greater inequality in the source country is associated with a rate of return to education in the origin area that exceeds the rate of return in the destination country, then $(\alpha_{1D} - \alpha_{1O}) < 0$. This implies that, as the education of the prospective migrant rises, the incentives to migrate tend to decline. Note also that under asymmetric information, employers in the destination region may not be able accurately to assess the skills of the migrants and they may offer lower

wages to the highly skilled migrants, when compared to employers in the source country, that are able to assess more accurately the potential migrants' skills and pay them wages more consonant with their skills. This will again reduce the rate of return to education received by the emigrants in the destination region relative to the source country ($(\alpha_{1D} - \alpha_{1O})$ will decline), shrinking the incentive to migrate of the highly skilled relative to the unskilled (see Stark and Taylor, 1991).

Some evidence appears to support the implications of the Roy model (see Borjas, 1987, 2006). For instance, recent research finds a negative selectivity in the migrant contingent from Mexico to the United States (that is, the emigrant group tends to have lower average skills than the population left behind), as would be expected from the relatively more unequal distribution of income in Mexico relative to the United States (see Fernandez-Huertas Moraga, 2007, for this analysis and Chiquiar and Hanson, 2005, for alternative results).

Income differences and international migration

Overall, the evidence on the importance of economic factors in motivating migration flows, as presented in equations (43.1)–(43.4) is extensive. Both documented and undocumented migration flows have been found to be strongly correlated to the relative economic conditions in recipient and source countries (see, for example, Adams, 1993; Hanson and Spillimbergo, 1999; Drinkwater, 2003; Castaldo et al., 2005).

But despite the widespread support for the hypothesis that increased income differentials between recipient and source countries stimulates migration, there is also ample support for the view that this connection does not always work and may actually hold in reverse (see Hatton and Williamson, 1998, and the review by Waddington and Sabates-Wheeler, 2003). For instance, in recent research seeking to determine the impact of differences in income per capita on migration flows in the world, Hatton and Williamson (2005, p. 240) find that in sub-Saharan Africa, increases in income at home increase migration. Indeed, the evidence suggests that the relationship between income per capita in source countries and emigration has an inverted-U shape. For poor countries, as income rises, migration actually increases. But as the wealth of a country grows, at some point, further increases in income per capita actually reduce emigration.

One explanation for this behavior is the fact that, at low levels of income per capita, a large part of the population just simply cannot afford the monetary costs of migrating. However, as per capita income in a country rises, this allows some people to save enough to pay for the transport and other costs of migration, thus increasing migration flows (see Hatton and

Williamson, 2005). An additional explanation is that the massive structural changes occurring in the early industrialization of an economy (the shift from agriculture to industry and services, from urban to rural areas, and so on) leads to a dislocation of the population that fosters international migration. As the economic development process matures, however, these changes diminish and migration declines. A third hypothesis for the paradoxical rise of migration flows as income increases in many economies is based on ‘relative deprivation theory’. It suggests that, as inequality rises in the early years of economic development (a trend first noted by economist Simon Kuznets), those who become relatively poor will become increasingly dissatisfied with their relative standing in the community. This will stimulate them to emigrate in order to improve their standard of living (see Stark and Taylor, 1991).

Families and the decision to migrate

Despite its powerful role in explaining migration flows, another problem of using a simple economic approach based on income differences across countries is that it cannot explain temporary migration. If there is a significant and persistent wage and employment differential between origin and destination regions, why do so many migrants wish to stay in the destination only for short periods of time?

One explanation is offered by the so-called new economics of labor migration (see Stark and Bloom, 1985). In this approach, it is understood that migration decisions are often made by families and households, not individuals. In contrast to the analysis in equations (43.1)–(43.4), the decision-making is not in the hands of an individual who maximizes his or her utility. Rather, migration decisions are made by families or households that maximize household utility over time. This utility, U , is derived from a stream of consumption by family members located in the home and/or destination regions, C_D and C_O , added all the way up to the time horizon of the family, T , and discounted to the present time, t :

$$U = \int_t^T \exp[-\rho(\tau - t)] \log U[C_D(\tau), C_O(\tau)] d\tau \quad (43.6)$$

where ρ is a rate of discount. Equation (43.6) is maximized subject to a budget constraint establishing that the present discounted value of income earned in the source and destination areas is equal to or greater than the present discounted value of family consumption spending.

Visualizing migration as an intertemporal family or household decision can help understand that the migration of some household members may be part of the savings and investment decisions of a family. The idea is that, if low-income households encounter capital market imperfections at home,

which exclude them from access to the financing of investments in housing, durable goods or in self-employed businesses, migration abroad may lead to the accumulation of remittances that can then be used to finance those purchases and investments. Migration becomes a short-term activity needed by households to raise funds in the absence of local financing. Another application of this approach is that the migration of one household member may reduce the costs of migration for other household members. This is what 'network or chain migration theory' suggests (see Piore, 1979; Bauer et al., 2000).

The political economy of immigration policy

The discussion so far has described the wide array of forces that may generate a supply of potential migrants. However, in an international context, destination-country governments exert an enormous force in restricting immigration flows. As a result, the volume of migration between developing countries and high-income economies is often determined by the immigration policies imposed by the latter.

Research on the determinants of immigration policies follows a growing literature in political economy that seeks to answer the question of how the policies of a country are generated through the interaction of economic, political and social forces (Mayda and Patel, 2006). A wide array of economic forces may lie behind the setting of immigration policies. For instance, if immigration reduces wages and raises the profits of employers and owners of capital, then persons who own firms or farms or own relatively large amounts of capital will be in favor of immigration, but those who do not have wealth and only have their labor will be against immigration (see Benhabib, 1996). But if the distribution of capital in a country is highly concentrated, with a great part of the workforce laboring at low wages, then there may be very few persons supporting immigration and many opposing it. If immigration policy is determined by influence of voters, immigration policy restrictions may be high. This force is magnified if the immigrants are unskilled since in this case the immigrants may be perceived as competing with the poor, generating stronger cries for immigration restrictions (Hatton and Williamson, 2005).

The available evidence confirms the role that economic forces play in affecting the immigration attitudes of natives in recipient countries (see Gang et al., 1999, 2002; Scheve and Slaughter, 2001; Mayda, 2006). But economic factors are not the only ones affecting attitudes towards immigrants. Social phenomena, such as xenophobia and bias against foreigners, can have a major influence on immigration policies (see Gang and Rivera-Batiz, 1994; and Gang et al., 2002, for analyses of the determinants of attitudes towards immigrants).

International migration and economic development

What are the economic consequences of mass migration for the developing world? What changes in the source countries are generated by migrants? This section examines the existing theory and evidence on how emigration affects developing nations.

The distributional and welfare impact of emigration

The simplest theoretical framework that can be used to examine the effects of international migration is one that focuses on the aggregate economy, within a simplified setting where there are only two inputs: capital and labor (see Bhagwati and Rodriguez, 1975; Borjas, 1999; Mishra, 2006). Figure 43.1 shows the labor market in the source country before and after emigration. The demand for labor (the economy's marginal value product curve) is L^D and the supply of labor before emigration is depicted by L^S , where it is assumed that there is an inelastic supply of labor given by the country's labor force. Under the assumption that the labor market is perfectly competitive and generates full-employment, the equilibrium before emigration is at point A , with a wage rate equal to W^* and employment E^* . If the number of emigrant workers is given by M , the labor force declines, shifting the aggregate supply of labor to the left, to $L^S - M$. The equilibrium after migration occurs at point B . The shortage of labor induced by emigration reduces employment from E^* to E^{**} , and raises

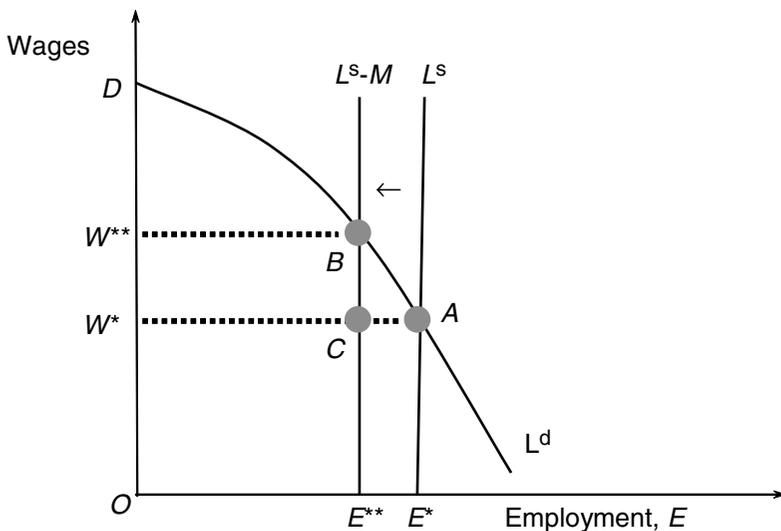


Figure 43.1 *The effects of emigration on the source country*

wages rates from W^* to W^{**} (this assumes that the domestic capital stock is unaffected by emigration).

The impact of emigration can be depicted in Figure 43.1 by noting first that national income is given by $Y = WL + rK$, with K equal to the economy's capital stock and r the rate of return to capital. Diagrammatically, after emigration, the income received by the workers that remain in the economy is equal to $W^{**}E^{**}$, as represented by the rectangle $W^{**}BE^{**}O$. This rises compared to the situation before emigration, when the income of these workers was W^*E^{**} . The gain in income by labor is $(W^{**} - W)E^{**}$. But although labor gains by emigration, capital is hurt by it. The income received by capital is given by $Y - WL$. Diagrammatically, the value of national income, Y , is equal to the area below the demand for labor curve up to the level of employment. After emigration, the value of national income is $DBCE^{**}O$. As a result, after emigration, the income received by capital is $Y - WL = DBCE^{**}O - W^{**}E^{**} = DBW^{**}$. The income derived by capital before the emigration was $DBACW^*$. Hence, the income of capital declines by $DBACW^* - DBW^{**} = W^{**}BAW^*$.

In this context, emigration results in a redistribution of income from capital to labor. But the emigration has also a net, overall negative impact on the income of those left behind. Adding the loss of capital and the gain to labor leaves a net loss represented diagrammatically by the area BAC . Algebraically, this loss can be approximated by:

$$(\Delta Y)/Y = - (1/2)(\Delta W)M = - (1/2)S_L\eta_{LL} (M/L)^2 \quad (43.7)$$

where ΔW is the change in wages ($\Delta W = W^{**} - W^*$), M is the loss of labor ($M = -\Delta E = E^* - E^{**}$), and where $\eta_{LL} = -(\partial W/\partial E)(E/W)$ is the negative of the elasticity of the labor demand curve with respect to wages. The effects on the income of labor and capital are then: $(\Delta WE)/Y = S_L\eta_{LL}(M/L)(1 - (M/L)) > 0$ and $(\Delta rK)/Y = -S_L\eta_{LL}(M/L)(1 - (M/2L)) < 0$.

Although in this simplified setting emigration results in a gain to labor (the mobile factor), a loss to capital (the fixed factor) and a net loss to the overall economy, these conclusions can be reversed in more complex frameworks. First of all, the model is a partial-equilibrium, closed-economy model but most modern economies are both highly diversified and engage in substantial international trade. In a general equilibrium framework with trade, such as the Hecksher–Ohlin–Samuelson (HOS) model, emigration has no lasting impact on the source country, whether on income distribution or in terms of the net impact on economic welfare. The explanation is that the emigrants themselves induce a reduction in the demand for labor in the source country. As the migrants leave, they do place upward pressure on wages, as noted in the earlier model. But this makes employment in

capital-intensive sectors, such as heavy manufacturing, more profitable. As production shifts away from labor-intensive products and into capital-intensive sectors, the overall demand for labor in the economy shrinks. This puts downward pressure on wages, reversing the initial impact of the emigration. This result, where an outflow of labor leads to no change in wages, is based on the Rybczynski effect in the trade literature (see Bhagwati and Rodriguez, 1975; Rivera-Batiz, 1983).

But the assumptions of the HOS model under which this result is derived are stringent. Besides the assumptions of perfect competition and no distortions, the framework assumes the absence of externalities and increasing returns. In addition, the model does not distinguish between skilled and unskilled labor emigration. Once these assumptions are relaxed, the theoretical analysis may yield complex and ambiguous effects of emigration (see, for example, Dutt, 2005). Furthermore, when considering costs and benefits from migration one should consider as well the overall, global impact of migration flows. If there are net world gains, there are then strong reasons for governments in recipient and destination areas to coordinate their migration policies, so that both developing countries and high-income economies can profit from the migration flows (see Pritchett, 2006; Kapur and McHale, 2005).

The empirical evidence on the welfare and distributional effects of emigration on source countries is scant (the analysis of the impact of immigration has been more extensive). Recently, however, some studies have utilized Census data over time to examine the issue. Mishra (2006) examines emigration from Mexico to the USA, concluding that the outflow of Mexican workers to the United States between 1970 and 2000 has increased worker earnings in Mexico by 5.9 percent of GDP and has reduced the income of the owners of fixed factors by 6.4 percent of GDP, with a small, negative net impact on overall economic welfare of those left behind (see also Borjas, 2006). On the other hand, there are other potential impacts of emigration to consider that are not examined in this literature, as discussed next.

Remittances and the impact of emigration

One of the most visible impacts of the migrants on source countries is connected to the remittances that they send back home. These flows of resources have grown exponentially in recent decades. Measured in 2000 international purchasing power parity (PPP) dollars, developing countries received \$50 billion in migrant remittances in 1980, but by 2005 this had multiplied to \$605 billion (World Bank, 2007). The regions receiving the most remittances were East Asia and the Pacific, and South Asia (each receiving about \$171 billion in 2005), followed by Latin America and the

Table 43.2 *Migrant remittances in developing countries, largest recipients, 2005*

Country/region	Remittances in current \$	Remittances in PPP-adjusted \$	Remittances as % of merchandise exports
Developing countries	179 425	605 678	6.0
Mexico	21 772	32 222	10.3
India	19 843	105 564	26.2
Philippines	11 634	45 605	29.2
Lebanon	5 722	5 493	335.3
Morocco	4 221	12 325	42.5
Serbia/Montenegro	4 129	9 868	103.8
Pakistan	3 955	14 277	29.5
Bangladesh	3 583	15 407	43.1
Brazil	3 540	7 682	3.0
Colombia	3 345	10 704	15.8
Egypt	3 341	8 018	43.5
Guatemala	2 592	5 962	88.2
El Salvador	2 564	5 718	77.8
Dominican Republic	2 471	6 671	43.0
Algeria	2 460	7 577	7.9
Jordan	2 288	5 010	58.9

Source: Data for remittances in current dollars are taken from World Bank (2007); other indicators are author's calculations using PPP adjustments and exports from World Bank (2007).

Caribbean (with \$86 billion), Europe and Central Asia (\$78 billion), the Middle East and North Africa (\$75 billion) and sub-Saharan Africa (\$24 billion).

The significance of remittances in many developing countries can be seen by comparing the value of remittances with the value of the merchandise goods exported by the source countries. Table 43.2 shows how significant migrant remittances can be, rising in some countries to over 100 or 300 percent of exports.

Remittances clearly constitute an improvement in the standard of living for family members who are recipients of such income. And recent evidence suggests that remittances are connected to lower poverty levels (see Acosta et al., 2006; Adams, 2007). Some questions have been raised as to the extent to which the remittances simply raise current consumption instead of stimulating investment and future economic growth. Recently, however, a number of studies have documented that, first, a significant portion of so-called consumption spending consists of household investments in

housing, automobiles and durable goods, whose long-term wealth-raising capacities are substantial; second, the use of remittances for community investment projects is not insignificant and also acts to stimulate local development; and, third, the multiplier effects of the increased consumption spending in generating local economic activity may be substantial (see, for example, de la Garza and Lowell, 2002; Adams, 2007).

On the other hand, remittances do tend to be spent largely on internationally non-tradable goods, which can result in rising prices of these goods compared to internationally traded goods, generating an appreciation in the value of the domestic currency, which adversely affects domestic export sectors (Rivera-Batiz, 1986). Evidence of this Dutch disease-type effect has emerged recently (see Amuedo-Orantes and Pozo, 2004; Acosta et al., 2007). A vicious cycle may develop, where emigration leads to remittances that then deteriorate domestic competitiveness and growth, leading to further emigration, and so on.

The impact of the brain drain

The contribution of migrant remittances for economic development must be weighted against any externalities generated by the migration flows. In countries where emigration leads to the loss of the most talented and skilled, the so-called brain drain, migration could result in significant negative externalities (see Bhagwati, 1979; Bhagwati and Rodriguez, 1975).

If the emigration of skilled labor is substantial and these workers are employed in local service sectors, the result can be acute shortages in the supply of essential services, from school teachers to professors and nurses. Note that the emigration of workers employed in sectors that produce exports and imports is not subject to these effects because local consumers can import these products from abroad when the laborers leave the country. But when the workers are employed in service sectors that produce internationally non-traded goods, the impact of emigration is more significant and potentially disastrous, because domestic consumers can only obtain those services locally (see Rivera-Batiz, 1982). If doctors and nurses emigrate, the supply of health services can collapse, resulting in higher prices and acute shortages. A brain drain can therefore reduce sharply the economic welfare of those left behind.

On the other hand, there may be positive externalities of skilled emigration on source countries. First of all, some researchers have recently suggested that a brain drain may actually raise the level of schooling of the population in the source country, at least in the long run. There are several reasons for this. One hypothesis is that the brain drain will raise the rate of return to education and, as a result, more young people in the country will decide to pursue higher education, thus raising educational attainment.

Indeed, the point that the emigration of skilled workers may lie behind the rising relative wages of skilled labor in many developing countries has frequently been made. However, evidence that these changes have stimulated local investments in human capital has not been produced yet (see Schiff, 2005).

Another possible positive externality associated with the brain drain relates to the fact that emigrants may generate international networks that could enhance the scientific and technological capacities at home. One example is the role that has been played by the Indian diaspora in Silicon Valley and elsewhere in the growth of India's information technology sector. By generating greater flow of skills and information among countries, and by raising the reputation of a domestic sector internationally, this type of emigration can have significantly positive effects at home (see Kapur and McHale, 2005).

Implications for future research

Both the theory and empirical evidence on the determinants and consequences of international migration have grown enormously in recent years. This research matches the growing importance of migratory flows. But despite the progress, the literature so far has generally failed to consider the dynamic effects of emigration, focusing instead on analyzing – and estimating – static models. This is an especially relevant issue to discuss in relation to the brain drain since there are a number of possible mechanisms through which the emigration of the skilled can affect a country's economic growth.

First of all, the mainstream analysis of economic growth, as developed by Solow and Swan, suggests that a drop in population growth should lead to an economic expansion for those left behind since the steady-state amount of capital per worker in the economy would rise, increasing per capita income. But there are a number of caveats to this story. First, the emigration of working-age people means that the dependency rate in the economy rises, which tends to absorb resources that would otherwise be dedicated to the accumulation of capital and economic growth. The working-age skilled emigrants are also more likely to have higher savings rates than the rest of the population. Indeed, the empirical evidence available is consistent with a negative impact of a brain drain due to these demographic effects (Bloom and Williamson, 1998).

Another issue is that the emigration of skilled labor may reduce the human capital available to those left behind. This can potentially have a devastating effect on economic growth. One of the main sources of economic growth is technological change, which depends to a large extent on human capital (Romer, 1990). But if human capital flees a country, then the

ability of those left behind to sustain innovation and technical change may be compromised, thus reducing an economy's economic growth (Rivera-Batiz, 1996).

On the other hand, a growing literature suggests that in countries where the quality of the public sector governance is low, many educated workers tend to be employed in activities that are not necessarily high-productivity activities. In other words, in economies where the government is highly bureaucratic, where corruption is rampant, and the rule of law does not apply, many highly educated workers will not be able to be gainfully employed (see Rivera-Batiz, 2002). In this case, the potentially negative impact of the brain drain on technological change and, therefore, on economic growth diminishes. It is a matter for future research to examine theoretically and empirically in greater detail how emigration is connected to technical change and economic growth.

References

- Acosta, P., C. Calderón, P. Fajnzylber and H. Lopez (2006), 'Remittances and Development in Latin America', *World Economy*, 30: 957–87.
- Acosta, P., E. Lartey and F. Mandelman (2007), 'Remittances and the Dutch Disease', Working Paper No. 2007-8, Atlanta: Federal Reserve Bank of Atlanta.
- Adams, R.H. (1993), 'The Economic and Demographic Determinants of International Migration in Rural Egypt', *Journal of Development Studies*, 30 (1): 146–67.
- Adams, R. (2007), 'International Remittances and the Household: Analysis and Review of Global Evidence', Policy Research Working Paper No. 4116, Washington, DC: World Bank.
- Amuedo-Dorantes, C. and S. Pozo (2004), 'Workers' Remittances and the Real Exchange Rate: A Paradox of Gifts', *World Development*, 32: 1407–17.
- Bauer, T.K., G.S. Epstein and I.N. Gang (2000), 'What are Migration Networks?', IZA Discussion Paper No. 200, Bonn: Institute for the Study of Labor.
- Benhabib, J. (1996), 'On the Political Economy of Immigration', *Economic European Review*, 40 (9): 1737–43.
- Bhagwati, Jagdish N. (1979), 'International Migration of the Highly Skilled: Economics, Ethics and Taxes', *Third World Quarterly*, 1 (3): 17–30.
- Bhagwati, Jagdish N. and Carlos Rodriguez (1975), 'Welfare-Theoretical Analyses of the Brain Drain', *Journal of Development Economics*, 2: 195–221.
- Bloom, David E. and Jeffrey Williamson (1998), 'Demographic Transitions and Economic Miracles in Emerging Asia', *World Bank Economic Review*, 12 (3): 419–55.
- Borjas, G. (1987), 'Self-Selection and the Earnings of Immigrants', *American Economic Review*, 77: 531–53.
- Borjas, G. (1999), 'The Economic Analysis of Immigration', in O. Ashenfelter and R. Freeman (eds), *Handbook of Labor Economics*, Amsterdam: North Holland, pp. 1697–760.
- Borjas, G. (2006), 'Labor Outflows and Labor Inflows in Puerto Rico', Working Paper, Cambridge, MA: Harvard University.
- Carrington, W. and E. Detragiache (1998), 'How Big is the Brain Drain?', IMF Working Paper No. 98/102, Washington, DC: International Monetary Fund.
- Castaldo, A., J. Litchfield and B. Reilly (2005), 'Who is Most Likely to Migrate from Albania? Evidence from the Albania Living Standards Measurement Study', Working Paper No. T-11, Sussex: Sussex Centre for Migration Research.
- Chiquiar, D. and G. Hanson (2005), 'International Migration, Self-Selection and the Distribution of Wages: Evidence from Mexico and the United States', *Journal of Political Economy*, 113 (2): 239–81.

- Chiswick, B. (1978), 'The Effects of Americanization on the Earnings of the Foreign-Born', *Journal of Political Economy*, 86: 897–922.
- Chiswick, B. (1999), 'Are Immigrants Favorably Selected?', *American Economic Review*, 89: 181–5.
- de la Garza, R. and B. Lindsay Lowell (2002), *Sending Money Home: Hispanic Remittances and Community Development*, London: Rowman & Littlefield.
- Docquier, Frederic and Abdeslam Marfouk (2006), 'International Migration by Educational Attainment: 1990–2000', in Caglar Ozden and Maurice Schiff (eds), *International Migration, Remittances and the Brain Drain*, Washington, DC: World Bank, pp. 151–99.
- Drinkwater, S. (2003), 'Estimating the Willingness to Move from Central and Eastern European Countries', Working Paper, Surrey: University of Surrey Department of Economics.
- Dutt, A.K. (2005), 'Globalization, South–North Migration and Uneven Development', in P. Gangopadhyay and M. Chatterji (eds), *Economics of Globalization*, Aldershot: Ashgate, 115–29.
- Fernandez-Huertas Moraga, Jesus (2007), 'New Evidence on Emigrant Selection', PhD dissertation, Department of Economics, Columbia University, New York.
- Fertig, Michael and Christoph Schmidt (2002), 'First and Second-Generation Migrants in Germany: What Do We Know and What Do People Think?', in R. Rotte and P. Stein (eds), *Migration Policy and the Economy: International Experiences*, Muenchen, Germany: Hanns-Seidel Foundation, pp. 179–218.
- Gang, Ira N. and Francisco L. Rivera-Batiz (1994), 'Unemployment and Attitudes Towards Foreigners in Germany', in Gunter Steinmann and Ralf Ulrich (eds), *The Economics of Immigration in Germany*, Berlin: Springer-Verlag, pp. 121–54.
- Gang, Ira N., Francisco L. Rivera-Batiz and Myeong-Su Yun (1999), 'Immigrants and Unemployment in the European Community: From the Eyes of Natives', IZA Discussion Paper No. 70, Bonn: Institute for the Study of Labor.
- Gang, Ira N., Francisco L. Rivera-Batiz and Myeong-Su Yun (2002), 'Economic Strain, Ethnic Concentration and Attitudes Towards Foreigners in the European Union', IZA Discussion paper No. 578, Bonn: Institute for the Study of Labor.
- Hanson, Gordon (2005), 'Emigration, Labor Supply and Earnings in Mexico', Working Paper No. 11412, Cambridge, MA: National Bureau of Economic Research.
- Hanson, Gordon (2006), 'Illegal Migration from Mexico to the United States', *Journal of Economic Literature*, 44: 869–924.
- Hanson, Gordon and Antonio Spilimbergo (1999), 'Illegal Immigration, Border Enforcement, and Relative Wages: Evidence from Apprehensions at the US–Mexico Border', *American Economic Review*, 89: 1337–57.
- Hatton, Timothy and Jeffrey G. Williamson (1998), *The Age of Mass Migration: Causes and Economic Impact*, New York: Oxford University Press.
- Hatton, Timothy and Jeffrey Williamson (2005), *Global Migration and the World Economy: Two Centuries of Policy and Performance*, Cambridge, MA: MIT Press.
- Karemera, D., V.I. Oguledo and B. Davis (2000), 'A Gravity Model Analysis of International Migration to North America', *Applied Economics*, 32 (13): 1745–55.
- Kapur, D. and J. McHale (2005), *Give Us Your Best and Brightest: The Global Hunt for Talent and its Impact on the Developing World*, Washington, DC: Brookings.
- Lucas, R.E.B. (1985), 'Migration among the Batswana', *Economic Journal*, 95 (June): 358–82.
- Mayda, Anna Maria (2006), 'Who Is Against Immigration? A Cross Country Investigation of Individual Attitudes towards Immigrants', *Review of Economics and Statistics*, 88 (3): 510–30.
- Mayda, Anna Maria and Krishna Patel (2006), 'International Migration Flows: The Role of Destination Countries' Migration Policies', Working Paper, Washington, DC: Center for Strategic and International Studies.
- McKenzie, David J. (2006), 'A Profile of the World's Young Developing Country Migrants', Policy Research Working paper 4021, Washington, DC: World Bank.
- Mishra, P. (2006), 'Emigration and Wages in Source Countries: Evidence from Mexico', Working Paper No. WP/86/06, Washington, DC: International Monetary Fund.

- Organisation for Economic Co-operation and Development (OECD) (2006), *Trends in International Migration*, Paris: OECD.
- Ortega, F. (2005), 'Immigration Quotas and Skill Upgrading', *Journal of Public Economics*, 89 (9–10): 1841–63.
- Parsons, Christopher, R. Skeldon, T. Walmsley and L. Alan Winters (2007), 'Quantifying the International Migration: A Database of Bilateral Movements of Migrant Stocks', World Bank Policy Research Working Paper 4165, Washington, DC: The World Bank.
- Piore, Michael J. (1979), *Birds of Passage: Migrant Labor and Industrial Societies*, New York: Cambridge University Press.
- Pritchett, Lant (2006), *Let Their People Come: Breaking the Gridlock on Global Labor Mobility*, Washington, DC: Center for Global Development.
- Rivera-Batiz, Francisco L. (1982), 'International Migration, Nontraded Goods and Economic Welfare in the Source Country', *Journal of Development Economics*, 11: 81–90.
- Rivera-Batiz, Francisco L. (1983), 'Trade Theory, Distribution of Income, and Immigration', *American Economic Review*, 73 (2): 183–7.
- Rivera-Batiz, Francisco L. (1986), 'International Migration, Remittances and Economic Welfare in the Source Country', *Journal of Economic Studies*, 13 (3): 3–19.
- Rivera-Batiz, Francisco L. (1996), 'The Economics of Technological Progress and Endogenous Growth in Open Economies', in H. Scharrer (ed.), *The Economics of High Technology Competition and Cooperation in Global Markets*, Hamburg: Institut für Wirtschaftsforschung, pp. 31–62.
- Rivera-Batiz, Francisco L. (2002), 'Democracy, Governance and Economic Growth: Theory and Evidence', *Review of Development Economics*, 6 (2): 225–47.
- Rivera-Batiz, Francisco L. (2007), 'How Do Migrants from Latin America and the Caribbean Fare in the US Labor Market?', *World Economy*, 30 (9): 1399–1429.
- Romer, P. (1990), 'Endogenous Technical Change', *Journal of Political Economy*, 98 (5): S71–102.
- Scheve, K. and M. Slaughter (2001), 'Labor Market Competition and Individual Preferences Over Immigration Policy', *Review of Economics and Statistics*, 83 (1): 133–45.
- Schiff, Maurice (2005), 'Brain Gain: Claims about its Size and Impact on Welfare and Growth are Greatly Exaggerated', in C. Ozden and M. Schiff, *International Migration, Remittances and the Brain Drain*, Washington, DC: World Bank, pp. 201–26.
- Sjaastad, Larry (1962), 'The Costs and Returns of Human Migration', *Journal of Political Economy*, 70 (5, Part 2): S80–S93.
- Stark, Oded and David E. Bloom (1985), 'The New Economics of Labor Migration', *American Economic Review*, 75 (2): 173–8.
- Stark, Oded and J. Edward Taylor (1991), 'Migration Incentives, Migration Types: The Role of Relative Deprivation', *Economic Journal*, 101 (408): 1163–78.
- United Nations (2007), 'International Migration Estimates', Demographic Estimates, Population Division, Department of Economic and Social Affairs, New York: United Nations.
- Waddington, Hugh and R. Sabates-Wheeler (2003), 'How Does Poverty Affect Migration Choice?', Development Research Centre on Migration, Globalization and Poverty Working Paper, Sussex: University of Sussex.
- World Bank (2007), *World Development Indicators*, Washington, DC: World Bank.
- Zimmermann, K. and A. Constant (eds) (2004), *How Labor Migrants Fare*, Berlin: Springer-Verlag.

44 International technology transfer: the role of foreign direct investment

Amy Jocelyn Glass and Kamal Saggi

Introduction

International technology transfer (ITT) refers to any process by which a party in one country gains access to technical information of a foreign party and successfully absorbs it into its production process. The importance of ITT for economic development is widely recognized and it has been argued that barriers to technology adoption help explain the income gap between developed and developing countries (Parente and Prescott, 1994). Such barriers include regulatory and institutional constraints that entrepreneurs must overcome as well as low levels of human capital. Furthermore, the technology frontier is a moving target – new technologies are continually being introduced. To close the technology gap, developing countries must adopt new technologies, at a faster rate than they are being created. Both market forces and government policies have an important role to play in accomplishing this formidable task.

At the heart of ITT is the exchange of information and knowledge. Technology may be codified (for example, in blueprints) or uncoded (for example, know-how of engineers). It may be embodied in products or people, or disembodied in ideas or services. ITT often occurs between unrelated partners in market-based transactions. However, information also flows internationally between related parties on a non-market basis, within the boundaries of firms and joint ventures. Given the multifaceted nature of technology transfer, there exist numerous channels through which technology flows across international boundaries. One major channel is trade in goods and services. All exports bear some potential for transmitting technological information. Trade in capital goods and technological inputs can directly improve productivity by being integrated into production processes. Another major channel of ITT is direct trade in knowledge via technology licensing, which may occur within firms, among joint venture partners or between unrelated firms. The focus of this chapter is the channel of foreign direct investment (FDI).

All these channels may facilitate imitation and reverse engineering. Because imitation does not require compensating technology owners, it can be an attractive option for developing economies. As Hoekman et al. (2005)

note, the temporary migration of students, scientists and managerial and technical personnel to universities, laboratories and conferences also plays an important role in encouraging ITT. Furthermore, ITT can also result from the temporary movement of professionals and other service suppliers who enter a developed country to perform specific services and in the process acquire additional knowledge and skills that are transferred back to the home country upon completion of the contract. While such channels of ITT are no doubt of crucial importance we do not discuss them here, in order to limit the scope of this chapter.

Although use of the word 'transfer' in the phrase 'international technology transfer' seems to suggest that the process of ITT is somehow smooth and automatic, nothing could be further from the truth. The fact that developing countries lag behind the technology frontier merely creates the potential for ITT. For ITT actually to occur, providers and acquirers of new technologies have to undertake deliberate and often significantly large investments.

Investment costs are not the only hurdle facing ITT. The market for technology is hampered by at least two significant market failures: the presence of asymmetric information and/or market power. In fact, the superior information possessed by sellers when protected by intellectual property rights is often what creates the market power. By keeping transactions within one firm, technology transfer via FDI can lessen some of the difficulties that confront arm's-length exchange of technology, yet many problems remain. Even within the same firm, Teece (1976) finds the costs of transferring technology to a foreign plant average 20 percent of the total investment required for the plant.¹

Fuller benefits for local technological capacity are realized if the technologies introduced from abroad diffuse locally. The first step – getting the technology into the country – is ITT, and the second step – getting the technology into the possession of local firms – is what we call technology diffusion. While the first step is typically a deliberate act, the second step can often be an unintended consequence. What makes the role of FDI especially important is that multinational companies dominate global research and development (R&D) and therefore act as important conduits of ITT. Yet multinationals are in the business of maximizing profit, not the economic development of host countries. However, available models and empirical studies argue that achieving a balance between the objectives of multinationals and host countries is not as difficult as it may appear on casual observation.

Foreign direct investment as a channel of international technology transfer

Today, intra-firm trade (that is, trade between subsidiaries and headquarters of multinational firms) accounts for roughly one-third of total world

trade, and sales of subsidiaries of multinational firms exceed worldwide exports of goods and services. Thus, FDI is the dominant channel through which firms serve customers in foreign markets. While much of FDI occurs between industrial countries, developing countries are becoming increasingly important host countries for FDI. Approximately 33 percent of the global stock of FDI today is in developing countries (UNCTAD, 2003).

FDI is growing in importance as a channel of ITT.² Multinational activity occurs primarily in industries that are characterized by a high ratio of R&D to sales and by large shares of professional, scientific and technical workers (Markusen, 1995). A basic tenet of the theory of the multinational firm is that such firms rely heavily on intangible assets, such as superior technology and well-established brand names, to offset the logistical and other disadvantages of operating in multiple countries as well as to compete successfully with local firms that are better acquainted with the host-country environment.³ In 1995, of all transactions in royalty and license fees, transactions within the same firm made up in excess of 80 percent, so most explicit trade in technology takes place within multinational firms (UNCTAD, 1997).

Virtually all empirical studies of FDI find that foreign-owned plants in developing countries are typically more productive than purely domestic ones. For example, a recent paper by Arnold and Javorcik (2005) provides direct evidence on the impact of FDI. Using data from Indonesia's manufacturing sector during the period 1983–86, the authors focus on the effects of acquisitions of local firms by foreign ones. They find that foreign ownership leads to significant improvements in acquired plants: after three years, the acquired plants outperform the control group in terms of total factor productivity by 34 percent.⁴

ITT through FDI can be either horizontal or vertical in form. When horizontal, FDI transfers the full technology needed to produce the good. When vertical, different stages of the production process are split across countries so only the technology for the stage (or stages) being produced in the host country is transferred. For developing countries, often the more labor-intensive stages are shifted abroad.

Of course, instead of opting for FDI, a firm may sometimes decide to contract with a local firm for production of components in a developing country rather than forming its own production subsidiary there. Due to the participation of local firms, licensing and joint ventures can yield both ITT and technology diffusion. As might be expected, it has been shown that costs of transferring technologies across countries work against FDI (and licensing) as mode choices. Norback (2001) confirms, using Swedish data, that high costs of technology transfer discourage production abroad in favor of exports.

Studies such as Mansfield and Romeo (1980) and Smarzynska (1999) have found that newer technologies are transferred through FDI, whereas older technologies are transferred through joint ventures and technology licensing. Mode choice may be one way that firms attempt to maintain their technological advantage by avoiding modes with high technology diffusion until technologies become somewhat dated. Or perhaps costs of transferring technologies via arm's-length channels are larger for newer technologies due to greater information asymmetries. Moral hazard considerations can also be important in this context. For example, Ramachandran (1993) has shown that subsidiaries receive greater resources than partially licensor-owned or independent firms once the incentives for both sides to invest in transferring technology are considered. Strategic incentives can also reinforce moral hazard and asymmetric information considerations: Fosfuri (2000) constructs a model with the feature that firms strategically use the vintage of technology to deter imitation by licensees, so that more recent technology is transferred to affiliates than to outsiders (see also Saggi, 1999).

Foreign direct investment and local technology diffusion: friends or foes?

An important consequence of FDI is that shifting production to a developing country can reduce technology adoption costs for indigenous local firms. The degree to which imitation costs are lowered by FDI might be higher for process than for product technologies. For product technologies, reverse engineering may be the main way that imitation costs are reduced. Since better process technologies tend to be difficult to deduce from inspection of the final good, first-hand experience with the technology may be required. Multinational firms bring production to the host country, providing workers with experience using the new technology. Workers then often leave to work for rival local firms or to start their own firms.⁵ Either way, such worker turnover generates knowledge flows that may lead to local firms adopting some aspects of the ways in which the new technology is better than the old. Also, any degree to which multinational firms adapt technologies to the local economic environment reduces costs of technology adoption for local firms.

Since technology advantages are often needed to survive as a multinational firm, why do multinationals not do anything and everything possible to curtail diffusion of their technologies to rival firms? As argued in Glass and Saggi (2002a), when the gains to local firms are great, the costs of preventing leakage of technologies to rivals (the wage premiums required to keep workers from leaving) are apt to be great as well. Additionally, the presence of multiple multinational firms in an industry likely leads to positive externalities among them: costly efforts undertaken

by any one multinational to curtail spillovers to local competitors would benefit all multinationals. A multinational firm that seeks to protect its technology through litigation, for example, bears the full cost but not the full benefits of its action. As a result, multinationals might very well underinvest in activities that curtail the local diffusion of their technologies.

Focusing on vertical technology transfer from a multinational to its suppliers, Pack and Saggi (2001) have shown that technology diffusion among suppliers can benefit foreign firms sourcing components. Thus, fully integrated multinational firms would be expected to be more averse to technology diffusion than firms involved in arm's-length production deals with local firms. Mexico's maquiladoras appear to have benefited from the transfer of sophisticated production techniques and backward linkages, especially in the automobile industry. Goh (2005) finds, however, that diffusion of knowledge to other potential suppliers can either encourage or discourage technology transfer depending on the incumbent supplier's cost of technological effort. Using firm-level data from Lithuania, Javorcik (2004a) finds evidence of spillovers from foreign affiliates to their local suppliers in upstream sectors, but only for projects with shared domestic and foreign ownership (not for fully foreign-owned investments).

Evidence regarding whether technologies transferred by multinational firms diffuse to competing local firms is mixed. Finding that sectors with more foreign involvement have higher productivity or faster productivity growth could stem from FDI being attracted to those sectors rather than FDI improving productivity or accelerating productivity growth. Plant-level studies are required to help alleviate any selection bias in industry-level studies. Haddad and Harrison (1993) find that foreign firms have higher levels of total factor productivity (TFP) but lower TFP growth than domestic firms in Morocco. A stronger positive effect of FDI in low-tech sectors than in high-tech sectors may indicate that local firms in high-tech sectors lack absorptive capacity. Or perhaps multinationals in high-tech sectors take more actions to preserve their technological advantages.

Aitken et al. (1996) provide an empirical assessment of the hypothesis that technology spillovers ought to increase the marginal product of labor, and this increased productivity should show up as higher wages. Their study employs data from manufacturing firms in Venezuela, Mexico and the United States. For both Mexico and Venezuela, a higher share of foreign employment is associated with higher overall wages for both skilled and unskilled workers. Furthermore, royalty payments to foreign firms from local firms are highly correlated with wages. Most importantly, the study finds no positive impact of FDI on the wages of workers employed by domestic firms. In fact, the authors report a small negative effect for domestic firms, whereas the overall effect for the entire industry is positive.

These findings differ from those for the United States, where a larger share of foreign firms in employment is associated with both a higher average wage as well as higher wages in domestic establishments. Putting Aitken et al.'s (1996) findings into the context of previous work, it is clear that wage spillovers (from foreign to domestic firms) are associated with higher productivity in domestic plants. Conversely, the absence of wage spillovers appears to accompany the existence of productivity differentials between domestic and foreign firms.

Using annual census data on more than 4000 Venezuelan firms, Aitken and Harrison (1999) provide a plant-level test of the spillover hypothesis. They find a positive relationship between foreign equity participation and plant performance, implying that foreign participation indeed benefits local plants that receive such participation. However, this own-plant effect is robust for only small plants, that is, those plants that employ fewer than 50 employees. For larger plants, foreign participation results in no significant improvement in productivity relative to domestic plants. More interestingly, they find that productivity in domestic plants declines with an increase in foreign investment – that is, they find evidence of negative spillovers from FDI. The authors suggest that these could result from a market-stealing effect: foreign competition may have forced domestic firms to lower output and thereby forgo economies of scale.⁶

However, the results of Haskel et al. (2007) contrast with those of Aitken and Harrison (1999). Haskel et al. (2007) use plant-level panel data for all UK manufacturing from 1973 to 1992 to re-examine the issue of spillovers from FDI. As the authors note, there can be little doubt that local firms in the UK possess sufficient absorptive capacity to benefit from the introduction of newer technologies by multinationals. So if spillovers do not materialize, they cannot be attributed to the limitations of domestic firms. Across a wide range of specifications, the authors find that there are positive spillovers from FDI at the industry level. More precisely, they find that a 10 percent increase in foreign presence in a UK industry raises the total factor productivity of that industry's domestic plants by about 0.5 percent. However, the authors also note that the large tax breaks and incentive packages given to multinationals seem out of proportion relative to the magnitude of spillovers they generate.

While some studies have cast doubt on the optimistic view that FDI generates positive spillovers for local firms, others have reached different conclusions. Regardless of one's view of these findings, it is worth stressing that domestic firms should be expected to suffer from an increase in competition that often results from FDI; in fact, part of the benefit of FDI is that it can help weed out relatively inefficient domestic firms. Resources released in this process will be put to better use by foreign firms with superior

technologies, efficient new entrants (both domestic and foreign) or some other sectors of the economy. However, such resource reallocation does not occur instantaneously. Existing studies of spillovers do not cover a long enough period to be able to determine accurately how FDI affects turnover rates (entry and exit). Furthermore, horizontal studies miss spillovers that may result from FDI in industries other than the one in which FDI occurs.

In a critical discussion of the plant-level studies of horizontal spillovers from FDI, Moran (2004) argues that there is a substantial difference in operating characteristics between subsidiaries that are integrated into the international sourcing networks of the parent multinationals, and those that serve protected domestic markets and are prevented by policy restrictions (such as mandatory joint venture and domestic content requirements) from being so integrated. These different operating characteristics include size of plant, proximity of technology and quality control procedures to industry best practices, speed with which production processes are brought to the frontier, efficiency of operations and cost of output. He argues that while the former have a positive impact on the host country, often accompanied by vertical backward linkages and externalities, the latter may actually have a negative impact. Drawing upon a wealth of case studies and econometric evidence, Moran (2004) argues that this contrast in performance holds across different industries, countries and time periods. He astutely notes that the failure to differentiate between export-oriented FDI and import-substitution FDI, or between foreign investors free to source from wherever they wish and foreign investors operating with domestic content requirements, or between foreign investors obliged to operate as minority shareholders and those with whole- or majority-ownership, accounts for the inability of earlier studies to isolate the influence(s) of FDI on host-country welfare.

We noted earlier that arm's-length technology transfer is usually of lower quality than its intra-firm counterpart. But is greater involvement of local firms, such as in the form of joint ventures, more likely to lead to diffusion? While this appears plausible, there is little empirical evidence in support of this idea. For example, Blomstöm and Sjöholm (1999) find that the degree of foreign ownership did not affect the productivity of local partners or spillovers to domestic firms in Indonesia for 1991. Yet having any foreign participation at all did matter: plants with no foreign participation were less productive. These findings could represent selection at the plant level – FDI is attracted to more productive plants; or a threshold effect – that foreign participation, not the degree of participation, is what matters most.

Although the extent of technology diffusion resulting from FDI is unresolved, that FDI stimulates economic growth in the host country enjoys strong empirical support. Balasubramanyam et al. (1996) find the

growth-stimulating effects of FDI are stronger for countries that pursue export promotion rather than import-substitution policies. So trade policy seems to affect the benefits of FDI, although trade orientation could proxy for other unmeasured differences across countries. For export-promoting countries, FDI stimulated growth more than domestic investment. Borensztein et al. (1998) find that FDI contributes more to economic growth than domestic investment for countries that have a sufficient stock of human capital. Countries with insufficient human capital presumably lack the ability to absorb technologies. Xu (2000) finds that countries need to achieve a minimum level of human capital in order for the technology transferred by US multinational firms to contribute to productivity growth, but most less-developed countries do not satisfy the required threshold.

Policy options for acquiring and absorbing new technologies

Separating the concept of ITT from that of local technology diffusion is especially important for analyzing policy choices. When assessing the impact of a policy option, the effects on both ITT and technology diffusion should be considered. Some policies might promote ITT but not technology diffusion. Others might promote technology diffusion but then adversely affect ITT through discouraging FDI. Multiple policy instruments are likely to be needed to achieve the ideal combination of ITT and technology diffusion. Too much technology diffusion, and there may not be much technology to diffuse since the potential for local spillovers may deter FDI. Too much ITT, and few of the advanced technologies may ever be used by indigenous firms.

Many countries such as Japan, South Korea and China have historically restricted FDI, often in favor of technology licensing or joint ventures. Foreign firms were often required to share technologies with local firms in order to conduct business in Japan. It is difficult to judge whether countries restricting FDI would have performed better or worse than if they had taken a more liberal approach, since the counterfactual is not observed. The findings that newer technologies are transferred through FDI rather than through joint ventures and technology licensing call into question the wisdom of policies that favor technology licensing or joint venture over FDI. Even if more technology diffusion results, the technology obtained may be far below the state of the art. It is possible that policy interventions could act to improve the terms of licensing contracts for local firms by removing alternatives (or making the alternatives far less desirable) for the foreign firm.

More recently, developing countries have become quite eager to attract FDI. Part of this eagerness may stem from enhanced awareness that FDI can serve as an important channel of ITT, although employment issues

surely also play a role. Many countries, both developed and developing, offer fiscal and financial incentives to attract FDI. Eliminating restrictions on FDI is likely to be beneficial, at least at the world level, since foreign firms would be freer to choose between modes without interference. However, when it comes to promoting FDI, a few words of caution may be in order. You can have too much of a good thing. Similar to free trade being best and export subsidies being distortionary, care is needed to avoid overstimulating FDI. Incentives could lead to FDI being attracted to the wrong countries – countries where costs will be higher than alternative locations. Excessive competition for FDI between locations could bid away all potential benefits.

As noted above, empirical evidence on technology spillovers from FDI is mixed, so benefits to local firms might not be realized. If governments use incentives to try to obtain the right kind of FDI, one has to question whether the government can indeed pick industries with the best potential for spillovers. Adequate human capital and infrastructure are needed for absorption, and any bureaucratic impediments to technology adoption should be minimized. So much focus on FDI risks overlooking opportunities to improve the diffusion of technologies.

Stronger protection of intellectual property (IP) has often been suggested as a means of attracting FDI. The thought is that firms will avoid FDI in favor of exports to countries with weak protection of IP (although they may also shift from licensing to FDI). Theoretical studies such as Glass and Saggi (2002b) and Glass and Wu (2007) based on the quality-ladder model of growth cast doubt on the idea that FDI rises with stronger IP protection, once the repeated nature of innovation is captured. Taylor (1993) has suggested that poor protection of IP may lead firms to mask their technologies in order to make them harder to imitate.⁷

However, in a recent paper Branstetter et al. (2006b) have argued that in a variety expansion North–South product cycle model with endogenous Northern innovation, Southern imitation and FDI, intellectual property rights (IPR) reform in the South leads to increased FDI from the North, as Northern firms shift production to Southern affiliates. Furthermore, this increased FDI drives an acceleration of Southern industrial development, as the South's share of global manufacturing and the pace at which production of more recently invented goods shifts to the South both increase. In addition, their empirical results show that US-based multinational corporations (MNCs) expand the scale of their activities in reforming countries after IPR reform, and this effect is disproportionately strong for affiliates whose parents rely strongly on patented intellectual property as part of their global business strategy. Furthermore, they also provide evidence from highly disaggregated trade data that suggests that the expansion of multinational activity leads to a higher net level of production shifting

to developing countries, more than offsetting any possible decline in the imitative activity of indigenous firms.⁸

Several other empirical studies find some support for IP protection encouraging FDI. Using data on US FDI, Lee and Mansfield (1996) find that a country's choice of IP protection influences the volume and composition of FDI it receives. Smith (2001) finds that stronger IP protection encourages affiliate sales and licensing for countries with imitative capacity. Javorcik (2004b) finds weak IP protection deters FDI in technology-intensive sectors for transition economies. Nunnenkamp and Spatz (2004) find evidence of IP protection spurring higher-quality FDI.

Stronger IP protection may be more important for technology licensing than for FDI due to the risk of opportunistic behavior and difficulty enforcing contracts. Yang and Maskus (2001) consider a model in which stronger IP protection increases the licensor's share of rents and reduces the costs of licensing contracts. Thus, better IP protection may indeed stimulate licensing and technology transfer.

Concluding remarks

International technology transfer is a complex, multifaceted phenomenon. In this short chapter, we have chosen to highlight the role foreign direct investment and multinational corporations play in the process of international technology transfer, paying little attention to international trade of goods and services, movement of workers and professionals, and other related phenomenon. While we do feel that FDI is the most important channel, we do not mean to imply that the other channels are not of considerable importance as well.

That multinational companies are pivotal in introducing new technologies to host countries is fairly well established. However, lagging countries have not only to obtain foreign technology but also to learn how to use it to its fullest potential. In this context, we have found it useful to make a distinction between initial international technology transfer and subsequent technology diffusion within host countries. This distinction can be important since some policies could promote technology transfer but deter technology diffusion, or promote technology diffusion but deter technology transfer. With respect to the contribution of FDI, there is some good news and bad news. First the bad news: multinationals will usually lose from further horizontal diffusion of their technologies and should be expected to take actions that thwart that process. The good news is that technology transfer to local suppliers is compatible with the motives of multinationals and a plethora of empirical evidence indicates that vertical linkages between multinationals and their local suppliers play a crucial role in the industrial development of host countries.

An important policy conclusion of this analysis is that host countries are better off facilitating processes that are compatible with the motives of multinationals. In other words, a developing country should perhaps be less concerned about being able to produce an automobile of its own and more concerned about developing a competitive network of suppliers that can serve (and gain from) well-established foreign firms. It is in this mutually beneficial exchange that the most productive policy intervention might lie. Of course, if both sides are indeed willing participants, policy intervention required would be 'light' as opposed to 'heavy'. Furthermore, it would not be targeted in nature. Instead it would ensure that local businesses have access to adequate infrastructure and skilled workers, and that their expansion or downsizing decisions are not hampered by burdensome regulations. In our view, this is another plus of pursuing policies that take proper account of the incentives multinational firms have (and do not have) to encourage industrial development in host countries.

Notes

1. See also Mansfield and Romeo (1980) and Ramachandran (1993).
2. That FDI leads to ITT may seem obvious, yet Glass and Saggi (1999) have argued that whether FDI creates ITT in aggregate depends crucially on whether substitute channels of ITT, such as imitation, exist. FDI could merely displace imitation that otherwise would have occurred, leaving ITT essentially unchanged. See also Glass and Saggi (1998) for a model in which narrowing the technology gap induces multinationals to transfer state-of-the-art technologies.
3. A nice feature of their approach is that they control for the self-selection problem – that is, while it is true that multinationals typically acquire firms that are relatively more productive, they further contribute to the future productivity of acquired firms.
4. See Saggi (2002) for an extensive discussion of this literature.
5. Cheng et al. (2005) examine the impact of the ability of workers to absorb foreign technologies on the extent of production by multinational firms.
6. Nevertheless, on balance, Aitken and Harrison (1999) find that the effect of FDI on the productivity of the entire industry is weakly positive. They also note that similar results are obtained for Indonesia, except that the positive effect on own plants is stronger, whereas the negative effect on domestic plants is weaker, suggesting a stronger overall positive effect.
7. See also Taylor (1994) for the effects of IP protection on technology transfer.
8. In a related paper, Branstetter et al. (2006a) have shown that there is a significant increase in technology transfer following reforms among affiliates of firms that make extensive use of the US patent system.

References

- Aitken, Brian and Ann E. Harrison (1999), 'Do Domestic Firms Benefit from Direct Foreign Investment?' *American Economic Review*, 89: 605–18.
- Aitken, Brian, Ann E. Harrison and Robert E. Lipsey (1996), 'Wages and Foreign Ownership: A Comparative Study of Mexico, Venezuela, and the United States', *Journal of International Economics*, 50: 345–71.
- Arnold, Jen M. and Beata S. Javorcik (2005), 'Gifted Kids or Pushy Parents? Foreign Acquisitions and Plant Productivity in Indonesia', Policy Research Working Paper 3193, World Bank.

- Balasubramanyam, Venkataraman N., Mohammed A. Salisu and David Sapsford (1996), 'Foreign Direct Investment and Growth in EP and IS Countries', *Economic Journal*, 106: 92–105.
- Blomström, Magnus and Fredrik Sjöholm (1999), 'Technology Transfer and Spillovers: Does Local Participation with Multinationals Matter?' *European Economic Review*, 43: 915–23.
- Borensztein, E., J. De Gregorio and J-W. Lee (1998), 'How Does Foreign Direct Investment Affect Economic Growth?' *Journal of International Economics*, 45: 115–35.
- Branstetter, Lee, Raymond Fisman and C. Fritz Foley (2006a), 'Does Stronger Intellectual Property Rights Increase International Technology Transfer? Empirical Evidence from US Firm-Level Data', *Quarterly Journal of Economics*, 121 (February): 321–49.
- Branstetter, Lee, Raymond Fisman, C. Fritz Foley and Kamal Saggi (2006b), 'Intellectual Property Rights, Imitation, and Foreign Direct Investment: Theory and Evidence', mimeo.
- Cheng, Leonard K., Larry D. Qiu and Guofu Tan (2005), 'Foreign Direct Investment and International Trade in a Continuum Ricardian Trade Model', *Journal of Development Economics*, 77: 477–501.
- Fosfuri, Andrea (2000), 'Patent Protection, Imitation, and the Mode of Technology Transfer', *International Journal of Industrial Organization*, 18: 1129–49.
- Glass, Amy J. and Kamal Saggi (1998), 'International Technology Transfer and the Technology Gap', *Journal of Development Economics*, 55: 369–98.
- Glass, Amy J. and Kamal Saggi (1999), 'Foreign Direct Investment and the Nature of R&D', *Canadian Journal of Economics*, 32: 92–117.
- Glass, Amy J. and Kamal Saggi (2002a), 'Multinational Firms and Technology Transfer', *Scandinavian Journal of Economics*, 104: 495–513.
- Glass, Amy J. and Kamal Saggi (2002b), 'Intellectual Property Protection and Foreign Direct Investment', *Journal of International Economics*, 56: 387–410.
- Glass, Amy J. and Xiaodong Wu (2007), 'Intellectual Property Rights and Quality Improvement', *Journal of Development Economics*, 82: 393–415.
- Goh, Ai-Ting (2005), 'Knowledge Diffusion, Input Supplier's Technological Effort and Technology Transfer via Vertical Relationships', *Journal of International Economics*, 66: 527–40.
- Haddad, Mona and Ann Harrison (1993), 'Are There Positive Spillovers from Direct Foreign Investment? Evidence from Panel Data from Morocco', *Journal of Development Economics*, 42: 51–74.
- Haskel, Jonathan E., Sonia Pereira and Matthew J. Slaughter (2007), 'Does Inward Foreign Direct Investment Boost the Productivity of Domestic Firms?', *Review of Economics and Statistics*, 89: 482–96.
- Hoekman, Bernard, Keith Maskus, and Kamal Saggi (2005), 'Transfer of Technology to Developing Countries: Unilateral and Multilateral Policy Options', *World Development*, 33: 1587–1602.
- Javorcik, Beata Smarzynska (2004a), 'Does Foreign Direct Investment Increase the Productivity of Domestic Firms? In Search of Spillovers through Backward Linkages', *American Economic Review*, 94: 605–27.
- Javorcik, Beata Smarzynska (2004b), 'The Composition of Foreign Direct Investment and Protection of Intellectual Property Rights: Evidence from Transition Economies', *European Economic Review*, 48: 39–62.
- Lee, Jeong-Yeon and Edwin Mansfield (1996), 'Intellectual Property Protection and US Foreign Direct Investment', *Review of Economics and Statistics*, 78: 181–6.
- Mansfield, Edwin and Anthony Romeo (1980), 'Technology Transfer to Overseas Subsidiaries by US Based Firms', *Quarterly Journal of Economics*, 9: 737–49.
- Markusen, James R. (1995), 'The Boundaries of Multinational Enterprises and the Theory of International Trade', *Journal of Economic Perspectives*, 9: 169–89.
- Moran, Theodore (2004), 'How Does Foreign Direct Investment Affect Host Country Development: Do We Already Know the Answer? Using Industry Case Studies to Make Reliable Generalizations', in Magnus Blomstrom, Edward Graham and Theodore Moran (eds), *The Impact of Foreign Direct Investment on Development: New Measures, New*

- Outcomes, New Policy Approaches*, Washington, DC: Institute for International Economics, pp. 281–314.
- Norback, Pehr-Johan (2001), 'Multinational Firms, Technology and Location', *Journal of International Economics*, 54: 449–69.
- Nunnenkamp, Peter and Julius Spatz (2004), 'Intellectual Property Rights and Foreign Direct Investment: A Disaggregated Analysis', *Review of World Economics/Weltwirtschaftliches Archiv*, 140: 393–413.
- Pack, Howard and Kamal Saggi (2001), 'Vertical Technology Transfer via International Outsourcing', *Journal of Development Economics*, 65: 389–415.
- Parente, Stephen L. and Edward C. Prescott (1994), 'Barriers to Technology Adoption and Development', *Journal of Political Economy*, 10: 298–321.
- Ramachandran, Vijaya (1993), 'Technology Transfer, Firm Ownership, and Investment in Human Capital', *Review of Economics and Statistics*, 75: 664–70.
- Saggi, Kamal (1999), 'Foreign Direct Investment, Licensing, and Incentives for Innovation', *Review of International Economics*, 7: 699–714.
- Saggi, Kamal (2002), 'Trade, Foreign Direct Investment, and International Technology Transfer: A Survey', *World Bank Research Observer*, 17: 191–235.
- Smarzynska, Beata K. (1999), 'Technological Leadership and the Choice of Entry Mode by Foreign Investors: Evidence from Transition Economies', mimeo.
- Smith, Pamela J. (2001), 'How Do Foreign Patent Rights Affect US Exports, Affiliate Sales, and Licenses?' *Journal of International Economics*, 55: 411–39.
- Taylor, M. Scott (1993), 'TRIPS, Trade, and Technology Transfer', *Canadian Journal of Economics*, 26: 625–37.
- Taylor, M. Scott (1994), 'TRIPS, Trade, and Growth', *International Economic Review*, 35: 361–81.
- Teece, David J. (1976), *The Multinational Corporation and Resource Cost of International Technology Transfer*, Cambridge, MA: Ballinger.
- UNCTAD (1997), *World Investment Report: Transnational Corporations, Market Structure, and Competition Policy*, New York: United Nations.
- UNCTAD (2003), *World Investment Report: FDI Policies for Development: National and International Perspectives*, New York: United Nations.
- Xu, Bin (2000), 'Multinational Enterprises, Technology Diffusion, and Host Country Productivity Growth', *Journal of Development Economics*, 62: 477–93.
- Yang, Guifang and Keith E. Maskus (2001), 'Intellectual Property Rights, Licensing, and Innovation in an Endogenous Product-Cycle Model', *Journal of International Economics*, 53: 169–87.

45 International institutions and development

Kunibert Raffer

Introduction

Those international financial institutions (IFIs) most important for development nowadays were not established for this purpose. Keynes drafted the Bretton Woods system as the allied counter-proposal to Nazi Germany's 'New Order' of a European Economic Community, basically structured like the present EU, even including the prospect of a European currency union, though no European 'Parliament' without real parliamentary powers (Raffer and Singer, 2001 [2002, 2004], pp. 1–2). The OECD (1985, p. 140) describes the initial tasks: 'The IBRD was there to guarantee European borrowing in international (North American) markets; the IMF was there to smooth the flow of repayments.' Because Southern delegations demanded resources for development, 'and Development' was added to 'International Bank for Reconstruction' (IBRD).

When the communist threat made European economic recovery a very important issue, other programs, mainly the Marshall Plan, took over. Apparently, IBRD loans were considered inappropriate for successfully reconstructing Europe, which raises questions whether they can develop much poorer countries. The addition 'and Development' allowed the IBRD to turn South (Caufield 1998, p. 56). Somewhat later the IMF also shifted totally to the South in spite of its more general mandate.

The International Trade Organization, the last pillar of the Bretton Woods system, did not come into being. Only the 'provisional' General Agreement on Tariffs and Trade (GATT) regulating trade in manufactures was established. Keynes's ideas of developmental interest, such as stabilizing commodity prices, were not implemented. The GATT's successor, the World Trade Organization (WTO) is characterized by a sharply different view on development, based on the Washington Consensus.

The Bretton Woods institutions (BWIs) and the WTO reflect global power asymmetries. Unlike the Inter-American Development Bank, where regional developing members must have 50.005 percent of total votes, industrialized countries (ICs) control the BWIs by clear voting majorities. The WTO's one-country-one-vote principle is in practice overturned by the consensus approach, bilateral pressure on developing countries (DCs), and

the so-called 'Green Room', the practice of backroom negotiations to which only a few countries are invited, whose results are then presented to the rest for 'consensus', usually under time and other pressure. At Seattle this triggered strong protests. Southern discontent has also surfaced during the Doha Round.

Shifting focus

The IBRD (2005) and its low-income-country window IDA (International Development Association) declare 'global poverty reduction and the improvement of living standards' as their 'mission'. The bank is not 'a bank in the common sense' (*ibid.*). These statements differ fundamentally from its initial businesslike approach precluding financing social activities. It not only refused 'messing around with education and health' or a water treatment plant, but even forced Columbia not to accept a French loan for waterworks (Caufield, 1998, p. 64). McNamara was the first president focusing on poverty. McNamara had the merit of giving credibility to the idea that helping the poor is not wasting resources, but makes economic sense. Brought about by Euromarket lending, the debt problem shifted the orientation of the BWIs fundamentally. The IBRD and IDA moved from project financing towards program lending, which should be exceptional pursuant to their Articles of Agreement.

The IMF was not designed as a development organization, but to enable members of the Bretton Woods system with short-term balance-of-payments problems to stay within the agreed parity bands. The demise of Bretton Woods left the IMF with very few remaining tasks, such as the Compensatory Financing Facility (CFF). These could have been transferred to another institution. Arguably, the IMF should have been dissolved. The debt crisis 1982 provided a new role.

Until the Cologne Summit entrusted both BWIs with the Highly Indebted Poor Country Initiative II (HIPC II), which explicitly includes anti-poverty measures, the IMF had usually and rightly stated that this was not its mandate. Turning debt managers, both BWIs gained strong control over development policies, characterized by the term 'conditionality'. Conditionality became part of the IMF's statutes as late as 1969. Until its introduction the IMF fulfilled a highly useful role of emergency lending. Forced to get the BWI 'seal of approval' in order to get urgently needed new loans, DCs in distress have to fulfill a wide range of conditions, not all visibly connected to economic necessities. BWI loans might carry over 100 conditions, which raises questions whether all can be complied with. All debt relief measures have increased their leverage.

Introduced to compensate export earnings shortfalls beyond the member's control, the CFF illustrates the strengthening of conditionality

over time. Initially, a statement sufficed to cooperate with the IMF where required. Eventually, the 'Fund has increasingly come to the realization that even though a country's export shortfall was both "temporary" and largely beyond its control the country might still have balance-of-payments difficulties attributable to inappropriate policies and that large amounts of unconditional credit might cause the country to delay adopting needed policy adjustments' (Polak, 1991, p. 9). Polak (1991, p. 12), an influential IMF theoretician, is outspoken: 'The purpose of the Fund's conditionality is to make as sure as possible that a country drawing on the Fund's resources pursues a set of policies that are, in the Fund's view, appropriate to its economic situation in general and its payments situation in particular' – even if the country's economic policy is not at all the reason for temporary problems.

The IBRD has never made unconditional loans. Conditions requiring policy changes have even been attached to projects (cf. Mosley et al., 1991, p. 27). 'Programme lending' increased conditionality: '[T]he Bank felt that it needed a place at the top policy-making table' (Mosley et al., 1991, p. 34) beyond what it could expect from project monitoring. Stern (1983, p. 91), the IBRD's Senior Vice-President, praised structural adjustment lending as enabling 'the Bank to address basic issues of economic management and of development strategy more directly and urgently', as a 'unique opportunity to achieve a comprehensive and timely approach to policy reform' (Stern, 1983, p. 104), the response to a 'feasible . . . call for increased sacrifices' (Stern, 1983, p. 91).

The IMF started adjustment measures in sub-Saharan Africa after 1973. After decades of adjusting debtor countries and 'appropriate' development strategies, no country regained economic viability. Attempts to prove success econometrically were given up long ago. Often no statistically significant difference between program and non-program countries was found. Khan (1990), an IMF econometrician, found significantly reduced growth in program countries; as Polak (1991, p. 42) points out, a predicted reduction in the growth rate of at least 0.7 percent of GDP each year a country had an IMF program. After years of 'Structural adjustment' the IBRD (1989, p. 6) found a lack of 'an integrated analytical framework to understand better the links between a program and its expected macroeconomic outcomes'. In spite of little success, official creditors have steadily increased the role of the BWIs.

Rodrik's (1996) analysis of neoliberal reforms might explain this behavior better. He sees the debt crisis as an opportunity seized for a 'wholesale reform of prevailing policies', offering the chance 'to wipe the slate clean and mount a frontal attack on the entire range of policies in use' (ibid., p. 17). A crisis brought about by overspending, overlending and the sudden

change of economic policy in ICs, which sent interest rates skyrocketing, was declared to stem from disliked policies in DCs.

The BWIs have strongly pushed for policy change while refusing to subject themselves to market mechanisms and basic legal principles. This produced an economically inverted incentive system absolutely at odds with market incentives. Although IFIs (co-)determine their clients' policies, they refuse to share the risks involved appropriately, insisting on full repayment, even if damages caused tortuously by their staff should have occurred (Raffer, 2004). Borrowers have to pay for such damages. IFIs may gain financially from tortuous behavior and errors by extending new loans necessary to repair damages done by prior loans. New and larger crises increase their importance. Grave negligence creating damages leads to new loans correcting such damages, increasing IFI incomes. The IMF's proposal of a Sovereign Debt Restructuring Mechanism would have perpetuated this situation, granting IFIs *de jure* preferred creditor status and increasing the Fund's role in debt management (cf. Raffer, 2005).

The WTO

The WTO (2005a) presents itself as a 'negotiating forum', 'a set of rules' and a help 'to settle disputes'. It sees its mandate in preventing 'self-defeating, destructive . . . protectionism' (WTO, 2005b). Theoretically it rests its case solely on comparative advantages (*ibid.*), although this theorem only works in a two-countries-two-goods world and constant returns to scale are necessary to guarantee welfare gains (cf. Viner, 1937, pp. 470–79).

Mattoo and Subramanian (2005, p. 19) argue that the WTO seems to be the 'best vehicle' for advancing Northern corporate interests, seeking 'the opening of markets in developing countries for manufactured goods'. While able to retain high protection where wanted, ICs have managed to restrict or outlaw protection where it could be in the interest of DCs. The principles of the WTO are in many important respects the very opposite of the ideas behind the creation of UNCTAD.

In spite of rhetoric, tariff escalation continues to exist: 'OECD tariffs on finished industrial products are about eight times higher than on raw materials . . . These barriers delay entry into the export-oriented industries, which are most accessible to developing countries' (OECD, 2000, pp. 31–2). Average tariffs on Southern manufactured exports continue to be a multiple of those on imports from other ICs. The Doha Development Round again pressures for more market access of ICs. Market access for non-agricultural products is, for example, hardly of overwhelming developmental interest. DC interests can hardly be identified except, arguably, in trace elements. Special and differentiated treatment practically abolished by the WTO is again discussed. The Doha Declaration calls for a review to

strengthen it, without practical noteworthy effects so far. The WTO's record has triggered proposals for making it more development-friendly, even from within the BWIs (Hoekman, 2005).

Heavy subsidizing of domestic Northern agriculture and agro-exports conform to WTO obligations. US exports are priced 65 percent below production costs. The EU exports sugar and beef at less than half their production costs. WTO cotton and sugar panels legally established that ICs had failed to abide by the loose rules on subsidies they had crafted during the Uruguay Round. 'Comparative access to subsidies, not comparative advantage' (Oxfam, 2005, p. 9) shapes 'world markets'. Institutions inter-link. In the name of economic efficiency the IBRD pressured Mali to pay local cotton producers this subsidy-determined 'world market price' in 2004. The government ultimately refused to bankrupt domestic peasants.

'Voluntary Export Restrictions' have become legalized. The Trade-Related Investment Measures (TRIMS) treaty restricts developmental options to industrialize. Politics successfully pursued by Asia's dragons are now outlawed. The Trade-Related Intellectual Property Rights (TRIPS) agreement does not strictly speaking protect intellectual property, because the host of tribal knowledge in many DCs remains unprotected. It 'increased the monopoly power of patent holders and limits the ability of generic producers to compete' (Mattoo and Subramanian, 2005, p. 20), enabling pharmaceutical companies to raise prices far above what many poor people can afford. DCs are pressured not to use those WTO safeguards to protect public health, which the USA or Canada have used. Over the years the USA has threatened trade sanctions against countries revising their legislation to incorporate TRIPS safeguards, such as Thailand and South Africa. Complaints were launched against Brazil and South Africa using their WTO rights to fight AIDS via affordable drugs. DC protests brought about change for the better. At Doha the right of WTO members to use, to the full, the provisions in the TRIPS agreement was 'reaffirmed'. This would have been unnecessary if DC rights under TRIPS had been respected.

Dispute settlement was one of the main chocolates on the tray to convince smaller and weaker countries to sign the WTO treaties, promoted as a rule-based system protecting the rights of the weak. It had been argued that the WTO would substitute bilateral (and GATT-violating) measures such as the US Super 301, a unilateral measure implemented in breach of international treaties. Meanwhile, the WTO accepted Super 301 (WTO, 2000, pp. 67–8). A review process had been agreed at Marrakesh, but ICs blocked any change. Although Doha is called a 'single undertaking' encompassing all issues, dispute settlement, a main concern of DCs, is not part of it.

Dispute settlement has no mandate to protect members' rights. Article 3.7 of the Understanding on Rules and Procedures Covering the Settlement of Disputes states: 'Before bringing a case, a Member shall exercise its judgment as to whether action under these procedures would be fruitful. The aim of the dispute settlement mechanism is to secure a positive solution to the dispute.' The probability of success, not the rule of law, is explicitly established as the guiding principle. This is a unique and unfortunate approach. Suing big guys is often fruitless, not least because of the way 'relief' is organized. There is no right to compensation for damages suffered by violations of contractual obligations. The winning party may be authorized to suspend WTO concessions subject to strict and constraining rules. After winning against the USA, Antigua, whose exports of Internet games were blocked in breach of contract, was authorized to sanction the USA. The positive side is that DCs can and do win cases and can often obtain relief.

Less agricultural export subsidies and reviewing dispute settlement are not the only unfulfilled promises made while the signature of DCs was coveted. Before Marrakesh, the countries of the Organisation for Economic Co-operation and Development (OECD) apparently perceived a need to assure DCs of relief measures against expected higher food prices. After ratification net importers were referred to existing BWI facilities. A list of net food-importing developing countries exists meanwhile, but being listed does not confer automatic benefits. Donors and international institutions want a role to play. Estimates of the WTO's benefits to DCs were 'exaggerated and its costs were underplayed', 'liberalization assumptions were disconnected from what the [Uruguay] Round actually achieved' (Mattoo and Subramanian, 2005, p. 21). Calling Doha a 'Development Round' continues this tradition.

The WTO is used to cleanse trade relations from disliked historical obligations. WTO compatibility is presently the EU's argument to remove those remnants of the Lomé framework that had been adopted in favor of and due to pressure by DC signatories. The Final Act of the Uruguay Round also increased IFI influence. IFIs are to cooperate more closely with the WTO to achieve greater coherence in economic policy. Mali's example above illustrates how this is done. Structural adjustment lending forced DCs to open and liberalize their economies to the extent of making the 'WTO process a "victim" of the success of the World Bank and the IMF' (Mattoo and Subramanian, 2005, p. 20). The WTO treaties are now useful to lock in liberalization, to restrict the options of DCs. Simultaneously, bilateral treaties try to gain further concessions, which are likely to become enshrined into WTO treaties at a later stage.

Unlike in other cases where signing means accepting all obligations of treaties, accession to the WTO means prior bilateral agreements with all

members. These are used to extract further, individual concessions. Small countries have little choice, unlike big ones. The EU demanded that Russia increase its domestic oil price as a precondition for being allowed to join the WTO. Russia declined. Apart from China, few DCs could defend issues important for them as easily.

Massive protests, the events of Seattle, and the forming of the G20 have improved the position of DCs somewhat. Especially Brazil and India have acquired new importance. But it remains to be seen how much influence any DCs will finally have.

Statutory rights and developing members

Critical minds point at the considerable difference between statutory rights and the policy space DCs enjoy *de facto*, as well as at the costs resulting from this discrepancy. These rights, the fear that bilateral arrangements might be worse, promises (such as large cuts in agro-export subsidies), and pressure seem to have enticed DCs into signing the WTO treaties. Once members, they are safe from bilateral accession agreements.

The WTO does not fully protect the rights of weaker members, as pharmaceuticals illustrate. Big players have a choice and cannot be forced to honor contractual obligations. When the EU complained against the Helms–Burton Act, the US observed that this would not lead to resolving the dispute but pose serious risks for the WTO. After agreeing bilaterally not to apply Helms–Burton to EU corporations the complaint was dropped. In a dispute with Brazil, Canada simply refused to provide information it was obliged to disclose promptly and fully pursuant to Article 13.1 of the Dispute Settlement Understanding. Expressly mentioning this and its potential to undermine the dispute settlement system, the WTO (2000, p. 59) found against Brazil, as Canada's WTO-violating behavior could not be proved because of Canada's additional violation of WTO rules on providing information.

DCs exercising contractual rights may raise the WTO's concern. Though 'not extensively used' in Asia after the 1997 crisis, selective tariff increases by some countries remaining 'within the flexibility allowed by bindings under the WTO agreements' gave 'cause for concern to the extent they may distort the pattern of production and trade' (WTO, 1998, p. 28). The WTO has never voiced similar concern on potential distortions regarding ICs, including agrarian subsidies or the long phasing-out of GATT-inconsistent restrictions.

Capital controls are a membership right pursuant to the IMF's constitution explicitly restricting the use of Fund resources to finance 'large and sustained' outflows. Even current transfers can be restricted with the Fund's approval. The IMF may, but is not obliged to, request controls. Its

statutes clearly show that it is not supposed to press for liberalization of capital movements. Asian countries had the right to control capital outflows – as the IMF admitted when Malaysia exercised it (cf. Raffer and Singer, 2001 [2002, 2004], p. 157). IMF programs financing large and sustained outflows by speculators violated the IMF's constitution, causing damages to DCs while increasing IMF drawings and thus earnings. Pressure to liberalize capital accounts has made increased stocks of international reserves necessary. These have become 'one of the widely-used targets of poverty reduction strategies in Africa' (UNCTAD, 2002, p. 31). This money is lost for poverty eradication, debt service or financing the Millennium Development Goals (MDGs). As DCs pay higher interest than they receive on reserves (often US Treasury Bonds), they face substantial annual costs which they could avoid by exercising statutory rights.

The IBRD's statutes demand debt relief in the case of default, without any conditions. Article IV.4.c confers a right onto members suffering from an acute exchange stringency to 'apply to the Bank for a relaxation of the conditions of payment'. Article IV.7 contains the obligation to reduce claims in the case of default. The statutes of the Asian and Inter-American Development Banks are similar. The African Development Bank's new statutes do not contain this clause. The European Bank for Reconstruction and Development writes off losses, proving that IFIs can survive while doing so. By simply refusing to acknowledge default, even if countries have not paid anything for six or seven years, the IBRD does not act according to its statutes. This creates damages by delaying solutions.

Attempts at reforms by the South

The preponderance of Northern interest in most international institutions triggered Southern attempts to create new global institutions more in line with developmental needs and less under Northern control. Success has remained limited.

Early attempts focused on mechanisms within the UN, the most prominent being SUNFED (Special UN Fund for Economic Development). Inspired by the generosity of the Marshall Plan, the 'wild men' at the UN advocated setting up a UN Fund to administer large-scale soft aid. Theoretically, this idea was based on the then generally accepted Keynesian consensus that capital availability determined growth, which in turn was needed to improve the lot of the poor. It was easy to argue that European countries had a moral obligation to help as they had been helped by the US. ICs opposed the idea of the UN administering large funds. Demanding harder terms, nearer to the market, the IBRD was particularly strongly opposed to the principle of soft financing on economic

grounds. Once it became clear that soft multilateral financing would be done by an institution administered by the IBRD, the Bank dropped all reservations. Helped by the Cuban revolution, which sparked a wave of US 'generosity', and US interests in disposing of embarrassingly high, practically useless holdings of inconvertible currencies, IDA was established in 1960, attached to the IBRD, which ICs control. ICs preferred IDA to any UN Fund, 'because the structure of the World Bank ensured weighted voting in their favour' (OECD, 1985, p. 141). The Marshall Plan's participatory and recipient-friendly approach was not to be repeated. The OECD (1985, p. 146) observed a certain differentiation regarding multilateral aid: 'by and large the largest donors have favoured the World Bank, while the smaller donors have favoured the United Nations'. The UN obtained two valuable consolation prizes: the UNDP and the World Food Programme.

The UN Economic Commission for Latin America (ECLA; Spanish: CEPAL) took the lead in defending Southern interests by publishing Raúl Prebisch's (1949) findings on the effects of trade on DCs, one pillar of the Prebisch–Singer thesis. This publication angered the USA to the point of attempting to close ECLA (Toye and Toye, 2003, p. 463). ECLA's economists, the 'Cepalistas', continued over decades to produce divergent views, a source of heterodox theories emanating from the South.

Prebisch was also the engine behind the establishment of the United Nations Conference on Trade and Development (UNCTAD) created in order to reduce or eliminate damaging effects of world trade on DCs. Its ideas on appropriate trade-related development policies differ pronouncedly from the WTO approach. Commodity agreements were established, all of which collapsed later on. In the 1970s the South demanded a Common Fund to stabilize commodity prices. Its 'Second Window' was to finance projects such as research and development (R&D) and economic diversification. After protracted bargaining the Common Fund was agreed on before the UNCTAD meeting of 1979. Voting shares were allocated to the North (42 percent), the Group of 77 (that is, the South, 47 percent), communist countries (8 percent) and China (3 percent). ICs (East and West) stipulated financial contributions higher than their shares in voting rights. It took roughly a decade until enough countries had ratified to bring the Fund into existence, although the Organization of Petroleum Exporting Countries (OPEC) bankrolled the contributions of poor DCs. Inadequate resources prevented the Common Fund from playing any role. Several of this Fund's ideas were taken up by an eminent person's report on commodities in 2003 (Khor, 2005).

A highly useful recent activity of UNCTAD is its Debt Management Financial Advisory Services (DMFAS) program helping DCs to improve

debt management. As effective debt management is part of good governance, and accounting errors by creditors to the detriment of DCs are documented, one might have expected the BWIs to help debtors in establishing appropriate controlling. They did not. Advice by UNCTAD is preferable because UNCTAD is, unlike the BWIs, not a creditor.

There exist other multilateral institutions, notably those established by OPEC or Arab countries, such as the OPEC Special Fund. Established in 1977 with strong political and financial support of OPEC to finance projects in favor of the rural poor, IFAD (International Fund for Agricultural Development) was a successful attempt to make multilateral structures more democratic (Maurizio, 1983). Active participation by the projects' beneficiaries was sought. Votes were equally split among the North, OPEC countries and other DCs, separating votes and contributions. This did not make IFAD particularly popular with ICs. The first replenishment led to quarrels. Soon ICs demanded a restructuring of IFAD so as to align votes and contributions. Meanwhile over 56 percent of all votes are distributed in accordance with cumulative convertible currency contributions. IFAD, a child of the South's drive for a New International Economic Order in the 1970s, was finally brought in line with present realities. At a time when the power of ICs has reached its apex since decolonialization, a multilateral institution where they have only one-third of the votes is simply unacceptable.

Before 1997, Asian countries wanted to establish an Asian Monetary Fund as a regional counterweight to the IMF. The Asian crisis put an end to this, at least for some time. It is interesting to note that the IBRD (1999, p. 2) knew years before that the liberalization policies it encouraged in Asia would lead to disaster (cf. Raffer and Singer, 2001 [2002, 2004], pp. 150–51). This helps us to understand why conspiracy theories abounded. Even as distinguished an economist as Bhagwati (1997) spoke of a 'Wall Street–Treasury complex' dictating the agenda.

Conclusion

Evaluating the record and policies of international institutions does not suggest that they fully support development. Attempts by DCs to establish more development-friendly international institutions corroborate this conclusion.

At present one sees an interlinking of important institutions. The BWIs force debtor DCs to liberalize quickly and strongly. The WTO then protects and perpetuates the results. DCs suffering from negative WTO effects may draw on IMF resources, thus increasing their dependence on the Fund. Bilateral treaties go beyond the WTO, using the greater leverage ICs have *vis-à-vis* most DCs, justifying DC fears that bilateralism might protect them even less. But bilateral treaties also prepare the ground for further

concessions by DCs that may eventually become WTO obligations. Typically, more recent treaties extend the definition of ‘investment’ to loans, thus increasing creditor leverage. The WTO process moves so quickly that small DCs simply have no chance to implement agreements before new changes are negotiated or even implemented. Unequal treatment of members depending on their political clout is another fact.

Chang (2002, p. 139) argues that ICs are ‘kicking away the ladder’ of development by ‘insisting that developing countries adopt policies and institutions that were *not* the ones that they had used in order to develop’. Chang feels this may be done ‘out of genuine (if misinformed) good will’, nevertheless with catastrophic results. He might as well have quoted List’s recommendation of North–South relations. Better known for his opposition to the ‘English philosophy’ of free trade as harmful to Germany in its early development stages, List (1920 [1841], p. 211) advocated joint exploitation of DCs as ‘promising much richer and more certain fruits than the mutual enmity of war and trade regulations’. International institutions seem to show such ‘neo-Listian’ (Raffer, 1987) tendencies.

Important international institutions are constructed in such a way that they cannot but implement the wishes of ICs. Within IFIs, ICs usually enjoy comfortable voting majorities. Within the WTO, unequal economic and political influence asserts itself in spite of ‘one-country-one-vote’. Even if they wanted to, international institutions could not act against their powerful members. ICs seem more interested in gaining economic and political advantage than in supporting development. As selective WTO liberalization proves, ICs do not wish to approximate free market conditions globally, but seek export and investment possibilities where this is in their interest. Thus, Northern FDI in the South is said to be good for anyone needing WTO protection. Chinese FDI in the USA, however, is not – and is treated differently.

Finally, international institutions have self-interests. They want to gain influence and enlarge their mandates. This seems to explain the BWIs’ role in debt management particularly well. Regarding the three most important institutions, these factors combine to produce effects that hinder rather than foster development.

References

- Bhagwati, Jagdish (1997), ‘Wall Street, not economics, dictates the agenda’, interview in *The Times of India*, 31 December, accessed 18 January 1998.
- Chang, Ha-Joon (2002), *Kicking Away the Ladder, Development Strategy in Historical Perspective*, London: Anthem.
- Caufield, Catherine (1998), *Masters of Illusion, The World Bank and the Poverty of Nations*, London: Pan.
- Hoekman, Bernard (2005), ‘Making the WTO More Supportive of Development’, *Finance and Development*, 42 (1): 14–18.

- IBRD (1989), *Project Performance Results for 1987*, Operations Evaluation Department, Washington, DC: IBRD.
- IBRD (1999), *1998 Annual Review of Development Effectiveness*, Washington, DC: OED.
- IBRD (2005), 'About Us', <http://web.worldbank.org/WBSITE/EXTERNAL/EXTABOUTUS/0,,contentMDK:20046292~menuPK:51123588~pagePK:50004410~piPK:36602~theSitePK:29708,00.html>
- Khan, Moshin S. (1990), 'The Macroeconomic Effects of Fund-Supported Adjustment Programs', *IMF Staff Papers*, 37 (2): 195–232.
- Khor, Martin (2005), 'The Commodities Crisis and the Global Trade in Agriculture: Present Problems and Some Proposals', in Helsinki Process Secretariat (ed.), *Helsinki Papers on Global Economic Agenda*, Helsinki: Foreign Ministry, pp. 167–89.
- List, F. (1920), *Das nationale System der politischen Ökonomie*, Jena: Fischer: first published (1841).
- Mattoo, Aaditya and Arvind Subramanian (2005), 'Why Prospects for Trade Talks are Not Bright', *Finance and Development*, 42 (1): 19–21.
- Maurizio, Roberto (1983), 'IFAD: A New Tool for Agriculture', *Politica Internazionale*, 3 (1): 105–18.
- Mosley, Paul, Jane Harrigan and John Toye (1991), *Aid and Power, The World Bank and Policy Based Lending*, Vol. 1, London and New York: Routledge.
- OECD (1985), *Twenty-Five Years of Development Co-operation: A Review, 1985 Report*, Paris: OECD.
- OECD (2000), *Development Co-operation, Efforts and Policies of the Members of the Development Assistance Committee, 1999 Report*, Paris: OECD.
- Oxfam (2005), 'A Round for Free, How Rich Countries are Getting a Free Ride on Agricultural Subsidies at the WTO', Briefing Paper no. 76.
- Polak, Jaques, J. (1991), *The Changing Nature of Conditionality*, Essays in International Finance No. 184, Princeton, NJ: Princeton University.
- Prebisch, Raúl (1949), 'El desarrollo económico de la América Latina y algunos de sus principales problemas', *El Trimestre Económico*, (63): 347–413; English translation (1950), *The Economic Development of Latin America and its Principal Problems*, New York: United Nations.
- Raffer, Kunibert (1987), 'Tendencias Towards a "Neo Listian" World Economy', *Journal für Entwicklungspolitik (JEP)*, 3 (3): 45–57.
- Raffer, Kunibert (2004), 'International Financial Institutions and Financial Accountability', *Ethics and International Affairs*, 18 (2): 61–78.
- Raffer, Kunibert (2005), 'The Present State of the Discussion on Restructuring Sovereign Debts: Which Specific Sovereign Insolvency Procedure?' in UNCTAD (ed.), *Proceedings of the Fourth Inter-regional Debt Management Conference and WADMO Conference*, Geneva and New York: UN, pp. 69–74; and <http://r0.unctad.org/dmfas/pdfs/raffer.pdf>.
- Raffer, Kunibert and H.W. Singer (2001), *The Economic North–South Divide: Six Decades of Unequal Development*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar; Paperback edn (2002, 2004).
- Rodrik, Dani (1996), 'Understanding Policy Reform', *Journal of Economic Literature*, 34 (1): 9–42.
- Singer, H.W. (1984), 'The Terms of Trade Controversy and the Evolution of Soft Financing: Early Years in the UN', in Gerald M. Meier and Dudley Seers (eds), *Pioneers in Development*, Oxford: OUP, pp. 275–304.
- Stern, Ernest (1983), 'World Bank Financing and Structural Adjustment', in John Williamson (ed.), *IMF Conditionality*, Washington, DC: IIF and MIT, pp. 87–108.
- Toye, John and Richard Toye (2003), 'The Origins and Interpretation of the Prebisch–Singer Thesis', *History of Political Economy*, 35 (3): 437–67.
- UNCTAD (2002), *Economic Development in Africa: From Adjustment to Poverty Reduction: What is New?* Geneva: UN.
- Viner, Jacob (1937), *Studies in the Theory of International Trade*, London: Allen & Unwin.
- WTO (1998), *Annual Report 1998*, Vol. 1, Geneva: WTO.
- WTO (2000), *Annual Report 2000*, Geneva: WTO.

WTO (2005a), 'Understanding the WTO: Basics – What is the World Trade Organization?'
http://www.wto.org/english/thewto_e/whatis_e/tif_e/fact1_e.htm.

WTO (2005b), 'Understanding the WTO: Basics – The Case for Open Trade',
http://www.wto.org/english/thewto_e/whatis_e/tif_e/fact3_e.htm.

46 North–South issues

*Amitava Krishna Dutt*¹

Introduction

It has become established practice in development economics to view the world as being divided into two parts, one comprised of rich, developed countries and the other comprising poor, less-developed countries. Although other terms have also been used to describe this division – including Center (or Core or Metropole) and Periphery, the West and the Rest, the First World and the Third World – the North–South terminology, despite its less than complete geographical accuracy, has become the most popular one, as reflected by its usage in North–South (or Brandt) Commission, ‘North–South’ negotiations in international organizations and other arenas, and the literature on ‘North–South’ models.

The relevance of the division between the North and South has sometimes been questioned because of the lack of homogeneity within the North and especially the South (as reflected, for instance, in the emergence of the newly industrialized countries or NICs from the ranks of the South, with characteristics very different from the least-developed and often stagnating economies of Africa and elsewhere) and because of divergent interests within them which has reduced their ability to bargain as unified blocs of countries. Despite this, however, the North–South dichotomy has remained an enduring one – in both descriptive and analytic senses – for a number of reasons. First, there are great differences in levels of economic development – however measured – between rich and poor countries, and it is of enormous ethical importance to examine whether such differences are persistent or not, and if they are, what can be done to remove them and bring about greater international equality. Second, one of the central issues in all of economics is why poor countries remain poor and what can be done to raise their levels of income and production, and it is possible that much can be learnt by comparing the experiences of rich and poor countries. Third, it is widely perceived that there are major asymmetries in the economic characteristics of rich and poor countries, and that this may well imply that trade, factor movements and other interactions between them will have asymmetric impacts on them. As the interaction between countries seems to be growing due to the widely noted process of globalization, it is important to explore the effects of such interaction on rich and poor countries. Fourth, there are many issues, including those of trade

policy, the activities of transnational corporations (TNCs), international labor migration, intellectual property rights and the global environment, on which the interests of rich and poor countries appear to be opposed, and on which cooperation between them is arguably required for their mutual benefit. All these issues can be examined and analyzed most simply by dividing the world into a rich and a poor part. It is little wonder that large and growing theoretical, empirical and policy-oriented literatures have developed, analyzing the world economy in terms of the North–South dichotomy.

This chapter provides a flavor of the main issues raised in these discussions.² First, it discusses the main findings on whether the gap in levels of development between the North and the South have been growing or falling. Second, it briefly reviews some contributions to the theoretical literature on North–South models. Following that, it examines in turn the implications of, and issues concerning, trade, capital flows, labor migration, technology transfers and environmental interactions between the North and the South. The final section mentions other North–South issues and some issues which are downplayed by dividing the world into a North and a South.

North–South inequality

A large body of empirical research has examined changes in the level of international inequality and the gap between rich and poor countries. This issue has been investigated in various ways, mostly using purchasing power parity-adjusted per capita real GDP figures. Many observers have employed standard indicators of inequality – such as the Lorenz curve, the Gini coefficient and the Theil index – to show that inequality across countries has increased; Sala-i-Martin (1996) shows that the standard deviation of the log of per capita real GDP for 110 countries increased more or less steadily between 1960 and 1990, implying what is called σ -divergence. An alternative technique regresses growth rates of per capita GDP for the 1960–90 period on the logarithm of initial level of per capita GDP to find a positive coefficient, implying that richer countries on average grow faster, so that there is β -divergence (Sala-i-Martin, 1996). Quadratic regression equations involving the same variables find an inverse U-shaped relationship, implying a positive relationship between starting income level and per capita growth for most of the sample, and a negative one for a small group of high-income countries. Quah (1993), examining the distribution of per capita GDP levels (relative to the world average), finds that the distribution tends over time to one with a thinning middle and accumulation at the two tails (the so-called twin-peaks phenomenon), and that countries very seldom move from low to high ends of the distribution. In sum, these

figures suggest growing inequality among countries, with convergence among a group of rich countries. Overall growth figures for groups of countries tell a similar story: according to World Bank data, the richest one-third of countries on average grew by an annual rate of 1.9 percent between 1970 and 1995, whereas the middle third grew by only 0.7 percent and the bottom third showed hardly any growth at all (Scott, 2001, pp. 162–3). Pritchett (1997), using different plausible estimates for initial levels of income for poor countries (for which hard data are not available), finds that there has been ‘divergence, big time’ between rich and poor countries over the last 150 years.

These findings, however, have been disputed by some analysts, who point out that by treating each country as one observation, most studies do not give adequate weight to the large low-income countries, China (which is sometimes left out of the sample due to lack of data) and India, which have experienced relatively high rates of growth in recent years (Dollar and Kraay, 2002; Bhalla, 2003). They argue that correcting the problem yields the result that poor countries – as a whole – are growing faster than rich countries. However, there may be a case for giving each country an equal weight, since each country follows a particular set of policies and is thus a single observation. Thus, if we are interested in the relative performance of countries (rather than individuals) it is more appropriate to give equal weights to countries. If we are interested in inequality among people in the world, population-weighted measures are appropriate, but because such measures implicitly assume that there is no inequality within countries, they have to be supplemented by data on within-country inequality to measure inequality among people (see Milanovic, 2005).

Even if the North and South are experiencing divergent patterns of development, it does not imply that this divergence is due to the nature of their interaction. Divergence could be occurring due to factors internal to them, as explored in models of poverty traps which formalize the earlier literature on vicious and virtuous circles in closed economies (see Azariadis and Stachurski, 2005). Since the empirical exploration of the relationship between North–South interactions and divergence is still in its infancy, we may turn to theoretical North–South models to examine the implication of such interaction.

North–South models

North–South models are not new. The classical economists, including Adam Smith (who discussed the importance of increasing productivity due to the division of labor as economies grow) and David Ricardo (who examined the role of trade in postponing the arrival of the stationary state in rich countries by enabling cheap food imports and in having the opposite effect

for poor countries), discussed North–South issues. Subsequently, Marxist writers on imperialism and dependency theorists, and development economists more generally, stressed the role of the South in providing markets and investment outlets for the North, and examined the problems of surplus transfers from the South, the deterioration of the Southern terms of trade, and of uneven development. For instance, Lewis's (1954), pioneering contribution on economic development with unlimited supplies of labor examined the determination of the terms of trade between rich and poor countries in a world with three goods: a Northern good, a Southern good and a non-traded good which both regions produce. Lewis showed that technological change in the Southern goods sector, low and stagnant productivity of the non-traded goods sector (food) in the South, and high and rising productivity in that sector for the North, have the effect of turning the adverse Southern terms of trade further against the South.

More mathematically explicit general equilibrium North–South models which endogenize capital accumulation in the two regions emerged from the early 1980s. Findlay (1980) examines capital accumulation in a global economy with the North growing with full employment as in Solow's model and the South with unlimited supplies of labor at a fixed real wage as in the Lewis model. Taylor (1983) allows for unemployment in the North as well, assuming that effective demand determines Northern growth as in the Kaleck–Keynes approach. These and other models can be thought of as special cases of a general framework (see Dutt, 1990) in which the North and South are completely specialized in the production of their good, where the Northern good is a consumption-cum-investment good, and the Southern a pure consumption good. The models embody specific behavioral and institutional assumptions for the North and the South, thereby highlighting their structural differences, and assume that fixed fractions of consumption expenditure in each region are spent on the two products (that is, preferences are homothetic), and trade between the two regions is balanced. The framework examines short-run equilibria with given stocks of capital with markets for both goods clearing, and long-run equilibria in which the stocks of capital in the two regions grow at the same rate due to saving and investment. The models are used to examine the effects of changes in such things as technology, consumption expenditure patterns and savings rates. Of particular interest are results which demonstrate that Southern growth depends on Northern growth, which is determined independently of the South (which imply that attempts by the South to grow faster by saving more merely imply a deterioration in its terms of trade, revealing its dependent status), the relation between the Southern terms of trade and Southern growth, and the possibility of uneven development (reflected by a rise in the relative capital stock of the North to the South)

due, for instance, to technological change, changes in consumer preferences and industrial concentration in the North (see Dutt, 1990). These results confirm some of the informal ideas of earlier writers on uneven development, but they depend on some of the specific assumptions made about the structures of the Northern and Southern economies. The models have also been extended to incorporate issues such as international capital flows and technology transfers.

The models stressing structural asymmetries between the North and the South do not explain why such asymmetries arise. Implicitly, they assume that events in the past, such as the Atlantic slave trade (Darity, 1992) or colonial domination and consequent policy regimes (Dutt, 1992a), related to the balance of military power (Findlay, 1992) create and lock in these structural differences. Models which assume identical structures for the two regions have also been developed to show how ‘small’ historical events can make one region (the North) end up exporting goods exhibiting increasing returns to scale and productivity-enhancing learning effects, while the other region (the South) becomes more oriented towards goods exhibiting constant returns and relatively technologically stagnant sectors, so that there is uneven development (Krugman, 1981, 1990).

The models discussed so far can be seen as reactions to the dominant neoclassical Heckscher–Ohlin–Samuelson (HOS) trade models which contain optimizing agents, are usually static in nature, and assume that markets clear so that labor and other resources are fully employed everywhere. However, the neoclassical approach has also contributed to the development of North–South models. Most of the North–South models from the neoclassical perspective, however, have followed the contributions of new growth theory, which emphasize the role of increasing returns and externalities in the growth process (see Darity and Davis, 2005). Many new growth theory models imply economic divergence between rich and poor countries due to economies of scale along Smithian lines even without any interaction between the two. These results often carry over to models with North–South trade, as we shall see below.

North–South trade

According to the canonical HOS theory of trade, with its assumptions of constant returns to scale and perfect competition, countries export goods which use their abundant factor intensively. It is generally assumed that the North is capital-abundant and the South labor-abundant, and trade with the North exporting capital-intensive goods and the South labor-intensive ones results in gains from trade through the reallocation of resources according to comparative advantage. With the addition of the assumption of identical technology across trading partners and some other conditions,

the theory also implies – according to Samuelson’s factor price equalization theorem – that trade equalizes factor returns across countries even without any factor mobility between them, since in effect abundant factors move abroad embodied in traded goods. These implications are frequently invoked as proof of mutually beneficial North–South trade, and of convergence. Modifications of the approach, which focus on skilled and unskilled labor as the two factors and assume that the North is skilled-labor abundant, have also been used to examine how the North can suffer increasing inequality and (with rigid wages) the unemployment of unskilled labor (Wood, 1994).

The HOS approach, however, does not imply that countries necessarily gain from trade or from moving to freer trade. If economies are ‘distorted’, for example because of the existence of production externalities or factor market rigidities, the economy may well lose from trade (see Bhagwati and Srinivasan, 1983). If such ‘imperfections’ are more prevalent in the South, as commonly supposed, North–South trade may not benefit the South. Dynamic extensions to the approach which incorporate some of the insights of new growth theory even imply uneven development. For instance, when trade liberalization leads to a rise in the wages of skilled workers or the return to capital in the North and reductions in the South, it can speed up human and physical capital accumulation in the North and slow it down in the South, resulting in a growing gap between the two regions (Baldwin, 1992; Stokey, 1991).

More drastic departures from the HOS approach may also imply divergent growth patterns. North–South models in which the North is specialized in goods with higher income elasticities than Southern goods imply that balanced growth for the two regions creates an excess supply of the Southern goods and a deterioration of the Southern terms of trade (Thirlwall, 1979). Long-run equilibrium in models with such a property – which seems empirically plausible – leads to uneven development (see Dutt, 2003). Models such as Krugman’s (1981 [1990]), in which the North specializes in the production of relatively sophisticated goods which exhibit increasing returns and dynamic learning and spin-off effects, while the South specializes in goods which do not have these properties, also imply uneven development and possible losses from trade for the South, a result which is found in a range of similar models (see Ethier, 1982; Boldrin and Scheinkman, 1988).

If insights such as these have any validity, the policy implication for the South is to attempt to change its pattern of specialization with industrial and trade policies towards goods with favorable demand and technological characteristics. Whether individual Southern countries can do so effectively is, of course, another matter. But the experiences of late industrializers in

the past, such as the USA and Germany (where such policies were espoused by Alexander Hamilton and Friedrich List) and Japan, and more recent success stories of South Korea and Taiwan, certainly point to their importance (see Amsden, 2001).

North–South capital movements

According to standard neoclassical theory the mobility of factors of production leads to production convergence. The argument can be made in its simplest form for a one-good, two-factor – capital and labor – world with diminishing returns to factors of production, perfect competition, flexible prices and profit-maximizing behavior for the case of capital movements (see Bhagwati, 1979). With the North being capital-abundant, the marginal productivity of capital is low compared to that of the South, where capital is scarce (assuming superior Northern technology does not outweigh the strength of diminishing returns). With the rental rate equal to the marginal product of capital, capital will move from the North to the South, increasing per capita production in the South and reducing it in the North, increasing production in the world as a whole, and increasing per capita income (taking into account payments to capital) in both regions.

In fact, capital does not move as much from the North to the South as is suggested by this theory, as shown by the fact that for substantial periods of time there has been a reverse transfer of financial resources from the South to the North. To the extent that capital does move from the North to the South, it moves to a small group of countries. Modifications of the neoclassical model provide reasons why this may be so. If we replace the assumption of diminishing returns by increasing returns, introducing external economies, or internal scale economies with imperfect competition, it is no longer the case that the capital-abundant North has a lower return to capital than the South, and capital will move in the reverse direction (Lucas, 1990). If lenders do not know exactly what borrowers do with borrowed funds and can only observe outcomes of their activity, while borrowers know what they are doing, we have the problem of asymmetric information, and lenders will require collateral to ensure that borrowers do not willfully default. The implication of this is that borrowers in rich countries who have higher initial endowments of capital will be able to borrow more than those in poor countries, because they can put up collateral to overcome moral hazard problems, while borrowers in poor countries are less able to do so (Gertler and Rogoff, 1990; Matsuyama, 2004). This may imply that capital will flow from poor to rich countries, making rich countries even richer, resulting in a process of uneven development.

If capital does move from the North to the South, the effects may not be as implied by the simple neoclassical theory. If we depart from the

one-good assumption, and allow the North and South to trade in different products, capital flows from the North to the South can turn the terms of trade against the South (Singer, 1950), and lead to Southern immiserization of the type discussed by Bhagwati (1956). These insights have been incorporated into some North–South models. Burgstaller and Saavedra-Rivano (1984) introduce capital mobility into Findlay’s model and find that with capital mobility (compared to the case without it) Northern per capita income will be higher, but Southern income per worker will be lower (because of the payments that have to be made for foreign capital), and the terms of trade of the South deteriorates. Relative Southern employment will also fall if the Northern propensity to spend on the Southern good is less than that out of Southern profits, since there is a reduction in the demand of the Southern good due to the redistribution of income caused by capital mobility. In Blecker’s (1996) model in which Northern growth is driven by aggregate demand, greater capital mobility leads to uneven development in the sense of an increase in the stock of Northern capital to Southern capital, due to Southern terms-of-trade deterioration. Such outcomes, however, depend on the assumption that capital flows increase the production of the Southern good, thereby resulting in a deterioration of its relative price. If, as pointed out in some empirical studies, capital flows in the form of foreign direct investment feature deep integration, involving the production of typically Northern goods in the South, capital flows could lead to more even development patterns as the North loses markets to the South and experiences greater excess capacity and unemployment and lower growth, and lower profits encourage more capital flows to the South (see Dutt, 1996). TNCs bring in not only capital, but also technology and exporting capabilities, which make these effects more likely. Such outcomes are more supportive of those who fear Northern stagnation due to capital outflows.

However, North–South capital flows may not bring about Southern development because they may not be productively invested. Borrowing by the South in the past, often due to loan-pushing by Northern banks, led to the accumulation of Southern debt but often went right back to the North through capital flight by corrupt elites (Darity and Horn, 1988). Debt-ridden Southern countries are then forced to make interest payments and repay loans by running current account surpluses and reducing growth. Default can threaten financial stability in the North as well. Portfolio flows and bank loans often finance short-term investment in the South in stock markets and real estate, leading to speculative bubbles which, when they burst, bring about sharp capital outflows, currency crises and macroeconomic contraction (see Stiglitz, 2002).

Labor migration

The standard one-good neoclassical model with diminishing returns and perfect competition discussed earlier also implies convergence of per capita production due to the movement of people from the labor-abundant South to the labor-scarce North. Such an approach, which implies that migration leads to a reduction in the Northern wage, can be used to explain some of the opposition to immigration in the North. It can also be used to explain income losses to those who are left behind in the South, because of the disappearance of the surplus produced by the workers who migrate (see Bhagwati, 1979).

Such a simple framework does not capture the complexities of real-world migration, however. Countries of the North heavily restrict the migration of unskilled workers from the South, allowing mainly the legal immigration of skilled workers. If one distinguishes between skilled and unskilled workers, and introduces scale economies and imperfect competition, uneven development due to brain drain from the South may occur. For instance, if unskilled workers and intermediate services are used in the production of the final good under conditions of constant returns to scale and perfect competition, while non-traded intermediate services are differentiated products, each produced by a monopolistic competitor under conditions of increasing returns with skilled labor as the only factor of production, the migration of skilled labor from the South to the North reduces the number of intermediate goods, quantity of the final good produced, and per capita income in the South, and has an opposite effect in the North, implying uneven development (Dutt, 2005). The endogenization of skilled labor supply, due to tax-financed government educational spending or the decision by workers to accumulate human capital, into new growth theory models can also imply divergence growth patterns (Haque and Kim, 1995; Wong and Yip, 1999).

Technology transfers

If knowledge is something that all countries can share, it may be supposed that the South will eventually catch up with the North. Lucas (2000) develops a simulation model in which countries take off in sequence (depending on their internal conditions) and in which latecomers grow faster than leaders because they have access to technology and policy experiences of the latter. His model implies that although international inequality increases initially, 'sooner or later everyone will join the industrial revolution . . . economies will grow at the rate common to the wealthiest economies, and . . . percent differences in income levels will disappear' (Lucas, 2000, p. 166).

Such an outcome, however, does not occur even in standard neoclassical models of technology transfer. In Krugman's (1979 [1990]) model of

product innovation which formalizes the product life cycle approach, and which assumes that Northern 'new' goods become 'old' goods produced by the South at a constant rate of 'radioactive' decay, the ratio of Northern to Southern income stabilizes in equilibrium, but does not become unity. In models of technology transfer involving process innovation, where the change in Southern productivity depends positively on the North–South productivity gap (reflecting more opportunities for transfer), a narrowing of the technology gap occurs if there is initially a large gap, but complete technological catch-up does not occur unless transfers occur at an infinite rate.

Models in which rates of technology transfer are either constant or monotonically related to the relative technology gap (a higher technology gap leading to a faster rate of technology transfer) may not capture the realities of the process of technology transfer. Since much technological knowledge is tacit, and requires constant modification and adaptation, the process of transferring technology is not so very different from that of innovation, and successful technology transfer requires the development of some amount of social and technological capability of the South (Abramovitz, 1986; Bell and Pavitt, 1993). This can have a number of consequences. First, if we measure (relative) technological capability by the South–North productivity gap, a large productivity gap may increase the potential for technology transfer, but reduce the ability of the South to do so successfully. In this case, even standard models will imply convergence of the technology gap if the gap is not too large, but divergence if the North and South are too far apart (Verspagen, 1991), which may explain the formation of convergence clubs of rich countries and the exclusion of many poor countries. Second, technology transfer may be linked to other aspects of North–South relations. To the extent that technological capability is enhanced by the presence of TNCs, FDI can speed up technology transfer. However, an adverse impact of indigenous technological development cannot be ruled out if competition between TNCs and domestic firms hurt the latter. Trade liberalization and its consequence on the composition of production may, as mentioned earlier, have adverse effects on learning by doing and slow down the accumulation of technological capability (see Van der Klundert and Smulders, 1996).

In addition to these issues, the protection of intellectual property rights (IPRs) can serve as a barrier to technology transfers. Indeed, the protection of IPRs has been a major source of conflict between the North and the South, as reflected in the Uruguay Round discussions of the WTO. Since most technological innovation occurs in the North, it is not surprising that the North, especially its innovating firms, has an interest in protecting IPRs internationally, while the South, which is more interested in the speedy

diffusion of technology from the North to the South, is against such protection. These insights are confirmed by simple partial equilibrium models of product and process innovations in terms of consumer and producer surpluses in the North and South (see Chin and Grossman, 1990; Deardorff, 1992).

Although the interests of the North and South may be in conflict, it is often argued that the international IPR protection may be good for the world as a whole, because in its absence there would be insufficient innovation, since innovating firms would receive smaller rewards for their research and development activity and therefore conduct less of it. However, the partial equilibrium models imply that the effect of international IPR protection on world welfare (measured by the total surplus accruing to the North and the South) is ambiguous: if the North is a large part of the economy and protects IPRs internally, there will be sufficient innovation in the North to make world welfare higher without international IPR protection than with it. This result can carry over to a dynamic general equilibrium setting: in Grossman and Helpman's (1991) model weaker protection of IPRs will not only speed up technology transfers, but also accelerate innovation in the North, because Northern resources are devoted more to innovation, rather than to production, which increasingly moves to the South.

While it is often the case that tighter international IPR protection slows down technology transfer, this may not be true under all circumstances. For instance, with IPR protection TNCs may more willing transfer better technology abroad, while without it they are likely to hold it back, fearing imitation (see Lai, 1998). However, Glass and Saggi (2002) find that although stronger IPR protection makes TNCs and Northern firms safer from imitation, the greater difficulty in imitation results in more resources being absorbed in imitative activity which reduces FDI, as well as Northern innovation.

Environmental issues

The literature on the environmental Kuznets curve suggests that the rich North is able to deal better with environmental problems than the South, where poverty, population growth and lax pollution control have an adverse environmental effect which can in turn have an adverse effect on growth and other development indicators, such as health. North–South interactions through trade and capital movements may exacerbate such problems, if they lead to the movement of the production of relatively dirty and resource-intensive goods from the North to the South, for instance, because of less-stringent environmental regulation; although, if capital-intensive Northern goods are more pollution-intensive, this may not necessarily occur (Copeland and Taylor, 2003).

Further North–South interactions regarding the environment occur because of the global public goods nature of several environmental issues, such as global warming, depletion of the ozone layer and loss of biodiversity (Sandler, 1997). The North blames the South for its increasing contribution to damaging the global environment, because of high population growth and pollution-intensive growth, while the South blames the North because of its high per capita contribution to this damage. The solution of global environmental problems requires such conflicts to be resolved by mutually beneficial international agreements between the North and the South which take into account the poverty of the South.

Conclusion

The North–South approach can be criticized for downplaying the differences within rich and poor countries, and the possibility that poor countries may grow and join the ranks of the rich. However, two-region models have been extended to include a third, consisting of newly industrialized countries, to explore the causes of its growth and to analyze whether its growth results in the end of uneven development or the exacerbation of the gap between the rest of the South and the North (see Dutt, 1992b). Moreover, multi-region models have been developed to show how the countries can endogenously sort themselves into Northern and Southern groups (see Matsuyama, 1996, 2004).

This brief review has confined attention to only a few of the major North–South issues, not examining other areas of interaction (for instance, due to cultural and political influences), conflict (such as those over labor standards) and cooperation (related to the existence of global public goods such as global health and security). Nevertheless, it suggests that the North–South framework remains a useful way of looking at many important and interesting development issues.

Notes

1. I am grateful to Jaime Ros for his helpful comments on an earlier draft of this chapter.
2. More details can be found in the other chapters in Part VI of this *Handbook*.

References

- Abramovitz, Moses (1986), 'Catching Up, Forging Ahead, and Falling Behind', *Journal of Economic History*, 46 (2): 385–406.
- Amsden, Alice H. (2001), *The Rise of the 'Rest': Challenges to the West from Late-Industrializing Economies*, Oxford and New York: Oxford University Press.
- Azariadis, Costas and John Stachurski (2005), 'Poverty Traps', in P. Aghion and S. Durlauf (eds), *Handbook of Economic Growth*, Vol. 1A, Amsterdam: Elsevier, pp. 295–384.
- Baldwin, Richard E. (1992), 'Measurable Dynamic Gains from Trade', *Journal of Political Economy*, 100 (1): 162–74.

- Bell, Martin and Keith Pavitt (1993), 'Accumulating Technological Capability in Developing Countries', *Proceedings of the World Bank Annual Conference on Development Economics 1992*, Washington DC: World Bank.
- Bhagwati, Jagdish (1956), 'Immiserizing Growth: A Geometrical Note', *Review of Economic Studies*, June: 201–5.
- Bhagwati, Jagdish (1979), 'International Factor Movements and National Advantage', *Indian Economic Review*, 14 (2): 73–100.
- Bhagwati, Jagdish and T.N. Srinivasan (1983), *Lectures on International Trade*, Cambridge, MA: MIT Press.
- Bhalla, Surjit S. (2003), *Imagine there's No Country: Poverty, Inequality and Growth in the Era of Globalization*, New Delhi: Penguin Books.
- Blecker, Robert (1996), 'The New Economic Integration: Structuralist Models of North–South Trade and Investment Liberalization', *Structural Change and Economic Dynamics*, 7: 321–45.
- Boldrin, M. and J.A. Scheinkman (1988), 'Learning by Doing, International Trade and Growth: A Note', in P.W. Anderson, K.J. Arrow and D. Pines (eds), *The Economy as an Evolving Competitive System*, Redwood City, CA: Addison-Wesley, pp. 285–99.
- Burgstaller, Andre and Neantro Saavedra-Rivano (1984), 'Capital Mobility and Growth in a North–South Model', *Journal of Development Economics*, 15 (1,2,3): 213–37.
- Chin, Judith C. and Gene M. Grossman (1990), 'Intellectual Property Rights and North–South Trade', in Ronald W. Jones and Anne O. Krueger (ed.), *The Political Economy of International Trade: Essays in Honor of Robert E. Baldwin*, Oxford: Blackwell, pp. 90–107.
- Copeland, Brian R. and Scott M. Taylor (2003), *Trade and the Environment*, Princeton, NJ: Princeton University Press.
- Darity, William A. (1992), 'A Model of Original 'Sin': Rise of the West and Lag of the Rest', *American Economic Review*, 82 (2): 162–7.
- Darity, William A. and Lewis S. Davis (2005), 'Growth, Trade and Uneven Development', *Cambridge Journal of Economics*, 29: 141–70.
- Darity, William A. and Bobbie L. Horn (1988), *The Loan Pushers: The Role of Commercial Banks in the International Debt Crisis*, Cambridge, MA: Ballinger.
- Deardorff, Alan V. (1992), 'Welfare Effects of Global Patent Protection', *Economica*, New Series, 59 (233): 35–51.
- Dollar, David and Aart Kraay (2002), 'Spreading the Wealth', *Foreign Affairs*, January–February: 120–33.
- Dutt, Amitava Krishna (1990), *Growth, Distribution and Uneven Development*, Cambridge: Cambridge University Press.
- Dutt, Amitava Krishna (1992a), 'The Origins of Uneven Development: The Indian subcontinent', *American Economic Review*, 82 (2): 146–50.
- Dutt, Amitava Krishna (1992b), 'The NICs, Global Accumulation and Uneven Development: Implications of a Three-Region Model', *World Development*, 20 (8): 1159–71.
- Dutt, Amitava Krishna (1996), 'Direct Foreign Investment, Globalization, and Northern Growth: Implications of a North–South Model', *Review of Radical Political Economics*, 28 (3): 102–14.
- Dutt, Amitava Krishna (2003), 'Income Elasticity of Imports, North–South Trade and Uneven Development', in A.K. Dutt and J. Ros (eds), *Development Economics and Structuralist Macroeconomics*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 307–35.
- Dutt, Amitava Krishna (2005), 'Globalization, South–North Migration and Uneven Development', in P. Gangopadhyay and M. Chatterji (eds), *Economics of Globalization*, Aldershot: Ashgate, pp. 115–29.
- Ethier, W.J. (1982), 'Decreasing Costs in International Trade and Frank Graham's Argument for Protection', *Econometrica*, 50: 1243–68.
- Findlay, Ronald (1980), 'The Terms of Trade and Equilibrium Growth in the World Economy', *American Economic Review*, 70 (3): 291–9.
- Findlay, Ronald (1992), 'The Roots of Divergence: Western Economic History in Comparative Perspective', *American Economic Review*, 82 (2): 158–61.

- Gertler, Mark and Kenneth Rogoff (1990), 'North-South Lending and Endogenous Domestic Capital Market Inefficiencies', *Journal of Monetary Economics*, 26 (2), October: 245-66.
- Glass, Amy Jocelyn and Kamal Saggi (2002), 'Intellectual Property Rights and Foreign Direct Investment', *Journal of International Economics*, 56, 387-410.
- Grossman, Gene and Elhanan Helpman (1991), *Innovation and Growth in the World Economy*, Cambridge, MA: MIT Press.
- Haque, Nadeem U. and Se-Jik Kim (1995), "'Human Capital Flight": The Impact of Migration on Income and Growth', *IMF Staff Papers*, 42 (3): 577-607.
- Krugman, Paul (1979), 'A Model of Innovation, Technology Transfer, and the World Distribution of Income', *Journal of Political Economy*, 87 (2): 253-66; reprinted in Paul Krugman (1990), *Rethinking International Trade*, Cambridge, MA: MIT Press, pp. 139-51.
- Krugman, Paul (1981), 'Trade, Accumulation, and Uneven Development', *Journal of Development Economics*, 8 (149-61); reprinted in Paul Krugman (1990), *Rethinking International Trade*, Cambridge, MA: MIT Press, pp. 93-105.
- Krugman, Paul (1990), *Rethinking International Trade*, Cambridge, MA: MIT Press.
- Lai, Edwin L.C. (1998), 'International Intellectual Property Rights Protection and the Rate of Product Innovation', *Journal of Development Economics*, 55: 115-30.
- Lewis, W. Arthur (1954), 'Economic Development with Unlimited Supplies of Labour', *Manchester School of Economic and Social Studies*, 22 (2): 139-91.
- Lucas, Robert E. (1990), 'Why Doesn't Capital Flow from Rich to Poor Countries?', *American Economic Review*, 80: 92-6.
- Lucas, Robert E. (2000), 'Some Macroeconomics for the 21st Century', *Journal of Economic Perspectives*, 14 (1): 159-68.
- Matsuyama, Kiminori (1996), 'Why are there Rich and Poor Countries? Symmetry-Breaking in the World Economy', *Journal of Japanese and International Economics*, 10 (4): 419-39.
- Matsuyama, Kiminori (2004), 'Financial market globalization, symmetry breaking and endogenous inequality of nations', *Econometrica*, 72 (3): 853-84.
- Milanovic, Branko (2005), *Worlds Apart. Measuring international and global inequality*, Princeton, NJ: Princeton University Press.
- Myrdal, Gunnar (1957), *Rich Lands and Poor Lands*, New York: Harper & Brothers.
- Pritchett, Lant (1997), 'Divergence, Big Time', *Journal of Economic Perspectives*, 11 (3): 3-17.
- Quah, Danny T. (1993), 'Empirical Cross-Section Dynamics in Economic Growth', *European Economic Review*, 37 (2-3): 426-34.
- Sala-i-Martin, Xavier (1996), 'The Classical Approach to Convergence Analysis', *Economic Journal*, 106: 1019-36.
- Sandler, Todd (1997), *Global Challenges*, Cambridge: Cambridge University Press.
- Scott, Bruce R. (2001), 'The Great Divide in the Global Village', *Foreign Affairs*, 80 (1), January-February: 160-77.
- Singer, Hans (1950), 'The Distribution of Gains between Investing and Borrowing Countries', *American Economic Review, Papers and Proceedings*, 4 (2): 473-85.
- Stiglitz, Joseph E. (2002), *Globalization and its Discontents*, New York: W.W. Norton.
- Stokey, Nancy J. (1991), 'Human Capital, Product Quality and Growth', *Quarterly Journal of Economics*, 106 (2): 587-616.
- Taylor, Lance (1983), *Structuralist Macroeconomics*, New York: Basic Books.
- Thirlwall, Anthony P. (1979), 'The Balance of Payments Constraint as an Explanation of International Growth Rate Differences', *Banca Nazionale del Lavoro Quarterly Review*, March (128): 45-53.
- Van der Klundert and Sjak Smulders (1996), 'North-South Knowledge Spillovers and Competition: Convergence versus Divergence', *Journal of Development Economics*, 50: 213-32.
- Verspagen, Bart (1991), 'A New Empirical Approach to Catching Up or Falling Behind', *Structural Change and Economic Dynamics*, 2 (2): 359-80.
- Wong, Kar-yiu and Chong-Kee Yip (1999), 'Education, Economic Growth and the Brain Drain', *Journal of Economic Dynamics and Control*, 23 (5-6): 699-726.
- Wood, Adrian (1994), *North-South Trade, Employment and Inequality: Changing Fortunes in a Skill-Driven World*, Oxford: Oxford University Press.

PART VII

DISTRIBUTION AND POVERTY

47 Measures of the distribution of income and their interpretation in developing countries

*Samuel A. Morley*¹

There is a rising level of concern about income inequality and poverty in developing countries. The development process has brought rising inequality within many countries and has also increased the gap between the poorest and richest countries. All this has led to an increasingly vocal demand for a more equitable style of development. Certainly no one would wish for more inequitable development, but it is still important to stop and ask oneself what distribution measures really measure, and what comparisons of these measures over time or across countries tell us. In particular, what do trends in inequality mean in developing countries? How are they affected by the development process? Are they something that governments in those countries can affect and should worry about?

I am going to address these questions by first considering at the individual or family level what should be included in a distribution measure. Once one has a satisfactory measure at the individual level, one needs a number or index which aggregates or summarizes all the individual distribution data across some reference group. The best-known is the Gini coefficient, but there are many others whose properties will be discussed further on. But in addition to this one has to decide the group over which to aggregate the individual distribution data. What group should that be? Should it be a country, a region within a country, the whole world, or should it be confined to a group such as an age cohort within a country? Responses to this question will have a big impact. My focus in this chapter is going to be on the distribution within countries rather than for the world as a whole, or between countries, because I believe that the national state is the basic economic and administrative unit which is able to influence the distributions that matter to its citizens.

Next I want to consider briefly some problems in the interpretation of the aggregate measures. What do changes over time or comparisons across countries tell us? Finally I will discuss several of the determinants of the distribution and the ways that they are likely to affect the distribution during the development process.

Measuring inequality at the individual or family level

Even though in some ultimate sense 'fairness' may imply an equal distribution of welfare across a population, I am going to limit my discussion to the distribution of goods or income produced by an economy. I am going to assume that a more equal distribution of goods or income is in the economic sense a fairer distribution, even though I have no way of knowing or of proving that a fairer economic distribution is also a more equitable distribution of welfare. I am bringing welfare into the discussion here because it will guide me in my choice of what to measure and, more importantly, what the reference group should be for my aggregate distribution statistics.

All of my statistics on distribution come from household surveys which are now available for a large number of countries. These measures can be based on income or on consumption. The World Bank favors consumption because it believes that it is both more accurately measured and a better proxy for lifetime income than observed income. If one thinks that people are concerned with their lifetime income rather than their income in any particular year, then the distribution measure should also be based on lifetime income. While I cannot measure lifetime income, it is likely that consumption is a better proxy than observed income in any single year. Families smooth out temporary fluctuations in income by either running down saving during recessions or adding to their assets when they have a temporary windfall. In some situations consumption is not a perfect measure, however. Suppose one is interested in the relationship between stabilization or recession and the distribution of welfare. Poor families may maintain their consumption by distress selling of assets or by borrowing at high interest rates. In either case the distribution of welfare is made less equal, even though the distribution of consumption may not show that, or may understate the rise in welfare inequality. If one is interested in short-run fluctuations and their distributional impact, it is probably better to use an income-based distribution measure. For longer-run questions, the consumption-based measure is better.

Every family receives income in kind. While the statisticians have learned how to include the imputed value of home-grown food or the rent on owned houses in the total value of family consumption, they do not generally include as consumption the value of government services such as public health care or education. These can have a fairly sizable effect on measures of the distribution.

For labor market questions, the distribution of earnings is relevant. However, since welfare is based on consumption, and since the family is the social unit which converts the earnings of some of its members into consumption for all its members, one should use the distribution of income among families. I can convert that into an individual distribution for

poverty analysis by dividing total family income or consumption by the number of family members.²

Aggregation problems

I now want to address several aggregation questions.³ The first is what group the distribution should be based on. Using welfare as our criterion for answering this question, the group should be the one to which the individual belongs and across which he or she makes comparisons. That turns out to have important implications. It is customary to calculate distributions at the national level at a point in time. That is reasonable if national relative position is an important consideration to individuals in a society. But in large countries it may well be that a regional distribution is more consistent with how people see their own welfare. If inequality indexes are thought to be good proxies for social tension or the likelihood of populist demands for redistribution, it is important that the aggregation reflects the interpersonal comparisons made by people in a society.⁴

Another possible sub-national group classification is by age or cohort. Suppose that people's primary concern is their relative position in their own age cohort. Young people presumably know that there is a positive age-earnings profile. Incomes rise with age and experience. If that is so, it would seem reasonable that purely age-based inequality would be less significant as a source of social tension than inequality within an age cohort. To take this to an extreme, suppose the individuals are concerned only with intra-cohort inequality. Now imagine an economy in which each age cohort has exactly the same expected lifetime income and age-income profile. In such an economy, there would be no income inequality provided that we were able to base the distribution measure on permanent income. That would not be true if the measure was based on consumption, if consumption, like earnings, has an inverted U-shape, rising through most of an individual's working life, and then falling in retirement. In that case any aggregate distribution measure based on observed income or consumption at a point in time will show a good deal of inequality even when there is no intra- or inter-cohort inequality. In any year the younger members of the society will be earning or consuming less than their elders. Observed inequality will be higher, the steeper the age-consumption profile, the greater the return to experience or the more credit-constrained the younger members of the society are. Apparent inequality will also be higher, the faster the rate of population growth, not because younger cohorts have lower expected lifetime incomes, but simply because there will be more of them at the bottom of the income pyramid at any point in time.

Statistically in this hypothetical economy there is inequality between the old and the young. But the significance or meaning attached to that

inequality depends on the attitudes or preferences of the members of that society. If the comparator group for each member is his or her own age cohort (or their own region of the country) then a national distribution of income or consumption will be a misleading indicator of what one might call 'socially relevant' inequality.

Aggregate measures of inequality

Once a choice has been made of what to measure, one is left with the problem of how to aggregate the information into a small number of understandable statistics. There are many ways of doing this, each of which implicitly assigns certain weights to each individual observation. The most straightforward statistic is the shares of total income going to different shares of the population such as the poor and the rich, or the bottom and top quintiles.

Perhaps the best-known measure is the Gini coefficient, now used around the world to compare distributions between countries or within countries over time. The Gini coefficient is the ratio of the area between a 45° line and a curve called the Lorenz curve showing the shares of total income accruing to different shares of the population. The Gini varies between zero and one. If there is perfect equality, each individual or family earns the same income, and the actual Lorenz curve overlays the 45° line. Here the Gini coefficient is zero since the gap between the lines is zero. Maximum inequality is when one individual or family owns all the income. In this case the Lorenz curve has a backward L-shape, and the Gini is equal to one since the gap between the Lorenz curve and the 45° line is equal to the entire area under the 45° line.

There are a number of alternative aggregate measures of the distribution found in the literature. One is the coefficient of variation (CV). It is defined as the square root of the variance of income divided by the mean to make it, like the Gini, independent of the level of income. The CV has the somewhat counterintuitive property that it gives equal weight to transfers between individuals regardless of their incomes. That is, inequality changes by the same amount if changes in income occur close to or further away from the mean, provided only that the changes are of the same size:

$$CV = Var^{1/2}/\mu$$

If one wishes to put more weight on what happens at the bottom of the distribution, one can transform income into the log of income and then take the standard deviation of this transformed income measure. Since the log of income falls rapidly for small values of income and since the standard deviation uses the square of the deviation from the mean, the

transformation will increase this measure of inequality when there are many people with levels of income far below the mean. The measure is defined as:

$$SDY = \{\Sigma(\log \mu - \log X_i)^2/n\}^{1/2}$$

Distribution comparisons among countries and over time

Two of the main uses of distribution statistics are to compare distributions across countries or to measure changes in a single country over time. Deininger and Squire (1996) at the World Bank recently collected a set of comparable national Gini coefficients for 108 countries around the world. Many of these countries have estimates going back over 30 years. Table 47.1 gives their estimates of the median Gini by region and decade.

Other than the very large differences between Latin America and most other regions of the world, what is striking is that there is no convergence over time especially in Latin America. Latin inequality fell slightly in the 1960s, but then it reversed after 1970. Meanwhile there was some improvement in the Middle East, Africa and South Asia, and worsening in Eastern Europe in the 1990s after the fall of communism. In the developed countries inequality has remained roughly constant at a low level. In the other developing regions where inequality was once high, it has fallen quite sharply. Only in Latin America has inequality remained near its high initial level.

The differences between regional Gini coefficients translate into large differences in the amount or share of income going to the rich and poor. In the 1990s on average in Latin America the top 5 percent of the population received 25 percent of total income while the bottom 30 percent got only 7.5 percent. In South-East Asia the top 5 percent received only 16 percent of income while the bottom 30 percent got 12.2 percent. The comparable

Table 47.1 Decadal median Gini coefficients by region

	1960s	1970s	1980s	1990s
E Europe	25.1	24.6	25.0	28.9
S. Asia	36.2	33.9	35.0	31.9
OECD & high-income	35.0	34.8	33.2	33.7
E. Asia & Pacific	37.4	39.9	38.7	38.1
M. East & N. Africa	41.4	41.9	40.5	38.0
Sub-Saharan Africa	49.9	48.2	43.5	46.9
Latin America	53.2	49.1	49.7	49.3

Source: Deininger and Squire (1996).

figures for Africa show the top 5 percent got 24 percent, the bottom 30 percent got 10.1 percent. In the developed countries the top 5 percent got only 13 percent of total income while the poorest 30 percent got 12.7 percent. These shares imply that in Latin America the average income of the richest 5 percent is 20 times that of the poorest 30 percent. In South-East Asia the richest 5 percent have average incomes only 7.9 times that of the poorest 30 percent.

Problems in comparisons across countries

One of the problems with interpreting differences in any of the aggregate measures between countries is that they force one to make value judgments about the weights that one gives to welfare or income at different income levels. One can see this most easily for the Gini coefficient. Lorenz curves for different countries may cross as may the curves for the same country over time. The Gini will say unambiguously which country has a more equal distribution. But even assuming that one thinks that more equal distributions are preferable, can one really be sure that one prefers the more equal distribution in all cases? Take the case of two countries with the same mean income. Suppose that in country A there are relatively few poor people because the middle class has been taxed to support a generous safety net. But there are a lot of rich people. In country B the poor have a much smaller share of income than in country A, but there is a big middle class and a relatively small group of the rich. Here the Lorenz curves of the two countries cross. The curve for country A lies above that of country B at the bottom of the distribution, but below it at the top of the distribution. If one is really concerned about the welfare of the poor, one may prefer the distribution of country A, even though its Gini is larger (more unequal) than the Gini of country B.

Large countries are likely to be more heterogeneous and have significant backward regions and significant differences in regional incomes. That will make their measured income distribution less equal than in smaller countries. Similarly, national distributions are generally less equal than purely urban distributions because on average, rural incomes are lower than urban. Countries with big indigenous populations are also likely to have more unequal distributions as will be seen below. In all these cases the reason for higher inequality is the existence of a large group within a national border which does not fully participate in the process which generates economic growth. Note that none of this would be particularly relevant to cross-country comparisons if comparator groups were truly national. As mentioned above, welfare statements about distributions are based on the position of individuals relative to others in their comparator group. If that group is local, or limited to members of an indigenous group

or the rural population, then comparisons of national distributions between countries may have misleading welfare implications.

Comparisons for the same country over time

Comparisons over time in single countries may have the same ambiguities and problems of interpretation that I noted above in the discussion of crossing Lorenz curves. They have additional interpretation problems in economies in which the population is growing. In that situation, by definition, over time the populations whose distributions are being measured are not the same and it is important to make a distinction between what I will call the base period population and the later group to which it will be compared. Apart from any questions about comparator groups, changes in the observed distribution will be affected by the earnings of the new entrants. Apparent earnings of groups such as the rich or the poor will appear to rise or fall even if the earnings of the base period population do not change. Distribution statistics are based on shares of total income received by different shares of the population. When there is an increase in the income-earning population, both income shares and population shares are affected. Total and average incomes are affected by where these new entrants come into the income pyramid.

When studies of the distribution report statistics on the incomes or income shares of the rich or the poor, they mean the group at the top or the bottom of the income pyramid at different moments. They do not mean the base period rich or poor. Where population growth is rapid or where comparisons are made over long periods of time, the distinction between the distribution or growth rate of income of the base period population and the observed population is significant (see Morley, 1981).

Determinants of the distribution

I turn now to a short discussion of the determinants of the distribution of income. One should distinguish between what I will call the primary or earnings distribution and the family distribution. This first is the distribution of income to the owners of the factors of production that produce it. The family distribution is derived from the primary distribution according to what factors of production each family owns and what each of those factors earns in economic activity. Economic factors mainly affect the primary distribution while demographic factors play a key role in the corresponding family distribution.

A useful abstraction here is the factor market where the demand and supply of each factor determines its earnings. For simplicity consider four factors, skilled and unskilled labor, capital and land. Like other economic markets, prices are determined in factor markets by the interaction of the

supply of each factor and the demand for factor services as well as institutional factors such as the minimum wage. Markets clear at a price at which someone is willing to use the available supply of each of the factors. That set of market-clearing factor prices determines the relative price of skilled and unskilled labor and the rate of return on capital and land. In the labor market, it may well be the case that demand is so low that many are forced to work part-time in the informal sector or are unemployed altogether because the legally prescribed minimum wage in the formal sector exceeds the marginal product of the fully employed labor force.

On the demand side, economic growth shifts out the demand for each of the factors, which tends to raise each of their prices. What happens to relative factor prices depends on the nature of the growth process. If it is skill-intensive, skill differentials widen. If it is led by sectors such as agriculture which use mainly unskilled labor, the reverse should occur. Structural reforms such as trade liberalization also affect factor demand by changing relative goods prices and the composition of output.

The supply side is critical to understanding the dynamics of the process. In the short run the supply of factors is fixed because each of the factors is a stock which produces a flow of services. That stock can change, but that takes time. Thus in the short run there is a fixed stock of factors determined by past investment decisions, and a set of factor demand curves. The two together produce the short-run set of factor prices. While this process sets the valuation of the factors of production, one needs to know the distribution of ownership of the factors of production in order to determine the primary earnings distribution, since it is the owner that receives the payment for the factor services. The primary distribution in the short run is determined jointly by the relative factor prices that come from the factor market and by the pattern of ownership of the factors of production.

The dynamics of the process come from investment. Investment increases the stock of physical capital, while the education system produces graduates who enter the labor force and change the supply of both skilled and unskilled labor. Training and experience also increase the stock of skilled labor or what I will call from here on, 'human capital'. Migration also changes the position of factor supply curves, and that could be for either unskilled or skilled labor.

It is important to remember that investment takes time and that the amount that can be added to the stock of any of the factors is relatively small over any short-run time period. Thus, in any period as short as say a year, it is impossible to shift the factor supply curves out by more than 3–5 percent. That means that in the short run, it is changes that come from the demand side that are most likely to explain observed changes in the primary

distribution. Over time these stocks change because of additions to the physical capital stock, education, retirements and so on.

An important feature of the distribution process is the dynamic feedback between factor prices and changes in factor supplies through investment. Relative prices or returns on physical capital determined in the factor market affect investment and education decisions. If there is a rise in the rate of return on capital, investment in physical capital increases. Similarly if there is an increase in the wage differential between skilled and unskilled labor, or between university and high school graduates, that will tend to increase the demand for university education. Fewer people will enter the labor force after high school. More will continue on to earn a university degree. Over time, those decisions will increase the supply of educated labor in the labor force. If there were no changes on the demand side, this increase in the supply of skilled or educated labor should reduce the skill differential. One could thus think of the initial rise in the skill differential or the return to capital as a dynamic signal which sets in motion an investment process which eventually adjusts factor supplies and relative factor prices to a long-run equilibrium. By definition that equilibrium is one in which additions to supply are just sufficient to offset changes in the demand for factor services coming from economic growth and technological change.

From the point of view of the primary earnings distribution, a rising skill differential or profit rate which tends to increase inequality in the short run is also a signal which expands the supply of educated labor or physical capital in the long run. These long-run changes on the supply side may well reverse the short-run rise in inequality that induced them since they tend to drive down the rate of return to capital, both human and non-human.

There is thus an important distinction or ambiguity between the short-run and the long-run meaning of a rise in the skill differential or the rate of return to capital. In the short run an increase in either of these two is almost surely regressive. But in the long run, so long as the supply side reacts positively to these changes in the rate of return, the change could be progressive either because of upward mobility, or because the increase in physical capital drives down the rate of return and raises the productivity and the average wage of workers.

The ambiguity I am discussing here is a specific example of the dual function of income in a market system. On the one hand relative income determines the distribution at each point in time. Any relative increase in the income of the rich is regressive. But on the other hand income is also the signal or incentive by which economic agents are encouraged to change their behavior. A rise in the skill differential induces socially desirable education investment. Similarly a rise in profits induces investment and a shift of productive resources from less-desirable to more-desirable uses. One

makes a serious error of interpretation if one concentrates only on the short-run regressive effect of changing factor returns without taking into account progressive long-run supply responses.

Inequality and economic growth

In a classic article, Simon Kuznets empirically analysed the relationship between growth and equity, using the historical experience of England and the United States (Kuznets, 1955). He found that over the course of the nineteenth century, inequality rose as both countries grew, but somewhere around the beginning of the twentieth century the trend reversed and inequality began to decline. This non-linear, inverted U-shaped relationship between income and inequality has been labelled the Kuznets curve and has been the subject of a vast empirical literature looking for similar relationships in other countries and in international cross-sections between countries.

Kuznets's rationale for the relationship is based on the important idea that growth starts somewhere specific, in either a region, a sector or a city. From that starting point its effects spread through a variety of linkages to the rest of the economy. In the first stage of the process, inequality rises. Later, as the effect of the growth stimulus spreads out in the population, inequality begins to decline.

In the Kuznets study, the growth process was the Industrial Revolution. Industrialization started in the urban centers of agricultural economies. Incomes in the small industrial sector were much higher than those in the agricultural sector. As industry expanded, inequality at first increased because the higher urban wages benefitted only a small fraction of the population. This is the stage when growth and inequality are positively related or when the Kuznets curve has a positive slope. At some point, when the urban sector got big enough, inequality trends were determined by those left behind in low-wage agriculture. At that point, further industrial urban growth began to reduce inequality because it reduced the population share in low-wage agriculture.

The point here is that the growth process starts somewhere specific, after which its effects spread by a variety of linkages to the rest of the economy. In the first stages of this process inequality is almost certain to rise. Later, as the effect of the growth stimulus spreads in the population, inequality will begin to decline. How long this takes to happen, or indeed whether the spread effect is big enough to offset the initial increase in inequality, depends on the strength of linkages. Linkages are the connections between other sectors or economic actors and the sector in which growth is occurring.

Linkages are a key determinant of the relationship between growth and inequality. They determine the 'spread effect' of growth in the economy.

The stronger they are, the faster and further the benefits of growth will spread out, and the more equitable growth is likely to be. This notion can be applied in a variety of important ways. In some economies there are big backward regions or indigenous populations which are only weakly or marginally connected to the modern, dynamic sector where growth is occurring. The areas themselves have a significant proportion of total population, which means that their relative income levels will have a noticeable effect on inequality. But for reasons that are not entirely understood, growth in the dynamic sectors does not induce much forward or backward linkage activity. As a result, when these countries grow, there is not much of a spread effect to their backward or poor regions. Growth under these conditions tends to be inequitable. One could say that in these countries inequality is high because of growth, in the sense that if the entire country had remained in the same conditions as its backward regions, inequality would be lower. Here, growth leaves behind significant proportions of the population. Growth always leaves some people behind, if one accepts the idea that growth starts in a particular sector or area of the country. The key thing here is that if the linkages are weak and the areas left behind are large, then the interval in which growth is inequitable is likely to be a long one.

For the same reason, inequality is likely to rise with growth in countries with large indigenous populations. Typically the links between indigenous people and the rest of the economy are weak. When growth occurs, it provides little stimulus to incomes of the indigenous. Conversely one could say that inequality is likely to be lower the smaller and more homogenous the economy. Small countries with homogenous populations are unlikely to have backward regions or groups which are disconnected from the modern economy. Examples are countries like Uruguay or Hong Kong where the bulk of the population lives in a small number of interconnected urban areas. When this sort of country grows, a greater share of the population benefits because most people are linked either directly or indirectly to the sector where the growth stimulus began.

Two examples illustrate the point I am making here. In Brazil growth was very rapid between 1960 and 1980, but it was centered in the south-east, both because of rapid industrialization in the area around Sao Paulo and also because of the expansion of the agricultural frontier. The north-east, which in 1960 contained 33 percent of the population, also grew, but far more slowly than the south and south-east. Partly as a result, the Gini for Brazil rose from 0.53 in 1960 to around 0.58 in 1980, one of the fastest increases in inequality observed anywhere. Similarly China has had a period of explosive growth since the mid-1980s. But that growth was concentrated on the coast, leaving the interior of the country relatively untouched. Between 1990 and 1999 average income in the coastal provinces

and in urban centers grew far faster than in the inland provinces and the rural sector. This was a major contributor to rising national inequality indexes. For the country as a whole the Gini increased from around 0.26 in the mid-1980s to 0.37 in 2000 (Kanbur and Zhang, 2005). Over one-half of that increase was caused by the increase in the gap between rural and urban incomes, and about one-third by the rising inland–coastal gap (Kanbur and Zhang, 2005).⁵

Despite the predictions of Kuznets that at some point in all countries inequality should decline, there seem to be at least two features of growth under current conditions that may reverse this trend. First, modern growth has become increasingly skill- and capital-intensive. That has raised the return to both capital and skilled labor. Wage differentials for the educated have widened considerably and profits have gained relative to labor. Second, and this is more controversial, removing the barriers to capital mobility, an important part of the current generation of economic reforms, has increased the bargaining power of capital and its ability to extract wage concessions under the threat of moving to lower-cost developing countries. At the same time trade liberalization has lowered the prices of simple manufactures produced in developing countries. That has also put pressure on the wage differential in developed countries. All of this may make the world distribution more equal by narrowing the wage differential between developing and developed countries, but at the same time the within-country distributions that I have been discussing may well get less equal in both the advanced and the developing countries.

Governments can do a number of things to make the distribution more equal. The burden of taxes can be shifted toward the more affluent. Government can spend more or subsidize goods and services such as education and health care that benefit the poor. It can create a safety net financed out of general revenue to reduce or even eliminate extreme poverty. It can sponsor public works programs to absorb the unemployed and build useful infrastructure. But probably the two most important things that government can do are to maintain a sustainable and stable growth rate for the economy and to reduce the ranks of unskilled labor by effective programs of education and training.

Conclusions

In this chapter I have discussed how to measure the distribution, how distributions are likely to change as countries develop, and what government can do to make the distribution more equitable. I argued that the effect of growth on the distribution is determined jointly by the skill-intensity of growth, the structure of the economy and its factor markets and by how region- and factor-specific it is. The stronger the linkages between the

leading sector, factor or region and the rest of the economy, the more equitable growth will be.

I paid particular attention to several problems of interpretation of changes in the distribution as indicators of social welfare either within countries over time or between countries. To be a socially relevant measure, it is important that the distribution be defined over the appropriate comparator groups – those with whom people actually make welfare comparisons. The typical aggregation is by country. But distributions by regions or by age cohorts may be better indicators of welfare, even at the national level.

Comparisons of distributions over time are particularly troublesome. First, the populations are different and the lifetime distribution of any cohort is likely to be very different from the national distribution. Second, the distributions that are observed are short-run and any growth process that drives up skill differentials and profit rates will be judged regressive in the short run. But it is unclear how to evaluate this short-run rise in inequality if it generates a progressive factor supply response in the long run by increasing either the supply of skilled labor or employment-creating investment.

Notes

1. The author would like to thank the editors for comments on a previous version of this chapter.
2. This measure is called family income or consumption per capita
3. For a fuller discussion of aggregation problems and measures see Sen (1973).
4. The distinction I am making here is closely related to the notion of horizontal equity discussed by Ravallion (2004). National distribution statistics implicitly assume that welfare is invariant with respect to where income is earned, or in other words that welfare is unaffected by equivalent transfers among groups with the same income, such as the rural and urban poor. A trade reform which lowers the price of food may leave aggregate national distribution statistics unchanged, by helping the urban poor and harming the rural poor. That could exacerbate social tensions even if the national distribution of consumption became more equal.
5. One could cite a third example. Bourguignon and Morrisson (2002) showed that there was a very large increase in world inequality between 1820 and 1950, almost all of which was caused by an increase in the gap between the fast-growing industrial economies in the USA and Europe and those in the rest of the world. Weak linkages between these economies and those on what many have called the periphery were presumably responsible for this result.

Bibliography

- Bourguignon, Francois and C. Morrisson (2002), 'Inequality Among World Citizens', *American Economic Review*, 92 (4): 727–44.
- Cowell, Frank (2000), 'Measurement of Inequality', in A.B. Atkinson and F. Bourguignon (eds), *Handbook of Income Distribution*, Amsterdam: North-Holland.
- Deining, Klaus and L. Squire (1996), 'A New Data Set Measuring Income Inequality', *World Bank Economic Review*, 10 (September): 565–91.
- Fields, Gary (2001), *Distribution and Development*, New York: Russell Sage.

- Kanbur, Ravi and X. Zhang (2005), 'Fifty Years of Regional Inequality in China: a Journey Through Central Planning, Reform and Openness', *Review of Development Economics*, 9 (1): 87–106.
- Khan, Azizur R. and C. Riskin (2001), *Inequality and Poverty in China in the Age of Globalization*, Oxford: Oxford University Press.
- Kuznets, Simon (1955), 'Economic Growth and Income Inequality', *American Economic Review*, 45 (1): 1–28.
- Milanovic, Branko (2002), 'True World Income Distribution, 1988 and 1993: First Calculation Based on Household Surveys Alone', *Economic Journal*, 112 (1): 51–92.
- Morley, Samuel A. (1981), 'The Effect of Changes in the Population on Several Measures of Income Distribution', *American Economic Review*, 71 (3): 285–94.
- Morley, Samuel A. (2000), 'The Effects of Growth and Economic Reform on Income Distribution in Latin America', *Cepal Review*, 71: 23–40.
- Morley, Samuel A. (2001), *The Income Distribution Problem in Latin America and the Caribbean*. Santiago: CEPAL.
- Ravallion, Martin (2004), 'Competing Concepts of Inequality in the Globalization Debate', World Bank Policy Research Working Paper 3243.
- Ravallion, Martin and S. Chen (1997), 'What Can New Survey Data Tell Us about Recent Changes in Distribution and Poverty?' *World Bank Economic Review*, 11 (1): 357–82.
- Sala-i-Martin, Xavier (2002), 'The Disturbing "Rise" of Global Income Inequality', National Bureau of Economic Research Working Paper 8904.
- Sen, Amartya (1973), *On Economic Inequality*, Oxford: Oxford University Press.

48 Income distribution: effects on growth and development

*Nancy Birdsall*¹

Introduction

Until the end of the Cold War, most development economists were not particularly concerned with the distribution of income, but instead with understanding growth and reducing absolute poverty in the developing world. For one thing, Kuznets (1955) had suggested that a deterioration in the distribution of income might be the natural outcome of the early stages of development, as people begin the shift from low-productivity subsistence agriculture to high-productivity sectors. And mainstream economists' starting assumption, rooted in the Smithian trade-off between efficiency and equity, was that in the other direction of causation, inequality resulting for example from increased security of property rights would enhance growth by encouraging investment and savings and creating a necessary incentive for individuals to work hard.²

But beginning in the 1990s, as panel data on changes in the distribution of income in developing countries became available, as mainstream development economists became more concerned with political economy analysis, and – perhaps – once the fall of the Berlin Wall liberated the mainstream from the taboo of Marxian analysis, economists became more interested in assessing the effects of income distribution on growth. In the last 15 years a major focus of new theoretical and empirical work has been the effects of income inequality on growth and other indicators of inequality in the developing world. Much of that work has been ably reviewed in major reports of the United Nations Development Programme (UNDP), the Inter-American Development Bank and the World Bank.³

Still there is no consensus among economists that income inequality matters, and little attention among development practitioners to policies to address inequality as opposed to growth and poverty reduction.⁴

Obviously if people care about their relative income status then *ipso facto* inequality matters. That they do, to some extent, has long been remarked; consider Adam Smith, who noted that a man to retain his dignity may in one society need enough income to buy a linen shirt, and Veblen (1970) who noted that the absolutely well-off worry about their status relative to the more absolutely well-off.⁵ Hirschman (1973) observed that people stuck in

a tunnel in a stopped lane of traffic eventually become deeply frustrated if the other lane, but not theirs, has been inexplicably (and presumably unfairly) moving – quite independent, to extend the metaphor, of the kind of car they drive. Easterlin (1995) noted that happiness (or subjective well-being, or utility, to use the economists' term) varies directly with one's own income and inversely with the income of others, that is, that relative as well as absolute income matters. He comes to that conclusion in part based on surveys of happiness within countries over time; the average level of happiness has not increased even where average incomes have increased substantially.⁶ It is possible in fact that inequality of income reduces the utility or happiness not only of the relatively poor but of the better-off, who may enjoy their own affluence less if others are visibly worse off.

In this review I focus, however, on the instrumental reasons why a highly unequal distribution of income matters in developing countries.⁷ I review a large body of work, primarily of economists, indicating that beyond some level inequality in developing countries matters because: (1) where markets are underdeveloped, high income inequality is likely to inhibit growth; (2) high income inequality can discourage the evolution of the economic and political institutions associated with accountable government (which in turn enable a market environment conducive to investment and growth); and (3) high income inequality can undermine the civic and social life that sustains effective collective decision-making, especially in multi-ethnic settings.

Theory and some empirical work suggest that inequality does not undermine growth directly. Instead it is the interaction of inequality with imperfect markets or with unaccountable or incompetent governments (increasingly labeled weak 'institutions' in the latest literature on growth – see for example Acemoglu et al., 2000) that harms growth. In addition, and conceptually different, inequality (that is high enough) may directly create conditions that lead to or exacerbate poor governance and thus poor economic policy, and/or weak social and economic institutions and thus ineffective implementation of stable and sound policies – reducing growth through the effect on economic, political and social institutions. Weak institutions broadly defined are increasingly viewed as the key cause of low growth in developing countries. Since weak markets, poor governance and underdeveloped institutions might be said to be the very characteristics that define a country as 'developing', it follows that inequality is a key factor in understanding the dynamics of growth and institutional development in the developing world.

The reader will note that I do not discuss the effect of growth on inequality, the subject of the Kuznets hypothesis, nor the evidence that inequality and growth may each be simultaneously affected (Lundberg and Squire, 2003), either similarly or differently, by still other economic and non-economic

variables such as inflation and increased access to education. Once panels of household data enabled analysis of changes in the distribution of income over time within countries, the existence of a stylized Kuznets effect was not supported by the evidence (for example, Deininger and Squire, 1996), almost certainly because so many other country-specific factors compound any fundamental relationship there might be.

Effect of inequality on economic growth and poverty: theory and evidence

Two stylized facts emerge from the growing literature on the effects of inequality on growth. First, the evidence suggests that inequality above some level is more likely to reduce growth. Second, theory and empirical work suggest that high levels of inequality are more likely to harm growth in developing than in developed countries.

Barro (2000), in a study of the determinants of growth, was among the first to report a structurally different relationship of inequality to growth in developing compared to developed countries. Across developed and developing countries combined, he found no clear effect of inequality on growth. However, dividing the sample into the two groups he found the relationship is structurally different. In higher-income developed countries inequality may indeed be associated with higher growth (as often referred to in contrasting the USA and countries of Western Europe). Below a certain income level (about \$2000 US 1985 dollars – equivalent to about US \$3200 in 2000 dollars), higher income inequality is associated with lower growth. (The simple relationship is illustrated for developed and developing countries in Figure 48.1.) Cornia et al. (2004), using data from a more comprehensive set of household surveys, tested the relationship between changes in inequality and growth over almost four decades for 25 countries. They report a positive effect on growth as the Gini coefficient increases from very low levels (from the .15 typical say of subsistence economies and of the former socialist economies to .30) and a negative effect as the Gini coefficient rises from .45 (typical in Latin America and sub-Saharan Africa) to higher levels.⁸

The specific thresholds should not be taken too seriously, given poor measurement particularly of the distribution of income. However, they allow for a rough assessment of how widespread across people and countries in the developing world the resulting vulnerability might be. The critical thresholds of a Gini at or above .45 and income per capita at or below \$3200 affect a significant number of countries and people in the developing world. Virtually no developing or transitional economies have income Gini coefficients below .30, though India and China did at about that level for much of the post-World War II period until the 1990s. About 15 percent of the population of the developing world currently lives in countries (33

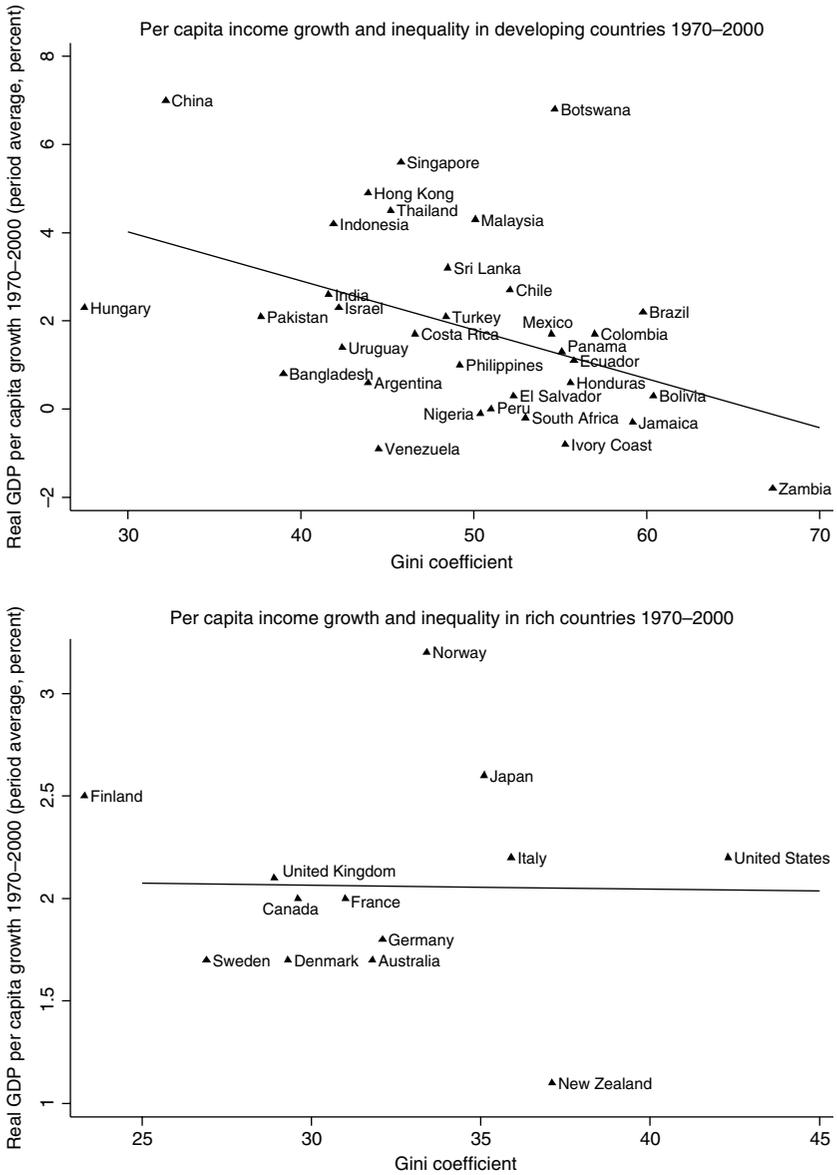


Figure 48.1 *Inequality and per capita income growth in developing and rich countries, 1970–2000*

countries) with reported Gini coefficients of .45 or higher and per capita income below \$3200 (in 2000 dollars), mostly in Latin America and sub-Saharan Africa. But that percentage mounts to 40 percent if China, whose reported 2003 Gini coefficient was 44.9, is included, and rises further to 44 percent if Brazil, whose per capita income now just exceeds \$3200, is included. Other countries with per capita income below the Barro threshold where the income Gini has risen in the last 15 years and is now above .40 are Bangladesh and Pakistan. In India and Vietnam, inequality has also risen rapidly since the 1990s but reported Gini coefficients are still below .40.⁹

These findings are broadly consistent with theory. Why might some level of inequality enhance growth? First, inequality can be too low, as when it was imposed in state-managed economies where planning and controls replaced price and other market signals, encouraging 'shirking' and free-riding. A certain degree of inequality may be necessary to permit the incentives that induce individuals to work hard, innovate and undertake risky but productive investment projects, resulting in higher output and productivity, and therefore higher average incomes and growth rates. (For economists, these incentive effects are the backbone of the moral hazard argument against tax-financed distribution; Okun, 1975). Second, some concentration of income could encourage growth if high rates of saving enable more investment, and if savings rates are greater where income is concentrated in the hands of the rich whose marginal propensity to save is higher than that of the poor (Galenson and Leibenstein, 1955; Kaldor, 1961). A related idea is that investments in infrastructure and industry critical to development are large and indivisible; in the absence of well-functioning capital markets, wealth and income need to be highly concentrated to generate the minimum required resources to undertake new investment projects.¹⁰ (Recent 'endogenous' models of growth, however, rely much more heavily on the incentive effects of institutions and policy than on high savings and investment as the keys to sustained growth.)

The incentive effects of inequality can be thought of as the outcome of 'constructive' inequality, that is, income inequality that reflects solely differences in individuals' responses to equal incentives or opportunities, and is thus consistent with efficient resource allocation.¹¹ In contrast would be 'destructive' inequality, reflecting inefficient privileges for the rich, social and economic discrimination which reduces incentives for effort, investment and innovation by some groups, and in general reduced potential for productive contributions of the already poor. In a kind of tautology, destructive inequality can be defined as that inequality which results in lower, rather than higher economic growth (Birdsall, 2001).

The idea of destructive effects of inequality is consistent with the empirical evidence noted above of lower growth at very high measured levels of

inequality. Theory suggests that inequality is also more likely to be destructive in developing countries (as Figure 48.1 suggests). The remainder of this section sets out why that is likely to be the case. In summary it is because inequality tends to undermine growth when it combines with or interacts with weak markets and poor government policy. In general in developing compared to developed countries, financial and other markets are less complete and public policy is less effective in addressing market failures and imperfections.

Imperfect credit and other markets

Benabou (1996) and Aghion et al. (1999) develop models in which inequality exacerbates the effect of capital and other market failures on growth. When creditworthy borrowers cannot borrow because they lack collateral to comfort lenders (given imperfect information, a market failure in itself), then their lack of income or wealth limits their ability to invest. In addition, given limited liability (the borrower cannot repay more than his or her net worth), borrowers with less wealth have less incentive to exert effort to ensure success of an investment since they must pay lenders a higher portion of their returns (a moral hazard effect). In this case redistributing wealth has no adverse incentive effects – on the contrary it creates a positive incentive – and will be growth-enhancing. Weak or non-existent insurance markets will also force those without assets to bypass high-return projects. Galor and Zeira (1993) and earlier Loury (1981) suggest that the distribution of wealth affects output due to the indivisibility of investments in human capital. When it is difficult to borrow, lack of liquidity limits investments in human capital despite prospective high returns; this obviously affects the poor but may also affect the large majority of middle-income people in developing countries with a high concentration of income at the top of the income distribution. Birdsall et al. (1998) note that even where the poor are credit-constrained, they can exploit an increase in the return to potential new investment (in education or their own farm or business) by increasing their work effort. They will do so as long as the returns to their labor are adequate – as was the case in Korea and Taiwan in much of the post-war twentieth century. If labor markets are functioning well, and returns to education or other investments are rising, the credit market may not matter as much. Or in those countries, lower overall inequality of wealth, income and land (well below the Gini of 0.45 on the distribution of income), may have minimized the negative effect on growth of an interaction between inequality and weak markets.¹²

Obviously weaknesses in capital markets are greater in developing countries, as are compensatory policies such as enforcement of creditor rights. They are also more likely the lower average income is and the higher the

proportion of poor people, making it difficult to distinguish empirically between the negative effect of inequality per se (whether of income, wealth, education or land) interacting with weak markets from the negative effects of high rates of poverty. In any event, whether or not because capital markets are weaker and more people are poorer, it is not surprising that inequality undermines growth in developing countries though not necessarily in developed countries.

In these models, it is not actually income inequality but inequality of financial wealth or other assets that interacts with weak capital markets to reduce growth. (Only recently has household level data on financial wealth of reasonable comparability across countries become available.)¹³ But increasing evidence suggests that other assets – land and education – tell the same story. Latin America still appears to bear the costs of its historic land inequality. Carter and Coles (1998) show that concentration of land ownership is associated over long subsequent periods with concentration of income, even in countries where the economic relevance of agriculture has declined. Birdsall and Londono (1997) show that across countries inequality in the distribution of education reduces growth, and that once inequality of land and education are accounted for, inequality of income washes out as a factor affecting growth. In that respect, market economies in Latin America compared to East Asia, discussed below, do not operate differently – it is just that they operate in a context of high concentration of land and education.

Ineffective or corrupt institutions of the state and resultant poor public policy

As with weak markets, weak governments and poor public policy are likely to exacerbate the effects of inequality (of income, assets, education and so on) on growth. Behrman et al. (2000) show that differences across countries in social mobility, measured by differences in the effect of parents' income and education on children's education, are robustly and systematically affected by differences in two factors: public spending on primary education and the depth of financial markets. Repressed interest rates and directed credit programs that end up limiting access to credit except for privileged insiders worsen the effect of inherently imperfect capital markets on growth. Lack of adequate public spending on basic health and education means that public policy is not correcting for the inherent inability of markets alone to compensate for differences across households in endowments of all kinds. Growth is then lower than it could be since aggregate accumulation of human capital is reduced.

If income inequality interacting with poor policy reduces growth, then it is implicated in reduced poverty reduction – given that empirically, growth

has seemed necessary if not sufficient for reducing poverty, and since whatever growth occurs will help the poor less in an accounting sense the less equal the distribution of income (Ravallion, 1997, 2001).¹⁴ There may also be a more substantive link of inequality to the persistence of poverty where state institutions and government policy fail to ensure equal opportunities for the poor, even when there is income growth on average. Birdsall and Londono (1997) report that across countries in the period 1960 to 1990 greater land and education inequality reduced the income growth of the poorest quintile about twice as much as they reduced average income growth for all quintiles.¹⁵ In the extreme, unequal distribution of land may cut off altogether the usual effect of growth in agriculture on reduction of rural poverty. Some evidence suggests that agricultural growth in Latin America in the 1970s and 1980s failed to reduce poverty at all (De Janvry and Sadoulet, 2000), as large landowners captured most of the benefits. In contrast, in Indonesia, where small farmers provide the bulk of agricultural production, growth was good for the rural poor even in the days of Sukarno, and still better in the days of Suharto (Timmer, 2006a, 2006b).

Political instability and social conflict

Initial theorizing put any negative effect of inequality on growth not on its interaction with weak markets or poor public policy, but through a direct effect in the political sphere, as Benabou put it: ‘where asset markets are complete and distributional effects arise solely from the balance of power in the political system’.¹⁶ Economists suggested that higher inequality causes lower growth because the median voter, who is relatively poorer where inequality is high, votes for inefficient redistribution financed by growth-reducing higher taxes (Persson and Tabellini, 1994; Alesina and Rodrik, 1994). Their cross-country tests were not, however, convincing. Moreover, the median voter theory did not square with anecdotal evidence that policies in unequal countries are often shaped not by the relatively poor median voter (even where there is democracy) but by a more politically influential elite,¹⁷ and with lack of any evidence that redistributive policies, measured in terms of the marginal tax rate, are associated with lower growth (Easterly and Rebelo, 1993).

An alternative political explanation blames political instability on ‘social discontent’ (associated with inequality among other things) (Alesina and Perotti, 1996). Socio-political differences that reduce the security of property rights and the expected return on investment thus reduce investment and subsequently growth. In a test of the determinants of growth collapses after 1975, Rodrik (1999) found that high inequality and the quality of institutions that manage conflict were key underlying factors – not the size nor the intensity of external shocks. He argued that with high inequality,

distributional conflicts would be more difficult to resolve, delaying fiscal and monetary adjustment and diverting productive resources to bargaining over distributional changes. Benabou (1996) notes that if the rich understand the implication for growth of rent-seeking in unequal societies and of populist revolts, it may be in their collective interests collectively to transfer wealth to the poor through land reform, education subsidies or trade protection.¹⁸ The problem may be (as experience in Latin America and Africa suggests) that for such transfers to be efficient and growth-enhancing requires effective institutions of the state.

Effect of inequality on political and economic institutions

A large literature is concerned with the importance of effective institutions for growth (for example North, 1990; Acemoglu et al., 2004). Does inequality (in some 'initial' state) contribute to the failure of effective institutions to emerge in some societies? Engerman and Sokoloff (1997, 2002) suggest that differences in the factor endowments of colonial North and South America contributed to differences in the concentration of income which in turn affected the evolution of different economic and political institutions. Abundant slave or indigenous labor, and soil and climate conducive to large plantation agriculture in the south, and the opportunities for extraction of mineral wealth, were conducive to the high concentration of income, human capital and political power. The elite in the south then tended to create and sustain institutional arrangements that protected their interests but did not encourage broad-based investment, for example in education or productive economic activity. In contrast were the smallholder farms of the north, where the soils and climate were conducive to wheat, for example, and cheap labor was not available. In these settings, more democratic institutions evolved, property rights were broad-based, and a thriving smallholder class supported public financing of education and in general created local governments that were accountable to most citizens.

Public-choice models similarly attribute poor public policy to government regimes in which bureaucrats and insiders face no real checks on the pursuit of their own interests (Buchanan and Tollison, 1984). If the rich favor public policy that preserves privileges independent of their economic efficiency, inequality may not only inhibit growth by interacting with government failure and poor public policy, as set out above, but may contribute to poor institutions and government failures in the first place. The problem seems especially great when concentration of income at the top is combined with substantial poverty at the bottom, and there is not a large middle class to demand accountability from government. Easterly (2001) and Easterly et al. (2006), use country-level data on size of the middle class (instrumented

by differences in commodities produced, recalling Engerman and Sokoloff), to study the determinants of good ‘institutions’ (measured in terms of survey results on accountability, corruption, property rights, and so on). They conclude that a small middle class is implicated in weak institutions, and through weak institutions in low growth.

An example is the apparent relationship between a high concentration of income in a society and differences across countries in the policy and institutional capacity that ensure access to education – as in the difference between East Asia and Latin America in educational opportunities for the poor (Birdsall et al., 1997). Supply of publicly subsidized education is likely to be limited where the rich resist a large tax burden to finance services which they can purchase privately. Targeting social services to the poor can help reduce the fiscal burden of greater public spending, but easily leads to loss of political support from the working and middle class. Without middle-class interest and pressure, the quality of public services deteriorates (and the middle class resorts to private services).¹⁹ Thus it is possible for income inequality to contribute to poor public policy and institutions even where there is little or no absolute poverty – for example in US cities.

It is also likely that high income inequality encourages rent-seeking by the rich through bribes and extortion in the political sphere, and populist and protectionist policies when those who feel disadvantaged acquire a political voice. Keefer and Knack (2002), like Easterly et al. (2006), find that income inequality is associated with weakening of the protection of property rights.²⁰

In short, not only does theory and some evidence suggest inequality harms growth in interaction with poor public policy, but it is plausible that high inequality more directly undermines good public policy by delaying or stalling the emergence of the political and economic institutions (property rights, an independent judiciary, accountability to voters and checks on abuse of privileges and power) – institutions that are increasingly viewed as fundamental to sustaining growth.

Effect of inequality on social institutions, social capital and collective decision-making

Amartya Sen places considerable emphasis on individuals’ ‘capability’ to participate in the life of the community as an aspect of development independent of any implications for economic growth (Sen, 1992, 1999). Participation in the life of the community suggests there are assets that are held not individually but only in relation to others; Putnam (1993) defines the asset of social capital in terms of trusts, norms and networks that can improve the efficiency of society, ‘facilitating coordinating actions’. Social capital may also have economic value to the extent that it reduces the cost

of transactions and of contract enforcement, and as Rodrik (1993, 1999) argues, reduces resistance of losing groups to political compromises.

There is good evidence from microeconomic analyses that income inequality adversely affects some of the inputs or correlates of social capital. In Tanzania, informal insurance is higher in communities where income inequality is lower (La Ferrara, 2000). Among sugar cooperatives in India, where land ownership is more unequal, cooperatives are less productive (Banerjee et al., 2001).²¹ The literature on local public finance addresses the same issue indirectly, in assessments of the link between income levels and the formation of communities with different amounts of heterogeneity. A typical finding is that the quality of publicly provided education is inversely related to income inequality, controlling for average income (Fernandez and Rogerson, 2003).

Finally there is the evidence from studies of crime and violence. Fajnzylber et al. (2002) assessed the impact of inequality on homicide rates in a cross section of 39 countries over the period 1965–95. Income inequality measured by the Gini coefficient had a significant and positive effect on homicide rates, robust to a variety of specifications. Ratios of income of contiguous quintiles starting with the second quintile (that is, third to second, fourth to third, and fifth to fourth) exacerbate crime, and at an increasing rate. In other words, it was not poverty nor inequality at the bottom that explained crime, but the disparity between the middle strata and their richer counterparts. It was not absolute but relative income that mattered.

It is difficult to distinguish conceptually between the effects of inequality on political and economic institutions and on such ‘social’ institutions as social capital and shared civic customs and habits. To some extent that may be because across societies such ‘institutions’ as broad-based property rights, democracy with checks on abuse of power, and ‘trust’ among citizens, tend to be correlated with each other. In any event, evidence suggests that in each category, such institutions have evolved less successfully where income inequality has been high.

Inequality and growth in East Asia versus Latin America

In 1960, average real per capita income in Latin America was higher than in East Asia. Since then, average per capita income has risen almost tenfold in East Asia whereas in Latin America it has less than doubled (Table 48.1). In 1960, income and land inequality were significantly higher and income concentration much more extreme in Latin America compared to East Asia. (Table 48.2); Taiwan and Korea both benefited from externally imposed land reform after World War II.²² The contrast over four decades between fast-growing East Asia, with its relatively low inequality in 1960

Table 48.1 *Inequality, Income and Growth in Latin America and East Asia, 1960 and 2000*

	Income Gini ³		Income share of poorest 10% of population ^{3,4} (%)		Income share of richest 10% of population ^{3,4} (%)		GDP per capita (constant 2000 US\$)		Average real GDP per capita growth rate (%)
	1960	2000	1960	2000	1960	2000	1960	2000	1960–2000
Latin America ¹	0.51	0.53	1.7	1.1	42.5	40.5	1 950	3 050	1.1
East Asia ²	0.42	0.43	2.6	2.1	32.4	34.4	1 300	11 740	5.7
China	0.32	0.39	n/a	n/a	n/a	n/a	100	860	5.6
Mexico	0.53	0.54	1.3	1.1	41.9	41.8	2 560	5 930	2.1

Notes:

All group averages are unweighted.
n/a indicates data not available.

1. Latin America includes Argentina, Bolivia, Brazil, Chile, Colombia, Costa Rica, Ecuador, El Salvador, Guatemala, Honduras, Nicaragua, Panama, Paraguay, Peru, Uruguay and Venezuela. 1960 and 2000 income Gini data not available for Ecuador, Guatemala, Nicaragua and Paraguay.
2. East Asia includes Hong Kong, Indonesia, Malaysia, Singapore, South Korea, Taiwan and Thailand, but excludes China.
3. 1960 data are for earliest year available for the period 1958–68. 2000 data are for latest year available for the period 1996–2001.
4. East Asia excludes Singapore, Latin America excludes Bolivia, Chile, Honduras and Peru.

Sources: WDI (2005) and WIDER (2005).

Table 48.2 *Income, Education and Land Inequality in Latin America and East Asia, 1960 and 2000*

	Income Gini ³		Education Gini ⁴		Land Gini ⁵	
	1960	2000	1960	2000	1960s	1990s
Latin America ¹						
mean	0.51	0.53	0.53	0.42	0.83	0.81
standard deviation	0.06	0.06	0.13	0.08	0.06	0.07
East Asia ²						
mean	0.42	0.43	0.58	0.35	0.47	0.42
standard deviation	0.04	0.08	0.10	0.10	0.10	0.07
China	0.32	0.39	n/a	0.38	n/a	n/a
Mexico	0.53	0.54	0.56	0.36	0.62	n/a

Notes:

All group averages are unweighted.

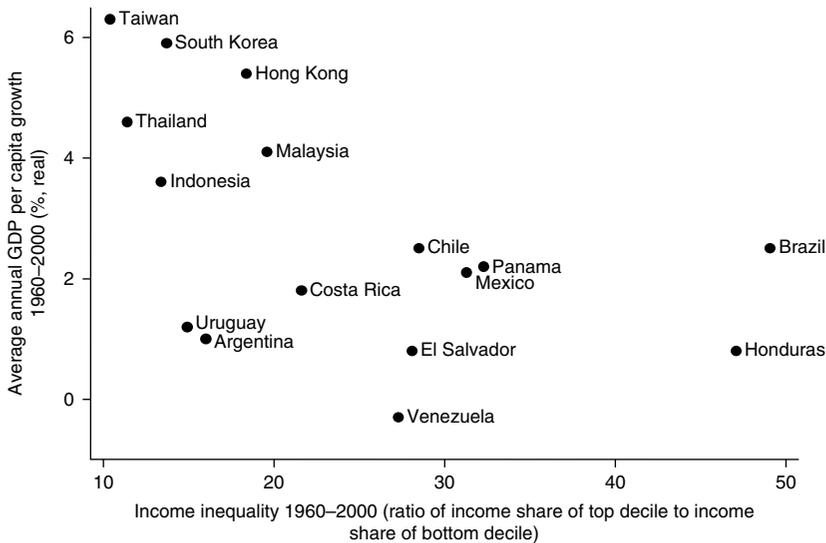
n/a indicates data not available.

1. Latin America includes Argentina, Bolivia, Brazil, Chile, Colombia, Costa Rica, Ecuador, El Salvador, Guatemala, Honduras, Nicaragua, Panama, Paraguay, Peru, Uruguay and Venezuela. 1960 and 2000 income Gini data not available for Ecuador, Guatemala, Nicaragua and Paraguay.
2. East Asia includes Hong Kong, Indonesia, Malaysia, Singapore, South Korea, Taiwan and Thailand.
3. 1960 data are for earliest year available for the period 1958–68 and 2000 data are for latest year available for the period 1996–2001.
4. Education Ginis for population aged 15 years and over.
5. East Asia average excludes Hong Kong, Singapore, Taiwan and Malaysia for which data are not available. Latin America average excludes Bolivia, Chile, Ecuador, El Salvador and Guatemala.

Sources: WDI (2005), WIDER (2005), Thomas et al. (2001), Deininger and Squire (1996) and FAO (2006).

(compared to Latin America, and in particular in fast-growing Korea and Taiwan compared to Thailand and Indonesia) and slow-growing Latin America, with its very high inequality (Figure 48.2), is consistent with the theory and evidence reported above: that high inequality in developing countries, where it is likely to be combined with imperfect and weak markets and poor government policy, reduces an economy's growth prospects; and that high 'initial' inequality puts at risk the development of the economic, political and social institutions that support deeper markets, better government and sustained growth.²³

Rapid growth in East Asia is associated with the region's early export push, supported by high savings and investment and healthy rates of total



Sources: WDI (2005), WIDER (2005), and authors' calculations.

Figure 48.2 Income Inequality and GDP per capita growth in East Asia and Latin America, 1960–2000

factor productivity growth in manufacturing (World Bank, 1993). Behind export success were other factors rooted in rapid changes in household decisions and behavior. Those other factors included unprecedented gains in small farmers' agricultural productivity, high demand for schooling including of girls, and declines in fertility far more rapid than and at lower income levels than had occurred in the industrialized economies (Birdsall and Sabot, 2002). Governments generally ensured that exchange rates were competitive and that fiscal discipline kept inflation low. Governments also favored public investment in basic (primary and secondary) education.

In Latin America, inflation and overvalued exchange rates penalized agriculture, and were combined with tariff and other protection of industry and subsidies to capital that may have reduced the demand for labor. Spending on education was comparable to that in East Asia but was much more concentrated on highly subsidized university education for a select few, responding to the demands of richer households. In the 1960s, educational attainment of the adult population was at roughly the same levels in East Asia and Latin America, and inequality of education (measured in terms of number of years of schooling achieved) was actually higher in East Asia. Since then, educational attainment has risen more quickly in

East Asia and education inequality has fallen faster (Birdsall and Londono, 1998). Broad-based investment in basic education in East Asian countries led to substantial growth of labor productivity and enabled firms to acquire and adapt new technologies and move up the value chain as increasingly skilled cohorts of workers became available (Schultz, 1961; Romer, 1994). The export-push, labor-demanding strategy chosen by East Asian countries generated the conditions for a savings and investment boom in middle-income and poor households and farms (Birdsall et al., 1998).

It seems plausible that one region's lower inequality compared to the other, among other things, affected the difference in the two regions' subsequent trajectories of growth, inequality and investment in human capital. The story is not straightforward. Latin America has a longer history of democracy, for example. But the differences do suggest that the potential negative effect of inequality – of income, land and other assets – on growth and on the evolution of institutions that support the development process, deserves continuing attention.

Notes

1. I am grateful to my former and current research assistants: Gunilla Pettersson, Christine Park, and Karelle Samuda; and to Lyn Squire, John Williamson and Amitava Krishna Dutt for their thoughtful comments on an earlier draft. This chapter was meant to be co-authored with Richard Sabot, with whom I enjoyed spirited discussions about its content and emphasis before his untimely death in 2005; this chapter reflects inadequately those discussions and the contribution he would have made.
2. For example, Finis Welch entitled his 1999 address to the American Economics Association 'In Defense of Inequality'. The reference to the Smithian trade-off is to Smith's *Theory of Moral Sentiments* published in 1759 (Smith, 1982 [1759]). Kaldor (1961) noted that a higher profit share would encourage savings on the assumption that capitalists have a higher propensity to save, from which it follows that when income is more concentrated, savings and investment and thus the equilibrium rate of growth will be higher.
3. Smith in *Wealth of Nations* published in 1776 (cited as Smith, 1981 [1776]). The 1998 *Human Development Report, Consumption for Human Development* of the UNDP, led by Richard Jolly, included analysis of consumption inequality. The 1998/1999 IPES of the IDB, led by Ricardo Hausmann, the Chief Economist, was entitled *Facing Up to Inequality in Latin America*. The 2006 *World Development Report* of the World Bank, led by Francisco Perreira and Michael Walton, is entitled *Equity and Development*. For a review from the perspective of new endogenous growth theories in economics, see Aghion et al. (1999).
4. Lyn Squire (personal correspondence; and see Lundberg and Squire, 2003) makes the point that policy recommendations for addressing inequality may not be much different from those meant to address poverty in a country with an egalitarian distribution of income. (An exception might be tax policy, which ideally might be more progressive in the former setting, if only to sustain politically open markets. In addition greater priority in the face of political and administrative constraints might go to anti-trust and anti-monopoly programs in high-inequality settings.) I do not try to address this point in this chapter since it is not focused on policy per se but on a review of the implications of inequality for the dynamics of growth in the developing world.
5. Graham and Felton (2006) provide a survey of recent studies linking measures of 'well-being' (or reported 'happiness' in surveys of individuals) to prevailing levels of inequality.

Results depend on setting, definition of reference group, and the particular measure of well-being. In Europe and the USA inequality has generally negative effects on reported measures of well-being.

6. Graham and Pettinato (2002) make the point that what is important is people's perceptions about their current and future income relative to others. Graham and Felton (2006) report based on happiness surveys that people in Nigeria are as happy as people in France despite the huge discrepancy in per capita incomes.
7. Much of what I say about income inequality applies to consumption inequality, and much theory reviewed below applies better to wealth than to income and consumption inequality. In principle 'income inequality' as I use it refers to 'permanent income', though in fact empirical work on income inequality is almost always based on current income, and sometimes on wages and other pre-tax income. Elsewhere I have used the term 'money inequality' to distinguish income and consumption inequality from inequality of 'opportunity' (which is difficult if not impossible to measure) and of land, education and other non-monetary assets. See Birdsall (2001).
8. See Chapter 47 in this volume for a discussion of the Gini coefficient and other measures of distribution. The studies referred to all use panels of country observations and employ country fixed-effects estimations, so that they are assessing changes over time within countries, not differences across countries.
9. Reported Gini coefficients are from the WIDER (WIID2a) database; see <http://www.wider.unu.edu/wiid/wiid.htm> and WIDER (2005). Income per capita is from the *World Bank World Development Indicators* (<http://www.worldbank.org/data>). For the statements in this paragraph, I used Gini coefficients from as many countries as possible. For some countries only Gini coefficients of the distribution of consumption are available. The distribution of consumption will be more equal than the distribution of income so that the number of countries and people in the categories I defined may be higher than stated here.
10. With this in mind, many developing countries embraced the need for the state to assume the commanding heights of the economy and used tax and donor resources to finance state-led industrial investments throughout much of the post-World War II twentieth century. This approach almost certainly, and ironically, led to increased concentration of income. Worse, in some countries the later privatization of those investments further increased income concentration, though there is also good evidence that privatization of water, electricity and other utilities has improved access to these services by the poor (Nellis and Birdsall, 2005).
11. Rawls (1971) argued that unequal systems of incentives and rewards may be justified if they improve the position of the least advantaged. His fundamental point was that an increase in inequality can only be justified if the outcome is an improvement in the welfare of the worst-off.
12. A closely related more Keynesian point is that greater inequality may depress aggregate demand, and thus investment incentives and growth – even where markets are otherwise functioning well. See Chapter 14 in this *Handbook*.
13. Davies et al. (2006).
14. Ravallion (2001) reported an average elasticity of poverty reduction with respect to growth of -2.5 , implying that for every 1 percent increase in the growth rate in average income, the proportion of the population living below \$1/day falls by an average of 2.5 percent.
15. See also Deininger and Squire (1996). These findings contrast with the conclusion of 'Growth is Good for the Poor' in which Dollar and Kraay (2002) find that average incomes of the poorest quintile rise proportionately with average incomes in a sample of 92 countries spanning the last four decades.
16. Benabou (1996).
17. De Mello and Tiongson (2006) find no evidence that governments of highly unequal countries are more likely to attempt to redistribute income.
18. Similarly it is often in the collective interest of an ethnic or racial majority to support anti-discrimination and other policies and programs to reduce horizontal inequalities, that is, inequalities among groups in political, economic and social dimensions, as these

- otherwise can provide the basis for inter-group animosity and fuel civil conflict (Stewart, 2001; Ostby, 2003). See Chapter 63 in this *Handbook*.
19. On the demand side, low public spending combined with pressure to maintain or expand enrollments has led to low-quality schools, reducing the economic returns to poor families of sending children to school who can otherwise help at home or work. In effect schooling could be analyzed in terms of a two-sector model, with poor families confined to one sector with low returns, and the rich going to the other sector where returns are high. The difference in returns between poor and rich would explain the high dropout rates throughout much of Latin America, even in the face of high returns on average to those who manage to complete secondary school (Behrman and Birdsall, 1983).
 20. The importance of institutions in development is discussed further in Chapter 61 in this *Handbook*.
 21. In the USA the percentage of households that participate in various membership organizations is higher in metropolitan areas with lower income inequality – controlling for racial and ethnic heterogeneity, income, education and other household characteristics. The effect is substantial. An increase in the Gini coefficient of inequality by one standard deviation leads to a reduction in the probability of participation of 24 percentage points – more than twice the effect on participation of an individual going from the status of high school dropout to high school graduate or higher (Alesina and La Ferrara, 1999).
 22. Land inequality is still extremely high in Latin America.
 23. Rapid growth in East Asia without accompanying increases in inequality also contravenes the pattern suggested by Kuznets. More recently in China, rapid growth has been accompanied by rising inequality.

References

- Acemoglu, Daron, Simon Johnson and James A. Robinson (2000), 'The Colonial Origins of Comparative Development: An Empirical Investigation', NBER Working Paper 7771, Cambridge, MA: National Bureau of Economic Research.
- Acemoglu, Daron, Simon Johnson and James A. Robinson (2004), 'Institutions as the Fundamental Cause of Long-Run Growth', NBER Working Paper 10481, Cambridge, MA: National Bureau of Economic Research.
- Aghion, Philippe, Eve Caroli and Cecilia Garcia-Peñalosa (1999), 'Inequality and Economic Growth: The Perspective of the New Growth Theories', *Journal of Economic Literature*, 37 (4): 1615–60.
- Alesina, Alberto and Roberto Perotti (1996), 'Income Distribution, Political Instability, and Investment', *European Economic Review*, 40 (6): 1203–28.
- Alesina, Alberto and Dani Rodrik (1994), 'Distributive Politics and Economic Growth', *Quarterly Journal of Economics*, 109 (2): 465–90.
- Alesina, Alberto F. and Eliana La Ferrara (1999), 'Participation in Heterogeneous Communities', NBER Working Paper W7155.
- Banerjee, Abhijit, Dilip Mookherjee, Kaivan Munshi and Debraj Ray (2001), 'Inequality, Control Rights, and Rent Seeking: Sugar Cooperatives in Maharashtra', *Journal of Political Economy*, 109 (1): 138–90.
- Barro, Robert, J. (2000), 'Inequality and Growth in a Panel of Countries', *Journal of Economic Growth*, 5 (1): 5–32.
- Behrman, Jere and Nancy Birdsall (1983), 'The Quality of Schooling: Quantity Alone is Misleading', *American Economic Review*, 73 (5).
- Behrman, Jere, Nancy Birdsall and Miguel Székely (2000), 'Intergenerational Mobility in Latin America: Deeper Markets and Better Schools Make a Difference', in Nancy Birdsall and Carol Graham (eds), *New Markets, New Opportunities? Economic and Social Mobility in a Changing World*, Washington, DC: Brookings Institution and the Carnegie Endowment for International Peace, pp. 135–67.
- Benabou, Ronald (1996), 'Inequality and Growth', in Ben Bernanke and Julio Rotemberg (eds), *NBER Macroeconomics Annual*, Cambridge, MA: MIT Press, pp. 11–74.

- Birdsall, Nancy (2001), 'Why Inequality Matters: Some Economics Issues', *Ethics and International Affairs*, 15 (2): 3–28.
- Birdsall, Nancy and Juan Luis Londono (1997), 'Asset Inequality Matters: An Assessment of the World Bank's Approach to Poverty Reduction', *American Economic Review*, 87 (2): 32–7.
- Birdsall, Nancy and Juan Luis Londono (1998), 'No Tradeoff: Efficient Growth Via More Equal Human Capital Accumulation', in Nancy Birdsall, Carol Graham and Richard H. Sabot (eds), *Beyond Tradeoffs: Market Reform and Equitable Growth in Latin America*, Washington, DC: Inter-American Development Bank, pp. 111–46.
- Birdsall, Nancy, Thomas C. Pinckney and Richard H. Sabot (1998), 'Why Low Inequality Spurs Growth: Savings and Investment by the Poor', in A. Solimano (ed.), *Social Inequality. Values, Growth, and the State*, Ann Arbor, MI: University of Michigan Press, pp. 68–81.
- Birdsall, Nancy, David Ross and Richard H. Sabot (1997), 'Education, Growth, and Inequality', in Nancy Birdsall and Frederick Jaspersen (eds), *Pathways to Growth: Comparing East Asia and Latin America*, Washington, DC: Inter-American Development Bank, pp. 93–127.
- Birdsall, Nancy and Richard H. Sabot (2002), 'Virtuous Circles: Human Capital and Labor Markets in East Asia', Center for Global Development, unpublished manuscript.
- Buchanan, James MacGill and Robert D. Tollison (eds) (1984), *The Theory of Public Choice*, Ann Arbor, MI: University of Michigan Press.
- Carter, Michael R. and Jonathan Coles (1998), 'Inequality-Reducing Growth in Agriculture: A Market-Friendly Policy Agenda', in Nancy Birdsall, Carol Graham and Richard H. Sabot (eds), *Beyond Tradeoffs: Market Reform and Equitable Growth in Latin America*, Washington, DC: Inter-American Development Bank, pp. 147–82.
- Cornia, Giovanni Andrea, Tony Addison and Sampsa Kiiski (2004), 'Income Distribution Changes and Their Impact in the Post-Second World War Period' in Giovanni Andrea Cornia (ed.), *Inequality, Growth, and Poverty in an Era of Liberalization and Globalization*, Oxford: Oxford University Press, pp. 26–56.
- Davies, James B., Susanna Sandstrom, Anthony Shorrocks and Edward N. Wolff (2006), 'The World Distribution of Household Wealth', World Institute for Development Economics Research of the United Nations University (UNU-WIDER).
- De Janvry, Alain and Elisabeth Sadoulet (2000), 'Growth, Poverty, and Inequality in Latin America: A Causal Analysis, 1970–94', *Review of Income and Wealth*, 46: 3.
- De Mello, Luiz and Erwin R. Tiongson (2006), 'Income Inequality and Redistributive Government Spending', *Public Finance Review*, 34 (3): 282–305.
- Deininger, Klaus and Lyn Squire (1996), 'A New Dataset Measuring Income Inequality', *World Bank Economic Review*, 10 (3): 565–91.
- Dollar, David and Art Kraay (2002), 'Growth is Good for the Poor', *Journal of Economic Growth*, 7 (3): 195–225.
- Easterlin, Richard A. (1995), 'Will Raising the Incomes of All Increase the Happiness of All?' *Journal of Economic Behavior and Organization*, 27 (1): 35–47.
- Easterly, William (2001), 'The Middle Class Consensus and Economic Development', *Journal of Economic Growth*, 6 (4): 317–35.
- Easterly, William and Sergio Rebelo (1993), 'Fiscal Policy and Economic Growth: An Empirical Investigation', *Journal of Monetary Economics*, 32 (3): 417–58.
- Easterly, William, Jozef Ritzen and Michael Woolcock (2006), 'Social Cohesion, Institutions, and Growth', *Economics and Politics*, 18 (2): 103–20.
- Engerman, Stanley L. and Kenneth L. Sokoloff (1997), 'Factor Endowments, Institutions, and Differential Paths of Growth Among New World Economies: A View from Economic Historians of the United States', in Stephen Haber (ed.), *How Latin America Fell Behind: Essays on the Economic Histories of Brazil and Mexico, 1800–1914*, Stanford, CA: Stanford University Press.
- Engerman, Stanley L. and Kenneth L. Sokoloff (2002), 'Factor Endowments, Inequality, and Paths of Development Among New World Economies', NBER Working Paper 9259, Cambridge, MA: National Bureau of Economic Research.

- Fajnzylber, Pablo, Daniel Lederman and Norman Loayza (2002), 'Inequality and Violent Crime', *Journal of Law and Economics*, 45: 1–39.
- Fernandez, Raquel and Richard Rogerson (2003), 'Equity and Resources: An Analysis of Education Finance Systems', *Quarterly Journal of Economics*, 111 (4): 858–97.
- Food and Agriculture Organization (FAO) of the United Nations (2006), FAOSTAT Database, available at <http://faostat.fao.org>.
- Galenson, Walter and Harvey Leibenstein (1955), 'Investment Criteria, Productivity, and Economic Development', *Quarterly Journal of Economics*, 69 (3): 343–70.
- Galor, Oded and Joseph Zeira (1993), 'Income Distribution and Macroeconomics', *Review of Economic Studies*, 60 (1): 35–52.
- Graham, Carol and Stefano Pettinato (2002), *Happiness and Hardship: Opportunity and Insecurity in New Market Economies*, Washington, DC: Brookings Institution Press.
- Graham, Carol and Andrew Felton (2006), 'Does Inequality Matter to Individual Welfare: Some Evidence from Happiness Surveys for Latin America', *Journal of Economic Inequality*, 4 (1): 107–22.
- Hirschman, Albert O. (1973), 'The Changing Tolerance for Income Inequality in the Course of Economic Development, with a Mathematical Appendix by Michael Rothschild', *Quarterly Journal of Economics*, 87: 544–66.
- Kaldor, Nicholas (1961), 'Capital Accumulation and Economic Growth', in Friedrich A. Lutz and Douglas C. Hague (eds), *The Theory of Capital*, New York: St Martin's Press, pp. 177–220.
- Keefer, Philip and Stephen Knack (2002), 'Polarization, Politics and Property Rights: Links Between Inequality and Growth', *Public Choice*, 111 (1–2): 127–54.
- Kuznets, Simon (1955), 'Economic Growth and Income Inequality', *American Economic Review*, 45 (1): 1–28.
- La Ferrara, Eliana (2000), 'Inequality and Group Participation: Theory and Evidence from Rural Tanzania', CEPR Discussion Paper 2433, London: Centre for Economic Policy Research.
- Loury, Glenn C. (1981), 'Intergenerational Transfers and the Distribution of Earnings', *Econometrica*, 49 (4): 843–67.
- Lundberg, Mattias and Lyn Squire (2003), 'The Simultaneous Evolution of Growth and Inequality', *Economic Journal*, 113 (487): 326–44.
- Nellis, John and Nancy Birdsall (eds) (2005), *Reality Check: The Distributional Impact of Privatization in Developing Countries*, Washington, DC: Center for Global Development.
- North, Douglass C. (1990), *Institutions, Institutional Change and Economic Performance*, New York: Cambridge University Press.
- Okun, Arthur M. (1975), *Equality and Efficiency: The Big Trade-Off*, Washington, DC: Brookings Institutions Press.
- Ostby, Gudrun (2003), 'Horizontal Inequality and Civil War', PRIO – International Peace Research Institute, Oslo, Norway.
- Persson, Torsten and Guido Tabellini (1994), 'Is Inequality Harmful for Growth', *American Economic Review*, 84 (3): 600–621.
- Putnam, Robert D. (1993), *Making Democracy Work: Civic Traditions in Modern Italy*, Princeton, NJ: Princeton University Press.
- Ravallion, Martin (1997), 'Can High-Inequality Developing Countries Escape Absolute Poverty?' *Economics Letters*, 56 (1): 51–7.
- Ravallion, Martin (2001), 'Growth, Inequality and Poverty: Looking Beyond Averages', Washington, DC: World Bank.
- Rawls, John (1971), *A Theory of Justice*, Cambridge, MA: Harvard University Press.
- Rodrik, Dani (1993), 'The Positive Economics of Policy Reform', *American Economic Review*, 83 (2): 356–61.
- Rodrik, Dani (1999), 'Where Did All the Growth Go? External Shocks, Social Conflict, and Growth Collapses', *Journal of Economic Growth*, 4 (4): 385–412.
- Romer, Paul M. (1994), 'The Origins of Endogenous Growth', *Journal of Economic Perspectives*, 8 (1): 3–22.
- Schultz, Theodore W. (1961), 'Investment in Human Capital', *American Economic Review*, 1 (2): 1–17.

- Sen, Amartya (1992), *Inequality Reexamined*, Cambridge, MA: Harvard University Press.
- Sen, Amartya (1999), *Development as Freedom*, New York: Knopf.
- Smith, Adam (1981), *An Inquiry into the Nature and Causes of the Wealth of Nations*, Indianapolis, IN: Liberty Classics; first published (1776); also available at <http://www.econlib.org/LIBRARY/Smith/smWN.html>
- Smith, Adam (1982), *The Theory of Moral Sentiments*, Indianapolis, IN: Liberty Classics; first published (1759); also available at <http://www.econlib.org/Library/Smith/smMS1.html>.
- Stewart, Frances (2001), *Horizontal Inequalities: A Neglected Dimension of Development*, UNU World Institute for Development Economics Research (UNU/WIDER).
- Thomas, Vinod, Yan Wang and Xibo Fan (2001), 'Measuring Education Inequality: Gini Coefficients of Education', World Bank Policy Research Working Paper No. 2525.
- Timmer, C. Peter (2006a), 'How Indonesia Connected the Poor to Rapid Economic Growth', in T. Besley and L. Cord (eds), *Operationalizing Pro-Poor Growth: Synthesis and Country Experiences*, Basingstoke: Palgrave Macmillan, pp. 22–58.
- Timmer, C. Peter (2006b), 'A Historical Perspective on Pro-Poor Growth in Indonesia', in M. Grimm, S. Klasen and A. McKay (eds), *Determinants of Pro Poor Growth: Analytical Issues and Findings from Case Studies*, Basingstoke: Palgrave Macmillan, pp. 164–90.
- Veblen, Thorstein (1970), *The Theory of the Leisure Class: An Economic Study of Institutions*, London: Allen & Unwin.
- WIDER (2005), 'World Income Inequality Database: User Guide and Data Sources', <http://www.wider.unu.edu/wiid/wiid.htm>
- World Bank (1993), *The East Asian Miracle: Economic Growth and Public Policy*, Washington, DC: World Bank.
- World Bank (2005), *World Development Indicators Database (WDI)*, Washington, DC: World Bank.

49 Measuring progress in reducing poverty

Lyn Squire¹

Understanding poverty in all its dimensions

The eradication of poverty is a fundamental goal of development. Ameliorating the lot of the poor is central not only in its own right but because improvements in their health, education and access to credit, infrastructure and markets unleash the untapped potential of large sections of the population, thereby contributing to overall growth and development. Progress in reducing poverty is thus a key indicator of development. *Webster's Dictionary* defines poverty as 'the state of having little or no money and few or no material possessions'. Taking this definition as a starting point, one person out of every six worldwide struggles to survive each day on less than the price of a cup of coffee in the United States. This statement contains two immensely important messages: a large part of the world's population is obliged to live on the merest pittance; and the disparities between the rich and the poor are truly staggering.

Prestigious though it undoubtedly is, *Webster's Dictionary* may not, however, capture current thinking well about the term 'poverty' nor its current usage in much of the professional literature. As a result of intensive research on the subject, economists and other social scientists have developed new and richer perspectives on the manifestations of poverty that better capture the dire straits that constitute daily life for so many people. For example, Amartya Sen has emphasized the importance of an individual's capabilities to determine their own lives (Sen, 1999), of which control over resources (income) is only one aspect. Others have explored the factors that empower the poor to lead full and productive lives (Stern et al., 2005). In addition to low incomes and inadequate consumption, poverty in these interpretations is understood to encompass lack of education, poor health, insecurity, violence, social and political exclusion, as well as absence of basic rights and human dignity. This broader conceptualization of poverty has pushed development experts to expand the conventional indicators of poverty to reflect today's more comprehensive characterization of the phenomenon. With this in mind, this chapter undertakes three tasks.

It first describes the attempts to measure poverty as the definition of the term has expanded from an initial focus on subsistence to a broader appreciation of the many elements constituting a person's well-being. As we shall see in the next section, even unidimensional measures of poverty based

exclusively on income or expenditure pose many conceptual problems as well as the ever-present difficulty of securing appropriate data. Moving to richer, multidimensional measures poses additional problems, among which the allocation of appropriate weights to the constituent elements is perhaps the most daunting.

The second task is the presentation of the best available evidence. Conceptual difficulties notwithstanding, statistics can be marshaled to demonstrate two important points. First and foremost, progress in reducing poverty on almost all fronts has been greater since 1980 to date than in any other similar period in the recorded history of mankind. And second, progress has been uneven in several respects: across countries; within countries; and across the various dimensions of poverty. Evidence on each of these points is presented in the third section. If progress has been uneven around the world – and it has – then it is of interest to explore why some countries or regions have done better than others and to share the lessons of their success with those that have fared less well. This, the third task, is undertaken in the fourth section. It concludes with some observations on what the available evidence suggests for future efforts to attack poverty.

From unidimensional to multidimensional poverty

A unidimensional focus on income

Consider first how to measure poverty when this is confined to meaning lack of income. For purposes of exposition assume that we have a universally accepted measure of income (as we shall see, matters are not so simple). Call this measure y . Let us also assume that we know the income of each individual in a population so that they can be ranked according to income beginning from the individual with the least income and proceeding in order to the individual with the greatest income. Finally, let y^* be the poverty line, that is, the income threshold below which people will be said to be poor.

Even with all the elements apparently in place to measure poverty, the analyst still faces the task of deciding how to present the evidence on poverty in a single indicator. Here I present three well-known indices to illustrate the range of choice and the issues to be considered. Perhaps the simplest is to count the number of individuals for whom $y < y^*$. Known as the ‘head count’, this indicator represents the number of individuals in poverty and is the most commonly used measure.

The head count, however, says nothing about the depth of poverty or the extent to which an individual’s income falls below the poverty line. To address this issue, the ‘poverty gap’ measures the average income needed to bring the incomes of all poor people up to the poverty line or $y^* - \bar{y}$ where \bar{y}

is the mean income of all individuals below the poverty line. The poverty gap however makes no allowance for the severity of poverty because it treats an extra dollar for the poorest person among the poor as equal in value to an extra dollar for the person just below the poverty line. The Foster–Greer–Thorbecke index overcomes this deficiency by appropriate weighting. Thus, the measure of severity for the i th individual is given by $(y^* - y_i)^\alpha$ where y_i is the i th individual's income and α is the weight. As α increases beyond unity, the weight of those whose incomes are well below the poverty line is magnified in the summary index for all individuals (Foster et al., 1984). This feature, however, introduces a degree of arbitrariness because the appropriate value of the weight is inevitably a matter of judgment.

While the preceding points should be born in mind, much of the effort to track progress in reducing poverty worldwide has focused on the head count, a practice that we continue in the third section. Before looking at the numbers, however, I will first examine some of the difficulties in arriving at a measure of income and deciding on a poverty line, difficulties I had assumed away in the forgoing discussion.

Measuring income and expenditure

Researchers seeking to measure income in developing countries face many daunting challenges. Some of the main difficulties include: how to capture non-marketed and non-priced subsistence output such as crops grown by households for their own consumption; how to allow for free social services and pure public goods; whether to use data on income or expenditure; how to accommodate differences in the demographic composition of households; how to distinguish between temporary and permanent poverty; how to manage seasonal variations in income; and so on (Kanbur and Squire, 2001; Lok-Dessalieu, 1999).

While there is no universally accepted procedure for dealing with each of these issues, in practice most analysts follow some conventions. Thus, expenditure is usually preferred to income since the former is thought to be a better indicator of permanent income. Since data are usually collected at the level of the household, most analysts present information on expenditure per household member by dividing household expenditure by the number of household members. Thus no allowance is usually made for the different consumption needs of children and adults although 'equivalence scales' are available and are used on occasion. Nor is any allowance made for distributional rules within the household based on gender, age or working status.

The researcher's ability to deal with these issues depends crucially on the quality of the data, which are usually obtained from household surveys. In

addition to the difficulties noted above arising from the number of activities, products and services that are unrecorded, estimates of household expenditure are also affected by the limitations of sample surveys (Deaton, 2003). Recall errors, short reference periods and the exclusion from the sampling frame of people in remote areas and other marginal groups who are most likely to be poor are common problems. The quality of enumerators and the consistency of surveys over time are other factors complicating the task of the researcher. That said, surveys of reasonable quality are now available for 97 countries covering 93 percent of the population in the developing and transition worlds (Chen and Ravallion, 2004).

Setting the poverty line

Poverty lines are commonly distinguished according to whether they are absolute or relative. While arguments can be advanced that absolute measures best capture the notion of individual capabilities incorporated in the broader definitions of poverty, the issue is more open in the case of poverty confined to income or expenditure. In this case, 'absolute poverty' refers to subsistence below some minimum, socially acceptable norm, usually established on the basis of nutritional requirements plus other essential goods. 'Relative poverty' compares the proportion of the population below a poverty line defined relative to mean income or some measure of the overall well-being of the population. Relative poverty lines thus adjust to changing circumstances. Absolute poverty lines on the other hand remain unchanged and are therefore better able to track changes in poverty over time and, where they are based on equivalent real baskets of goods and services, can also be used to aggregate across countries to track global poverty. For this reason, the remainder of this chapter focuses on measures of absolute poverty.

The absolute poverty line currently used to track global poverty is the well-known figure of a \$1 a day. Introduced first in the 1990 *World Development Report* (World Bank, 1990), this figure is based on actual poverty lines then prevailing in some of the poorest countries in the world (World Bank, 1990). These measures typically are based on the cost at local prices of a diet providing the minimum nutritional requirements for subsistence, plus some allowance for other basic needs, usually confined to clothing and shelter. To translate these figures into dollars, purchasing power parity (PPP) conversion factors are used. Although not designed to capture the purchasing power of the poor, they remain the best instrument for translating local currencies into dollar equivalents for purposes of international comparisons. This weakness notwithstanding, the comparison revealed a marked bunching of these national lines around the dollar-a-day mark and resulted in its selection as a reasonable indicator of absolute

poverty, at least as perceived by those in the countries suffering most acutely from low incomes. For the remainder of this chapter, we will use this poverty line to indicate extreme poverty but rely as well on a higher figure – roughly \$2 a day – to provide a measure of poverty that is more than mere subsistence.

Multidimensional measures of poverty

Estimates of poverty based exclusively on income or expenditure fail to capture significant aspects of deprivation experienced by the poor. Hence there is a need to go beyond traditional methods of measuring poverty and address the broader multidimensional nature of poverty by including measures of health status, educational attainment, political voice and social inclusion as well as measures of control over material resources. While conceptually straightforward, this extension encounters several practical problems of implementation (Falkingham and Namazie, 2001; Boltvinik, 1999).

Data issues A major difficulty is that many of the required measures are not readily available or, where they are available, suffer from problems of interpretation. Consider political voice or social inclusion. Neither has been measured on a routine basis or according to a widely endorsed procedure. While data collection efforts will undoubtedly continue to improve, it is currently difficult to track progress in all dimensions, especially for historical periods. Even where measures are available on a regular basis and are collected according to generally accepted conventions, quality remains an issue. For example, educational attainment is often represented by the net primary school enrollment rate but such measures are silent on the quality of education actually provided.

Moreover, many of the measures of interest are reported only as national aggregates with no distributional breakdown. Consequently, it is not possible to develop measures of who is poor or deprived and who is not for many of the non-income indicators. At best, one can report national averages. Increasingly, however, household surveys are incorporating questions about health status, educational attainment, and so on. A common result emerging from these surveys is that the poor in one dimension are, as one might expect, often the poor in other dimensions as well. For example, school attendance rates for 6- to 17-year-olds for the poorest decile of the population ranked by income were 31 percent in Pakistan, 32 percent in Nepal and 51 percent in Vietnam, compared with rates of 71 percent, 88 percent and 84 percent for the richest decile in the same countries (Appleton and Song, 1999). Evidence on health status tells a similar story. Stunting, as measured by height-for-age, affected 52 percent of pre-schoolers in the poorest decile in Pakistan, 55 percent in Nepal and 58

percent in Vietnam. In contrast, the corresponding rates for the richest segment were 32 percent, 27 percent and 23 percent (Appleton and Song, 1999).

While quantitative information on health status and educational attainment by households ranked according to income or expenditure is beginning to appear, more qualitative approaches are required to capture political voice, social exclusion, measures of self-respect, and so on. Interviews and focus-group discussions with villagers, urban slum dwellers and marginalized communities remain the best means of securing the fullest and most complete understanding of the plight of the poor (Narayan et al., 2000).

To aggregate or not Putting the data issues to one side, the focus on multiple dimensions of poverty raises the question of the most useful form of presentation. One possibility is to combine the various individual indicators into a single index. Thus, if there are now y_n indicators with corresponding weights represented by α_n , then $\sum \alpha_n y_n$ provides an overall measure of well-being. In principle, households could then be ranked according to their value of this measure, and then those falling below some minimum could be identified as the poor, in the same manner as with the single indicator. In fact, this is never done because as noted above distributional data do not exist for many of the dimensions of interest. Nevertheless, attempts have been made to combine income and social indicators at the national level to provide a more complete measure of well-being.

The best-known such index is the Human Development Index introduced by the United Nations in 1990 (UNDP, 1990) and its subsequent extension, the Human Poverty Index. The latter aggregates three separate measures: longevity (the percentage of people expected to die before age 40); literacy (the percentage of adults who are illiterate); and deprivation in overall economic provisioning both public and private (the percentage of people without access to water and health services and the percentage of underweight children below the age of five). The basic problem with such aggregates is that there is no satisfactory way of arriving at the weights (Srinivasan, 1994). The United Nations Development Programme (UNDP) has assumed weights of one-third for each of the three measures without any real justification other than convenience. That said, composite indices are considered valuable at the global level for advocacy purposes, but less so for country-specific policy-making purposes since they effectively conceal knowledge on the individual measures through the process of aggregation. For this reason, I report an array of indicators in the remainder of this chapter to capture the main dimensions of poverty but stop short of using an aggregate index.

Reducing poverty: the record since 1980 to date

Progress in the aggregate: a success story

The previous section has introduced the basic proposition that poverty should be understood as a multidimensional phenomenon: it is the failure to meet minimally acceptable standards, not only with respect to income or expenditure, but also with respect to health, education, and social and political values. This richness carries a cost: it is difficult to arrive at simple statistics that fully capture all aspects of such a broad concept. We have also seen that attempts to aggregate individual indicators into composite measures pose insuperable problems and in fact conceal information. Accordingly, I present in Table 49.1 several indicators, each of which reflects an important dimension of poverty.

The table reveals two decades of genuine progress. Perhaps the most dramatic improvement is the virtual halving of the percentage of the population living in poverty in low-income and middle-income countries (following World Bank classifications). In Table 49.1, the poor, defined as those subsisting on less than \$1 a day, accounted for 40 percent of the population in 1981 but only 21 percent in 2001.² In terms of absolute numbers, however, the outcome is not so impressive. Because of population growth (roughly an increase of 1.5 billion people), the number of poor only fell by 390 million. On the other hand, had the percentage remained at its 1981 level, the number of poor would have increased to over 2 billion, implying that as many as 1 billion people, or one-fifth of the population, avoided poverty relative to what might have happened based on an extrapolation of 1981 circumstances.

Table 49.1 also presents the progress of low- and middle-income countries in reducing poverty in its non-income dimensions and in improving the

Table 49.1 Measures of aggregate well-being in low-income and middle-income countries

	1981	1990	2001
% of extreme poor	40.4	27.9	21.1
Number of extreme poor (billion)	1.48	1.22	1.09
Life expectancy (years)	60	63	65
Under-5 mortality rate per '000	131	103	87
Literacy (%)	61	68	78
Net primary school enrollment (%)	78	95	97

Source: Chen and Ravallion (2004) and World Bank (various issues).

welfare of their people. It shows an increase of four years in life expectancy at birth for this group of countries during the two decades to 2001. To see an increase of one year in life expectancy every five years is remarkable progress. While child mortality rates also witnessed significant progress, they remain unacceptably high. The literacy rate also shows steady improvement in this group of countries and net primary school enrollment is now almost 100 percent.

Failures and setbacks

The progress in aggregate does not mean that everyone advanced and in all dimensions. Indeed, more disaggregated evidence reveals that some groups not only failed to participate in the overall progress, but they saw their situation worsen in one respect or another. At least four groups can be identified where progress in one or more dimension of poverty departs significantly from the overall progress revealed by Table 49.1: those in Africa; those in marginalized regions or social groups despite country-wide progress; those just above the poverty line; and those where indicators for health or education have improved while poverty as measured by expenditure has worsened, or vice versa. I briefly consider each in turn.

The overall progress revealed in Table 49.1 notwithstanding, one region of the world – Africa – has seen poverty worsen. Recall that the number of extremely poor in China fell by 420 million during the two decades of the 1980s and 1990s. This is more than the worldwide decline in the extremely poor – 380 million. Thus, in the rest of the developing and transition world the number of poor increased slightly during this 20-year period, and if some regions saw declines in poverty, as South Asia did, then other regions must have experienced substantial increases in extreme poverty. This happened in sub-Saharan Africa, and dramatically so: the number of extremely poor almost doubled from about 160 million in 1981 to 315 million in 2001 (Table 49.2). Sub-Saharan Africa is the only region in the developing world to see the head count index increase – from 41.6 percent in 1981 to 46.9 percent in 2001.³ As a result, Africa's share of the world's extremely poor increased from 11 to 29 percent during this period. Worse yet, not only has the number of Africans in extreme poverty increased, but the severity of their poverty has also increased. While the rest of the world's poor saw their daily expenditure increase from \$0.70 in 1981 to \$0.77 in 2001, those in Africa saw theirs fall from \$0.64 to \$0.61. Poverty is worsening in Africa and becoming more concentrated in Africa (Chen and Ravallion, 2004).

Uneven progress is apparent among social indicators as well. Take for example life expectancy, perhaps the most fundamental measure of overall well-being. Table 49.1 reveals steady progress. This has been true of almost all countries, but 26 countries saw life expectancy fall during the 1990s and

Table 49.2 *Extreme poverty in sub-Saharan Africa*

	1981	1990	2001
% of extreme poor	41.6	44.6	46.9
Number of extreme poor (million)	163	227	316

Source: Chen and Ravallion (2004).

in some cases precipitously so. For example, life expectancy fell by 17.8 years in Botswana, 16.3 years in Zimbabwe and 14.1 years in South Africa (World Bank, 2004). These figures depict a rapid undoing of steady progress over many years. Nineteen of the 26 countries are in Africa, and four are in the republics of the Former Soviet Union. While causes vary across countries, the advent of the AIDS crisis has played a major role, especially in Africa. In several countries the prevalence of HIV among the 15–49 age group in 2001 exceeds one-fifth: Botswana (38 percent), Lesotho (29.6), Namibia (21.3), South Africa (20.9), Swaziland (38.2) and Zimbabwe (24.9) (UNAIDS, 2004).

This phenomenon of uneven progress across regions of the world is also evident within countries. A second group therefore comprises those living in certain areas or belonging to certain ethnic groups who have not participated fully in their country's overall progress. Consider the case of India. Using the official poverty line, the number of Indian poor fell from 320 million in 1993–94 to 260 million in 1999–2000. But not all states prospered equally. Indeed, two states – Madhya Pradesh and Orissa – saw the numbers in poverty increase, albeit only slightly (Government of India, 2002). Groups within states can also experience different outcomes. For example, the incidence of poverty in the coastal districts of Orissa declined from 46 percent to 32 percent during this period. The experience of the southern districts, however, was the reverse: the incidence of poverty increased from 43 percent to 50 percent. And the districts with the highest initial incidence of poverty – the northern districts – experienced a significant increase from 66 percent to 81 percent (de Haan, 2004). Isolation in economic terms, political terms, social terms and racial terms underlie many of these examples of lagging regions or lagging social groups.

A third, major group to witness a deterioration in their well-being is the less poor, or those just above the extreme poverty line. The dollar-a-day poverty line is usually characterized as depicting extreme poverty; it allows mere subsistence at best. Consequently, a higher poverty line – \$2 a day – is also frequently employed.⁴ This provides important new information. Using this measure of poverty, the number of poor actually increased from

2.4 billion in 1981 to 2.7 billion in 2001 (see Table 49.3). Given that the number below the dollar-a-day line fell, the net result is a significant increase in the number of people surviving on expenditure per day of between \$1 and \$2. In fact, the number increased from about 1 billion in 1981 to 1.7 billion in 2001. This is significant for at least two reasons. It suggests that the factors – be they growth, redistributive policies, or whatever – that produced the decline in extreme poverty have been less successful in helping those just above the extreme poverty line. And it suggests that a large number of people remain highly vulnerable to droughts, economic crises, conflicts or other calamities that could easily push them into the ranks of the extremely poor.

The final group comprises those who have experienced uneven progress among indicators. While Table 49.1 shows that in aggregate the various indicators all move in a positive direction, in some cases progress in reducing poverty as measured by increases in expenditure has co-existed with deteriorating outcomes for the social indicators. Thus, eight of the 19 African countries that saw life expectancy fall in the 1990s enjoyed increases in real gross domestic product (GDP) per capita. Since HIV was prevalent in all eight countries, it appears that the disease strikes life expectancy more or sooner than GDP. In still other cases, real GDP per capita has declined while life expectancy has continued to increase. Indeed, life expectancy increased in just over half of the 48 countries that saw real GDP per capita fall in the 1990s (World Bank, 2004). While many factors come into play, it is noteworthy that 22 out of the 26 countries that saw continued progress in life expectancy spent 2 or more percent of GDN on health. The often referred-to cases of Sri Lanka and the state of Kerala, India, are powerful illustrations of how sustained investment in nutrition and health can result in unusually long life expectancies at relatively low levels of income. Thus, life expectancy in Sri Lanka was 73 years in 2000 despite its relatively modest level of GDP per capita.

Table 49.3 'Vulnerable' population

	1981	1990	2001
% of poor	66.7	60.8	52.9
Number of poor (billion)	2.45	2.65	2.74
% of 'vulnerable'	26.3	32.9	31.8
Number of 'vulnerable' (billion)	0.97	1.43	1.64

Source: Chen and Ravallion (2004).

Learning from the past

The review of progress in reducing poverty in all its dimensions presented in the previous section revealed substantial but uneven progress. Overall the well-being of the poor has improved, but many have been bypassed or marginalized and others have seen their situation worsen. This suggests two broad questions: What is driving the general progress? And what can be done to ensure that progress is more widespread and inclusive?

In answering these questions, the starting point for most analysts is the role of GDP growth. Table 49.4 illustrates the association between growth in GDP per capita and selected indicators of poverty. For the 13 countries that enjoyed growth rates of at least 2.3 percent a year in GDP per capita in the 1980s and 1990s, the head count index fell by seven percentage points, illiteracy by six percentage points, and life expectancy increased by more than three years. Thus, the people in these 13 countries saw average conditions improve in three important dimensions. At the other extreme, those living in the 39 countries that experienced low growth in GDP per capita saw the head count index increase, illiteracy fall by over seven percentage points, and life expectancy increase by 1.4 years.

The following interpretation of these aggregate trends draws on a wide range of other material including country studies and case studies as well as cross-country analyses that cannot be reported here because of space limitations. Table 49.4 plus the evidence of many other studies indicates a strong but imperfect relationship between growth in average incomes and the incomes of the poor. This reflects the tendency for inequality in national

Table 49.4 Development outcomes in the 1980s and 1990s, by growth class (unweighted means)

Change in indicator: comparing 1980s and 1990s	Unit	Period	Moderate or		
			High growth	improved growth	Low growth
Head Count Index	% with less than US\$1 a day	1990s	24.1	31.4	36.9
		1980s	31.0	32.1	30.2
Illiteracy	%	1990s	17.2	31.2	31.4
		1980s	22.9	37.6	38.8
Life expectancy	Years	1990s	70.0	62.9	59.8
		1980s	66.8	60.6	58.4
Number of countries			13	53	39

Source: World Bank (2000).

income to change only slowly (see Li et al., 1998) implying that incomes of the poor increase more or less by the same percentage as the incomes of everyone else (Dollar and Kraay, 2002). The strong influence of growth is seen clearly in Table 49.4. Results of this sort, however, say nothing about policy or causality. Thus, the observed transformation of growth into a reduction in the head count index may or may not require a wide range of redistributive policies (Kanbur, 2003). The evidence presented thus far is silent on this issue. And with respect to causality, both growth and inequality are outcomes of the same economic system and all the factors that influence that system (Lundberg and Squire, 2003). Thus, growth and poverty as measured by the head count index are jointly determined.

Reductions in illiteracy, on the other hand, are not obviously correlated with growth and are therefore presumably driven by other factors. For countries with universal enrollment, reductions in illiteracy are largely determined by each country's population dynamics and the share of school-age children in the population. And for those without universal enrollment, the capacity of the school system is the decisive factor. The key policy instrument in these circumstances is the provision of adequate support from the public budget to maintain schools, improve quality and expand capacity where necessary. While funding alone is not sufficient – many other factors influence the delivery of public services (World Bank, 2004) – it is nevertheless an essential ingredient. As long as countries can maintain budget outlays for schooling, we should not therefore expect to see a strong relationship between movements in the growth rate and reductions in illiteracy. In the extreme, however, a collapse in GDP as has occurred in failed states or countries in major transitions will inevitably undermine the educational system. For example in countries like Albania, Bosnia, Congo, the former Yugoslavia, Rwanda, and so on, various shocks and ethnic conflicts have resulted in a colossal destruction of human and social capital (World Bank, 2000).

Improvements in longevity, however, do appear to be linked with growth in GDP per capita (see Table 49.4). Nutritional intake, a key factor influencing morbidity and longevity, depends to a considerable degree on the incomes available to households to buy food of increasing variety and quality. At the same time, life expectancy, like literacy, reflects the quantity and quality of public spending on, in this case, health services. Extraordinary examples like Sri Lanka and the Indian state of Kerala reveal the power of intensive and sustained public support for health services. By the same token and as we have seen above, major new diseases like HIV/AIDS can quickly undermine years of progress in extending people's lives.

The key question facing the policy-maker is, of course, the policies and programs required to achieve growth in GDP and to ensure that all, and

especially the poor, benefit. While it is not possible to go into detail, the broad elements of a strategy consistent with the existing evidence can be sketched. In short, the evidence from the success stories suggests a 'two-part strategy' that on the one hand harnesses market incentives, social and political institutions, infrastructure and technology to promote growth, and on the other hand, supports the poor through the provision of health, education and other basic services (World Bank, 1990). The first part of the strategy provides income-earning opportunities for the poor, while the second part strengthens the capabilities of the poor to take full advantage of those opportunities. In this sense, the two parts of the strategy are mutually reinforcing.

A recent assessment of development experience that takes a very broad definition of poverty as its starting point has elaborated and refined this 'two-track' approach (Stern et al., 2005). The first pillar of this strategy calls for creating a supportive investment climate to encourage firms and farms, small and large, to invest, create jobs and increase productivity. The investment climate is determined by several factors that can be classified under three broad groups: macroeconomic and trade policies, infrastructure, and governance and institutions. The second pillar involves empowering and investing in poor people by enhancing their health, education and security and by fostering mechanisms for them to participate in the growth process of the economy. This approach is broadly consistent with the evidence reported here.

In conclusion it is worth making two cautionary remarks. First, useful though it is to have some general view on strategy, country circumstances vary so much that careful and possibly major tailoring may be required to arrive at an appropriate national strategy. For example, many developing countries have suffered from ethnic strife. In Sierra Leone, the prolonged strife has exacted a heavy human and economic toll. Other countries have been ravaged by the AIDS epidemic. Any national poverty-reduction strategy would have to recognize and respond to such local circumstances. Second, special actions may be required for groups which may otherwise be excluded from the national strategy's reach such as ethnic minorities, HIV/AIDS-positive individuals and those damaged by war, or which have suffered temporary setbacks arising from price fluctuations, unemployment or natural disasters.

Notes

1. The valuable research assistance of Partha P. Sahu, Intern, global development network (GDN), is acknowledged with much appreciation.
2. The exact figure is \$1.08. The original figure of \$1 was recalculated using the new PPP of 1993 (Chen and Ravallion, 2004).
3. Eastern Europe and Central Asia also saw the head count index rise, but it remains negligible; the paucity of survey data for this region in the 1980s should not be forgotten.

Thus the estimates are heavily based on interpolations, which do not allow for any changes in distribution (Chen and Ravallion, 2004).

4. The exact figure is \$2.15. The original figure of \$2 was recalculated using the new PPP of 1993 (Chen and Ravallion, 2004).

References

- Appleton, Simon and Lina Song (1999), 'Income and Human Development at the Household Level', Background Paper for the *World Development Report 2000/2001*.
- Boltvink, J. (1999), 'Poverty Measurement Methods: An Overview', UNDP Social Development and Poverty Elimination Division Poverty Reduction Series.
- Chen, Shaohua and M. Ravallion (2004), 'How Have the World's Poorest Fared since the Early 1980s?' *World Bank Research Observer*, 19 (2): 141–70.
- Deaton, Angus (2003), 'Measuring Poverty in a Growing World (or Measuring Growth in a Poor World)', Research Program in Development Studies, Woodrow Wilson School, Princeton University.
- de Haan, A. (2004), 'Disparities within India's Poorest Regions: Why Do the Same Institutions Work Differently in Different Places?' Background Papers, Equity and Development, *World Development Report 2006*.
- Dollar, David and Aart Kraay (2002), 'Growth is Good for the Poor', *Journal of Economic Growth*, 7 (3): 195–225.
- Falkingham, Jane and Ceema Namazie (2001), 'Identifying the Poor: A Critical Review of Alternative Approaches', paper commissioned by DFID.
- Foster, J., J. Greer and E. Thorbecke (1984), 'A Class of Decomposable Poverty Measures', *Econometrica*, 52: 761–6.
- Government of India (2002), *National Human Development Report: 2001*, New Delhi: Planning Commission.
- Kanbur, R. (2003), 'Conceptual Challenges in Poverty and Inequality: One Development Economist's Perspective', paper prepared for WIDER Conference, Inequality, Poverty and Human Well-being, Helsinki, 30–31 May.
- Kanbur, R. and Lyn Squire (2001), 'The Evolution of Thinking about Poverty: Exploring the Interactions', in Gerald Meier and Joseph Stiglitz (eds), *Frontiers of Development*, New York: Oxford University Press.
- Li, Hongyi, Lyn Squire and Heng-fu Zou (1998), 'Explaining International and Inter-temporal Variations in Income Inequality', *Economic Journal*, 108 (January): 26–43.
- Lok-Dessalieu, R. (1999), 'Review of Poverty Concepts and Indicators', UNDP Social Development and Poverty Elimination Division Poverty Reduction Series.
- Lundberg, M. and Lyn Squire (2003), 'The Simultaneous Evolution of Growth and Inequality', *Economic Journal*, 113 (487): 326–44.
- Narayan, Deepa with Raj Patel, Kai Schaft, Anne Rademacher and Sarah Koch-Schulte (2000), *Can Anyone Hear Us? Voices of the Poor*, New York: Oxford University Press for World Bank.
- Sen, A. (1999), *Development as Freedom*, New York: Oxford University Press.
- Srinivasan, T.N. (1994), Human Development: A New Paradigm or Reinvention of the Wheel? *American Economic Review*, 18 (2): 238–43.
- Stern, N., Jean-Jacques Dethier and F. Halsey Rogers (2005), *Growth and Development: Making Development Happen*, Cambridge: MIT Press.
- UNDP (1990), *Human Development Report 1990*, New York: Oxford University Press for United Nations Development Programme.
- UNAIDS Joint United Nations Programme on HIV/AIDS (2004), *Report on the Global AIDS Epidemic: 4th Global Report*, Geneva: UNAIDS.
- World Bank (1990), *World Development Report 1990: Poverty*, New York: Oxford University Press.
- World Bank (2000), *The Quality of Growth*, New York: Oxford University Press.
- World Bank (2004), *World Development Indicators*, Washington, DC: World Bank.
- World Bank (various issues), *World Development Report*.

50 Gender issues in development

Diane Elson

Introduction

Two questions are at the heart of economic analysis of gender equality issues in development:

1. Are the fruits of economic growth and development in developing countries fairly shared between women and men, girls and boys?
2. Does gender inequality promote or hinder economic growth and development in developing countries?

There is a vast literature addressing the first question, which will be considered in the next section; and a much smaller but growing literature addressing the second question, which will be considered subsequently. The challenges of designing macroeconomic policies to promote gender-equitable growth and development will be briefly considered in the final section.

Gender and the distribution of the fruits of growth and development

A gender analysis of the distribution of the 'goods' produced by development (not only income and wealth but also capabilities) goes beyond a focus on women as an isolated group and beyond mere disaggregation by sex. It situates distribution of these 'goods' in the context of the social construction of gender (that is, the social construction of norms of masculinity and femininity), which shape the choices made by people; and the consequences of those choices. It recognizes not only differences between the sexes, but also structural inequality between the sexes, embedded in institutions, including not only families but also markets, businesses and states (see for example, Sen, 1983; Folbre, 1986; Sen, 1996; Agarwal, 1997; Harriss-White, 1998; Elson, 1999).

Prior to the 1970s, there was little discussion of gender issues in economic development. If the topic was mentioned, there was a tendency to assume that women were an underutilized factor of production which could be mobilized for structural transformation and economic growth, and that economic growth and structural transformation would in turn be good for women. An example is the work of Arthur Lewis. In his famous model of 'economic development with unlimited supplies of labour', he identified 'the wives and daughters of the household' (in the subsistence

sector), as a source of labour for the modern 'capitalist' sector, arguing that this would lead to gains for women: 'because most of the things which women otherwise do in the household can in fact be done much better or more cheaply outside, thanks to large scale economies of specialization, and also to the use of capital (grinding grain, fetching water from the river, making cloth, making clothes, cooking the midday meal, teaching children, nursing the sick etc.)' (Lewis 1955, p. 404). In his book on the theory of economic growth, Lewis was in no doubt about the benefits to women: 'Women benefit from growth even more than men . . . Woman gains freedom from drudgery, is emancipated from the seclusion of the household, and gains at last the chance to be a full human being, exercising her mind and her talents in the same way as men' (Lewis, 1955, p. 422).

This optimistic view was challenged by Ester Boserup (1970), who argued that modernization of economies in Africa and Asia had marginalized women. In the agricultural sector, she argued, women had been deprived of access to training, land rights, education and technology, by both colonial and post-colonial administrators, who could not conceive of women being farmers in their own right, even though in much of sub-Saharan Africa and South-East Asia women enjoyed a significant autonomous role in traditional agricultural production. This lack of access to resources meant that while men's productivity in farming increased, women's productivity did not.

In the industrial sector, she argued, women accounted for a much lower percentage of the industrial labour force in large-scale modern factories than they did in home-based handicraft manufacturing. She pointed to obstacles on the demand side, including labour market regulations, and employers' prejudiced perceptions of women's capacities and work commitment; and on the supply side, she suggested that women had difficulties combining work in the modern sector with their family responsibilities, and were hindered by the view that work outside the home was not proper for women. Above all, women were hampered by their lack of appropriate skills, stemming from their lack of formal education. As a result of all these factors, women had been left marginalized and excluded from development. Boserup's remedy for this was investment in more and better education and training for women – planners must change their view that women were primarily housewives, and train women to compete equally with men in the marketplace, so that women could be included in economic modernization.

As Naila Kabeer (1994) points out, Boserup's book laid the foundations for a large body of 'Women-in-Development' literature, and a large number of policy initiatives aimed at 'integrating women into development'. Irene Tinker (1990), in describing the making of the field of 'Women-in-

Development', calls Boserup's book 'the fundamental text for the UN Decade for Women' (1975–85) (Boserup, 1970). Boserup's 'marginalization' thesis found support from other authors, such as Saffiotti (1978), who examined the implications for women's employment of import-substitution industrialization in Brazil and found that during the 1950s and 1960s, while women's industrial employment increased overall, their share relative to men in the industrial labour force declined.

Addressing women's marginalization through education has been a constant theme in the 'Women-in Development' literature. There have been significant increases in girls' enrolment in primary school since the 1970s, and by 2000 almost all girls (and boys) were enrolled in primary school in developing countries in East Asia and the Pacific, Europe and Central Asia, and Latin America and the Caribbean (UNESCO, 2004). However, in the Middle East and North Africa, South Asia and sub-Saharan Africa, significant numbers of children remained out of school, of which 54 per cent were girls. In South Asia, the gender gap was particularly wide, with girls constituting two-thirds of out-of-school children (UNESCO, 2004). Even in regions with parity in enrolment, girls drop out of school more than boys, so that, with the exception of Latin America and the Caribbean, boys' completion rates remained higher than those of girls.

At the secondary level, by 2000, no region had succeeded in enrolling all children, and girls' enrolment lagged that of boys in all regions, with the exception of Latin America and the Caribbean, where the reverse was true. Moreover, girls' secondary enrolment rates remained substantially lower than their primary enrolment rates. In sub-Saharan Africa and India, girls' secondary enrolment was particularly low, at 29.7 per cent and 47.1 per cent respectively. The comparable figures for boys were 35.6 per cent and 53.7 per cent (UNESCO, 2004).

Boserup's concern about the lack of visibility of women farmers and their lack of resources, including training, has been another continuing theme in the 'Women-in-Development' literature. The importance of agriculture as a source of livelihood has declined in most developing countries, though it remains high in some countries and regions. The proportion of the economically active population in agriculture in 2000 was, on average, 21 per cent in Latin America and the Caribbean, 59 per cent in Africa and 36 per cent in Asia. In the latter region there was wide variation, with only 10 per cent of the economically active population in agriculture in South Korea, while the figure for China was 67 per cent and for India, 60 per cent (UNRISD, 2005, p. 91).

There is still a lack of reliable data on women's share of the economically active population in agriculture, since women's work is still undercounted in censuses and labour force surveys, despite some improvements (Beneria,

1992, 2003). In particular, women's 'own-account' or 'self-employed' farming is undercounted, and women are more likely to be enumerated as 'unpaid family workers' contributing to farms managed by their husbands, and 'agricultural wage workers', working for larger-scale commercial farms, than as farmers in their own right.

In much of sub-Saharan Africa, women and men in the same household farm and manage separate plots, while at the same time supplying labour inputs to each other's plots, so that many women are both 'own-account farmers' and 'unpaid family workers'. Moreover, migration, war and HIV/AIDS have reduced rural male populations in sub-Saharan Africa, and about one-third of all rural households are headed by women, leading to the 'feminization of agriculture', according to the UN Food and Agriculture Organization (FAO, 2005). In many Latin American countries, there has been a feminization of agriculture, as men have migrated from rural areas in search of better incomes, and women have taken over the management of family farms, and do the bulk of farm labour (UNRISD, 2005, p. 96). In Asia, similar factors have led to more women in China, India and South-East Asia taking on the management of farming activities, though there is not full agreement on whether farm management could be said to be 'feminized' (UNRISD, 2005, p. 97). In India, for instance, 48 per cent of self-employed farmers are women; and in dairying and animal husbandry, women farmers far outnumber men (Ministry of Agriculture, Government of India, 2005).

Farming women throughout the developing world have less rights over land than do men, as a result of a variety of factors, such as unequal inheritance practices, registration of land titles only in the name of male household heads, and land reforms that are biased against women (Grown et al., 2005, p. 75). Comprehensive data on the size of the gender gap in land ownership are not available. A study covering five Latin American countries found that women account for between 11 and 27 per cent of landowners (Deere and Leon, 2003). Women farmers in sub-Saharan Africa lack secure rights to the land they farm, and frequently lose their land when they are widowed or divorced (World Bank, 2001, pp. 121–2). In South Asia, few women own agricultural land; and of those that do, few exercise full control over it (Agarwal, 1994).

Since the early 1990s there have been some improvements in women's formal land rights in Latin America and sub-Saharan Africa. Joint titling of land to couples has been introduced in Brazil, Colombia, Costa Rica, Honduras and Nicaragua (Deere and Leon, 2003). In Uganda, Kenya, Tanzania, South Africa and Rwanda women's formal statutory land rights have been strengthened – but customary law is still a powerful obstacle, generally preventing women from owning or inheriting land in their own name

(Grown et al., 2005, p. 81). Moreover, in other parts of sub-Saharan Africa, women's customary use rights over land have been weakened through the introduction of individual ownership of land (Lastarria-Cornhiel, 1997). Women's land rights were strengthened by new legislation in India in 2005 but many challenges of implementation remain (Agarwal, 2005).

Formal ownership rights do not necessarily ensure gender equality in livelihoods. For instance, land markets themselves are not gender-neutral: women buyers have lower bargaining power than men, and face discrimination (Deere and Leon, 2003). Moreover, women farmers still receive less technical support from agricultural extension workers, and only a tiny percentage of extension workers are women (World Bank, 2001, p. 52). In the case of women smallholders, many aspects of the disadvantages depicted by Boserup still persist. But efforts to promote gender equality are unlikely to yield substantial gains in terms of improved standards of living for women smallholders without more effective strategies for the agricultural sector as a whole (UNRISD, 2005, p. 104).

Some women have been moving into new areas of 'modern' commercial agricultural production. In the production of flowers, fruit and vegetables for export from Latin America and sub-Saharan Africa, women comprise between 45 and 90 per cent of contract workers (UNRISD, 2005, p. 98). However, women are mainly employed in temporary or seasonal low-paid jobs, with long hours of work, poor health and safety conditions and no social protection (UNRISD, 2005, p. 98). This illustrates a key shortcoming of the Women-in-Development (WID) 'marginalization' thesis: it identifies gender issues in development in terms of women's exclusion from development, whereas the problem is increasingly the terms and conditions of women's integration in development.

This type of criticism of the WID approach was initiated in the early 1980s, especially by women scholars from developing countries who argued that women were subordinated to development rather than marginalized from it (see, for example, Beneria and Sen, 1981). The disadvantage experienced by women in the process of development was not, they argued, primarily the result of the persistence of 'traditional' cultural practices and prejudices, but of the way in which development has created new job structures, in which gender inequality is embedded. For instance, women are concentrated at the lowest rungs, with lower pay, and less job security and social protection than men; and are expected to combine paid work with ongoing responsibility for the unpaid work of caring for families and communities, reproducing both the labour force and the social fabric. This approach argued for changing development strategies to make structural transformation more egalitarian (Sen and Grown, 1987).

This critique was reinforced by the emerging evidence that, unlike import-substitution industrialization, export-oriented industrialization did not marginalize women, but actually increased their share of manufacturing employment. Women's share of employment in the growing service sector also increased. By 2003, women's share of wage employment in non-agricultural sectors in Latin America and the Caribbean had reached 43.5 per cent, approaching that of the developed regions (46.4 per cent). In Eastern Asia, women's share was 40 per cent, in South Eastern Asia 38.6 per cent and in sub-Saharan Africa 35.8 per cent. Lower shares prevailed in Southern Asia (18 per cent), Northern Africa (21.5 per cent) and Western Asia (aka the Middle East) (20.2 per cent) (data from UN Statistics Division, Millennium Indicators Database).

The rising trend in women's share of employment in industrial and service sectors has been accompanied by a lively debate on the extent to which such employment has improved women's lives and reduced gender gaps in well-being (see, for instance, for an early contribution, Elson and Pearson, 1981; and for a recent contribution, Kabeer, 2000; for a survey of the arguments, see Razavi, 1999). Gender inequalities persist in pay and conditions of employment, with most women segregated in a few occupations in which the vast majority of those employed are women (Anker, 1998). At the end of the twentieth century, on average, the hourly wages of women in developing countries were 73 per cent of those of men (compared to 77 per cent in developed countries). In both developed and developing regions, more than 80 per cent of the gender wage gap could not be explained by measurable differences in workers' characteristics, and probably indicates discrimination in the labour market (World Bank, 2001, pp. 55–6). Women workers in developing countries are more concentrated than men in informal employment that lacks social protection; and within informal employment, in the more precarious types, with lower incomes (Chen et al., 2005). Informal employment has been growing as a share of total employment, as labour markets have become both more flexibilized and more feminized (Standing, 1999).

The conventional wisdom is that, despite problems of the quality of women's paid employment, their earnings give them greater bargaining power within their households, and more influence over how household resources are allocated (see for example, Kakwani and Son, 2006). However, this is not universally the case (Elson, 1999). Women in Uganda report that when a woman starts to earn an income of her own, her husband is liable to withdraw his financial support, and shift to her the responsibility for paying school fees, medical bills, and buying food and clothing and other necessities for the whole household (Ellis et al., 2006, p. 24). Moreover, in most parts of the world, the division of unpaid domestic work between men

and women in the household is rarely renegotiated when women start earning (Kabeer, 2005). Case studies reveal that it matters where women obtain their earnings: for instance, in Ahmedabad, India, home-based paid work does not give women as much say in household decisions as employment outside the home (Kantor, 2003).

Women's entry into the labour market was accelerated in the 1980s and early 1990s by stabilization and structural adjustment policies (Çağatay and Ozler, 1995). It has been suggested that this acceleration was widely linked to 'distress sales', as women are forced to try to make good shortfalls in household income following the loss of employment by male household members (for example, Moser, 1989; Gonzalez de la Rocha, 2000).

Elson (1991) argued that stabilization and structural adjustment policies implicitly assumed unlimited supplies of female labour, available to make good through unpaid work in families and communities any shortfalls in provision of public sector non-tradeable services (such as health, education, water and sanitation), and to increase production of exports, while at the same time maintaining household food security and the social fabric of family and community networks. Moreover the theory ignored the gender norms that structure the division of labour, and mean that men's labour tends not to be reallocated to 'women's work', where there is a decrease in what is considered to be 'men's work' (for example construction) and an increase in what is considered to be 'women's work' (for example garment-making, unpaid care work). Instead, a more likely outcome is unemployment and underemployment for men (who do less paid work but little or no more unpaid work), and overwork for women (who do additional paid work as well as unpaid work). Failure to take this into account in designing adjustment policies, argued Elson, results in extra burdens for women, and risks deterioration in health, nutrition and education.

A number of case studies provide empirical backing for Elson's argument (for example Moser, 1989; Lim, 2000; Tanski, 1994), but lack of data and methodological problems hinder definitive conclusions about whether women and girls have, on average, borne a greater share of the costs of adjustment (Haddad et al., 1995). Nevertheless, there is widespread empirical support for the conclusion that women are disadvantaged in the adjustment process, unless specific measures are taken to address pre-existing gender inequalities (Haddad et al., 1995).

One of the important gaps in data is comprehensive information on the extent of men's and women's unpaid work. Time-use surveys have been used for more than two decades in developed countries to remedy this. In the late 1990s such surveys began to be introduced in a growing number of developing countries. They confirmed what small-scale case studies had suggested: women and girls spend more time on unpaid work than men and

boys; and when both paid and unpaid work is taken into account, women and girls have a longer total working day than men and boys. For example, in South Korea, in 1999, the total working time for males was almost six hours a day on average, while for females it was almost seven hours a day. Males spent on average only 50 minutes per day on unpaid work, while females spent almost four hours (Tae-Hong, 2001, p. 8). A similar picture was revealed by a time-use survey for six states of India in 1998–99: on average female total working time was just over 7.5 hours per day, while for males it was 6.5 hours per day. Females spent almost five hours a day in unpaid work, and males spent only about 30 minutes (calculated from Chakraborty, 2005, Table 3). Charmes (2006, Table 3.2) shows a similar picture for sub-Saharan African countries. For example, in Benin in 1998 the total working time for females was on average almost 7.5 hours a day, while for males it was about five hours a day; females spent almost 3.5 hours per day on unpaid work, while males spent just over one hour. In Madagascar in 2001, the total working time for females was on average almost 6.5 hours per day, while for males it was almost 5.5 hours per day; females spent just over 3.5 hours on unpaid work, while males spent almost 50 minutes. In Mexico, in 1995, the total working time for females was on average just over 8.5 hours a day, while for males it was almost 7.5 hours a day. Females spent four hours in unpaid work, while males spent just over 1.5 hours (calculated from Elson, 2000, p. 102). Very few developing countries have conducted time-use surveys at regular intervals, so it is not possible to examine trends. Cross-country analysis that includes both developed and developing countries shows that higher gross domestic product (GDP) per capita is associated with a decline in time spent on unpaid work, and smaller gender gaps in total hours worked and in time spent in unpaid work (World Bank, 2001, p. 185). A key factor in this is investment in infrastructure and public services. Nevertheless, development does not eliminate unpaid work, as people value time to care for their family and friends (Folbre, 2001).

By the beginning of the twenty-first century, the context for discussion of gender issues in development had become the acceleration of globalization, the growth of income inequality between countries and between households within countries, and the growing differences in the trajectories of developing countries, with some, such as China and India, experiencing rapid structural transformation and growth, others, such as many countries in Latin America and the Caribbean, growing only slowly, and yet others, especially in sub-Saharan Africa, suffering deteriorating conditions.

Millions of women in China and India were not, however, benefiting from the rapid increase in national income because they were ‘missing’, in the sense of either having died prematurely, or not having been born at all.

This phenomenon was first brought to public attention by Amartya Sen (Sen, 1990a). More recent data (China 2000 and India 2001, see Klasen and Wink, 2003) confirms its persistence, and also its prevalence in middle-income developing countries such as Taiwan and Korea (Klasen and Wink, 2003). Preference for sons has not been reduced by rapid development. Indeed, in India the sex ratio imbalance is higher in the higher-income states than in the lower-income states.

Despite rapid growth in some developing countries, millions of people throughout the developing world continue to live in poverty. There has been considerable debate on whether poverty is 'feminized' in the sense of females being disproportionately represented among the poor (Çağatay, 1998; World Bank, 2001, p. 63). Measurement of poverty is dominated by a focus on consumption poverty, calculated on a household basis. This has led to a preoccupation with comparing the poverty rates of male- and female-headed households. The evidence is mixed: in some countries female-headed households are disproportionately represented in households below the poverty line; in others they are not (World Bank, 2001, p. 64). An alternative comparison is between proportions of the male and female populations that are in households below the poverty line. Again the evidence is mixed. In some countries a higher proportion of the female population than of the male population lives in households below the poverty line, but not in others (see for instance, ECLAC, 2002, Tables 6a and 6b). However, there is general support for the view that women are more vulnerable to poverty in old age than are men, as many more women are widows than men are widowers (World Bank, 2001, p. 67).

Moreover, women are certainly over-represented among the adult population who have no income of their own, because their participation in paid work remains lower than that of men. This limits their bargaining power within households and their capability to live a life of dignity, even if the household in which they live has an average income above the poverty line (Sen, 1990b). There is also plenty of evidence that poverty is differently experienced by males and females, and that gender gaps in education, health and work burdens tend to be larger in households in poverty (World Bank, 2001, pp. 61, 66).

Some studies have found that economic growth narrows gender gaps. For instance, Dollar and Gatti (1999) found that a higher level of per capita GDP was associated with greater gender equality in secondary schooling, in life expectancy and in representation in parliaments. However, when the labour market is brought into the picture the results are different for different groups of countries. Seguino used a composite gender equality indicator that includes relative labour force participation rates and female share of technical, professional and administrative positions, as well as the

indicators used by Dollar and Gatti, and clustered countries into four groups, ranging from poorest to richest in per capita terms. She found that in the highest and third-highest groups, there was a positive relation between growth and gender equality, but there was a negative relation in the lowest and second-highest income groups. Her conclusion was that economic growth is not sufficient by itself to achieve gender equality (UNRISD, 2005, p. 58).

These conclusions are reinforced by two regional studies of the impact of growth on gender equality. In Asia, in the period 1970–90, gender equality (assessed using a composite quality-of-life indicator that includes allowance for ‘missing women’) was highest in those countries that grew slowest (Seguino, 2002). In Latin America and the Caribbean, for the period 1970–2000, economic growth was found to have little beneficial effect on closing gender gaps in well-being, though the share of the manufacturing sector in GDP and the share of government expenditure in GDP were positively related to the reduction of gender gaps (Seguino, 2007).

Attention is now beginning to be focused on inequalities between women in developing countries. If inter-household inequality is rising, then it seems likely that inequality between women is also rising. There is as yet no comprehensive study on this topic. A recent study of maternal mortality and poverty in ten developing countries, based on demographic and health surveys, found that the proportion of women dying from maternal causes increases with the poverty of the households in which they live. For instance, in Indonesia, the probability of maternal death was three to four times greater in the poorest than in the richest quintile. In the Philippines and Tanzania the probability was two to three times greater (Graham, 2004). More research needs to be done on inequalities between women.

Gender and the determinants of economic growth and development

In this section, we examine whether gender inequality hampers or promotes economic growth and development. The 1990s saw the emergence of a growing literature on this topic, though it is still very small compared to the literature on the distribution of the fruits of growth and development. A number of cross-country econometric studies relating gender inequality to economic growth in developing countries is now available. They have produced a variety of results, depending on the dimensions of inequality considered, the selection of countries and the specification of the equations (World Bank, 2001).

Barro and Lee (1994) found a negative relationship between female secondary schooling and economic growth, though the relation between male secondary schooling and growth was positive. However, subsequent studies have found a positive relation between economic growth and gender

equality in education (for example Hill and King, 1995; Dollar and Gatti, 1999; Esteve-Volart, 2000; Klasen, 1999, 2002). Dollar and Gatti (1999) used data from over 100 countries covering three decades, and found that an increase in one percentage point in the proportion of adult women who have secondary education is linked to an increase in growth rates of per capita income of 0.3 percentage points per year. Klasen (2002) finds that the higher gender gaps in education in sub-Saharan Africa, compared to East Asia, and their slower reduction, accounted for 0.6 percentage points in the 3.5 percentage points difference in the growth rates in the two regions in the period 1960–92. Closing the gender gap in education enrolment by 2005 is a Millennium Development target, adopted by the UN General Assembly at the Millennium Summit in 2000. An estimate of the impact on the economic growth of countries that were not on track to meet this target found that they would have grown faster by about 0.1 to 0.3 percentage points if they had been on track to close the gap (Abu-Ghaida and Klasen, 2004).

Increasing the level of education of girls may increase growth in per capita incomes directly, by increasing the participation of women in the labour market and the productivity of women's labour, and indirectly, by facilitating a transition from a high to a low rate of fertility (since educated women have fewer children). During the fertility transition, the working-age population grows at a higher rate than the dependent-age population. Bloom and Williamson (1998) refer to this as a 'demographic gift' and argue that it raises the rate of growth of per capita income during the transition (provided policies are in place to employ the working-age population productively). They estimate that it accounts for between 1.4 and 1.8 percentage points of growth in per capita income in East Asia, in the period 1965–90. Sub-Saharan Africa has yet to experience the transition. Its fertility remains high and its working-age population has not grown faster than its total population. A recent study of Uganda estimates that a period of sustained fertility decline could boost medium-term per capita growth rates by 0.5 to 0.6 percentage points a year (Klasen, 2005).

Nevertheless, questions remain about the direction of causation between the education of girls and economic growth. For instance, Robbins (1999) argued, in a study of six Latin American countries, that causation goes from increases in growth to increases in education of girls, rather than vice versa. He found that economic growth leads to rising educational attainment by drawing more women into the labour force, increasing the opportunity cost of women's time, and thus reducing fertility and leading families to invest more in the education of their (fewer) children, girls as well as boys.

There is a general agreement that increasing the level of girls' education improves outcomes for their children. Studies based on household data

show that the more educated are mothers, the lower their children's mortality, controlling for household income and other aspects of socio-economic status; and child immunization rates rise with mother's education (World Bank, 2001, pp. 79, 80). Cross-country regression analysis indicates that higher levels of female enrolment in school, and lower gender education gaps, are associated with lower rates of infant mortality (Hill and King, 1995). Increases in women's education accounted for 43 per cent of declines in child malnutrition in 1970–95 (Smith and Haddad, 2000). A recent estimate of the costs of failing to achieve gender parity in educational enrolment by 2005 found that by 2015, such countries would have on average 15 per 1000 higher rates of under-five mortality and 2.5 percentage points higher prevalence of underweight children under five (Abu-Ghaida and Klasen, 2004).

Focusing on labour markets, rather than education, produces mixed results: greater gender equality in participation in the labour market seems to promote faster growth, but greater equality in wages does not. Recent studies on the Middle East and North Africa (Klasen and Lamanna, 2003) and India (Esteve-Volart, 2004) suggest that growth would be higher if the gender gap in labour market participation were reduced (through more women entering the market). However, cross-country regression analysis of growth and the gender wage gap in 20 semi-industrialized economies (such as those in East Asia) in the period 1975–95 found that, controlling for gender differences in educational attainment, gender wage inequality was a stimulus to growth. A 0.1 percentage point increase in the gap between female and male returns per year of secondary education is associated with a 0.1 percentage point increase in the growth of per capita GDP (Seguino, 2000). There is a strong inverse relation between the gender wage gap and the educational attainment gap in semi-industrialized countries (Seguino, 2005, Table 2). This implies that although narrowing gender gaps in education (through increasing the education of girls) tends to raise productivity, women's lack of bargaining power in the labour market holds their wages down. Seguino argues that this stimulates growth through higher profits in female labour-intensive manufactured exports, higher investment and higher foreign exchange earnings. Of course, if high gender wage gaps held back household investment in girls' education, the dynamic would be different, but this does not seem to have happened in semi-industrializing countries, especially in East Asia (Seguino, 2005, p. 22). Instead the combination of educated but cheap female labour has been critical to the expansion of exports of manufactures. Seguino has extended her analysis to compare the impact of different kinds of inequality on growth in 37 semi-industrialized countries in the period 1975–99. She finds that whereas income inequality between households is negatively associated with

growth, the gender wage gap is positively associated with growth (Seguino, 2005, p. 23).

However, in agrarian economies, in which self-employment and family labour are more important than wage labour, it seems likely that gender inequality in production can hamper economic growth. Lack of aggregate data means that cross-country regression analysis has not been used to investigate this relationship, but micro-level studies suggest that gender inequality in access to and control of assets and income is likely to hamper growth. For instance several studies of smallholder agriculture found that agricultural productivity could be increased if female farmers had the same levels of inputs (such as fertilizer, land and labour) and education and training as male farmers (World Bank, 2001, pp. 85–6). The following examples are highlighted by Blackden and Bhanu (1999). In Burkino Faso, output could be increased by 10–20 per cent by shifting resources from men's plots to women's plots within the same household. In Kenya, increasing the education and input levels of female farmers to those of male farmers could increase yields obtained by women farmers by as much as 22 per cent. In Tanzania, reducing the amount of unpaid work that women have to do could increase household cash incomes of smallholder coffee and banana growers by 10 per cent, labour productivity by 15 per cent and capital productivity by 44 per cent. In Zambia, if women farmers enjoyed the same level of investment in agricultural inputs as men farmers, agricultural output could increase by up to 15 per cent.

Gender inequality in control of resources and division of responsibilities within households can hamper the expansion of agricultural exports, including non-traditional agricultural exports, by smallholders. In sub-Saharan Africa, the production of export crops often requires inputs of women's labour on plots controlled by their husbands, while the cash income from this production is controlled by their husbands. The Poverty and Social Impact Assessment of Uganda's Strategic Exports Initiative (Booth et al., 2003) found that such intra-household inequalities limit the export supply response in Uganda because women prefer to put more of their labour into producing crops on their own plots for household use and sale in local markets, rather than into producing export crops controlled by their husbands. This is corroborated by other studies on Uganda (for example Muhereza, 2001), and there is evidence of similar effects in other countries, such as Zambia (Wold, 1997) and Burkino Faso (Smith and Chavas, 1999).

It is important not to overstate the role of household inequalities in constraining export expansion and productivity in smallholder economies (Whitehead, 2005). There are many extra-household gender inequalities that are also important, including unequal access to extension services,

markets, transport and credit, and unequal laws and social norms. Such inequalities also constrain the contributions to growth made by self-employed women in non-agricultural activities. This is important for future development, since the share of agriculture in employment and income generation has been declining and is expected to decline further. Self-employment is particularly important in sub-Saharan Africa. In this region (excluding South Africa), informal employment comprises 78 per cent of non-agricultural employment and self-employment represents 70 per cent of informal employment (ILO, 2002).

Many hopes have been invested in new forms of microfinance as a means of enabling self-employed women to both increase their own incomes and contribute to economic growth. During the 1990s there was a large expansion in poor women's access to small loans from microfinance institutions, many of them based on the social collateral of group liability, rather than the economic collateral of individual assets, such as land. Optimism about the impact of microfinance reached a high in the 1997 Microcredit Summit, which issued a Declaration stating that:

empirical evidence has shown that women, as a group, are consistently better in promptness and reliability of repayment. Targeting women as clients of micro-credit programs has also been a very effective method of ensuring that the benefits of increased income accrue to the general welfare of the family, and particularly the children. At the same time, women themselves benefit from the higher status they achieve when they are able to provide new income (quoted in Mayoux, 2000, p. 3)

However, detailed evaluations of a range of programmes in Africa and Asia have shown that there are a number of limitations. Mayoux (2000, pp. 12–13) summarizes them as follows: women may not have control over the loans they get, but act as *de facto* intermediaries between male family members and microfinance institutions; even if women control the loans, they may not generate significant increases in income, because women are crowded into a narrow range of traditionally female low-return activities; even if there is an increase in women's income, men may control its use, or men may reduce their own contributions to meeting household expenses. Kabeer (2005) finds that microfinance institutions that combine financial services with other forms of support, and which build the organizational capacity of poor women, are more likely to be empowering. Poor women need more than just loans and savings accounts.

To make a significant impact on their own incomes and on the national economy, women need to be able to move beyond informal micro-enterprises and develop small businesses that are registered and eligible for loans from the formal financial sector, and for support from government

agencies. There remain many legal and social barriers which prevent women, more than men, from formalizing and growing their businesses, as shown in detail in the case of Uganda by Ellis et al. (2006).

Facilitating women's increased participation in the market economy, either in self-employment or in wage employment, will only promote economic growth if there is sufficient demand for their products and their labour. This raises the question of appropriate macroeconomic policies, to which we turn briefly in the next section.

Macroeconomic policy challenges for gender-equitable development

At the sectoral level, there is a great deal of knowledge about economic policies that work to improve the position of poor women, to reduce gender inequality and to promote growth and development (Grown et al., 2005). Less attention has been paid until recently to appropriate macroeconomic policies. The importance of examining macroeconomic policies from a gender perspective was emphasized in contributions to two special issues of *World Development* (Çağatay et al., 1995; Grown et al., 2000). However, the World Bank paid little attention to macroeconomics in its 2001 policy research report, *Engendering Development*.

In this final section we briefly consider some aspects of monetary and fiscal policy and trade and investment policy that need to be addressed if the relations between growth, development and gender equality are to become mutually reinforcing.

Removing deflationary bias in monetary and fiscal policy is one important challenge. Employment has ceased to be a goal of macroeconomic policy. Instead, the focus is solely on financial variables, such as inflation, the fiscal deficit and debt-to-GDP ratios. Rates of inflation have been brought down to much lower levels than in the 1980s, but in many regions, this has been at a huge sacrifice in public investment, economic growth and decent jobs (for evidence, see, for example, UNRISD, 2005, p. 30). Deflationary bias in macroeconomic policy was identified as an important issue for women in the UN *World Survey on the Role of Women in Development* (UN, 1999), which argued that women disproportionately bear the costs of this bias. Much of the research and policy development on gender equality in employment has been focused on measures to enable women to compete with men on an equal basis. These measures are important, but they are not sufficient. To the extent that they are successful, they will simply redistribute some jobs from men to women. This will reduce gender gaps, but not in a way that provides 'full and productive employment and decent work for all', as called for by the UN Secretary-General (UN, 2006, p. 6). In order for gender equality to be realized in ways that 'equalize up', rather than 'equalize down', there needs to be an expansion

of the total number of decent jobs, as well as an improvement of women's access to them.

Women are particularly likely to be disadvantaged by deflationary bias because it interacts with, and reinforces, other policy biases, such as male breadwinner bias, the assumption that men are more deserving of decent jobs because they are assumed to be the principal economic support of families, while women's incomes are wrongly perceived to be merely supplementary, and not essential to family well-being (Elson and Çağatay, 2000, pp. 1354–56). Seguino (2003) finds this to be important in explaining why women are much more likely to be unemployed than men in Barbados, Jamaica and Trinidad and Tobago, using data from the period 1980–99. This inequality cannot be explained in terms of women being less educated than men, since women have a higher unemployment rate than men with the same education; nor in terms of the different sectoral distributions of men's and women's employment. While both male and female unemployment rates fell in economic upturns, male rates fell more than female rates. Male workers were the first to be hired in economic upturns, even in the female-intensive service sector.

Global unemployment rose from 5.6 per cent in 1993 to 6.2 per cent in 2003, and the female unemployment rate was slightly higher than the male rate; among young people, the gender gap was bigger (ILO, 2004). There were some regional differences: in the Middle East and North Africa, and Latin America and the Caribbean, the female unemployment rate was higher than the male, while in sub-Saharan Africa and East Asia the reverse was true (ILO, 2004, p. 2). However, in many countries, female unemployment rates are likely to underestimate the true extent of women's unemployment because women are more likely to be 'discouraged workers' who respond to their failure to find jobs by ceasing to search for one actively, although they would like to have a job if one were available. The discouragement and underemployment of women appear to have been significant in the aftermath of the Asian financial crisis in 1997–98. For instance, in South Korea, the rate of job loss for women was higher than for men, but subsequently, male unemployment rates appeared to be higher than female rates, while a higher proportion of women than before were employed in various types of informal employment (UNRISD, 2005, p. 42).

In a pioneering study, Braunstein and Heintz (2006) investigated the link between monetary policies and gender equality in employment in 17 low- and middle-income countries in the period 1970–2003. They identified episodes when monetary policy led to contractionary inflation reduction, when the growth of employment fell below its long-run trend; and episodes when it led to expansionary inflation reduction episodes, when the growth of employment was faster than its long-run trend. (The contractionary

episodes were associated with real interest rates being maintained above their long-run trend – an indicator of deflationary bias.) The study found that in 67 per cent of the contractionary inflation reduction episodes, the female-to-male employment ratio fell below its long-run trend, indicating that women were disproportionately affected by the slowdown in employment. However, in expansionary inflation reduction episodes, there was no clear disproportionate effect on either women or men. The female-to-male employment ratio increased faster than the trend in 53 per cent of cases, and at or below the trend in 47 per cent of cases.

Braunstein and Heintz concluded that a policy of responding to inflationary pressures by raising positive real interest rates above their long-run trend, and reducing real money supply below its long-run trend, tended to be associated with a greater loss in female than in male employment (relative to long-run trends in both). They noted that in 33 per cent of contractionary inflation reduction episodes, women's employment was not disproportionately affected by deflationary policies. They found that in these episodes, the real exchange rate either depreciated or remained at its long-run trend. They concluded that 'maintaining a competitive exchange rate may offset some of the gender bias observed during contractionary inflation-reduction' (Braunstein and Heintz, 2006, p. 12).

Inadequate levels of employment and decent work are also related to an emphasis on managing government budgets by cutting expenditure rather than raising tax revenues. Globalization has made it harder for governments to raise tax revenue. Çağatay (2003) summarizes the key aspects of this fiscal squeeze. Trade liberalization cuts import duties and export taxes, key sources of revenue in many poor countries. Competition to attract multinational corporations and their highly paid executives leads to cuts in corporation and capital gains taxes, and tax holidays and other exemptions and to cuts in top rates of income tax. Development cooperation grants have fallen as trade is supposed to replace aid. Governments have been encouraged or pressured into turning to sales taxes like value-added tax (VAT) to raise revenue, but such taxes fall most heavily on poor people and worsen the distribution of income. Revenue has also been raised through the sale of public enterprises and other public assets, but this only gives a one-time boost to revenue, and may result in costs for services, like water, that poor people cannot afford.

With revenue limited and debt burdens rising, the pressure has been on governments to make their budgets sustainable by cutting back on expenditure. This pressure has come from the public international financial institutions like the International Monetary Fund (IMF) and the World Bank, and also from private investors, who have seen budget deficits as harbingers of inflation, signals that the value of their assets would be eroded. In order

to build a reputation for 'sound finance' in financial markets, many governments have enacted legislation (such as balanced budget laws) that severely limits the fiscal space (Bakker, 2002).

More room for a variety of fiscal policies would not by itself ensure that fiscal policy is used to promote gender equality. Recognizing this, since the late 1990s, a series of gender budget initiatives (GBIs), in all parts of the world, have sought to improve the distribution, adequacy and impact of government budgets at national, regional and local levels; and to secure greater transparency in the use of public money; and to secure greater accountability to women as citizens. A number of tools have been developed for analyzing the gender dimensions of government budgets (Budlender and Sharp, 1998; Elson, 1998). By 2002, up to 50 countries in all parts of the world had hosted some kind of gender budget initiative (Budlender and Hewitt, 2002, p. 8). There is no one template: GBIs have taken place at all levels of government, involving regional and local government budgets as well as national budgets. Moreover, a multiplicity of actors have been involved: government ministers and officials (especially women's ministries, sometimes Ministries of Finance), parliamentarians, women's organizations and academic researchers (Budlender et al., 2002; Budlender and Hewitt, 2002). An important area of concern has been whether budgetary policies are reducing or increasing the amount of unpaid domestic work that poor women have to do; and whether they make it easier or more difficult for people (both men and women) to combine paid work and caring for their families without enduring excessive hours of work.

The most effective GBIs have produced some institutionalization of gender equality concerns in one or more stages of the budget cycle in one or more ministries, or have resulted in an ongoing public scrutiny of the budget from a gender equality perspective. Their aims have included: raising awareness and understanding of gender impacts of budgets and the policies they fund; making governments accountable for their budgetary and policy commitments; and changing and refining government budgets and policies to promote gender equality. Many examples of success in achieving the first two goals have been identified, though fewer in achieving the third (Sharp, 2002)

As well as the challenge of securing more gender-equitable public finance, there is also the challenge of securing policies on industry, trade and finance that enable increases in women's wages without jeopardizing economic growth. This challenge is analysed by Seguino and Grown (2006), who identify the need for an approach they label 'industrial policy under conditions of strategic openness'. This entails an industrial policy that promotes a shift of female employment to high-quality exports with a low price elasticity (so that higher wages for women will be less likely to have a

negative effect on exports). This should be complemented by policies that slow down the mobility of foreign direct investment, providing incentives for firms to respond to higher wages by investment in technological improvements, rather than by relocation. In addition, the maintenance of exchange rates should be maintained at competitive levels. Such a strategy would require a greater latitude for special and differential treatment of developing countries in international trade agreements.

It is also important to consider the impact of trade policies on unpaid domestic work. Pioneering research has been done on this topic by Fontana (2003), who used a computable general equilibrium model to simulate the impacts of trade liberalization on the paid and unpaid work of women and men in Bangladesh and Zambia. This technique is able to show the quantitative implications of the expansion of paid employment of women for the time they spend in unpaid domestic work and leisure, and the conditions under which there may be some redistribution of unpaid work between women and men.

The policy challenges identified above were considered by the United Nations Research Institute for Social Development in its comprehensive report on gender equality (UNRISD, 2005). The report concluded that to achieve greater gender equality, a new package of macroeconomic policies is required that puts more emphasis on redistributive taxation, gender-responsive public spending, the creation of decent work, universal social protection schemes, and policies to enable people more easily to reconcile their paid and unpaid work responsibilities, all in the context of a more just and equitable system of international economic relations (UNRISD, 2005). The challenge for the future is to ensure that not only micro and sectoral policies take account of gender issues, but also macroeconomic policies. This will require a more systematic incorporation of the unpaid work of the household sector into macroeconomic analysis, alongside the paid work of the public and private sectors.

References

- Abu-Ghaida, D. and S. Klasen (2004), 'The Costs of Missing the Millennium Development Goals on Gender Equity', *World Development*, 32 (7): 1075–1107.
- Agarwal, B. (1994), *A Field of One's Own: Gender and Land Rights in South Asia*, Cambridge: Cambridge University Press.
- Agarwal, B. (1997), "'Bargaining" and Gender Relations: Within and Beyond the Household', *Feminist Economics*, 3 (1): 1–51.
- Agarwal, B. (2005), 'Landmark step to gender equality', *Hindu*, 25 September.
- Anker, R. (1998), *Gender and Jobs: Sex Segregation of Occupations in the World*, Geneva: ILO.
- Bakker, I. (2002), 'Fiscal Policy, Accountability and Voice: The Example of Gender Responsive Budget Initiatives', Human Development Report Office Occasional Paper, New York: UNDP.
- Barro, R. and J-W. Lee (1994), 'Sources of Economic Growth', *Carnegie-Rochester Conference on Public Policy*, 40:1–46.

- Beneria, L. (1992), 'Accounting for Women's Work: the Progress of Two Decades', *World Development*, 20 (11): 1547–60.
- Beneria, L. (2003), *Gender, Development and Globalization*, New York: Routledge.
- Beneria, L. and G. Sen (1981), 'Accumulation, Reproduction and Women's Role in Economic Development', *Signs: Journal of Women in Cultural and Society*, 7 (2): 279–98.
- Blackden, C.M. and C. Bhanu (1999), 'Gender, Growth and Poverty Reduction', Special Program of Assistance for Africa, 1998 Status Report on Poverty, World Bank Technical Paper No. 428, Washington, DC: World Bank.
- Bloom, D. and J. Williamson (1998), 'Demographic Transitions and Economic Miracles in Emerging Asia', *World Bank Economic Review*, 12: 419–55.
- Booth, D., D. Kasente, G. Mavrotas, E. Mugambe and A. Muwonge (2003), 'Ex Ante Poverty and Social Impact Analysis: Ugandan Demonstration Exercise', Kampala: Department for International Development and Ministry of Finance and Economic Development.
- Boserup, E. (1970), *Women's Role in Economic Development*, New York: St Martins Press.
- Braunstein, Elissa and James Heintz (2006), 'Gender Bias and Central Bank Policy: Employment and Inflation Reduction', Working Paper, project on Alternatives to Inflation Targeting, Political Economy Research Institute, University of Massachusetts.
- Budlender, D., D. Elson, G. Hewitt and T. Mukhopadhyay (2002), *Gender Budgets Make Cents*, London: Commonwealth Secretariat.
- Budlender, D. and G. Hewitt (eds) (2002), *Gender Budgets Make More Cents*, London: Commonwealth Secretariat.
- Budlender, D. and R. Sharp (1998), *How to Do a Gender-Sensitive Budget Analysis*, London: Commonwealth Secretariat.
- Çağatay, N. (1998), 'Gender and Poverty', Working Paper No 5, UNDP Social Development and Policy Division, New York: UNDP.
- Çağatay, N. (2003), 'Gender Budgets and Beyond: Feminist Fiscal Policy in the Context of Globalization', *Gender and Development*, 11 (1): 15–24.
- Çağatay, N., D. Elson and C. Grown (eds) (1995), Special Issue on 'Gender, Adjustment and Macroeconomics', *World Development*, 23 (11).
- Çağatay, N. and S. Ozler (1995), 'Feminization of the Labor Force: The Effects of Long-Term Development and Structural Adjustment', *World Development*, 23 (11): 1883–94.
- Chakraborty, L. (2005), 'Public Investment and Unpaid Work in India: Selected Evidence from Time Use Data', paper presented to UNDP/Levy Economics Institute Global Conference on Unpaid Work and the Economy: Gender, Poverty and the Millennium Development Goals, Levy Economics Institute, October.
- Charmes, J. (2006), 'A Review of Empirical Evidence on Time Use in Africa from UN-Sponsored Surveys', in M. Blackden and Q. Wodon (eds), 'Gender, Time Use and Poverty in Sub-Saharan Africa', World Bank Working Paper No. 73, Washington, DC: World Bank.
- Chen, M., J. Vanek, F. Lind, J. Heintz, R. Jhabvala and C. Bonner (2005), *Women, Work and Poverty*, New York: UNIFEM.
- Deere, C. and M. Leon (2003), 'The Gender Asset Gap: Land in Latin America', *World Development*, 31 (6): 925–47.
- Dollar, D. and R. Gatti (1999), 'Gender Inequality, Income and Growth: Are Good Times Good for Women?', Background Paper for World Bank Policy Research Report, *Engendering Development*, Washington, DC: World Bank.
- ECLAC (2002), *Demographic Bulletin*, July, Santiago: ECLAC.
- Ellis, A., C. Manuel and M. Blackden (2006), *Gender and Economic Growth in Uganda*, Washington, DC: World Bank.
- Elson, D. (1991), 'Male Bias in Macroeconomics: The Case of Structural Adjustment', in D. Elson (ed.), *Male Bias in the Development Process*, Manchester: Manchester University Press, pp. 164–90.
- Elson, D. (1998), 'Integrating Gender Issues into National Budgetary Policies and Procedures: Some Policy Options', *Journal of International Development*, 10 (7): 929–41.
- Elson, D. (1999), 'Labour Markets as Gendered Institutions: Equality, Efficiency and Empowerment Issues', *World Development*, 27 (3): 611–27.
- Elson, D. (ed.) (2000), *Progress of the World's Women 2000*, New York: UNIFEM.

- Elson, D. and N. Çağatay (2000), 'The Social Content of Macroeconomic Policies', *World Development*, 28 (7): 1347–63.
- Elson, D. and R. Pearson (1981), 'The Subordination of Women and the Internationalisation of Factory Production', in K. Young, C. Wolkowitz and R. McCullagh (eds), *Of Marriage and the Market: Women's Subordination in International Perspective*, London: CSE Books, pp. 144–66.
- Esteve-Volart, B. (2000), 'Sex Discrimination and Growth', IMF Working Paper WP/00/84, Africa Department, Washington, DC: IMF.
- Esteve-Volart, B. (2004), 'Gender Discrimination and Growth: Theory and Evidence from India', STICERD Development Economics Papers, Suntory and Toyota International Centres for Economic and Related Disciplines, London: London School of Economics.
- Folbre, N. (1986), 'Hearts and Spades: Paradigms of Household Economics', *World Development*, 14 (2): 245–55.
- Folbre, N. (2001), *The Invisible Heart: Economics and Family Values*, New York: New Press.
- Fontana, M. (2003), 'Modeling the Effects of Trade on Women at Work and at Home: A Comparative Perspective', Trade and Macroeconomics Division Discussion Paper No.110, Washington, DC: International Food Policy Research Institute.
- Food and Agriculture Organization (FAO) (2005), Fact File on Feminization of Agriculture, www.fao.org/Gender/en/agrib2-e.htm, accessed 12 December 2005.
- Gonzales de la Rocha, M. (2000), 'Private Adjustments: Household Responses to the Erosion of Work', Conference Paper Series No. 6, New York: Bureau for Development Policy, UNDP.
- Graham, W. (2004), 'Exploring the Link Between Maternal Death and Poverty', *In Focus, Bulletin of the UNDP International Poverty Centre*, May, pp. 6–8.
- Grown, C., D. Elson and N. Çağatay (eds) (2000), Special Issue on 'Growth, Trade, Finance and Gender Inequality', *World Development*, 28 (7).
- Grown, C., G. Rao Gupta, and A. Kes (2005), *Taking Action: Achieving Gender Equality and Empowering Women*, Task Force on Gender Equality, Millennium Project, London: Earthscan.
- Haddad, L., L. Brown, A. Richter and L. Smith (1995), 'The Gender Dimensions of Economic Adjustment Policies: Potential Interactions and Evidence to Date', *World Development*, 23 (6): 881–96.
- Harriss-White, B. (1998), 'Female and Male Grain Marketing Systems: Analytical and Policy Issues for West Africa and India', in C. Jackson and R. Pearson (eds), *Feminist Visions of Development: Gender Analysis and Policy*, London and New York: Routledge, pp. 189–213.
- Hill, A. and E. King (1995), 'Women's Education and Economic Well-being', *Feminist Economics*, 1 (2): 1–26.
- ILO (2002), *Women and Men in the Informal Economy: A Statistical Picture*, Geneva: ILO.
- ILO (2004), *Global Employment Trends for Women 2004*, Geneva: ILO.
- Kabeer, N. (1994), *Reversed Realities: Gender Hierarchies in Development Thought*, London: Verso.
- Kabeer, N. (2000), *The Power to Choose: Bangladeshi Women and Labour Market Decisions*, London: Verso.
- Kabeer, N. (2005), 'Is Microfinance a "Magic Bullet" for Women's Empowerment? Analysis of Findings from South Asia', *Economic and Political Weekly*, 29 October: 4709–18.
- Kakwani, N. and H. Son (2006), 'Women's Earning Power and Well-being', Brasilia: UNDP International Poverty Centre.
- Kantor, P. (2003), 'Women's Empowerment Through Home-Based Work: Evidence From India', *Development and Change*, 34 (3): 425–45.
- Klasen, S. (1999), 'Does Gender Inequality Reduce Growth and Development? Evidence from Cross Country Regressions', Background Paper for World Bank Policy Research Report *Engendering Development*, Washington, DC: World Bank.
- Klasen, S. (2002), 'Low Schooling for Girls? Slower Growth for All?', *World Bank Economic Review*, 16: 345–73.
- Klasen, S. (2005), 'Population Growth, Economic Growth, and Poverty Reduction in Uganda: Theory and Evidence', Discussion Paper No. 125, Department of Economics, University of Goettingen.

- Klasen, S. and F. Lamanna (2003), 'The Impact of Gender Equality in Education and Employment in the Middle East and North Africa', background paper for report on Gender and Development in the Middle East and North Africa, Washington, DC: World Bank.
- Klasen, S. and C. Wink (2003), 'Missing Women: Revisiting the Debate', *Feminist Economics*, 9 (2–3): 263–99.
- Lastarria-Cornhiel, S. (1997), 'Impact of Privatization on Gender and Property Rights in Africa', *World Development*, 25 (8): 1317–33.
- Lewis, W. Arthur (1954), 'Economic Development with Unlimited Supplies of Labour', *Manchester School*, 22 (2): 139–91.
- Lewis, W. Arthur (1955), *The Theory of Economic Growth*, London: Allen & Unwin.
- Lim, J. (2000), 'The Effects of the East Asian Crisis on the Employment of Women and Men: the Philippine Case', *World Development*, 28 (7): 1285–306.
- Mayoux, L. (2000), 'Microfinance and the Empowerment of Women: A Review of Key Issues', Social Finance Unit Working Paper No. 23, Geneva: ILO.
- Ministry of Agriculture, Government of India (2005), 'Concept Note on Gender Resource Centre', www.apeda.com/compendium/MoA/CONCEPT_NOTE_ON.htm, accessed 12 December 2005.
- Moser, C. (1989), 'The impact of recession and adjustment at the micro level: low income women and their households in Guayquil, Ecuador', in UNICEF, *The Invisible Adjustment: Poor Women and the Economic Crisis*, New York: UNICEF.
- Muhereza, F. (2001), 'Ranches Restructuring and Changing Gender Relations in Pastoral Households in the Former Ankole Ranching Scheme', Kampala: Centre for Basic Research.
- Razavi, S. (1999), 'Export-oriented Employment, Poverty and Gender: Contested Accounts', *Development and Change*, 30: 653–83.
- Robbins, D. (1999), 'Gender, Human Capital and Growth: Evidence from Six Latin American Countries', OECD Development Centre, Working Paper No 151, Paris: OECD.
- Saffiotti, H. (1978), *Women in Class Society*, New York: Monthly Review Press.
- Seguino, S. (2000), 'Gender Inequality and Economic Growth: A Cross Country Analysis', *World Development*, 28 (7): 1211–30.
- Seguino, S. (2002), 'Gender, Quality of Life, and Growth in Asia 1979–1990', *Pacific Review*, 15 (2): 245–77.
- Seguino, S. (2003), 'Why are Women in the Caribbean so much More Likely than Men to be Unemployed?', *Social and Economic Studies*, 52 (4): 83–120.
- Seguino, S. (2005), 'All Types of Inequality are Not Created Equal: Divergent Impacts of Inequality on Economic Growth', Levy Economics Institute Working Paper No. 433, Levy Economics Institute, Annandale-on-Hudson.
- Seguino, S. (2007), 'The Great Equalizer? Globalization Effects on Gender Equity in Well-Being in Latin America and the Caribbean', in A. Shaikh (ed.), *Globalization and the Myth of Free Trade*, London: Routledge, pp. 177–214.
- Seguino, S. and C. Grown (2006), 'Gender Equity and Globalization', *Journal of International Development*, 18 (8): 1081–104.
- Sen, A.K. (1983), 'Economics and the Family', *Asian Development Review*, 1 (2): 14–26.
- Sen, A.K. (1990a), 'More than a Million Women Are Missing', *New York Review of Books*, 20 December.
- Sen, A.K. (1990b), 'Gender and Co-operative Conflicts', in I. Tinker (ed.), *Persistent Inequalities. Women and World Development*, Oxford: Oxford University Press, pp. 123–49.
- Sen, G. (1996), 'Gender, Markets and States: A Selective Review and Research Agenda', *World Development*, 24 (5): 821–9.
- Sen, G. and C. Grown (1987), *Development, Crises and Alternative Visions: Third World Women's Perspectives*, New York: Monthly Review Press.
- Sharp, R. (2002), 'Moving Forward: Multiple Strategies and Guiding Goals', in United Nations Fund for Women (UNIFEM), *Gender Budget Initiatives. Strategies, Concepts and Experiences*, New York: United Nations Development Fund for Women.
- Smith, L. and J-P. Chavas (1999), 'Supply Response of West African Agricultural Household: Implications of Intra-household Preference Heterogeneity', *Food Consumption and*

- Nutrition Division Discussion Paper No. 69, Washington, DC: International Food Policy Research Unit.
- Smith, L. and L. Haddad (2000), 'Explaining Child Malnutrition in Developing Countries: A Cross Country Analysis', Research Report 111, Washington, DC: International Food Policy Research Institute.
- Standing, G. (1999), 'Global Feminization Through Flexible Labor: A Theme Revisited', *World Development*, 27 (3): 583–602.
- Tae-Hong, K. (2001), 'Economic Evaluation of Unpaid Work in Republic of Korea', paper presented at Workshop on Integrating Paid and Unpaid Work in National Systems, United Nations Economic and Social Commission, Bangkok.
- Tanski, J. (1994), 'The impact of crisis, stabilization and structural adjustment on women in Lima, Peru', *World Development*, 22 (11): 1627–42.
- Tinker, Irene (1990), 'A Context for the Field and the Book', in I. Tinker (ed.), *Persistent Inequalities. Women and World Development*, Oxford: Oxford University Press, pp. 27–53.
- UN (1999), *World Survey on the Role of Women in Development*, Division for the Advancement of Women, Department of Economic and Social Affairs, New York: United Nations.
- UNESCO (2004), *Education for All Global Monitoring Report 2003/4*, Paris: UNESCO.
- United Nations (2006), Report of the Secretary-General on the work of the organization. General Assembly, Sixty-first Session, Supplement No. 1 (A/61/1), New York: United Nations.
- United Nations Research Institute for Social Development (UNRISD) (2005), *Gender Equality: Striving for Justice in an Unequal World*, Geneva: United Nations Research Institute for Social Development.
- Whitehead, A. (2005), 'The Gendered Impacts of Liberalization Policies on African Agricultural Economies and Rural Livelihoods', background paper for UNRISD Report on *Gender Equality: Striving for Justice in an Unequal World*, Geneva: UNRISD.
- Wold, B. (1997), 'Supply Response From a Gender Perspective: The Case of Structural Adjustment in Zambia', Oslo: Statistics Norway and Zambia Central Statistical Office.
- World Bank (2001), *Engendering Development*, New York: Oxford University Press.

51 Children and development

Paul Glewwe and Amy Damon

Introduction

In developing countries, 30 percent of the population is less than 15 years old. Thus any discussion of the well-being of the population in any developing country must examine the welfare of children. In general, the relationship between children and economic and social development is primarily one of the impact of development on children, rather than the impact of children on development. Yet today's children will become adults in one or two decades, and events during their childhood will have a strong impact on their lives as adults, including their contribution to economic and social development.

This chapter summarizes recent research by economists on the status of children in developing countries. It begins by examining the status of children in those countries, and proceeds by assessing the impact of economic growth on children. It then presents policies that are likely to improve children's health, education and employment status. The final section presents some concluding remarks.

Child welfare in developing countries

By virtually any measure, the welfare of children in developing countries has improved markedly over the past half-century, not only for developing countries as a whole but for all regions as well. This section reviews this progress and presents the current status of children with respect to their health, educational status and employment status.

Health

In almost every dimension the health status of children is improving in nearly all developing countries. Progress in reducing infant and child mortality rates is shown in Tables 51.1 and 51.2. These rates show how many children died, out of every 1000 born, before their first and fifth birthdays, respectively. For developing countries as a whole, the infant mortality rate has dropped dramatically, from 126 in 1960 to 53 in 2000, and the child mortality rate has declined sharply, from 197 to 78, over the same time period. This decline has occurred in all regions of the developing world.

Despite this progress, infant and child deaths are still common in developing countries. The situation in sub-Saharan African countries is especially

Table 51.1 Infant mortality rate (per 1000 live births)

Region	Statistic	1960	1980	2000
East Asia	mean	116	65	43
	<i>n</i>	14	17	22
Europe/Central Asia	mean	65	46	28
	<i>n</i>	19	26	27
Latin America/Caribbean	mean	95	57	27
	<i>n</i>	25	26	32
Middle East and North Africa	mean	151	83	39
	<i>n</i>	13	13	12
South Asia	mean	166	119	61
	<i>n</i>	8	8	7
Sub-Saharan Africa	mean	162	118	94
	<i>n</i>	38	45	47
All developing countries	mean	126	82	53
	<i>n</i>	117	135	147

Source: World Bank (2002b).

worrisome. In particular, progress in reducing infant and child mortality was much slower between 1980 and 2000 than it was from 1960 to 1980. While rates were similar in sub-Saharan Africa and South Asia in 1960 and 1980, and both made good progress in reducing rates in those two decades, South Asia was able to reduce infant and child deaths much more quickly between 1980 and 2000. The most obvious explanation is the advent of AIDS in sub-Saharan Africa, but slow economic growth is also likely to have played a role.

Education

Schooling is another area in which child welfare has increased in the last half-century, as seen in Tables 51.3 and 51.4. These tables present gross enrollment rates, which are defined as the number of children enrolled as primary (or secondary) students divided by the number of children in the age range associated with that level of schooling. Note that it is possible for these enrollment rates to exceed 100 because it is common for 'over-age' children to be in a particular level because of delayed initial enrollment or grade repetition.

In developing countries as a whole, the primary school gross enrollment rate increased from 68 percent in 1960 to 99 percent in 2000. These

Table 51.2 Child mortality rate (per 1000 live births)

Region	Statistic	1960	1980	2000
East Asia	mean	175	94	56
	<i>n</i>	14	17	22
Europe/Central Asia	mean	88	57	34
	<i>n</i>	19	26	27
Latin America/Caribbean	mean	139	77	34
	<i>n</i>	25	26	32
Middle East and North Africa	mean	237	117	49
	<i>n</i>	13	13	12
South Asia	mean	265	182	83
	<i>n</i>	8	8	7
Sub-Saharan Africa	mean	268	191	152
	<i>n</i>	38	45	47
All developing countries	mean	197	124	78
	<i>n</i>	117	135	147

Source: World Bank (2002b).

Table 51.3 Primary school gross enrollment rate

Region	Statistic	1960	1980	2000
East Asia	mean	77	102	105
	<i>n</i>	8	12	13
Europe/Central Asia	mean	101	98	99
	<i>n</i>	21	19	19
Latin America/Caribbean	mean	89	101	112
	<i>n</i>	23	25	26
Middle East and North Africa	mean	60	86	90
	<i>n</i>	8	9	8
South Asia	mean	33	79	88
	<i>n</i>	6	7	5
Sub-Saharan Africa	mean	42	75	90
	<i>n</i>	37	38	36
All developing countries	mean	68	89	99
	<i>n</i>	103	110	107

Table 51.4 Secondary school gross enrollment rate

Region	Statistic	1960	1980	2000
East Asia	mean	15	41	64
	<i>n</i>	8	11	13
Europe/Central Asia	mean	52	91	85
	<i>n</i>	21	20	16
Latin America/Caribbean	mean	18	47	75
	<i>n</i>	23	25	25
Middle East and North Africa	mean	12	37	59
	<i>n</i>	8	9	6
South Asia	mean	10	22	44
	<i>n</i>	6	7	4
Sub-Saharan Africa	mean	4	17	36
	<i>n</i>	36	36	27
All developing countries	mean	19	42	61
	<i>n</i>	102	108	91

increases occurred in all regions of the developing world (except in Europe and Central Asia, where the enrollment rate was already 101 percent in 1960). Increases in the secondary school gross enrollment rate are even more striking, increasing from 19 percent in 1960 to 61 percent in 2000. These rates also increased in all regions.

Despite these gains, there are some troubling patterns. First, primary school enrollment rates are still below 100 percent in sub-Saharan Africa, South Asia and the Middle East and North Africa. Indeed, heavy grade repetition exaggerates the percentage of children who are actually in primary school. Sub-Saharan Africa and South Asia also lag behind in their secondary school enrollment rates, although they are increasing rapidly over time. Second, there is evidence that children in developing countries learn much less than children in developed countries for a similar amount of time in school. These issues are discussed in detail in Glewwe and Kremer (2006).

Employment

School-age children work in many poor countries, sometimes so much so that they leave school. Most children who work in developing countries work alongside of their parents in agricultural activities, but some work in factories or in other activities in urban areas. Concerns regarding child

labor have received increased attention in policy circles in the last one to two decades.

While child labor has received increased attention, child labor itself has decreased in almost all developing countries. The school enrollment trends discussed in the previous subsection suggest that this is the case, and the data in Table 51.5 confirm this. In developing countries as a whole, the labor force participation of children aged 10–14 has been reduced by almost half from 1960 to 2000, from 24.4 percent to 13.5 percent. Yet it is still high in two regions, South Asia and sub-Saharan Africa, which are the two regions with the lowest school enrollment rates. This may reflect lower economic growth in those two regions, an issue that is explored further in the next section.

The impact of economic growth on child welfare

Economic development has often been equated with income growth, but development includes not only income growth but also better health outcomes, higher levels of education, better housing (including potable water and hygienic sanitation conditions), and perhaps even democracy and respect for human rights. While it is very rare for income growth not to be accompanied by these other aspects of the quality of life, the nature of economic growth can determine how quickly economic growth leads to

Table 51.5 Labor force participation of children, aged 10–14 (% of age group)

Region	Statistic	1960	1980	2000
East Asia	mean	28.4	21.3	12.8
	<i>n</i>	15	15	15
Europe/Central Asia	mean	3.8	1.1	0.3
	<i>n</i>	27	27	27
Latin America/Caribbean	mean	14.5	9.8	5.5
	<i>n</i>	26	26	26
Middle East and North Africa	mean	17.2	10.9	2.8
	<i>n</i>	13	13	13
South Asia	mean	40.0	31.5	22.3
	<i>n</i>	8	8	8
Sub-Saharan Africa	mean	40.6	35.2	27.7
	<i>n</i>	45	45	45
All developing countries	mean	24.4	19.3	13.5
	<i>n</i>	134	134	134

improvements in other areas. This section presents evidence on the potential role that economic growth can play in leading to improvements in child welfare, first by showing that per capita income is positively correlated with indicators of child welfare and that income growth is correlated with improvements in child welfare, and then citing several studies that examine more carefully the causal impact of income on child welfare.

Correlation between per capita income and child welfare

Nations with higher per capita incomes have, on average, higher levels of child welfare, as seen in Table 51.6. More specifically, in the year 2000 in low-income countries (those with per capita income levels of \$755 or less), 79 children out of every 1000 born died before their first birthday, and another 42 died before their fifth birthday. In contrast, in middle-income countries (those with per capita income levels above \$755 but below \$9266) only 27 out of 1000 children died before their first birthday, and only another seven died before their fifth birthday.

Middle-income countries also have much lower rates of child labor force participation than do low-income countries: 4 percent of children aged 10–14 work in middle-income countries, but 22 percent work in low-income countries. Primary and secondary school (gross) enrollment rates are also higher in middle-income countries; the rates for those countries are 110 percent and 77 percent, respectively; while the rates in low-income countries are much lower, namely 88 percent and 40 percent.

Not only is child welfare higher in countries with higher income, but improvements in child welfare are also positively correlated with the rate of

Table 51.6 Income levels and child welfare

Income level	Statistic	Child mortality				
		Infant mortality rate (per 1000 live births)	Child mortality rate (per 1000 live births, < 5 yrs)	Child labor (children 10–14)	Primary school gross enrollment rate	Secondary school gross enrollment rate
Low income	mean	79	121	22	88	40
	<i>n</i>	72	72	69	48	38
Middle income	mean	27	34	4	110	77
	<i>n</i>	72	72	59	56	51
Low and middle income	mean	53	77	13	100	61
	<i>n</i>	144	144	128	104	89

Table 51.7 *Changes in child welfare by rate of per capita income growth*

Growth rate	Statistic	% Δ Child labor	% Δ IMR	% Δ CMR
Slow growth (bottom 25%)	mean	-42.7%	-23.0%	-23.6%
	<i>n</i>	25	25	25
Medium growth (25th–75th percentile)	mean	-49.3%	-40.5%	-43.2%
	<i>n</i>	46	51	51
Fast growth (top quartile)	mean	-55.0%	-47.9%	-51.9%
	<i>n</i>	19	19	19

income growth. This is seen in Table 51.7. Data from about 90 countries are used to divide countries into the 25 percent that had the slowest rates of growth from 1980 to 2000, the 25 percent that had the fastest economic growth, and the 50 percent with ‘intermediate’ rates of economic growth. Child labor and infant mortality drop faster in countries with higher economic growth. For example, from 1980 to 2000 labor force participation of children aged 10–14 dropped by 55 percent in countries with high economic growth but only by 43 percent in countries with low economic growth. Similarly, the infant and child mortality rates decreased by almost half in countries that had high economic growth while decreasing only by about one-quarter in countries with low economic growth.

The causal impact of income on child welfare

The correlations shown in the previous subsection are consistent with the hypothesis that income growth causes improvements in child welfare, but they do not constitute proof of a causal relationship. This subsection briefly reviews several recent studies that use microeconomic (household survey) data to provide more convincing evidence of a causal relationship.

Income growth can lead to improvements in child welfare, and in social welfare more generally, by two distinct pathways. The first, and most obvious, pathway is that households with higher incomes can purchase goods and services that improve children’s health and education outcomes, and improve child welfare in other ways. The theoretical literature has developed formal models that show how these income effects can occur (see, *inter alia*, Basu, 1999; Glewwe, 2002). Second, households with higher incomes, and more generally higher-income economies, generate more tax revenue (via either direct or indirect taxes) that governments can use to provide health, education and other services. Household survey data can be used to search for causal relationships that operate through one or both of these pathways.

Four recent studies using data from Vietnam show causal impacts of household income and/or provision of education and health services on child health, education and child labor outcomes. Glewwe and Jacoby (2004) present a dynamic model of school attainment that focuses on the role played by household wealth. They use panel data from the 1990s to show that increases in household wealth over time lead to a substantial and statistically significant increase in years of schooling, even after controlling for changes in the quality of schooling, the rate of return to education and the opportunity cost of (child) labor.

A study by Glewwe et al. (2004) using the same data examines the impact of economic growth on children's nutritional status, as measured by their height for age. Unlike the case with education, household income by itself has little impact on children's nutritional status. This suggests that improvements in the nutritional status of Vietnamese children in the 1990s were primarily due to improvements in health care services. Unfortunately, there are no data available on changes in the quality of health care services over time, so the paper cannot present strong evidence in favor of this conjecture. However it does present some evidence that higher-quality health care facilities lead to improvements in children's nutritional status.

Research by Wagstaff and Nguyen (2004) examines the factors that affect child mortality rates in Vietnam. The authors find that access to safe drinking water, vaccination campaigns and access to trained medical personnel during childbirth reduce child mortality in that country. In contrast, but consistent with Glewwe's result for child nutritional status, they find no impact of household income on child mortality.

Finally, Edmonds (2005) shows that children in better-off households in Vietnam are less likely to work, and that households whose incomes increase over time are more likely to keep their children in school and less likely to put them to work. This pattern is seen in other countries as well, as discussed in Edmonds and Pavcnik (2005).

Policies to promote child welfare

The evidence from the previous section demonstrates that both increases in households' disposable incomes and increased government spending on social services lead to improvements in child welfare. This section reviews what developing countries can do to improve child welfare using policies that work through both types of causal pathways.

Promote economic growth

Macroeconomists and other economists have had, and continue to have, long debates on the best way to promote economic growth. For recent summaries of the evidence, see Easterly (2005), Rodrik (2005) and the World

Bank (2002a). Unfortunately, there is disagreement on what has been learned, but there are some areas of agreement. This chapter is too brief to delve into the details of the best policies to promote economic growth, but there is broad agreement among economists that economic growth is necessary for large, sustained improvements in child (and adult) living standards in developing countries.

Health policies

A recent book by Wagstaff and Claeson (2004) provides a detailed assessment of the effectiveness of health policies in developing countries. Policies that are effective in raising children's nutritional status include improved hygiene and sanitation, dietary supplements that provide iron and vitamin A (for both mothers and children), and provision of deworming drugs to school-age children. To reduce infant and child mortality, the authors recommend improved hygiene sanitation, provision of mosquito nets that are treated with insecticide, child immunizations, dietary supplements of zinc and vitamin A, and improved facilities for childbirth.

Finally, it is important to realize that child health has large implications for children's educational outcomes. Several recent studies have found sizeable and statistically significant positive impacts of child health on education outcomes. Thus there is growing evidence of a causal impact of child health on education. Note as well that there is no clear evidence of large gender differences in the impact of child health on education outcomes. For details on the relationship between health and education, see Glewwe and Miguel (2008).

Education policies

This subsection summarizes recent research on policies that lead to increases in the years children spend in school and in the skills learned while in school. For details, see Glewwe and Kremer (2006).

Several education policies have been found to be effective in increasing the number of years that children spend in school, as well as their daily attendance during those years. More specifically, programs that reduce the costs of schooling faced by parents or provide incentives for daily attendance (either explicitly or implicitly through school meals) have sizable impacts on school enrollment and attendance. Randomized evaluations of school-based health programs, for example a deworming program in Kenya, suggest that, in some situations, these programs can be an extraordinarily cost-effective means of increasing the amount of time that children in developing countries are in school.

Evidence concerning the impact of education policies, such as provision of textbooks and additional training for teachers, on the skills that children

acquire in school is more mixed. In general, studies based on cross-sectional data suggest that most education policies have had limited impacts on the academic skills of schoolchildren in developing countries. Evidence from recent 'natural experiments' in middle-income countries suggests that reducing class size can raise academic achievement but that providing computers has little effect. Recent randomized trials conducted in low-income countries provide a more mixed picture.

The evidence suggests that the most effective forms of spending on education are likely to be those that respond to inefficiencies in schooling systems. Providing textbooks written with atypical students in mind will benefit only atypical students, whereas remedial education may be extremely effective in an environment in which many students fall behind and are no longer able to follow their teachers' lessons. Providing radio mathematics education or computer-based education may be effective when teachers attend irregularly.

Schools in developing countries face significant institutional problems: distortions in education budgets often result in inefficient allocation and spending of funds; weak teacher incentives lead to problems such as high rates of teacher absenteeism; and, given the difficulties faced by these school systems, curriculums are often inappropriately matched with the level of the typical student. Yet reform initiatives can easily have unintended consequences. The details of these programs are critical for determining their effects on the incentives faced by teachers and others (principals, parents, and so on). Governance reforms and allowing school choice appear to hold more promise than simply providing monetary incentives to teachers based on test scores, but much more empirical evidence is needed on the impact of these reforms as well.

Employment policies

Of particular interest in recent years has been the issue of child labor. Clearly, children who work long hours cannot attend school, and there are many children who do work that could be directly harmful to their welfare. On the other hand, most children who work are working for their parents either on the family farm or on a rented plot of land, and their contribution towards their families' income may have important welfare benefits for them and for other family members. This raises the issue of when, if ever, it is appropriate for governments to oppose choices that parents make for their children.

Two related policies that should reduce child labor without attempting to interfere with the choices that parents make for their children are to improve school quality and reduce the cost of attending school. Better and less-expensive schools will make schooling more attractive and thus should

persuade parents to allocate more of their children's time to schooling and less to child labor. In contrast, Edmunds and Pavcnik (2005) argue that attempts to ban child labor are likely to be ineffective and, if effective, may harm poor families. For further discussion, see Edmunds and Pavcnik (2005) and the references cited in that paper.

Concluding comments

Child welfare has improved in all regions of the developing world since the 1960s, and continued improvements in the future are likely. Even so, there are still opportunities to increase children's schooling, raise their health and nutritional status, and reduce the amount of time they spend working to support their families. Policies to promote economic growth are one general avenue to accelerate progress, but there are other policies that are also effective, as briefly summarized in the preceding section. Efforts should focus on South Asia and sub-Saharan Africa, where the welfare of children is lowest.

A final important point is that improved child welfare today translates into better economic and social development in the future. Children who have higher levels of education and better health are more productive workers and thus contribute more not only to their own welfare as adults, but also to overall economic growth. Lastly, more-educated and healthier children will also be better able when they are adults to provide a better life for their own children.

References

- Basu, Kaushik (1999), 'Child Labor: Cause, Consequence and Cure, with Remarks on International Labor Standards', *Journal of Economic Literature*, 37 (3): 1083–1119.
- Easterly, William (2005), 'National Policies and Economic Growth: A Reappraisal', in P. Aghion and S. Durlauf (eds), *Handbook of Economic Growth*, North-Holland.
- Edmonds, Eric (2005), 'Does Child Labor Decline with Improving Economic Status?' *Journal of Human Resources*, 40 (1): 77–99.
- Edmonds, Eric and Nina Pavcnik (2005), 'Child Labor in the Global Economy', *Journal of Economic Perspectives*, 19 (1): 199–220.
- Glewwe, Paul (2002), 'Schools and Skills in Developing Countries: Education Policies and Socioeconomic Outcomes', *Journal of Economic Literature*, 40 (2): 436–82.
- Glewwe, Paul and Hanan Jacoby (2004), 'Economic Growth and the Demand for Education: Is there a Wealth Effect?', *Journal of Development Economics*, 74 (1): 33–51.
- Glewwe, Paul and Michael Kremer (2006), 'Schools, Teachers, and Education Outcomes in Developing Countries', in E. Hanushek and F. Welch (eds), *Handbook of the Economics of Education*, North-Holland.
- Glewwe, Paul, and Edward Miguel (2008), 'The Impact of Child Health and Nutrition on Education in Less Developed Countries', in T. Paul Schultz and John Strauss (eds), *Handbook of Development Economics, Volume 4*, North-Holland.
- Glewwe, Paul, Stefanie Koch and Bui Linh Nguyen (2004), 'Child Nutrition, Economic Growth, and the Provision of Health Care Services in Vietnam', in P. Glewwe, N. Agrawal and D. Dollar (eds), *Economic Growth, Poverty and Household Welfare in Vietnam*, Washington, DC: World Bank, pp. 315–89.

- Rodrik, Dani (2005), 'Growth Strategies', in P. Aghion and S. Durlauf (eds), *Handbook of Economic Growth*, North-Holland.
- Wagstaff, Adam and Mariam Claeson (2004), *The Millennium Development Goals for Health: Rising to the Challenges*, Washington, DC: World Bank.
- Wagstaff, Adam, and Nga Nguyet Nguyen (2004), 'Poverty and Survival Prospects of Vietnamese Children', in P. Glewwe, N. Agrawal and D. Dollar (eds), *Economic Growth, Poverty and Household Welfare in Vietnam*, Washington, DC: World Bank.
- World Bank (2002a), *Globalization, Growth and Poverty*, Washington, DC: World Bank.
- World Bank (2002b), *World Development Indicators*, Washington, DC: World Bank.

52 Ethnicity and economic development

William A. Darity Jr and Russell E. Triplett

The economic geography of ethnic and racial inequality

Virtually all countries have ethnic/racial divisions. Where those divisions are present they are universally characterized by dense gaps in economic status between the respective groups. Intergroup division and inequality occur in countries at all levels of per capita income or development more broadly construed. To the extent that intergroup inequality is significantly driven by discriminatory practices, those practices have a persistence that defy the standard prediction of conventional economics that discriminators will be driven from the marketplace because their actions are unprofitable. Furthermore, despite the strong and widely popularized claims of some scholars, there is no clear general relationship between ethnic/racial division within a country and its overall economic performance. Nor is there any sound reason to believe that causation is typically unidirectional from ethnic division to economic performance. But high rates of economic growth do not necessarily erode ethnic division.

Research that one of the authors of this chapter has undertaken with other collaborators (Darity and Deshpande, 2000; Darity and Nembhard, 2000) demonstrates the international prevalence of ethnic/racial differentiation and the strong correspondence between such differentiation and economic disparity. Darity and Nembhard (2000), in particular, demonstrate that intergroup differentiation and inequality is present in countries with large and small populations (for example the USA and Belize), countries with high and low average income levels (for example Japan and India), countries with a recent experience of rapid economic growth and slow economic growth (for example Malaysia and New Zealand), and countries with comparatively high degrees and low degrees of general inequality (for example Brazil and Australia). Indeed, even countries that are more inclusive toward women in their national political processes or that display greater gender equality do not consistently display reduced levels of ethnic/racial inequality.

Moreover, inequalities between ascriptively differentiated groups can be detected using a variety of measures of disparity. Intergroup gaps exist in income, occupational status, access to quality education, access to quality health care, wealth, and related measures of well-being. With respect to wealth, for example, in the USA, one of the world's most affluent countries,

blacks and Latinos have approximately a mere one-tenth of the net worth of whites at the median (Kochhar, 2004).

Ethnic and racial inequality and the Human Development Index

In the 1994 edition of the United Nations Development Programme's (UNDP) *Human Development Report* ethnic/racial inequality was described across a number of countries by calculating disaggregated values of the UNDP's Human Development Index (HDI). The HDI is a measure of well-being that includes not only per capita income as a component but also indicators of educational attainment and health status in a population. Again, cross-national comparisons suggest that intergroup differentiation and inequality is evident in countries with scores at all levels of the HDI (Darity, 2002).

Within countries intergroup disparity measured by the HDI is widely evident as well. In 1992, toward the end of the apartheid era, South Africa's overall HDI score was 0.65. However, the score estimated separately for white South Africans was 0.88, a score that would have placed white South Africa 24th out of approximately 180 countries worldwide (at the same level as Spain). In contrast, blacks in South Africa would have had an HDI score of only 0.46, placing them 123rd out of 180 countries, ranking slightly above the Congo.

Disaggregated estimates for Brazil were provided by region rather than by race or ethnicity. But the specific regions utilized in the UNDP, the south and the north-east of Brazil, provide information about racial/ethnic inequality there. The population of the south is disproportionately white, consisting largely of Euro-Brazilians, while the north-east is disproportionately black and mulatto or Afro-Brazilian. The overall HDI for Brazil as a whole was 0.76 in 1992, placing the country 63rd internationally. However, southern Brazil had a score of 0.84 which would have placed that region 43rd in the HDI rankings. North-eastern Brazil's score was 0.55; the region would have ranked in the lower half of the world on the basis of HDI.

Similarly for Nigeria the *World Development Report* did not provide direct information on HDI scores by ethnic group. Scores were provided for each of the 19 Nigerian states. This afforded indirect information about intergroup inequality in Nigeria because each of the states are relatively ethnically homogeneous internally. It is striking that while Nigeria's overall HDI was a very low 0.348 in 1992, ranking the country 139th in the world, the state of Bendel had a score of 0.67, higher than Sri Lanka or Cuba, while the state of Borno had an HDI of only 0.16, beneath any country in the world.

Canada's HDI in 1992 of 0.93 was the highest in the world. Nevertheless, even there the evidence of intergroup disparity was dramatic. While it was

not possible to construct separate HDIs for Canada's ethnic/racial groups with the available data, the data 'do show that the "aboriginals" (the Indians, the Inuit, and the Metis, constituting 2.3% of the population) have a life expectancy 5.6 years lower than the rest of the population, and their real income is one-third less' (UNDP, 1994, p. 100). Indigenous people were far more likely to be subjected to violence, experience depression and undergo unemployment. Their unemployment rate of 20 percent was twice the national average in Canada at the time (UNDP, 1994, pp. 25–6, 32).

The record also indicates that discrimination generally plays an important role in maintaining conditions of economic disparity, particularly discrimination in employment and housing. Plus, despite the theoretical presumption of conventional economics, the historical record gives little reason to believe that such discriminatory practices will disappear with the passage of time, even in market-based economies. In some countries, for example Brazil, there has been evidence of an increase in discriminatory differentials against blacks and mulattos over time (Darity, 1998). In others, like the USA and South Africa, where legal regimes of segregation have been overturned, there is evidence of a decline in discrimination in the period immediately following the regime change, but still high levels of discrimination persisting thereafter (Darity and Deshpande, 2000; Darity, 2002). Statistical estimates actually indicate that measured discrimination increased against Puerto Rican, Mexican and native American ancestry men in the USA between 1980 and 1990 (Darity, 2002).

Ethnic conflict, neoinstitutionalism and economic growth

A provocative and influential paper by William Easterly and Ross Levine (1997) oriented the discussion of the role of ethnicity toward an examination of its impact on economic development, rather than its impact on intergroup disparity. Easterly and Levine argued that the low rates of economic growth characteristic of African nations' economies since the 1970s are attributable to the high levels of ethnic diversity there. Ethnic variation in a country, in and of itself, would lead to rent-seeking practices that would prove to be predatory on effective governance and contribute directly to slower growth. In the most recent versions of the argument (Easterly, 2001; Easterly et al., 2006), taking a tone very similar to Robert Putnam's (2000) approach to conditions for community health, unity and participatory democracy, ethnic diversity becomes a critical factor undermining 'social cohesion' and thereby undermining the quality of 'institutions' that might otherwise promote economic growth.

One implication of the Easterly and Levine position is African governments would be larger – inefficiently larger – relative to the scale of their respective economies than governments in parts of the world with less

ethnic diversity. But Dani Rodrik (2000) finds that the public sectors in African countries are not generally comparatively larger by international standards, and he shows that there is at most a weak correlation between the size of the public sector and the magnitude of rent-seeking activity and corruption in a country.

Easterly and Levine focus on ethnic diversity – simply the magnitude of ethnic variation that exists in a country. They give considerably less attention to the forms of ethnic antagonism (for a catalogue of these forms see Bardhan, 1997) that might play out in different settings and their consequences. One can easily conclude that racial/ethnic differentiation that results in genocidal violence and/or conditions of civil war will not be good for economic growth. But this requires more than mere diversity; it requires high levels of group consciousness and high levels of between-group conflict. Indeed, in the context of a model utilizing evolutionary game theory it can be demonstrated that the continuation of group consciousness is dependent upon unequal resources being associated with identification with and membership in each group (Darity et al., 2006).

During the midst of the Burundian genocide of the early 1990s – in this case directed by the Tutsis against the Hutus – Leonce Ndikumana (1993, p. 30) described the country as possessing ‘a rigid ethnic stratification and unequal distribution of power along ethnic lines. While the Hutu make up the majority of the population (about 85 percent), the minority Tutsi (about 14 percent) control the government, the military, and the economy. The third ethnic group, the Twa, has assumed a role of second-class citizens with little integration in the economic and political system’. The Burundian genocide at that time was associated with an ‘annual decline in agriculture (value added) [that] reached –10.5% in 1994, exceeding the decline experienced after the 1972 massacre (–10%), [and] total production (GDP) . . . declined at an annual rate of over 6% in 1993 and 1994’ (Ndikumana, 1993, p. 30). Between 1960 and 1998 Nkurunziza and Ngaruko (2002) estimate that Burundian per capita income fell from \$620 to \$370. Without the civil wars they estimate that per capita income would have been \$667 in 1997 instead of \$397.

In neighboring Rwanda, five years of civil war culminated in the 1994 genocide directed by the majority Hutus against the minority Tutsis. Gross domestic product (GDP) fell in three of the five years; in 1994 there was an astronomical fall in GDP of 40 percent. Only the infusion of massive amounts of foreign aid led to a 9 percent growth rate in 1995 (US Department of State, 2007).

Burundi and Rwanda are extreme cases. Such extreme forms of inter-ethnic conflict are not unique to the African continent. Inter-ethnic strife leading to genocidal violence has occurred in the former Yugoslavia, in

Indonesia in 1997 and 1998, and in Cambodia under the Khmer Rouge, all contributing to economic decline. The UNDP (1994, p. 47) identifies Turkey, the United Kingdom (especially Northern Ireland), Iraq, Iran, Israel, Lebanon, Colombia, Guatemala, Bangladesh, India, Laos, Myanmar, Pakistan, the Philippines, Sri Lanka and Tajikistan as non-African nations where ethnic conflict has turned violent with adverse consequences for economic performance. France's ethnic violence in late 2005 also fits the model.

Generally, Robert Bates (2000) argues that the level of political violence represented by riots, demonstrations, revolts and assassinations is lower in African countries than might be expected, given the level of ethnic division there. While the short-term effects on growth of political violence, particularly genocidal violence, are negative, the long-term effects on economic growth are ambiguous:

Even a genocidal process of ethnic homogenization of population or ethnic homogenization of control over a nation's resources is not inimical to prosperity, at least for the 'winners' and their descendants. Indeed, wealth seizures in the form of conquest of native peoples and appropriation of their lands, coupled with the use of captive and slave labor, laid the basis for the affluence of today's richest nations, for example, the United States, Australia, Britain, France, Belgium and the whites of southern Africa.

Theft via conquest has long constituted an effective mechanism for achieving redistribution of wealth among groups. Industrialization, by destruction of lives of indigenous peoples, has been a commonplace event during the past half millennium. Violence is the historic adjunct to compulsory wealth redistribution across racial or ethnic lines. (Darity, 2002, p. 133)

Even in the near term, if the population being exterminated is located on an 'undeveloped' frontier, it may even be possible for economic activity to proceed in the 'developed' region of a nation undisturbed by the genocide. Indeed, do the genocidists even count the persons who are being eliminated as part of their relevant national population when they are computing the economic consequences of their actions? It is also possible that if the genocidal violence reduces total population sufficiently, in principle, it could offset a decline in national output sufficiently to produce an increase in per capita income. This perverse possibility reinforces the importance of Amartya Sen's (1997) warning that one should not confuse increases in income with improvements in human well-being.

The significance of forced intergroup wealth redistribution for the long history of economic growth and uneven development is the core theme of Eric Williams's (1994 [1944]) now classic study *Capitalism and Slavery*. Acemoglu et al. (2001) and Acemoglu (2003) edge onto this story but back away with a neoinstitutionalist explanation of the divide in the world

between rich and poor nations. In his paper 'Root Causes', Acemoglu (2003, p. 27) proposes that there are two principal explanations of 'the fundamental causes of prosperity between countries . . . geography and institutions'. For Acemoglu (2003, p. 27):

Good institutions [that promote economic development] have three key characteristics: enforcement of property rights for a broad cross section of the society, so that a variety of individuals have incentives to invest and take part in economic life; constraints on the actions of elites, politicians, and other powerful groups, so that these people cannot expropriate the incomes and investments of others or create a highly uneven playing field; and some degree of equal opportunity for broad segments of society, so that individuals can make investments, especially in human capital, and participate in productive activities.

Presumably, one must assume that industrialization in the United States in the midst of the consolidation of a regime of legal segregation was not a sufficient violation of the institutional conditions that Acemoglu lauds to invalidate his hypothesis.

Precisely why he settles on these two as the central explanations is not made clear. In the process Acemoglu eliminates from consideration the Williams perspective that emphasizes the hothouse effects on European economic development produced by the Atlantic slave trade and the slave plantation system in the Americas. Williams's perspective places the stress on the role of colonialism in explaining variations in prosperity in the world economy today. There is at least one additional explanation which, thankfully, Acemoglu does not invoke as an option – variations in national cultures.

When all is said and done, Acemoglu (2003, pp. 27, 29) contends that institutional strength trumps geography in the following key passages:

if you look around the world today you'll see almost no wealthy country achieves this position without institutions protecting the property rights of investors and imposing some control over government and elites.

geography neither condemns a nation to poverty nor guarantees its economic success. If you want to understand why a country is poor today, you have to look at its institutions rather than its geography.

Note that the 'equal opportunity' feature of 'good institutions' in Acemoglu's comment about 'look[ing] around the world today' is notably absent.

But what determined whether a country developed 'good institutions' on two out of three dimensions? For those regions of the world that underwent the process of colonialism, Acemoglu argues that it is the type of colonialism they experienced that set the path for the quality of their institutions. Colonies where Europeans established 'extractive societies'

tend to have a poor institutional framework today, while colonies where Europeans established 'settler societies' tend to have a positive institutional framework. Hence the places where resource extraction took place, but Europeans did not migrate in large numbers, tend to be poor while the places where Europeans relocated tend to be more affluent. Certainly this would be fuel for the Eurocentric cultural determinist, but Acemoglu avoids that trap. Instead he argues that the form that colonialism took shaped the incentives faced by local elites for institutional development that pushed the two types of colonial systems on distinct long-run paths.

From Eric Williams's perspective the distinction between colonies of extraction and colonies of settlement is artificial. In the latter there consistently were forms of extraction that took place – expropriation of the land by the settlers, the exploitation of native and slave labor, as well as other forms of coercion. The division of the world into rich and poor countries went hand in hand with the racialization of the colonial process. The colonizers were enriched whether they remained in their home country or moved to a site of settlement; the colonized were impoverished. These are the parallel elements of uneven development – development for some and underdevelopment for others. Simultaneously, ethnic divisions were crystallized and sustained to make it possible for the 'winners' of the colonial game to continue winning. The fascinating sixth chapter of Williams's (1942) *The Negro in the Caribbean*, entitled 'The Middle Class and the Racial Problem', provides a rich explication of the emergence of the interaction of color and class stratification as a consequence of racialized colonialism.

In another paper Acemoglu et al. (2005, pp. 546–7) highlight the centrality of Atlantic trade in the period between 1500 and 1800 as the foundation for European economic development. They contend that countries best able to take advantage of the growth opportunities afforded by the Atlantic economy were those 'with relatively nonabsolutist initial institutions, most notably in Britain and the Netherlands [i]n contrast [with] countries where the monarchy was highly absolutist, such as Spain and Portugal'. According to Acemoglu et al., non-absolutist states that did not experience rapid growth, like Venice and Genoa, did not have adequate physical access to the Atlantic to gain from the cross-oceanic trade.

This argument should imply that Britain, in particular, should have demonstrated rapid economic growth from the point at which significant constraints were imposed on absolutism, the signing of the Magna Carta in 1215, a full three centuries before the interval that Acemoglu et al. (2005) identify as the period of the 'rise of Europe'. An alternative to the neo-institutionalist account is the argument that the British and the Dutch were simply the winners of the game of mercantilist rivalry (Darity, 1990).

Indeed, many of the pecuniary benefits of the Portuguese and Spanish colonial systems were transferred to Britain via intra-European trade (Darity, 1990). For example, it has been estimated that during the eighteenth century trade surpluses with Portugal brought 50 000 pounds of bullion into London weekly (Birnie, 1935, pp. 175, 180).

Acemoglu et al. (2005) are aware that their analysis touting the importance of the Atlantic trade for European economic development could connect directly with arguments, like Williams's, about the importance of 'the associated profits from colonialism and slavery'. But they immediately seek to sever the connection with the following observation:

It is undoubtedly true that colonial relations with the New World and Asia contributed to European growth. Nevertheless, quantitative analyses, for example, Engerman (1972), Engerman and O'Brien (1991) [sic: O'Brien and Engerman (1991)], O'Brien (1982), and Bairoch (1993, ch. 5), suggest that the volume of trade and the profits generated by the Atlantic trade appear to be too small to account for much of European growth directly. (Acemoglu et al., 2005, p. 562)

Instead, they characterize the institutionally promoted gains from trade for European development as indirect effects: 'the rise in Atlantic trade enriched and strengthened commercial interests outside the royal circle and enabled them to demand and obtain the institutional changes necessary for economic growth (Acemoglu et al., 2005, p. 550). They dismiss the direct effects by invoking what one of the co-authors of this chapter has dubbed the 'small ratios' argument.

They do not appear to be aware that the 'small ratios' argument has been critiqued in a counter set of quantitative analyses to such an extent that the opposite position can be sustained. The volume of trade and profits from the Atlantic economy, particularly the slave trade itself, were enormous by comparative historical standards (Solow, 1985; Bailey, 1986; Darity, 1990; Cuenca Esteban, 1997; Inikori, 2002). Even O'Brien and Engerman (1991) concede that the available trade statistics indicate that colonial trade was of paramount importance for England, at least in the eighteenth century, if not the seventeenth century as well.

Easterly et al. (2006) also subscribe to a variant of neoinstitutionalism, but offer a characterization of 'good institutions' closer to Putnam's notions of social cohesion, and a quite different set of factors as contributors to 'good institutions'. One of the factors that they hypothesize will make for 'bad institutions' is, again, ethnic diversity. However, in the earlier paper on Africa, Easterly and Levine (1997) attribute the extreme ethnic diversity that they claim is present there to colonialism – perhaps providing a basis for rapprochement with Williams. However, they never explore or consider the direct enrichment effects on Europeans of the colonial

process, the task which Williams pursued in depth, from the crucible of the Atlantic slave trade to the mid-twentieth century.

Measuring ethnic diversity or ethnic polarization?

Easterly and Levine (1997) presume that ethnic diversity necessarily has a negative impact on economic growth. In contrast, Alesina and La Ferrara (2005) propose that there are ways in which ethnic diversity might prove beneficial for economic growth. They acknowledge that there are potential costs to diversity including ‘[c]onflict of preferences, racism, and prejudices . . . lead[ing] to policies that are at the same time odious and counterproductive for society as a whole [and] [t]he oppression of minorities may[be] lead[ing] to political unrest or even civil wars’ (Alesina and La Ferrara, 2005, p. 762). But they also say that ‘a diverse ethnic mix also brings about varieties in abilities, experiences and cultures that may be productive and may lead to innovation and creativity’ (Alesina and La Ferrara, 2005, p. 762). Robert Bates (2000) has also argued that ethnic group social capital can promote human capital formation among their own to levels that would not have taken place in the absence of group identification and membership – with potential positive effects on the society as a whole. Paul Collier (2000) has suggested that the effects of ethnic diversity on growth are negative in non-democratic societies and positive in democracies. Ultimately, Alesina and La Ferrara (2005, p. 763) conclude that the impact of diversity on economic performance is an empirical question since, at the theoretical level, it is unclear whether ‘the benefits of “diversity”’ will outweigh ‘the costs of heterogeneity of preferences’.

To address this question, an appropriate measure of ethnic diversity is needed. This is the point where a ferocious debate has raged among social scientists examining the relationship between ethnicity and economic development. The debate is simultaneously about which measure is most appropriate from a conceptual standpoint and which measure has the best technical attributes.

Easterly and Levine (1997) initiated the empirical work in this area by deploying the ethno-linguistic fractionalization (ELF) index. Using the Herfindahl concentration formula, a country’s ELF score represents the likelihood that any two people chosen at random from the population will speak different languages. The likelihood has a minimum value of zero when there is no ethno-linguistic variation in a country and a maximum value of 100 when the odds are perfect that any two persons drawn at random will speak a different language. They find that the ELF index (expressed as a percentage) is positively and statistically significantly related to their measures of unproductive macroeconomic policies.

Still, some peculiarities are immediately evident with this measure; the extent of ethnic tension and violence definitely is not reflected in the index score. Haiti's ELF score, for example, is 1 in Easterly and Levine's (1997) study. This is due to the high degree of linguistic homogeneity, providing no indication of the high level of conflict conducted by the social elite and the military toward the majority of civil society (UNDP, 1994, pp. 41–2). Burundi's ELF score of 4 is attributable to the near universal use of French. The index could not have been used effectively to forecast the long cycle of genocidal violence there. The same is true of Rwanda.

Daniel Posner (2004) criticizes both the design of the ELF index as a measure of ethnic diversity and its use in growth regressions. The ELF index was built on coding decisions made by a Soviet ethnographer in the 1960s. Insofar as ethnic identities undergo some fluctuation in importance and intensity, definitions now a half-century old may be misleading. More significant, the identification of ethnic groups solely on the basis of linguistic differences ignores other critical markers of distinction, such as phenotype, religion, attire and cultural practices. Furthermore, the ELF index does not take into account the spatial distribution of ethnic groups within a country, nor does it incorporate any information that would capture the depth of ethnic division.

Nonetheless, Posner's primary objection to the ELF index is the absence of consideration of political organization and contestation by the relevant racial/ethnic groups in a society. The original Easterly and Levine (1997) hypothesis that greater ethnic diversity leads to slower economic growth rests on an intermediate causal step involving the formation of macroeconomic policies. Greater ethnic diversity is supposed to increase interest group polarization, motivating rent-seeking, overspending and financial repression and subverting the provision of public goods. Posner argues that the ELF index is an inappropriate measure for testing this hypothesis because there is no necessary relationship between the existence of ethnic divisions and the forms of political competition that lead to an institutional environment that inhibits growth. He contends that many ethnic groups do not operate as independent political actors because they lack the political strength either to influence policy directly or to mobilize as part of larger coalitions and parties.

Posner proposes a revised measure of ethnic diversity called the Politically Relevant Ethnic Groups (PREG) index. Using secondary sources, Posner identifies those ethnic groups that have been politically active and have engaged in competition over social policy in each of four decades. He applies the Herfindahl formula to his new measure of ethnic groups and develops index values for 42 African countries in each decade. Although correlated with other comparable indices, the PREG index

generates important differences in the rank ordering of the most- and least-fractionalized countries.

The most common technique for analyzing the economic effects of ethnic diversity, in the aftermath of the Easterly and Levine (1997) study, is to include a measure of ethnic fractionalization as an explanatory variable in a cross-country growth regression. Posner compares the performance of the ELF index with the PREG index in explaining variations in economic growth across the African countries. First, a series of policy measures are regressed separately on the ELF and PREG indices. Surprisingly, although the ELF index has a statistically significant effect on many of the policy variables in the world sample, the effect disappears when the analysis is restricted to the African sub-sample. In contrast, the PREG index has a statistically significant relationship with the black market premium and with fiscal surpluses in the anticipated direction.

Posner then regresses growth rates separately on the ELF and PREG indices. The early Easterly and Levine (1997) results can be replicated with the global sample, but it appears to have no effect on the growth performance of the countries in the all-African subsample. The PREG index does have a statistically significant effect on African growth rates in the expected direction, that is, the higher the PREG value, the lower the country's growth rate.

Finally, Posner regresses rates of growth separately on the ELF and PREG indices with other policy variables included as controls. Peculiarly, the PREG index loses significance in the presence of these additional controls, while the ELF index now has a significant effect, even in the African subsample. Posner's interpretation is that the PREG index successfully captures the indirect effect of ethnic diversity on economic growth via macroeconomic policies but does not have a direct effect on growth. The ELF index, in contrast, has a direct effect on growth independent of other policy practices. Posner speculates that the ELF index may be picking up some other mechanism through which ethnic diversity affects economic development. In the end, Posner concludes that ethnic fractionalization is negatively and strongly related to economic growth in Africa, and offers his results as evidence that macroeconomic policies are an important channel through which ethnicity influences economic development.

James Fearon's (2003, p. 198) construction of an alternative measure to the ELF index attempts to locate ethnic groups based upon 'what people in the country identify as the most socially relevant ethnic groupings'. As Alesina and La Ferrara (2005, p. 792) point out, Fearon's basis for development of his index – Fearon's determination of what people in a given country decide are 'the most socially relevant ethnic groupings' – depends significantly upon Ted Gurr's (1996) Minorities at Risk Project at the University of Maryland.

And, indeed, the criteria that Fearon utilizes predicated upon Gurr's database comes far closer to capturing inter-ethnic tension-cum-violence than does the ELF index. Gurr (1996) defines a 'minority' (not necessarily a numerical minority but any group that is less than 100 percent of the population) at risk as a communal group that: (1) faces political and/or economic discrimination; and (2) acts on its own behalf collectively in the political process. The entire population consists of minorities at risk in Burundi, Chad and South Africa on the Gurr criteria. In a global sample Gurr finds that the African continent has the largest share of population comprised of minorities at risk, a result potentially inconsistent with Bates's observations about the levels of political violence in Africa relative to the levels of ethnic differentiation.

Alesina et al. (2003) show that the Fearon index, based largely upon the Gurr scale, is closely correlated with their version of the ELF index, an extended version that includes ethnic groups defined by other characteristics such as skin color. The 'more comprehensive' version of the ELF index captures salient ethnic differences in Latin America that the language-only index would not capture:

In [Latin America], the language index shows more homogeneity because the language of the former colonizers (Spanish, Portuguese, English) is often spoken by most, but the index based on skin color or ethnic origin (say black, mulattos, white, mestizos, Indian, etc.) shows more heterogeneity. (Alesina and La Ferrara, 2005, p. 792)

Alesina and La Ferrara seem to find the extended ELF and the Fearon indices both to be quite satisfactory, although quite different in design.

Jose Montalvo and Martha Reynal-Querol (2005) have proposed the polarization index as an alternative to the ELF measure that is conceptually quite distinct from the others discussed here. The ELF index presupposes that a country is more fractionalized, the greater its number of separate groups. A country with two similarly sized groups facing each other in a cauldron of hostility would not be depicted as highly fractionalized by the ELF index. The Montalvo and Reynal-Querol index reaches its peak value when a country consists of two equally sized groups, and then declines in value as the number of groups increases, departing from the half-and-half split. Alesina and La Ferrara (2005, p. 793) describe the comparative assessment of the polarization index with the ELF index as follows:

[Montalvo and Reynal-Querol] show that this index is highly correlated with ethno-linguistic fractionalization (ELF) at low levels of ELF, uncorrelated at intermediate levels, and negatively correlated at high levels. In a cross-country

regression analysis, they find that ethnic polarization has a positive impact on the likelihood that a civil war occurs and a negative effect on a country's growth rate. They do not find an independent effect of ethnic fractionalization. Using a different data set, Alesina et al. (2003) compare the results of the polarization index RQ and the fractionalization index ELF, and find that fractionalization works slightly better as a determinant of policies and economic outcomes. While the apparent inconsistency between the two sets of results may be due partly to different parameterization and partly to different data sources, it is between the two measures at low levels of fragmentation.

Tade Okediji (2005) focuses his criticism on the dimensional limitations of the ELF index. He argues that linguistic differences are only one of several possible cleavages that are associated with ethnic division. Racial and religious identities also form the basis of ethnic differentiation, and the ELF index neglects such groupings if they do not coincide with linguistic differences. Moreover, many countries are characterized by complex interactions of racial, religious and linguistic fractionalization, leading to overlapping identities and variation in the salient factors that distinguish one group from another. To counter these limitations of the language-only ELF index, he advances an alternative index that is quite similar conceptually to the more comprehensive ELF index developed by Alesina et al. (2003). But he then performs a comparison with the language-only ELF index that is quite original and has interesting implications.

Okediji proposes the Social Diversity Index (SDI), a measure of fractionalization intended to capture the multidimensional nature of ethnicity – a sort of Human Development Index for ethnicity. Also using secondary sources, he partitions each country into primary racial, religious and linguistic groups. His measure enables him to classify each individual in a country according to all three characteristics, although the person can only belong to one category within each characteristic. When compared with the ELF index, the SDI has a higher mean and a lower standard deviation across the same sample of countries. Okediji concludes that ethnically diverse societies are far more common than the ELF index would suggest, and that the variation in the degree of ethnic fractionalization across countries is much smaller than suggested by Easterly and Levine in 1997. Thus, Okediji returns the discussion to the point of origin of this chapter – the near universal presence of salient ethnic/racial differentiation across the nations of the world and the near universal presence of ethnic/racial inequality.

The research on the relationship between ethnic fractionalization and economic performance using cross-country regressions invariably treats ethnic fractionalization as driving economic performance. But it is quite plausible that the overall economic performance in an economy might

affect ethnic antagonisms positively or negatively. Perhaps conditions of slow economic growth can inflame inter-ethnic tensions while prosperity may relax them? Developments in Malaysia and in Indonesia seem to connect economic crisis to increased communal violence (Darity, 2002, p. 135). The possibility of simultaneous causation has not informed research in this area to any significant degree, although Alesina and La Ferrara (2005, pp. 772–3) are well aware that any number of the ‘independent’ variables used in regressions of this type are subject to the endogeneity complaint.

Similarly, it can be asked whether the general level of inequality in a society drives the level of intergroup inequality, rather than the level of intergroup inequality shaping the general level of inequality (Darity and Deshpande, 2000). There may be no general answer; indeed, the particular answer is probably contingent on the particular structure and history of each country or region.

And what about the fundamental relationship between ethnic/racial conflict and economic inequality between groups? At the very start of his 1985 monograph, *Racial Conflict and Economic Development*, W. Arthur Lewis raised the following question:

Is economic equality necessary for social peace? From one standpoint every so-called racial conflict is sustained (or even initiated) by an economic conflict, covert or open. What poses as a conflict between a dominant and a subordinate group, we are told, is really only a way that an exploiting minority recruits supporters of its case from people with whom it has only racial ties. The proposition is somewhat doubtful. It is probably true that every dispute, racial or not, has, or acquires some economic edge, but this is not the same as saying that all disputes originate in economic conflict. If the economic conflict were mitigated by movement toward equality, would the racial conflict be lessened automatically?

The difficulty is that, far more often than not, effective movement toward economic equality requires disrupting the economically privileged position of the socially dominant group. Its members will resist or act to destabilize the policies that have been adopted to push the society toward greater intergroup equality, whether it is affirmative action, school desegregation or a program of reparations. The very effort to achieve greater intergroup equality will exacerbate racial conflict – a backlash effect – from the group that sees its position of privilege as being threatened. The threat arises precisely because the members of that group have a material benefit from maintaining their identity as a group. To eliminate racial conflict, the benefits associated with racial division would have to be removed. But to remove those benefits typically unleashes racial conflict. In that sense, economic equality is necessary for social peace, but the process of achieving economic equality between ethnic and racial groups is invariably far from socially peaceful.

References

- Acemoglu, Daron (2003), 'Root Causes: A Historical Approach to Assessing the Role of Institutions in Economic Development', *Finance and Development*, June: 26–30.
- Acemoglu, Daron, Simon Johnson and James A. Robinson (2001), 'Colonial Origins of Comparative Development: An Empirical Investigation', *American Economic Review*, 91 (December): 1369–1401.
- Acemoglu, Daron, Simon Johnson and James A. Robinson (2005), 'The Rise of Europe: Atlantic Trade, Institutional Change and Economic Growth', *American Economic Review*, 95 (3), June: 546–79.
- Alesina, Alberto, Arnaud Devleeschauwer, William Easterly, Sergio Kurlaut and Romain Wacziarg (2003), 'Fractionalization', *Journal of Economic Growth*, 8: 155–94.
- Alesina Alberto and Eliana La Ferrara (2005), 'Ethnic Diversity and Economic Performance', *Journal of Economic Literature*, 43 (3): 762–800.
- Bailey, Ronald W. (1986), 'Africa, the Slave Trade, and the Rise of Industrial Capitalism in Europe and the United States: A Historiographic Review', *American History: A Bibliographic Review*, 2: 2–91.
- Bardhan, P. (1997), 'Method in the Madness? A Political-Economy Analysis of Ethnic Conflicts in Less Developed Countries', *World Development*, 25 (9): 1381–98.
- Bates, Robert H. (2000), 'Ethnicity and Development in Africa: A Reappraisal', *American Economic Review*, 90 (2): 131–4.
- Birnie, A.W. (1935), *An Economic History of the British Isles*, London: Methuen.
- Collier, Paul (2000), 'Ethnicity, Politics, Economic Performance', *Economics and Politics*, 12 (3): 225–45.
- Cuenca Esteban, Javier (1997), 'The Rising Share of British Industrial Output, 1700–1851', *Journal of Economic History*, 57 (4): 879–906.
- Darity, William, Jr. (1990), 'British Industry and the West Indies Plantations', *Social Science History*, 14 (1): 117–49.
- Darity, William, Jr. (1998), 'Intergroup Disparity: Economic Theory and Social Science Evidence', *Southern Economic Journal*, 64 (4): 805–26.
- Darity, William, Jr. (2002), 'Racial/Ethnic Disparity and Economic Development', in Paul Davidson and Jan Kregel (eds), *A Post Keynesian Perspective on 21st Century Economic Problems*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 126–36.
- Darity, William, Jr. and Ashwini Deshpande (2000), 'Tracing the Divide: Intergroup Disparity Across Countries', *Eastern Economic Journal*, 26 (1): 75–85.
- Darity, William, Jr., Patrick L. Mason and James B. Stewart (2006), 'The Economics of Identity: The Origin and Persistence of Racial Identity Norms', *Journal of Economic Behavior and Organization*, 60: 283–305.
- Darity, William, Jr. and Jessica Nembhard (2000), 'Racial and Ethnic Inequality: The International Record', *American Economic Review*, 90 (2): 308–11.
- Easterly, William (2001), 'Can Institutions Resolve Ethnic Conflict?' *Economic Development and Cultural Change*, 49 (4): 687–706.
- Easterly, William and Ross Levine (1997), 'Africa's Growth Tragedy: Policies and Ethnic Divisions', *Quarterly Journal of Economics*, 112 (4): 1203–50.
- Easterly, William, Josef Ritzén and Michael Woolcock (2006), 'Social Cohesion, Institutions, and Growth', *Economics and Politics*, 18 (2): 103–20.
- Fearon, James D. (2003), 'Ethnic Structure and Cultural Diversity by Country', *Journal of Economic Growth*, 8 (2): 195–222.
- Gurr, Ted Robert (1996), 'Minorities at Risk Dataset', University of Maryland at College Park.
- Inikori, Joseph (2002), *Africans and the Industrial Revolution in England: A Study of International Trade and Economic Development*, Cambridge: Cambridge University Press.
- Kochhar, Rakesh (2004), 'The Wealth of Hispanic Households: 1996–2002', Pew Hispanic Center Report, Washington, DC: Pew Hispanic Center.
- Lewis, W. Arthur (1985), *Racial Conflict and Economic Development*, Cambridge, MA: Harvard University Press.

- Montalvo, Jose Garcia and Martha Reynal-Querol (2005), 'Ethnic Diversity and Economic Development', *Journal of Development Economics*, 76: 293–323.
- Ndikumana, Leonce (1993), 'Institutional Failure and Ethnic Conflicts in Burundi', *African Studies Review*, 41 (1): 29–47.
- Nkurunziza, Javier P. and Floribert Ngaruko (2002), 'Explaining Growth in Burundi: 1960–2000', CSAE WPS/2002-2003, April.
- O'Brien, P.K. and S.L. Engerman (1991), 'Exports and Growth of the British Economy from the Glorious Revolution to the Peace of Amiens', in Barbara Solow (ed.), *Slavery and the Rise of the Atlantic System*, Cambridge: Cambridge University Press, pp. 177–209.
- Okediji, Tade O. (2005), 'The Dynamics of Ethnic Fragmentation: A Proposal for an Expanded Measurement Index', *American Journal of Economics and Sociology*, 64 (2): 637–50.
- Posner, Daniel N. (2004), 'Measuring Ethnic Fractionalization in Africa', *American Journal of Political Science*, 48 (4): 849–63.
- Putnam, Robert (2000), *Bowling Alone: The Collapse and Renewal of American Community*, New York: Simon & Schuster.
- Rodrik, Dani (2000), 'What Drives Public Employment in Developing Countries', *Review of Development Economics*, 4 (3), October: 229–43.
- Sen, Amartya K. (1997), 'From Income Inequality to Economic Inequality', *Southern Economic Journal*, 64 (2): 384–401.
- Solow, Barbara (1985), 'Caribbean Slavery and British Growth: The Eric Williams Hypothesis', *Journal of Development Economics*, 17, January–February: 99–115.
- UNDP (United Nations Development Programme) (1994), *Human Development Report 1994*, New York: Oxford University Press.
- US Department of State (2007), 'Background Note: Rwanda', Bureau of African Affairs, June.
- Williams, Eric E. (1942), *The Negro in the Caribbean*, Washington, DC: Associates in Negro Folk Education.
- Williams, Eric E. (1994), *Capitalism and Slavery*, Chapel Hill, NC: University of North Carolina Press; first published (1944).

PART VIII

THE STATE, INSTITUTIONS AND DEVELOPMENT

53 The role of the state and markets in development

Louis Putterman

The proper roles of states and markets in fostering and sustaining economic prosperity comprise one of the oldest and most debated topics in the history of economic thought. A focus of Adam Smith's attack on mercantilism in the late eighteenth century, the roles of states and markets were debated by the historical, institutional and early neoclassical schools of economics a century later. More recently, the issue enjoyed a place of prominence in late twentieth century debates over reasons for the economic growth of Japan and other East Asian economies, the most notable success stories of that period.

It should not surprise us, perhaps, to find disagreement over the roles of states and markets in the economy, because the two institutions have histories of both synergy and rivalry (Putterman and Rueschemeyer, 1992) stretching back to ancient times. Some 4000 years ago, populous societies marked by increasingly complex divisions of labor gave birth to the first city states and empires, and while those societies probably saw unprecedented expansions of market activity, their economies were also in some cases state-dominated, and all featured attempts by the new states to enrich themselves and to support larger armies and coteries of officials by controlling sources and flows of wealth. The interplay between traders and officials differed from one society to another, and in given societies across periods of time. While extractive, centralized states may have constrained economic prosperity in some instances, in others the absence of central authority constrained it. For example, the tenuousness of law, order and safe commercial routes in much of Europe following the collapse of Roman rule probably contributed to economic stagnation. It seems noteworthy that modern capitalism and the nation state arose in tandem in Europe after 1500. Since then, no prosperous modern economy has emerged in a society lacking a well-ordered state.

The roles of markets and states

Despite controversy, most economists agree that markets, permitting competition among independently managed enterprises free to select their products and methods of production, have been central to economic

progress and crucial to the growth of productivity, technological know-how and living standards since the Industrial Revolution. Independent enterprises can thrive, these economists would argue, only when free to interact with suppliers and customers in relatively uncontrolled labor, raw material, capital goods and product markets. At the same time, most economists also agree that market systems function poorly if at all without the protection of property rights, rule of law and availability of a stable currency, providing which have been central economic roles of states. Further, economists recognize domains in which competing enterprises cannot be expected to bring about optimal results, most prominently the cases of market power, public goods (including some key trade-facilitating infrastructures), and environmental externalities. Where debate exists is over the scope of the set of public goods (for example, whether they include health care and education), and over the degree to which government remedies (for instance, in the case of monopoly) tend to ameliorate rather than worsen unregulated outcomes. Full agreement is also lacking as to the macroeconomic responsibilities of governments, and whether inequalities in the distribution of income and wealth should be viewed as market failures, also calling for government interventions. A point of particular relevance to this chapter is whether governments can promote economic development by formulating strategies to promote growth, or whether the ideal role of government is simply to create a stable institutional environment and then allow the market to 'work its magic'.

Although properly speaking markets are the loci of interactions between buyers and sellers, sometimes although not always associated with specific institutional frameworks or locations, economists use the phrase 'the market' to refer not only to the settings or sets of those interactions but also to the decentralized economic mechanism as a whole, including the presence of autonomous enterprises that purchase inputs and transform them into goods and services. A 'market economy' (sometimes called 'the market') in this broader sense is said to allocate scarce resources efficiently among competing needs and wants, and to stimulate technological progress, for several reasons. First, the market mechanism causes both final consumers and intermediate users of goods, services and resources to signal the values they attach to those goods, and so on, as a by-product of the exchange of offers to buy and sell. Second, the pressure that enterprise owners feel to maximize net returns, lest they be driven from business or at least suffer serious financial losses, induces them to attempt to produce goods of maximum value to consumers using the least-cost combinations and quantities of resources, while also striving to satisfy the needs of buyers in terms of quantity, quality and variety. Third, individual workers, including prospective managers and specialists of other kinds, are motivated to

invest in demanded skills due to the higher earnings which market competition assigns to those with scarce capabilities. Fourth, owners of non-labor resources can be expected to steer those inputs toward the uses most valued by society, responding to price signals which indicate, for example, the value of a parcel of land as an orchard versus its value as a grain field or a parking lot. Finally, the rents that accrue to innovation, and the fear of falling behind competitors, are viewed as major causes of the high rates of technological change observed in market capitalist economies.

But markets do not exist in social and political vacuums. Market interactions are embedded in social systems (Granovetter, 1985) and many trades would be impossible without social norms that increase trust and facilitate contract enforcement (Putnam, 1993; Greif, 1994). Although state power and the formal legal systems to which states lend their force may offer protections of final recourse, market interactions rely more directly and extensively on widespread adherence to such norms by ordinary members of society. Recent research suggests considerable cross-country variability in trust and social capital, and this variability correlates with differences in levels of economic development and growth (Fukuyama, 1995; Knack and Keefer, 1997). It can be argued that the existence of a stable political order, honest administration and a non-corrupt judiciary may be important facilitators of social trust and norm abidance (although the relationship probably goes both ways). State actions that contain the extremes of inequality, poverty and neglect of worker health and safety may, while intervening in the full expression of the logic of competition, end up helping the market by helping to stabilize the polity and society and to increase the social acceptability of leaving most economic coordination to the market.

Lessons from planned economies

While economists' convictions about the efficacy of markets derive in part from classical and neoclassical theory as well as from observation of market economies, many draw lessons also from experience with alternative economic systems and policies. The premier experiment in operating large modern economies with almost no state-tolerated role for markets is the one that began in Russia in 1928 and lasted into the 1980s there and in numerous other countries eventually ruled by communist parties. In those centrally planned economies, prices were set administratively and the allocation of resources and determination of production plans was to take place not under the influence of market forces, but rather under the aegis of a planning bureaucracy directly weighing leaders' political goals and perceptions of societal needs. Planners ignored notions of comparative advantage and were guided instead by the goal of building industrial

economies, which they hoped to achieve by directing huge investments into the capital goods sectors, ordering low-price crop deliveries from farmers, and exploiting natural resources without regard for opportunity cost and environmental impact. Planners paid only limited attention to consumer goods production, and even less to the provision of services other than health care and education.

According to observers (for example, Nove, 1983; Kornai, 1992), planners in the Soviet-type economies struggled with the problem of providing effective incentives to, and the need to elicit information from, enterprise managers. Quality, variety, spare parts and maintenance were perennial problems. Enterprises integrated vertically to avoid relying on the planning bureaucracy for the inputs they needed. Considerable amounts of resources were diverted into black market activities. Innovation proved difficult to engender at levels comparable to those of industrialized market economies, except perhaps in the military sector. Considerable activity took place outside of the approved plans, with some observers going so far as to argue that the claim that such economies were primarily plan-based is inaccurate.

For a time, rapid structural change and achievements in health and education sectors made the model attractive to some outsiders, especially in the developing world. The Soviet Union's industrial output growth rate exceeded that of the United States during the 1930s and again from World War II until the late 1970s. Most estimates suggest that China achieved a higher rate of industrial growth than India from the late 1940s to the late 1970s, despite ideological excesses and political upheavals. Life expectancy in China exceeded that in India by some 12 years in 1978, and in general planned economies achieved higher life expectancy, lower infant mortality and higher literacy rates than non-communist countries at similar income levels. However, the curtailment of individual freedoms and comparisons with incomes and consumer good availability in neighboring countries like West Germany and Taiwan bred dissatisfaction with the system, ultimately leading to the system's demise in both Europe and Asia. In addition, the planned economies' growth was unbalanced, and much of the capacity put in place by their economic system had little value when the countries in question adopted market-oriented reforms and became more open to international trade.

State roles elsewhere

Less extreme in their departures from free market principles are the numerous cases in which market pricing and exchange were permitted but with key government interventions 'distorting' the price system. Most developing countries of the late twentieth century controlled foreign exchange transactions, usually overvaluing their country's currency. By doing so,

they inadvertently discouraged exporting and necessitated combinations of import restrictions, tariffs and borrowing to deal with trade imbalances. Interest rates on bank loans were often subject to regulatory caps, and import licensing and tariff arrangements made some capital goods less expensive, creating artificially low prices for certain producer goods while the cost of capital to small-scale borrowers, including farmers, remained high. These and other interventions, described by some economists as 'getting prices wrong', led to limited growth of bank deposits and other forms of financial mediation ('shallow finance'), excessive capital intensity and limited job creation in a few modern sector activities, and capital starvation and underemployment in other parts of the economy. By discouraging exports and domestic savings while encouraging imports, they also contributed to the growth of unsustainable burdens of debt at the national level.

Although these examples convince many of the virtue of markets, pure free market economies are textbook abstractions to which no modern national economy adheres in reality. After the 1930s, the ideal of a fully self-regulating economy was abandoned by most economists and politicians in industrialized countries, with macroeconomic stabilization being considered a responsibility of states. Other government roles were also growing. The proportion of national income used by governments to pay civil servants and to support various functions grew steadily until close to the end of the twentieth century. Governments were asked to respond to market failures, for example to set and police environmental standards, and to monitor the safety of workplaces, foods and pharmaceuticals. Governments invested in roads, bridges, maintenance of waterways and harbors, and rail lines. Due to some combination of market failure and distributive concerns, governments also provided unemployment benefits, food and health subsidies, pensions and other social benefits. The notion that the market and autonomous firms could be engines of production and technological progress, but that the distribution of social benefits could be partly separated from that of market rewards, was mainstream in the politics if not in the economics professions of the world's most prosperous and technologically advanced nations after World War II.

At the same time as the role of the state was growing in industrialized mixed economies, economists were beginning to re-examine their depiction of government as a benevolent agent that could be counted upon to follow the prescriptions of normative economic theory regarding the correcting of market failures. Mainstream political economy viewed governments as being composed of individuals who might promote the well-being of citizens if imbued with social concerns or held accountable by an engaged public, but who might also be as motivated by self-interest as other

individuals. If monitoring by the public is costly and if formally democratic political institutions can be captured by groups with concentrated interests in particular areas, state officials and politicians might be poor servants of the public as a whole. Grievous policy errors could also result from simple misunderstanding of the effects of instruments like exchange controls and interest rate ceilings. The idea of ‘government failure’ entered the lexicon of economics, alongside the term ‘market failure’, and the possibility was raised that even when there exist imaginable state interventions that can increase social welfare if effected, actual government involvement might worsen rather than ameliorate some market failures. Efforts to help the poor might also have effects at odds with that aim in the long run if they resulted in reduced incentives to invest in human and physical capital and thus lower rates of growth.

The special problems of less-developed countries

Some arguably distinct aspects of the state–market relationship in those countries lagging far behind the most industrialized economies have been the subject of separate discussions at various points in time. In the nineteenth century, development strategies were proposed and to some degree adopted by national governments in then-lagging countries including the United States, Germany, Japan and Russia, usually including tariff protection against a range of importable manufactured goods and government assistance or active participation in the accumulation of investment funds. The so-called Great Industrialization Debate in the Bolshevik-ruled Russia of the 1920s would be re-examined by the advisors to leftist Third World governments like those of Mozambique and Tanzania as late as the 1970s. The post-World War II era in which the Bretton Woods institutions were formed to help manage the economic problems of less-developed and especially newly decolonized countries saw the growth of a new literature of development economics.

Early post-World War II writers such as Ragnar Nurkse (1953), W. Arthur Lewis (1954) and W.W. Rostow (1960) argued that the central problem of a developing economy was to raise the share of capital formation in national product to a level sufficient to fuel the growth of modern-sector activities and, in the formulation of Lewis and of Fei and Ranis (1964), to ‘drain off’ the pool of surplus labor underemployed in the traditional, mainly agricultural, sector by absorbing it into modern employment. As a result of such thinking, government development plans identifying the gap between domestic savings and investment targets were formulated and used as bases for seeking investment financing from international financial institutions and foreign governments. During the 1960s and 1970s, there was still considerable tolerance among Western development specialists and

advisors for using inflation as a tax to finance government investment, using tariffs to protect domestic industries deemed promising, and adopting other measures that would even then have been considered inadvisable in a developed-country context.

Governments were also thought to have a role to play as coordinators of the overall push for development. Although enterprises were mainly privately owned and prices determined by supply and demand, it was argued by some that national economic planning could still play a crucial coordinating role. The theory espoused by advocates of planning in post-World War II France, for example, was that firms might be reluctant to invest if they could not be confident that complementary investments were being made by others up and down the relevant production streams. Complementary investments in infrastructure and manpower training might also be called for. The market might be a suitable coordinator of short-term production decisions, the theory went, but the large-scale investment decisions needed to bring about true structural change might be too lumpy to overcome the hurdles of uncertainty in an unplanned economy. The argument was stretched by some to the point of suggesting that government itself had to do the investing, even in industrial enterprises and mines, to overcome private sector hesitation. But for the most part, it was used in favor of a government coordinating and facilitating role, not state ownership. The role of governments in Japan and later Korea were often understood in this light, but so too were the national plans typical in many other developing countries.

The 'international division of labor' had a central place in many discussions. Free trade among nations, it was argued, was more beneficial to rich than to poor countries. Europe's poor former colonies in Asia, Africa and Latin America had been brought into the world economy for the benefit of their colonizers as sources of cheap raw materials and foodstuffs and as markets for European and North American manufactured goods. It was in the 'core' or 'metropolitan' countries' interests that the 'peripheral' countries' comparative advantage remain one based on unskilled labor and raw materials. 'Free trade' would perpetuate this because manufactured goods would remain less expensive for poor countries to import than to produce, so they would have no chance to learn by doing and to move towards the international frontier of industrial capability.

In response to such concerns, economic moderates called for the leveling of the playing field by reducing discrimination against developing-country agricultural exports, devising mechanisms to stabilize and maintain the prices of tropical commodities like sugar cane and coffee, and fully opening developed-country markets to Third World manufactures – a 'New International Economic Order'. Believing that positive steps were required to foster structural change away from the old reliance on primary product

exports, many also believed that selected industries in poor countries should be protected from foreign competition by tariffs of sufficient magnitude to let domestic manufacturers obtain a foothold – the groundwork of the import substitution industrialization approach. More radical commentators called for de-linking developing economies from the world trading system, substituting (especially for smaller economies) links with neighboring countries and/or with Communist states (for a discussion, see Diaz-Alejandro, 1978). In either case, national strategies, which only governments could put forth and implement, were seen as requirements for escaping the self-perpetuating status of underdevelopment. Thus, a key role was assigned to the state; leaving things to market forces would only perpetuate underdevelopment and dependency.

More recent discussion

The 1980s were a watershed decade for policy and professional opinion on the problems of economic development. The decade was marked by slowdowns of growth in the industrialized market economies, a still more pronounced slowdown in productivity growth in Communist countries, the accumulation of unsustainable debt levels by many middle- and low-income developing countries, and increased international recognition of the remarkable growth achievements of a number of East Asian economies. These developments helped to fuel a conservative backlash against the ‘welfare state’ and state-owned industries in the West; the beginnings of radical economic reform in China and a last decade of reform experimentation in the Soviet bloc; initiation of structural adjustment programs in developing countries in the wake of their debt crises; and the fall from respectability of import substitution industrialization as a policy approach. Moved partly by the necessity of accepting International Monetary Fund (IMF) conditions for urgently needed loan programs, partly by the perception of the relative merit of East Asia’s more outward-focussed orientation, most developing-country governments devalued their currencies, reduced spending, began reversing the trend of nationalization, and attempted to make their economies attractive to foreign investors. Going into the 1990s and the start of the twenty-first century, increasing flows of foreign direct investment and international bank lending, growth of trade volumes, and vigorous participation in world trade by China, India and other developing economies, became hallmarks of the intensification of international trade, investment, knowledge and cultural flows that was dubbed ‘globalization’.

Even though policies were trending in this period towards liberalization in comparison with prior decades, it would be incorrect to describe the approaches of most developing-country governments as *laissez-faire*. Nor would it be accurate to suggest that the prevailing view of the state’s role in

development had become the minimalist one favoring protection of property rights and of a stable currency as the only valid economic functions of government. The 1997 *World Development Report* of the World Bank stated that: '[a]n effective state is vital for the provision of the goods and services – and the rules and institutions – that allow markets to flourish and people to lead healthier, happier lives. Without it, sustainable development, both economic and social, is impossible.' Government investments in the health and education sectors were viewed as important both to immediate well-being and to economic growth. The role of governance, especially the rule of law and the absence of corruption, was accorded considerable importance in World Bank publications and was the focus of a number of studies including Kaufmann et al., 2000.

China's quarter-century growth spurt after 1979 occurred in an economy in which, until the mid-1990s, most industrial enterprises were owned by local or higher levels of governments. More importantly, even the China of the early 2000s, when the private sector had come to play a more important role, resembled more the Japan of the 1950s and 1960s with its government-guided industrial policy, including special subsidies and incentives to sectors accorded important roles, than it did *laissez-faire*. Until its admission to the World Trade Organization (WTO) in 2001, China maintained a dual trade regime, with substantial tariff protection of most domestic industry but also a more liberal regime for imported inputs applying only to export-oriented industries originally restricted to special economic zones and coastal 'open cities' (Naughton, 2007). These distinctions were only gradually dismantled during the early WTO years.

Opinion regarding the role of the state was also strongly influenced by the experience of countries in the former Soviet bloc transitioning from state socialism to market capitalism. Whereas initially the emphasis of economic advisors was on ending governments' involvements in the economies of the countries concerned, it was soon recognized that healthy market economies could not exist without strong institutional supports, including legal protection of property rights, low tolerance for corruption and adequate monitoring of financial market institutions. Furthermore, the advantages exhibited by China since 1978, a little earlier by Taiwan, Korea and Singapore, and more recently by India, were seen to be in part the result of substantial public investments in education, transportation, communications and health. Comparative political stability has also been a major asset of these countries. Although the relatively closed nature of the Chinese and Indian economies before the 1980s has often been viewed as an error that retarded development, it is difficult to prove that those initial closed periods were not of some benefit to their economies, creating a breathing space during which indigenous capacities and skills could be incubated.

Efficient states, old states

With or without agreement on what governments need to do to facilitate development, there is evidence of a general correlation between more efficient and capable government and better economic outcomes. A number of studies have found a correlation between measures of government quality and rates of economic growth. Mauro (1995) found that countries with more corrupt governments had lower rates of investment and economic growth. Evans and Rauch (1999) found that developing countries whose government administrative structures exhibited more of the classic bureaucratic features of meritocratic recruitment and predictable long-term careers achieved higher rates of economic growth even after controlling for initial GDP and human capital. Kaufmann et al. (2000) found several governance measures to be correlated with rates of economic growth.

The fact that countries with relatively capable and stable governments have better growth records than others does not prove that good government is an independent cause of economic growth. The same propitious factors may give rise both to economic growth and to a well-ordered state. Interestingly, one of those factors may be a long history of large-scale political organization. Recent studies suggest that regions that saw earlier formation of kingdoms, states or empires, especially ones not subsequently disrupted by large population shifts due to colonization, enjoyed faster growth between 1960 and 2000 (Bockstette et al., 2002; Chanda and Putterman, 2007). The countries hosting old states also tend to have better scores on commonly used measures of institutional quality. Old states are associated with early transitions to agriculture, and one study suggests that the time of agricultural transition is a strong predictor of current level of development (Hibbs and Olsson, 2004).

Conclusion

Although most economists agree that private enterprises disciplined by market competition are more efficient providers of most goods and services than are government-owned enterprises, there is also general agreement that neither a prosperous national economy nor a sustained process of economic growth are achievable in the absence of a well-functioning state. States are needed to secure property rights, manage currencies and provide the civil order without which commerce and investment become excessively risky propositions. States help to create competitive markets by regulating and breaking up monopolies and promulgating rules for the operation of banking systems and financial markets. States help to address market failures in such areas as environmental quality and workplace safety. And states can pursue macroeconomic policies that contain fluctuations in price and employment levels.

More controversial is whether state guidance or strategic planning is either necessary or desirable for a country to be launched on a path of economic growth. Some governments, especially in Asia, appear to have enjoyed success with government activism in this respect, but there is as yet no consensus about these cases, and misguided interventions can be harmful. Building state capacity while focusing on education, health, transportation, and a legal and tax environment conducive to investment may be tall enough orders for some governments. These tasks, in any case, constitute a starting point that all governments should strive for, and that the people of every country should actively demand of their governments.

References

- Bockstette, Valerie, Areendam Chanda and Louis Putterman (2002), 'States and Markets: The Advantage of an Early Start', *Journal of Economic Growth*, 7: 347–69.
- Chanda, Areendam and Louis Putterman (2007), 'Early Starts, Reversals and Catch-up in the Process of Economic Development', *Scandinavian Journal of Economics*, 109 (2): 387–413.
- Diaz-Alejandro, Carlos (1978), 'De-linking North and South: Unshackled or Unhinged?' in Albert Fishlow (ed.), *Rich and Poor Nations in the World Economy*, New York: McGraw Hill, pp. 85–162.
- Evans, Peter and James Rauch (1999), 'Bureaucracy and Growth: A Cross-National Analysis of the Effects of "Weberian" State Structures on Economic Growth', *American Sociological Review*, 64 (5): 748–65.
- Fei, John C.H. and Gustav Ranis (1964), *Development of the Labor Surplus Economy: Theory and Policy*, Homewood, IL: Irwin.
- Fukuyama, Francis (1995), *Trust: the Social Virtues and the Creation of Prosperity*, New York: Free Press.
- Granovetter, Mark (1985), 'Economic Action and Social Structure: The Problem of Embeddedness', *American Journal of Sociology*, 91 (3): 481–510.
- Greif, Avner (1994), 'Cultural Beliefs and the Organization of Society: A Historical and Theoretical Reflection on Collectivist and Individualist Societies', *Journal of Political Economy*, 102 (5): 912–50.
- Hibbs, Douglas A. and Ola Olsson (2004), 'Geography, biogeography, and why some countries are rich and others are poor', *Proceedings of the National Academy of Sciences*, 101: 3715–20.
- Kaufmann, Daniel, Aart Kraay and Pablo Zoido-Lobaton (2000), 'Governance Matters: From Measurement to Action', *Finance and Development*, 37 (2): 10–13.
- Knack, Stephen and Philip Keefer (1997), 'Does Social Capital have an Economic Payoff? A Cross-Country Investigation', *Quarterly Journal of Economics*, 112: 1251–88.
- Kornai, Janos (1992), *The Socialist System: the Political Economy of Communism*, Princeton, NJ: Princeton University Press.
- Lewis, W. Arthur (1954), 'Economic Development with Unlimited Supplies of Labors', *Manchester School of Economic and Social Studies*, 22: 139–91.
- Mauro, Paulo (1995), 'Corruption and Growth', *Quarterly Journal of Economics*, 110 (3): 681–712.
- Naughton, Barry (2007), *The Chinese Economy: From Transition to Development*, Cambridge, MA: MIT Press.
- Nove, Alec (1983), *The Economics of Feasible Socialism*, London: Allen & Unwin.
- Nurkse, Ragnar (1953), *Problems of Capital-Formation in Underdeveloped Countries*, New York: Oxford University Press.

- Putnam, Robert D. (1993), *Making Democracy Work: Civic Traditions in Modern Italy*, Princeton, NJ: Princeton University Press.
- Putterman, Louis and Dietrich Rueschemeyer (eds) (1992), *State and Market in Development: Synergy or Rivalry?* Boulder, CO: Lynne Rienner.
- Rostow, W.W. (1960), *The Stages of Economic Growth: A Non-Communist Manifesto*, Cambridge: Cambridge University Press.
- World Bank (1997), *World Development Report 1997: The State in a Changing World*, New York: Oxford University Press.

54 Monetary policy

*Paul R. Masson*¹

Any survey of the issues confronting monetary policy in developing countries must first address several basic questions. First, should the central bank target development as one of its objectives, or more broadly, can it contribute to development indirectly, through for instance maintaining financial stability, ensuring international competitiveness or delivering low inflation? Second, is there any scope for, and value to, monetary independence, either because monetary policy is under the thumb of fiscal policy, or because the country concerned is small and open, or because a foreign currency circulates domestically and currency substitution is rampant? Finally, assuming that there is some scope for an independent monetary policy, what should be the operational guide for policy-setting, and is it likely to be different depending on the degree of financial development or other structural features of the economy?

This chapter will be mainly about the third question, namely the way monetary policy should be set for the short to medium run, and the institutions and policy regimes that support and implement that policy, while recognizing that developing countries are very diverse and that 'one size does not fit all'. In discussing the issue, the position will be taken that monetary policy cannot be separated from exchange rate policy – the two have to be considered together. This is not to say that there may not be instruments that under some conditions have differential effects on domestic monetary conditions and the exchange rate (open-market operations and sterilized intervention, for instance). However, for many countries – in particular the 'emerging economies' with access to world capital markets and few controls on capital flows – the scope for differential effects is small (for instance, because of constraints on the cost and effectiveness of sterilized intervention). Over time, the extent of capital controls has declined, as their effectiveness has been blunted and their distorting effects on economic decisions have become manifest. Thus, countries are ill advised to use monetary policy to target domestic objectives while targeting a fixed value for the nominal or real exchange rate.

As for the first question, the position taken here is that development should not be an explicit goal of monetary policy. Development is a longer-term, structural issue relating (primarily) to the real economy while monetary neutrality precludes long-run effects of the money supply on the level

of real output. Forty years ago, this position would not command a consensus; indeed many then advocated deliberate use of monetary policy to keep interest rates low (below world real rates), to channel credit to particular sectors, to undervalue the real exchange rate in order to stimulate exports, or to redistribute income (via inflation) to those with higher saving rates. However, a policy of deliberately keeping interest rates low to stimulate investment is likely in all but financially repressed economies to produce accelerating inflation rather than higher growth, and inflation quickly gets anticipated, blunting any favorable real effects. Following McKinnon (1973) and Shaw (1973), the pervasive inefficiency of financially repressed economies has been recognized. Long-run monetary neutrality does not preclude monetary policy from contributing to short-run stabilization, or the long-run level of inflation from having real effects, but the time horizons for the effects on development are so long and complicated that they make targeting development an inadequate guide for monetary policy setting.² This leaves open which of the various possible intermediate targets for monetary policy (the exchange rate, the rate of inflation, nominal income or external competitiveness, among others) is most appropriate and will most contribute to development and welfare. This is issue three above.

Turning to the second question, for a discussion of monetary policy to be interesting, it must be assumed that there is some scope for monetary independence. At the same time, it will be noted below that in some developing economies the limits on that independence are tight indeed, and this may influence the choice of operational targets for monetary policy. In that respect, the size of countries and their structural characteristics matter in considering their appropriate monetary policy. Thus, monetary policy questions for developing countries are somewhat distinct from those for developed countries, for which there is an extensive literature. There are reasons to expect that countries with higher per capita gross domestic product (GDP), more developed financial sectors and stronger institutions face different policy trade-offs. In what follows we shall pay considerable attention to a policy regime – inflation targeting (IT) – that is increasingly finding favor among both industrial and emerging market economies, while considering whether the two sets of countries differ with regard to IT's suitability as a monetary policy regime and its implementation.

Interdependence of monetary and exchange rate policies

The tight linkage between monetary and exchange rate policy is most clearly expressed in the doctrine of the 'impossible trinity': a country cannot at the same time maintain a fixed exchange rate and an independent monetary policy in a context of perfect capital mobility. Put another way,

a credible peg would not allow a country to run different interest rates from those in the anchor country, provided arbitrage was free to operate and unlimited in quantity. In practice, of course, capital is not perfectly mobile, leaving some scope for monetary policy independence even with an exchange rate target. Moreover, as Frankel (1999) has persuasively argued, countries can trade off some exchange rate fixity for some monetary independence, consistent with adopting an intermediate exchange rate regime.

The constraints on countries operating a pegged exchange rate have increased as capital has become more mobile. Increased capital mobility has occurred for essentially two sets of reasons. Increasing financial sophistication, technological advances and increased financial wealth have vastly increased the volume of capital flows and hence the resources that can be marshaled to speculate against a currency peg. And a growing consensus over the last three or four decades that liberalization increases economic efficiency (subject to the cautions expressed by Diaz-Alejandro, (1985), which were confirmed by the emerging market crises of the 1990s) has meant that few countries attempt to maintain tight government regulation of capital flows.

In contrast, during the early post-World War II Bretton Woods period of fixed but adjustable exchange rates there were pervasive controls on capital flows as well as extensive domestic financial regulation, or even 'financial repression'. In this context, monetary conditions could be set with little concern for external consequences and interest rates were kept low to stimulate investment. Some envisaged the deliberate use of inflation to raise the relative price of capital goods, lower real wages, shift income to those with higher saving propensities (that is, the rich), and call forth increased output for development. For instance, a respected text on development, Higgins (1959), discusses this strategy, advocated by Martin Bronfenbrenner; Higgins accepts the argument in principle, but argues that the optimal rate of inflation is likely to be less than 5 percent per year. Curiously, the incompatibility with the prevailing system of fixed exchange rates is not mentioned, nor does 'exchange rate' appear at all in the index of this comprehensive text. At most, there is a reference to the fact that inflation would make export industries increasingly unprofitable, aggravating balance-of-payments difficulties (Higgins, 1959, p. 464). The context was one where constraints on monetary policy came not from the fledgling capital account but rather through the competitiveness of the trade account.

Targets for monetary policy

Corden (2002) has provided a useful taxonomy of what he calls 'approaches' for exchange rate policy, which for the reasons above we will

adapt for our discussion of monetary policy: the 'real targets' approach, the 'nominal anchor' approach, and the 'exchange rate stability' approach. He distinguishes these approaches from what he considers to be the three (not two) polar regimes: an absolutely fixed exchange rate, pure floating, and the fixed but adjustable exchange rate regime (FBAR), which was the regime under the Bretton Woods period. Thus, his classification of regimes neatly sidesteps the bipolar classification of proponents of a 'hollowing-out' of intermediate regimes. But Corden maintains that the FBAR is not simply a compromise between the other two regimes. Because of its credibility problems – it involves an explicit or implicit commitment to a peg, without a corresponding assignment of the instruments needed to ensure maintenance of the peg – it differs from other intermediate regimes which do not promise so much, such as managed floating or crawling pegs or bands.

The 'real targets' approach presupposes nominal wage rigidity, so that in the short run, if not in the long run, expansionary monetary policy (or nominal exchange rate depreciation) can affect real output, employment and the real exchange rate. The value of using monetary policy in this fashion is greatest when the economy faces negative real shocks and fiscal policy is prevented from operating in a stabilizing way. However, the assignment of monetary policy to real targets suffers the disadvantage that it does not provide an anchor for the price level or the rate of inflation. Moreover, it relies on a degree of money illusion that is endogenous, and is likely to shrink drastically if monetary policy systematically tries to exploit it.

Hence the increasing emphasis, among central banks of the world, on 'nominal anchors' for monetary policy – a rigid link or a target for a nominal quantity or price which is intended to prevent the economy's overall price level or rate of inflation from wandering off. Nominal anchors can be divided into domestic variables (a monetary aggregate, nominal GDP or the rate of inflation itself), and external anchors, in particular a peg to another currency or to a world commodity (such as gold). In principle, either a domestic or foreign nominal variable can anchor the price level and produce long-run nominal stability, but different targets can yield different short-run outcomes for output and inflation. They also differ in terms of institutional requirements and expose the economy to a crisis in varying degrees.

The 'exchange rate stability' approach is to be distinguished from the desire to put in place a nominal anchor: it postulates that exchange rates left to themselves simply add noise to the world economy, perhaps because of self-fulfilling expectations and destabilizing speculation (Williamson, 2000). Though in some cases exchange rate flexibility could facilitate adjustment, movements in exchange rates are dominated by short-run volatility unrelated to economic fundamentals and by medium-term misalignments. In this

view, a system of credibly fixed rates would clearly improve welfare compared to exchange rate flexibility. Williamson has argued that an intermediate regime with an explicit exchange rate target (for example, a band-basket-crawl – or BBC – regime) would have some of the same advantages of anchoring expectations and taming volatility.

The exchange rate as a nominal anchor

Exchange rate-based stabilization (ERBS) has been used, with varying degrees of success, in reducing a high initial rate of inflation. The advantage of the exchange rate as nominal anchor is that it is visible, easily explained to the public and requires little institutional credibility – hence its attraction for countries suffering from high or hyperinflation, as Chile in 1979, Argentina in 1991 and Brazil in 1994. In these and other similar cases, the only way to achieve a modicum of monetary policy credibility is to tie the hands of the central bank, since unlike the inflation rate or a monetary aggregate, the exchange rate is a variable which the public observes directly and continuously. However, the principal drawback of this strategy is that it requires an eventual exit from the peg unless the economy is to undergo a severe deflation to remove loss of competitiveness resulting from the accumulated inflation. Indeed, because of the stickiness of inflation, bringing it down to industrial-country levels in the space of even a few years still leaves embodied in the price level the integral of the inflation gaps incurred during that time. Eliminating them would require an extended period of deflation, involving output losses that few governments would willingly incur. Therefore, pegs associated with ERBS become increasingly non-credible over time, even in countries which are successful in achieving low inflation.

The problem is compounded in the presence of sufficient capital mobility that investors can take positions against the currency large enough to exhaust the authorities' foreign exchange reserves. Thus, a speculative attack could force the authorities to devalue or float. Paradoxically, then, the attempt to gain credibility by using an external anchor sows the seeds of its own downfall. The trick is to ride the wave long enough to benefit from initial credibility gains without getting locked into a strategy that will eventually throw you up on the beach. Unfortunately, while the strategy is successful, there is little pressure on politicians to change, while when the peg is under attack, it is too late: exits in a crisis usually have dire consequences (Eichengreen et al., 1999).

There is a category of countries, however, for which monetary policy independence is of little use and hence a hard peg is credible. These are small, open countries with a high export concentration on a single commodity or service priced in an international currency, or a dominant trade

partner. In particular, countries in the Caribbean which rely heavily on tourism and banking services have long-standing and credible pegs to the US dollar. Other countries in this category are the small neighbors of South Africa – Lesotho, Namibia and Swaziland – which participate in the Common Monetary Area. This arrangement allows the countries to have their own currencies, exchangeable at par with the rand. Nepal is an example in Asia; that country pegs its currency to the Indian rupee. Finally, some countries – among which are Ecuador and Panama – have simply adopted a foreign currency, that is, have a regime of official dollarization.

It is relevant to examine the empirical determinants of changes in regimes. In Masson (2001) and Masson and Ruge-Murcia (2005), exchange rate regimes are divided into fixed, intermediate and flexible. Using data from 1975–97 for as many as 168 countries, the probability of changing regimes was related to macroeconomic determinants, foreign exchange reserves divided by GDP, and trade openness. The probability of abandoning any of the regimes was greater, the higher the rate of inflation and the lower real GDP growth (Masson and Ruge-Murcia, 2005). The intuition is clear: regimes are abandoned in bad economic times, not good ones. For instance, high inflation makes a peg increasingly precarious, but also makes it more likely that countries that are floating will use ERBS in order to reduce it.

Analysis of exchange rate regime transitions also permits testing formally the ‘hollowing-out’ hypothesis (Eichengreen, 1994). Hollowing-out requires transitions away from intermediate regimes, but not towards them from the poles of hard fixes and free floats. Using a constant transition matrix, that hypothesis can be rejected (Masson, 2001). Moreover, the existence of continuing negative shocks that produce high inflation and slow growth suggests that there will be continued cycling among regimes when the transition probabilities are endogenous, as described above. The idea that hard fixes such as currency boards are immune from crisis was decisively proved wrong by Argentina’s abandonment of its Convertibility Law in January 2002, and the floating of the peso. Despite having a credible ERBS based on institutional guarantees, the severe recession suffered by Argentina during 1998–2001 made maintaining the exchange rate strait-jacket difficult. Argentina also illustrates the need for support from fiscal policy to make any monetary regime successful – but especially so for a fixed rate regime. Earlier fiscal adjustment would have allowed Argentina to avoid a debt crisis and would have helped maintain international competitiveness and current account balance.

Domestic nominal anchors

The principal choice of domestic nominal anchor is between a monetary aggregate and inflation targeting. Monetary aggregates have the advantage

of being relatively easy to measure and of being a financial variable that, at least in the era of highly regulated financial systems, was relatively easy to control.³ It was postulated that there was a stable relationship between those financial and real variables, principally taking the form of a stable (and simple) money demand equation.

However, there is accumulating evidence that money demand is unstable. Greater access to other financial assets associated with liberalization has changed its nature, and probably increased the interest elasticity of demand for non-interest-bearing deposits and cash. Technological changes have also allowed greater opportunities for conserving on transactions balances. Finally, partial dollarization – the circulation of a foreign currency – provides another source of instability in the demand for the domestic currency. As a result of all these factors, targets for monetary aggregates, abandoned by almost all industrial countries, are now also increasingly being abandoned by developing countries.

Emerging-market countries have in a number of cases adopted inflation targeting. Chile was the precursor, announcing a target for consumer price index (CPI) inflation in 1991, albeit accompanied initially by an exchange rate target band. Other developing countries having forms of inflation targeting regimes include Brazil, Colombia, the Czech Republic, Israel, Korea, Mexico, Peru, Poland, South Africa and Thailand (see Mishkin and Schmidt-Hebbel, 2002, for details).

IT in developing countries: prerequisites and experience

Inflation targeting as a monetary policy regime has to be distinguished from a situation in which the central bank merely expresses a desire to lower the rate of inflation to a particular level or maintain it there. Price stability in some form is always part of a central bank's mandate, but IT aims to enhance the credibility of the central bank's commitment to price stability by improving its accountability. Announcing targets which are not met because they have no effect on policy or because they are over-ridden by other objectives does nothing to improve the credibility or effectiveness of monetary policy.

Two basic prerequisites for putting in place an IT regime are that the central bank, which is charged with implementing monetary policy, be given a reasonable degree of 'instrument' independence to carry out that task; and the absence of commitment to a target for another nominal variable (Masson et al., 1997). Hybrid regimes are of course possible (see below) in which countries have targets for both the rate of inflation and the money supply or the exchange rate; in practice, several developing countries have operated such a regime for a transitional period. In this sense, the advocates of IT who deride a 'prerequisite approach' to IT (for example,

Sterne, 2002) have a point: one can start doing ‘baby steps’ as a way of learning how to walk. However, from the standpoint of clarity it is important to be clear what the ultimate objective should be, so as to put in place the capabilities needed to achieve it.

Subject to the above two prerequisites, the IT regime needs to involve the following elements of a framework for monetary policy: quantitative targets for the rate of inflation, over a specified horizon; a commitment to those targets as overriding objectives for policy; a clear methodology for making inflation forecasts; and a transparent way of translating the possible expected deviations from target into changes in the instruments of monetary policy.

How successful are developing countries in meeting the prerequisites? First, IT is likely to be a candidate regime primarily for middle-sized or large middle-income countries – roughly speaking, the ‘emerging market economies’. Smaller, very open economies may well choose to peg – a credible monetary regime for them; and countries with low incomes would typically not have the financial development or institutional capacity to implement inflation targeting. Second, countries differ greatly as to the degree of central bank independence. In particular, where there is fiscal dominance, an independent monetary policy is impossible. Compared to industrial countries, seigniorage – used to finance fiscal deficits – in many developing countries is high. Third, *de jure* independence may not guarantee that the central bank is able to carry out its mandate in the face of lack of public support. A constituency in favor of low inflation has not developed in many emerging market economies. And *de jure* independence has sometimes been overridden, as in Argentina, where a central bank governor was summarily replaced for disagreeing with the Minister of Finance. Fourth, administered price changes and centralized negotiations (for example the *Pacto* in Mexico) that determine a large fraction of the economy’s annual rate of wage increase may interfere with the central bank’s ability to control inflation unless they are coordinated with the inflation target. Finally, forecasting inflation is difficult in many emerging-market countries because of highly unstable macroeconomies and lack of solid econometric relationships (due, for example, to an insufficiently long or homogeneous data sample).

These obstacles are not insurmountable, however. The experience of emerging-market economies that practice IT is generally favorable, since countries have typically met their inflation targets. Moreover, the inflation-targeting countries have had a better experience of avoiding the balance-of-payments crises than other emerging-market countries since 1994. Thus, taking as given that emerging-market countries face a more challenging environment for monetary policy, this does not imply that the choice of

regime should necessarily be biased away from IT. Instead, Fraga et al. (2003) suggest that IT be operated somewhat differently than for industrial countries, perhaps by using wider and 'softer' bands around the inflation target, higher targets so as to accommodate bigger shocks without risking deflation, and adjusting targets more flexibly in response to shocks (while making special efforts to communicate the reasons to the general public).

Another feature of the early years of inflation targeting in developing countries has been its association with exchange rate targets. Thus, Chile, Israel and Poland, for instance, for a time targeted a band for the exchange rate (crawling, with the band width also adjusted depending on circumstances). Such a policy had the advantage of easing into the IT framework, allowing experience with it to be built up while retaining a backstop which might prevent instability should the inflation target give the wrong signal. The disadvantage of combining the two targets, as described in Israel's case by Bufman and Leiderman (2000), was lack of transparency and the danger of conflicting signals which could add to the public's uncertainty about monetary policy. In practice, given the greater immediacy of the exchange rate variable (continuously observable and widely publicized), it tended to dominate the inflation target if they conflicted, and this then required an explicit change in the exchange rate band if inflation was to be given its proper weight. Over time, the bands for the exchange rate were widened, as inflation declined and confidence with the new regime increased. Israel now has no exchange rate target. A similar progression occurred in Chile's case. At the time of the Russian crisis in August 1998 concern about external developments led the central bank to increase the short-term interest rate, causing the economy to go into a severe recession (Morandé, 2002). Sole emphasis on the inflation target would have allowed some easing of monetary policy. The exchange rate target was abandoned in September 1999, and the Bank of Chile at the time of writing in 2006 sets its monetary policy to maintain inflation within a 2–4 percent target range.

Unresolved questions concerning inflation targeting

While the verdict so far in emerging-market countries is positive, the track record is quite short. For most countries, it dates from the end of the 1990s. The environment has been relatively benign, with low inflation and low interest rates prevailing in the industrial world; this has permitted all developing countries, whether inflation targeters or not, to reduce inflation. Moreover, this period has not seen any major contagion from emerging-market currency or balance-of-payments crises.⁴ It remains to be seen whether a major world inflation shock would be weathered well by the IT regimes in place. As noted above, monetary and exchange rate

regimes have been periodically adopted and then abandoned – will the same be true of IT? Since IT is a less rigid regime – often characterized as ‘constrained discretion’ – it is likely to be less fragile than strict exchange rate targets. The danger remains however that the credibility gains that have accrued to IT central banks may be dissipated by persistent overshoots or indications that inflation does not provide an effective guide for month-to-month policy-setting. Thus, IT will face several challenges going forward.

An important question that should condition views of the advantages of IT is whether in fact it is subject to speculative attack. Kumhoff (2002) argues that the regime can be attacked, and that moreover it behaves much more like a fixed exchange rate regime than is often claimed. In the face of an unsustainable fiscal policy, it might have to be abandoned. Favero and Giavazzi (2004) provide a formal model in which such a fiscal policy could lead, through increasing default risk, to severe constraints on monetary policy’s ability to deliver on its inflation commitment. Contrary to the analysis of Fraga et al. (2003), they conclude that for at least a short period in 2002, Brazil’s economy might have tipped into a regime of fiscal dominance that, had it continued, would have doomed the IT regime. Thus, they are not convinced that the regime was ‘stress tested’.

The continued spread and popularity of IT may be affected by a parallel trend toward monetary unions, following in the wake of the successful creation of the Eurozone. Already, European integration has eliminated two inflation targeters, Finland and Spain, through their joining the Eurozone; and several among the new EU members, in particular the Czech Republic, Hungary and Poland, are expected to join the Eurozone within a few years. Sweden and the United Kingdom, prominent inflation targeters, could also conceivably join. While reducing the number of central banks implementing inflation targeting, of course should the European Central Bank (ECB) clearly adopt that regime, then the economic area it applied to might increase.

Turning to other continents, Africa, the Middle East and Asia are also considering regional monetary integration that might lead to a common currency. In Masson and Pattillo (2004) the argument is made in the context of Africa that the European example does not translate well to other regions with less broad-based integration projects and less strong regional solidarity. Thus, the success of an African single currency seems doubtful; instead, greater monetary integration could arise around regional poles such as South Africa, which already has a successful IT monetary regime. Since the Eurozone, the United States and Japan are likely to want to retain their monetary independence and exchange rate flexibility, regional currency blocs, if they are created, are unlikely to anchor their

exchange rate to any single reserve currency. On balance, then, inflation targeting among emerging market (EM) countries is likely to remain the regime of choice, whether or not regional integration proceeds.

When considering implementation of inflation targeting, a still unresolved issue is the weight to give to other variables, in addition to the inflation forecast. While of course other variables may influence the inflation forecast (for instance, the current output gap or the actual exchange rate) it seems that most countries give some additional weight to those variables – for instance, lowering interest rates if activity is weak, provided the inflation forecast is within the target range. Thus, the influence of other variables on policy may be asymmetric, but non-zero; this is consistent with Mervyn King's dictum that central banks are not 'inflation nutters'. Developing countries in particular are loath to ignore the exchange rate, and as Calvo and Reinhart (2002) show, exhibit a 'fear of floating'. While intervention may smooth some exchange rate fluctuations, central banks may also want to use interest rate policy for that purpose – subject to the caveats evoked by Chile's experience in 1998.

Finally, should the target for the (long-run) inflation rate be higher in emerging markets, as suggested by Fraga et al. (2003), because of larger shocks facing those countries? More empirical evidence is needed to resolve this issue. While the Balassa–Samuelson effect would suggest higher average inflation in faster-growing developing countries (provided one wants to avoid falling tradable goods prices in domestic currency), the argument for avoiding deflation really requires a greater understanding of the nature of downward rigidities. And the cross-country literature on costs of inflation is not very precise on when inflation becomes costly. These remain important issues for monetary policy in developing countries, whether they have adopted IT or not.

Notes

1. I am grateful to Max Corden, Frederic Mishkin and Miguel Savastano for comments on this chapter.
2. Even long-run neutrality does not command unanimity, however, and recent models suggest that transitory effects may last longer than previously thought (for example, Mankiw and Reis, 2001).
3. Programs supported by the International Monetary Fund typically included targets for the central bank's net domestic assets.
4. However, Fraga et al. (2003) consider that Brazil 'stress-tested' its IT regime during 2002, when it faced a negative capital account swing of about 6 percent of GDP, missing its inflation target but not suffering a permanent loss of credibility for the regime.

References

- Bufman, Gil and Leo Leiderman (2000), 'Monetary Policy and Disinflation in Israel', in L. Mahadeva and G. Sterne (eds), *Monetary Frameworks in a Global Context*, London: Routledge, pp. 484–524.

- Calvo, Guillermo and Carmen Reinhart (2002), 'Fear of Floating', *Quarterly Journal of Economics*, 117: 379–408.
- Corden, W. Max (2002), *Too Sensational: On the Choice of Exchange Rate Regimes*, Cambridge, MA: MIT Press.
- Diaz-Alejandro, Carlos (1985), 'Good-Bye Financial Repression, Hello Financial Crash', *Journal of Development Economics*, 19: 1–24.
- Eichengreen, Barry (1994), *Exchange Rate Regimes for the 21st Century*, Washington, DC: Brookings.
- Eichengreen, Barry, Paul Masson, Miguel Savastano and Sunil Sharma (1999), 'Transition Strategies and Nominal Anchors on the Road to Greater Exchange-rate Flexibility', *Essays in International Finance*, No. 213, Princeton University.
- Favero, Carlo and Francesco Giavazzi (2004), 'Inflation Targeting and Debt: Lessons from Brazil', NBER Working Paper 10390.
- Fraga, Arminio, Ilan Goldfajn and André Minella (2003), 'Inflation Targeting in Emerging Market Economies', in Ben Bernanke, Mark Gertler and Kenneth Rogoff (eds), *NBER Macroeconomics Annual 2003*, Vol. 18, Cambridge, MA: MIT Press, pp. 365–400.
- Frankel, Jeffrey (1999), 'No Single Currency Regime is Right for All Countries Or At All Times', *Essays in International Finance*, No. 215, Princeton University.
- Higgins, Benjamin (1959), *Economic Development: Principles, Problems, and Policies*, New York: Norton.
- Kumhoff, Michael (2002), 'A Critical View of Inflation Targeting: Crises, Limited Sustainability, and Aggregate Shocks', in Norman Loayza and Raimondo Soto (eds), *Inflation Targeting: Design, Performance, Challenges*, Santiago: Central Bank of Chile, pp. 349–94.
- Mankiw, N. Gregory and Ricardo Reis (2001), 'Sticky Information: A Model of Monetary NonNeutrality and Structural Slumps', NBER Working Paper 8614.
- Masson, Paul R. (2001), 'Exchange Rate Regime Transitions', *Journal of Development Economics*, 64: 571–86.
- Masson, Paul R. and Catherine Pattillo (2004), *The Monetary Geography of Africa*, Washington, DC: Brookings.
- Masson, Paul R. and Francisco Ruge-Murcia (2005), 'Explaining the Transition between Exchange Rate Regimes', *Scandinavian Journal of Economics*, 107: 261–78.
- Masson, Paul, Miguel Savastano and Sunil Sharma (1997), 'The Scope for Inflation Targeting in Developing Countries', IMF Working Paper WP/97/130, International Monetary Fund.
- McKinnon, Ronald I. (1973), *Money and Capital in Economic Development*, Washington, DC: Brookings.
- Mishkin, Frederic and Klaus Schmidt-Hebbel (2002), 'A Decade of Inflation Targeting in the World: What Do We Know and What Do We Need to Know?' in Norman Loayza and Raimondo Soto (eds), *Inflation Targeting: Design, Performance, Challenges*, Santiago: Central Bank of Chile, pp. 171–219.
- Morandé, Felipe (2002), 'A Decade of Inflation Targeting in Chile: Developments, Lessons, and Challenges', in Norman Loayza and Raimondo Soto (eds), *Inflation Targeting: Design, Performance, Challenges*, Santiago: Central Bank of Chile, pp. 583–626.
- Shaw, Edward S. (1973), *Financial Deepening in Economic Development*, New York: Oxford University Press.
- Sterne, Gabriel (2002), 'Inflation Targets in a Global Context', in Norman Loayza and Raimondo Soto (eds), *Inflation Targeting: Design, Performance, Challenges*, Santiago: Central Bank of Chile, pp. 23–77.
- Williamson, John (2000), 'Exchange Rate Regimes for Emerging Markets: Reviving the Intermediate Option', *Policy Analyses in International Economics*, No. 60, Washington, DC: Institute for International Economics.

55 Fiscal policy

*Raghbendra Jha*¹

Introduction

Fiscal policy plays an increasingly important role in many developing countries. Decisions on fiscal policy, especially if properly synchronized with monetary policy, can help smooth business cycles, ensure adequate public investment and redistribute incomes.

The four main components of fiscal policy are: (1) expenditure, budget reform; (2) revenue (particularly tax revenue) mobilization; (3) deficit containment and financing; and (4) determining fiscal transfers from higher to lower levels of government.² Fiscal policy works through both aggregate demand and aggregate supply channels. Changes in total taxes and public expenditure affect the level of aggregate demand, whereas the structures of taxation and public expenditure affect, among others, the incentives to save and invest (at home and abroad), take risks, and export and import goods and services.

This chapter is organized as follows. It first analyses some basic fiscal issues confronting developing countries. Then it considers budgetary deficits and evaluates norms for tax and expenditure reforms. The final section concludes.

Tax and expenditure profiles of developing countries

Developing-country fiscal systems have three basic characteristics: (1) low tax–GDP and expenditure–GDP ratios compared to developed countries, even though developing countries need more public expenditure; (2) the fiscal stance is often procyclical; and (3) tax resources are more volatile than those of developed countries.

There are manifold pressures for high and growing government expenditure in developing countries. Because of their low per capita incomes and high poverty, developing countries face an urgency to raise growth rates. This places a heavy burden on policy whereas, concurrently, the limited efficacy of policy instruments and governance inadequacies constrain policy. Pressures for populism through price controls and the like are considerable. The state in many developing countries is politically weak and beset with lack of consensus on what constitutes a sound fiscal policy compared to most developed countries (Heady, 2004).

Table 55.1 Total government tax revenue as a percentage of GDP, 1990–95 and 1996–2002 (domestic prices, number of countries and median values of the simple averages)

	1990–95		1996–2002		Direction of change	
	Countries	Median	Countries	Median	Down	Up
Separate samples	56	31.9	53	33.8		
Combined sample	48	33.5	48	33.8	20	28
Developed countries	21	37.8	21	40.1	3	18
Transitional economies	14	34.7	14	31.4	12	2
Developing countries	13	18.7	13	19.2	5	8

Note: Tax revenue is computed as the sum of revenue accruing to central and local governments. For each country in each category average tax revenue as a percentage of GDP is computed. For each category of country the median value of this average tax revenue is reported in the chapter.

Source: UNPAN Statistics.

Table 55.1 shows government tax revenues in developed, transition and developing countries for two time periods, 1990 to 1995, and 1996 to 2002. In the median developing country the tax–GDP ratio was below 20 per cent whereas in the median transition economy it was 30 per cent and in developed countries 40 per cent. Unsurprisingly, on average, governments in many developing countries face a severe resource crunch.

Only three out of 21 developed countries and two out of 14 transition countries had revenues falling over the two time periods, whereas the corresponding magnitude for developing countries was five out of 13. The share of distortionary commodity and trade taxes in total central government revenue is higher in developing countries (Tables 55.2 and 55.3).

Table 55.3 shows that in the richest countries personal income taxes are the most significant and contribute more than half (54.3 per cent) of tax revenue. Next are various commodity taxes and then the corporation tax. Border taxes and seigniorage revenue are low, reflected in the low value for inflation. The informal sector is small. With falling gross domestic product (GDP) per capita tax revenue as a percentage of GDP drops and corporate taxes as a percentage of revenue rise. Income taxes remain at about 30 to 35 per cent of revenue and commodity taxes are high. High rates of inflation reflect high values of seigniorage revenue. The poorest among the developing countries raise almost a quarter of their revenue through seigniorage. The informal sector in developing countries is about twice that in developed countries.

Table 55.2 *Central government revenue by type of tax, 1990–2002 averages (% of total tax revenue, median value of simple averages)*

	No. of countries	Direct taxes	Payroll taxes	Sales taxes	Trade taxes
Complete sample	139	27.1	5.9	34.8	14.6
Developed countries	24	34.8	28.5	28.1	0.5
Transitional economies	23	17.7	33.3	38.9	6.2
Developing countries	92	27.6	0.7	33.0	24.9
Africa	32	27.2	0.2	30.7	33.0
Latin America & Caribbean	27	22.1	5.1	38.9	13.7
Asia & Oceania	33	34.1	0.0	34.8	25.6

Source: UNPAN Statistics.

Auriol and Warlters (2005) argue that the informal sector in developing countries is large because of the higher costs of entry into the formal economy. By keeping barriers to entry into the formal economy high, those firms and individuals who make it into the formal economy acquire large rents and hence may be easier to tax than a diffused set of small taxpayers. If this argument is correct then encouraging large formal sectors should be part of a government strategy to increase tax revenue. Data for 64 countries indicate that this is indeed the case, particularly in Africa, for example, 0.4 per cent of taxpayers account for 61 per cent of total domestic tax collection in Kenya and 57 per cent in Colombia. As general policy, Auriol and Warlters argue that developing countries should lower entry barriers and raise the size of the formal sector to raise tax revenues.

Further, rapid globalization, technological advancement and the accompanying movement of factors of production across national boundaries, and the emergence of multinational corporations as major actors have eroded many developing countries' tax bases. Taxpayers can more easily raise income outside of conventional channels (Lao-Araya, 2003).

With inflexible public expenditures and low tax revenues government finances in developing countries are weak, with high deficits, debts and debt-servicing obligations. Consolidated figures for the finances of central and local governments together are not readily available but Table 55.4 presents these for central governments. Typically, developing countries' revenues and expenses are lower and interest payments higher than in developed countries, although government consumption in developing countries is lower than that in developed countries (Table 55.5).

Table 55.3 Sources of Government Revenue (1996–2001)

GDP per capita	Tax revenue (% of GDP)	Income taxes (% of revenue)	Corporate income tax (% of income taxes)	Consumption & Production taxes (% of revenue)	Border taxes (% of revenue)	Inflation rate (%)	Seigniorage Income (% of revenue)	Informal economy (% of GDP)
<\$745	14.1	35.9	53.7	43.5	16.4	10.6	21.8	26.4
\$746–2975	16.7	31.5	49.1	51.8	9.3	15.7	24.9	29.5
\$2976–\$9205	20.2	29.4	30.3	53.1	5.4	7.4	6.0	32.5
All developing	17.6	31.2	42.3	51.2	8.6	11.8	16.3	30.1
>\$9206	25.0	54.3	17.8	32.9	0.7	2.2	1.7	14.0

Source: Gordon and Li (2005).

Table 55.4 *Finances of central governments for country groups*

Country group	Revenue (% of GDP)		Expense (% of GDP)		Cash surplus or deficit (% of GDP)		Net incurrence of liabilities (% of GDP)				Debt and interest payments			
							Domestic		Foreign					
	1995	2004	1995	2004	1995	2004	1995	2004	1995	2004	Total debt as % of GDP	2004	Interest payment as % of revenue	2004
Low income	13.5	13.0	15.5	15.5	-2.6	-3.2								
Middle income	17.3						1.1		0.8				9.1	
Lower middle income	16.7						0.9		1.1				8.5	
Upper middle income							2.9		0.6				10.5	
East Asia & Pacific	8.4	11.5	12.0			-2.1							7.6	
Europe & Central Asia		31.0	31.1			-1.2		0.9		0.4			3.5	
Latin America & Caribbean	20.9		23.0		-0.4		1.0		2.3				11.9	
Middle East & North Africa	28.3		23.5		0.0									
South Asia	13.2	12.4	15.4	15.1	-2.7	-3.1	3.8	1.3	1.1	1.1	1.1	65.8	16.4	
High Income		26.0	28.9		-2.8		1.2						6.0	
Europe EMU	36.3	35.7	38.8	38.6	-2.3	-2.3	1.1						6.4	

Source: World Bank (2006).

Table 55.5 Government consumption as percentages of GDP, 1990, 1996, 2002 (domestic prices, median values)

	Number of Countries	1990	1996	2002 ¹
Complete sample	114		15.3	15.7
less transitional economies	101	5.2	14.3	15.6
Developed countries	24	18.9	19.4	19.1
Transitional economies	13	20.0	18.0	
Developing countries	77	14.2	12.7	14.0
Africa	26	15.1	12.8	14.7
Latin America & Caribbean	25	12.9	13.4	14.6
Asia & Oceania	26	12.2	11.7	13.0

Note: 1. Or latest data (2000, 2001).

Source: UNPAN Statistics.

Jha (2006) reports that the unweighted average of tax buoyancy (defined as *Percentage change in tax revenue/Percentage change in tax base*) for several developing countries is larger than one, indicating that an expansion of income would lead to an increase in the tax–GDP ratio. Gordon and Li (2005) argue that taxation, by its very nature, must depend on the formal economy since bank records are needed to identify taxable activity. In rich countries the intermediary services provided by the financial sector are considerable, so there is a high cost of abandoning it and conducting business in the informal sector. However, this is not the case in developing countries. Further, their tax base is likely to be narrow (biased towards capital income) and cover mostly capital-intensive firms that need the financial sector the most, and tariffs are used to protect the capital-intensive sectors and shortfalls in revenue (from public expenditures) would often be met through seigniorage.

Another important characteristic of fiscal variables in developing countries is their instability. Table 55.6 reports on key fiscal variables in 13 Latin American developing countries and 14 industrialized countries. In terms of all categories and in both nominal and real terms, computed coefficients of variation are much higher for Latin American developing countries than for industrialized countries.

Fiscal variables in many developing countries move in a procyclical fashion. Standard Keynesian models require that fiscal policy should be countercyclical, that is, during recessions taxes should be lowered and public expenditure hiked whereas during good times, taxes are raised and public expenditures lessened to reduce chances of overheating of the

Table 55.6 *Coefficients of variation of key fiscal variables*

	Nominal		Real	
	Industrialized Countries	Latin America	Industrialized Countries	Latin America
Total revenue	0.15	0.55	0.05	0.14
Current revenue	0.15	0.56	0.05	0.14
Non-tax revenue	0.19	0.58	0.11	0.24
Tax revenue	0.15	0.56	0.05	0.17
Total expenditure	0.16	0.55	0.05	0.14
Current expenditure	0.16	0.55	0.05	0.13
Government consumption	0.15	0.54	0.05	0.13
Interest payment	0.22	0.63	0.13	0.28
Transfers	0.17	0.58	0.07	0.20
Capital formation	0.17	0.57	0.14	0.22

Notes:

1. There are 13 Latin American and 14 industrialized countries.
2. In the sample the Gavin and Perotti (1997) database is used for the analysis.

Source: Bertin-Levecq (2000).

economy. In contrast the ‘Ricardian equivalence’ hypothesis suggested by Barro (1979) suggests that since rational economic agents make decisions based on perfectly anticipated tax and expenditure policies of the government, fiscal policy should remain neutral over the business cycle and respond only to unanticipated changes that affect the government’s budget constraint.

Using a sample of 56 countries (20 developed and 36 developing) Talvi and Vegh (2005) show that in G7 countries fiscal policy follows Barro, whereas for developing countries it has been procyclical. Two plausible explanations for this phenomenon exist. The first is that tax bases are so narrow and public expenditure so inelastic in developing countries that tax revenues and expenditures rise during expansions, whereas during recessions revenues and expenditures both decline for similar reasons. Second, as Talvi and Vegh (2005) argue, since fluctuations in the tax base are much larger in developing countries than in developed countries, full tax smoothing would require large surpluses during good times which is not possible since public expenditures are inelastic and resources may be wasted in enhanced public expenditures on public sector undertakings and subsidies, instead of retiring of debt as full tax smoothing would require.

Fiscal deficit issues

The exercise of fiscal policy in developing countries has its limits. The combination of low revenues and inelastic expenditures means that expenditures routinely, and even increasingly, outpace revenues. Jha (2004) argues that there is considerable heterogeneity in experience with respect to the fiscal deficit, between the middle- and low-income country categories and even within the low-income category countries. Indeed, the poorest among the least-developed countries are caught in an insidious resource trap and the average least-developed country economy has, since the 1970s, been exposed to adverse external trade shocks with an impact, in the worst years, approximately double the average of other developing countries (UNCTAD, 2000).

External finance is limited, especially for the poorest countries, although large, stable economies attract considerable capital inflows. Official aid has been falling and private equity flows go to the best-performing developing and transition economies. Private loans, as Harberger (1985) notes, are available at increasingly difficult terms since the domestic resource cost (often underestimated) of servicing these increases with additional borrowing. Other reasons for differences across developing countries include continuity and stability of policy regimes: Zambia, with several policy reversals, will be associated with greater risks than Mauritius, which has had a credible and stable policy regime.

Given financing constraints many developing countries have to opt for some non-bond (monetary) financing of the deficit. This establishes a direct link between fiscal policy and the monetary base of the central bank, blurs the distinction between fiscal and monetary policy, and compromises central bank independence.³ If bond financing is chosen, private investment may get crowded out.

Jha (2004), shows that in the long term public revenue and public expenditure are unrelated in many developing countries so that any excess of expenditure over revenue cannot be financed by generating budgetary surpluses over a long enough time horizon. Thus fiscal deficits are unsustainable in many developing countries. Mendoza and Ostry (2007) argue that whereas fiscal policy in most countries is responsive to budgetary deficits, high-debt countries do run a risk of having an unsustainable fiscal stance.⁴

Jha (2004) also shows that current account deficits are unsustainable in many developing countries. The fact that external sustainability conditions are hard to meet would imply the need for continual capital inflow in order to keep the balance of payments in equilibrium, necessitating the maintenance of a substantial rate-of-return wedge between domestic and foreign rates of return. This raises domestic interest rates substantially above global interest rates and acts as a drag on higher growth, making debt servicing harder, and exacerbates the fiscal deficit.

However, public expenditure could be productive, so whether public deficits impede or spur economic growth becomes an empirical question. In this context Adam and Bevan (2005) examine the relation between fiscal deficits and growth for a panel of 45 developing countries over 1970–99. Public expenditure is permitted to be both growth-enhancing as well as growth-inhibiting and distortionary taxes exist and fiscal deficits are permitted. They show that the impact of the deficit depends upon the mode of financing it. Deficits can be growth-enhancing if financed by limited seigniorage, growth-inhibiting if financed by domestic debt, and have opposite flow and stock effects if financed by external loans at market rates. These opposite effects define a threshold effect, before attaining which fiscal deficit has growth-enhancing effects and after which the effects of fiscal deficits are growth-inhibiting. Adam and Bevan find this threshold figure to be around 1.5 per cent of GDP after grants.

Norms for tax and expenditure reforms in developing countries

One of the principal aims of a meaningful tax and expenditure reforms policy would be to bolster the savings and investment rates in the economy in order to raise growth rates. A higher growth rate, it is widely accepted, is the best way to lower poverty over the medium term. Loayza et al. (2000a, 2000b) indicate that the most important determinant of savings, across both developed and developing countries, is the level of per capita income and the rate of economic growth. Thus the higher the rate of savings, the higher the economic growth rate and the higher the growth rate, the higher the rate of savings at least at low absolute levels of per capita income. Their results also point to the possibility of incomplete Ricardian equivalence, that is, a given rise in public savings is accompanied by a less than commensurate drop in private savings.

The gap between the real rate of return on savings and the discount rate is critical. Savers who are liquidity constrained may be more sensitive to such differentials compared to those who are not. As financial deepening takes place and fewer consumers remain liquidity constrained, this responsiveness may drop. However, as consumers become less liquidity constrained they might also become less risk averse and opt for investments with higher returns, boosting the savings rate. Thus the impact of the tax structure on savings is of critical importance, and distorting differences in effective tax rates across sectors and assets and tax-induced distortions that create inefficiencies and lower the potential rate of economic growth should be eliminated. This would be an important component of tax reform, the basic tenets of which are well known and briefly summarized below.

As an economy develops, reliance on indirect taxation for revenue should decline. This is because indirect taxes typically have an excess burden

associated with them (Jha, 1998). Furthermore efficient indirect taxation (one that minimizes excess burden to the representative consumer) can be quite regressive.⁵ Indirect taxes can be made redistributive by sacrificing some efficiency, but the extent of this redistribution is limited (Sah, 1983).

If, however, indirect taxes can be levied on final consumption alone, tax-induced changes in relative prices that characterize production taxes such as excise duties could be avoided. Then, if consumer utility functions are weakly separable between consumption and leisure, a uniform tax on final consumption goods (say a value-added tax – VAT) would approximate a lump-sum tax.⁶ This tax, with only few exemptions (for items consumed in disproportionately large amounts by the poor), harmonized across levels of government in federal countries and few rates, is recommended. These could be supplemented with excise duties on environmental bads or ‘luxuries’. Peak tariff duties and effective rates of protection should be reduced gradually. If the tax base admits few exemptions and there are fewer rates, costs of compliance and monitoring will fall. But the VAT requires the netting out of input costs and the exemption of exports from the tax base. This, in turn, needs sophisticated account keeping which may be absent in many developing countries. The credibility of the tax regime is also important and tax reforms should aim for a stable tax environment and be well coordinated and, at all times, be simple. Tariff cuts should be accompanied by an upward revision of VAT rates to compensate for tax revenue.

However, Emran and Stiglitz (2005) show that the standard prescription of reducing trade taxes with revenue-compensating upward revision of the rate of a broad-based VAT is welfare-improving only in an economy with no informal sector, with all production and exchange activity in the tax net – conditions typically not satisfied in developing countries.⁷ When only the formal sector can be taxed, the introduction of a VAT (or a hike in its rate) may end up creating a distortion between the formal and informal sectors. Even broadening the VAT base to include more of the informal sector may reduce welfare (Piggott and Whalley, 2001). Similarly Bibi and Duclos (2007) show that for indirect tax reform to be poverty-reducing it must: (1) not remove all subsidies; (2) in some cases increase taxes on already taxed commodities and, concurrently, increase subsidies on already subsidized commodities; (3) not exclusively follow efficiency considerations, since redistribution may still play an important role in poverty reduction; and (4) concentrate on reform rather than removal of subsidy. Jha (2006) presents a taxonomy of the extant literature’s view on how to fine-tune the aforementioned tax reforms in order to make the resulting tax structures distributionally sensitive.

Tax structures in developing countries are not particularly progressive. Thus Chu et al. (2004) find that: (1) only 13 of the 36 overall tax systems

surveyed by them are progressive, seven are proportional, seven are regressive and the rest neutral or insignificant; (2) income taxes were progressive in 12 of the 14 cases studied whereas indirect taxes were broadly regressive. The progressivity of direct taxes declined over time in eight cases. This needs attention.

Another principle of tax reform is that the share of direct taxation in overall tax revenue should rise. Within direct taxation, reliance has to be shifted from corporate to income taxes. Since corporate profits are taxed at the level of personal income anyway, the rationale for separate corporate taxes is rather weak. There are only two arguments in favor of corporate taxes: (1) as a tax on foreigners' incomes; and (2) as a tax on non-competitive profits. Within the sphere of income taxation, the rate and exemptions structures need to be rationalized. The number of tax brackets should be small, the degree of progression mild with the top marginal tax rate low. Tax reform theory advocates taxation of 'full income' the Haig-Simons definition of which is 'all increases in human and physical capital during a period of time'. One cannot pick and choose the types of income one would like to tax.

Another area of importance for taxation is the conduct of commerce over the internet (e-commerce). Although e-commerce is a nascent industry it should be taxed since it would be inefficient as well as inequitable to tax goods traded through bricks and mortar stores and not tax e-commerce. A commodity that is sold in a bricks and mortar store and, therefore, subject to taxation would be deemed to be different if sold through e-commerce, and escape taxation. Further, those buying through e-commerce are likely to be rich. This exacerbates inequity. There is a rationale for zero customs duties on e-commerce in line with arguments for free trade, but not for zero taxes. A policy of not taxing e-commerce would provide another avenue for tax evasion as some US evidence shows. Further, given its projected phenomenal rate of growth, if e-commerce is not taxed there will be sharp erosion of the tax bases of governments that primarily levy sales taxes.

Another issue is the presence of tax havens. The Organisation for Economic Co-operation and Development (OECD) estimates that during 1985-94 foreign direct investment (FDI) by the G7 countries in some tax havens in the Caribbean and South Pacific increased more than fivefold to more than US\$200 billion – an increase well in excess of the growth of total outbound FDI. These concerns extend to transition and developing economies and have probably worsened in recent years. 'A race to the bottom' may ensue with national and/or state governments using tax incentives competitively to attract FDI. Such incentives interact dynamically with the existing avenues for tax evasion (for example because some

incomes are not taxed) to reduce current tax revenues and prospects for higher future tax revenues. In the face of this tax reform, particularly direct tax reform, should have a considerable element of international cooperation.

A related issue is service taxation. Services have become the dominant sector in many developing countries but are hard to tax. Not taxing services is inefficient as well as inequitable: inequitable because it discriminates between providers of goods and services; inefficient because it has the potential of creating several distortions, thus increasing non-labour costs.

Expenditure reform

Tax reforms should be complemented with appropriate adjustment of government expenditures. Typically this calls for reduction of current subsidies and augmentation of subsidies for well-managed capital projects. The impact of public expenditure is usually ascertained through an *ex post* incidence analysis but we must evaluate not what does exist but what might exist – the theme of benefit incidence analysis. Such analysis is marginal (to capture differences from the status quo) and behavioural (to generate counterfactuals) and is difficult to conduct in many developing countries.

Delineating expenditure adjustments according to their effects on the poor cannot await the development of *ex ante* analysis. A good rule of thumb is to delay or reduce cuts in public expenditure on goods and services that are directly or indirectly of high importance in the poor's budget, for example, coarser types of food, fuel and agricultural subsidies.

Within the broad category of basic services the selection of programmes needs to be sensitive to the type and severity of deprivation. If malnutrition is widespread, a programme of subsidized nutritional supplements would be more effective than an elementary education scheme. Rudra (2004) establishes that only the education component of public expenditure reduces income inequality in the face of globalization. Thus, when high inequality is a concern, expenditure on education should not be cut (van de Walle and Nead, 1995).

Conclusions

The role of fiscal policy in developing countries is as important as it is complex. Developing countries face the unenviable task of accelerating economic growth to reduce poverty in a short span of time even as they face greater uncertainty, in the face of globalization, about key elements of their fiscal policy such as the tax base. Furthermore, the exercise of fiscal policy is often circumscribed by increasing pressures from regulatory and exchange rate regimes in place, and subject to considerable pressure from external parameters such as competing countries' tax rates; for example, it

would be difficult for a given developing country to have corporate tax rates very different from its competitors or to burden monetary policy with high fiscal deficits which could lead to sharp depreciation of the exchange rate.

This chapter has outlined some of the major challenges that developing countries face in some key areas of fiscal policy, particularly tax and expenditure. Even here the treatment has been selective (for example, there has been little discussion of corporate taxation and indirect tax harmonization) to provide an overview of the issues involved and an introduction to the literature on these topics.

Notes

1. I am grateful to Amitava Krishna Dutt and Jaime Ros for helpful comments on an earlier draft of this chapter. The usual caveat applies.
2. The rationale for the existence of multi-tiered governments owes much to the classic statement by Oates (1972), and has been extensively reviewed (for example, Jha, 1998). Intergovernmental fiscal relations are surveyed, among others, by Fjeldstad (2001) and Bird and Smart (2002). Fiscal federalism is not considered in this chapter.
3. This may also lead to an exacerbation of inflation as de Haan and Zelhorst (1990), Easterly and Schmidt-Hebbel (1993) and Buffie (1999) show.
4. They find these countries to be Malaysia, Hungary, Ecuador, Morocco, Panama, Philippines, Indonesia, Bulgaria, Côte d'Ivoire, Egypt, Israel, Jordan, Lebanon, Nigeria and Pakistan. Clearly both transition and developing economies belong to this group.
5. Efficient indirect taxation calls for tax rates to vary inversely with compensated elasticity of demand making them regressive.
6. Separability of the utility function between goods and leisure would indicate that taxation of goods would have no implications for the labour-leisure choice.
7. CSO (2000) notes that in 1999–2000 as much as 60 per cent of India's GDP came from the unorganized sector and this sector employed 92 per cent of the labour force.

References

- Adam, C. and D. Bevan (2005), 'Fiscal Deficits and Growth in Developing Countries', *Journal of Public Economics*, 89 (3): 571–97.
- Auriol, E. and M. Warlters (2005), 'Taxation Base in Developing Countries', *Journal of Public Economics*, 89 (3): 625–46.
- Barro, R. (1979), 'Are Government Bonds Net Wealth?' *Journal of Political Economy*, 82: 1095–1117.
- Bertin-Levecq, S. (2000), 'Determinants of Fiscal Volatility: An Empirical Investigation of Latin American and Industrialized Countries', New York University, Department of Economics.
- Bibi, S. and J. Duclos (2007), 'Poverty-Decreasing Indirect Tax Reforms: Evidence from Tunisia', *International Tax and Public Finance*, 14 (2): 165–90.
- Bird, R. and M. Smart (2002), 'Intergovernmental Fiscal Transfers: International Lessons for Developing Countries', *World Development*, 30 (6): 899–912.
- Buffie, E. (1999), 'Public Sector Wage-Cycles and the Co-Movement of the Fiscal Deficit and Inflation in Less-Developed Countries', *Journal of Money, Credit and Banking*, 31 (4): 785–810.
- Chu, K., H. Davoodi and S. Gupta (2004), 'Income Distribution and Tax and Government Social Spending Policies in Developing Countries', in A. Cornia (ed.), *Inequality, Growth, and Poverty in an Era of Liberalization and Globalization*, New York: Oxford University Press, pp. 249–70.
- CSO (Central Statistical Organisation) (2000), National Account Statistics, New Delhi.

- de Haan, J. and D. Zelhorst (1990), 'The Impact of Government Deficits on Money Growth in Developing Countries', *Journal of International Money and Finance*, 9 (3): 455–69.
- Easterly, W. and K. Schmidt-Hebbel (1993), 'Fiscal Deficits and Macroeconomic Performance in Developing Countries', *World Bank Research Observer*, 8 (2): 211–37.
- Emran, M. and J. Stiglitz (2005), 'On Selective Indirect Tax Reform in Developing Countries', *Journal of Public Economics*, 89 (3): 599–623.
- Fjeldstad, O. (2001), 'Intergovernmental Fiscal Relations in Developing Countries: A Review of Issues', Working Paper 2001:11, Chr. Michelsen Institute, Development Studies and Human Rights.
- Gavin, M. and R. Perotti (1997), 'Fiscal Policy in Latin America', in B. Bernanke and J. Rotemberg (eds), *NBER Macroeconomics Annual 1997*, Cambridge, MA: MIT Press, pp. 11–61.
- Gordon, R. and W. Li (2005), 'Tax Structures in Developing Countries: Many Puzzles and a Possible Explanation', Working Paper 11267, National Bureau of Economic Research.
- Harberger, A. (1985), 'Lesson for Debtor-Country Managers and Policymakers', in G. Smith and J. Cuddington (eds), *International Debt and the Developing Countries*, Washington, DC: World Bank, pp. 236–57.
- Heady, C. (2004), 'Taxation Policy in Low-Income Countries', in T. Addison and A. Roe (eds), *Fiscal Policy for Development: Poverty, Reconstruction and Growth*, Basingstoke: Palgrave Macmillan, pp. 130–48.
- Jha, R. (1998), *Modern Public Economics*, London and New York: Routledge.
- Jha, R. (2004), 'Macroeconomics of Fiscal Policy in Developing Countries', in T. Addison and A. Roe (eds), *Fiscal Policy for Development: Poverty, Reconstruction and Growth*, Basingstoke: Palgrave Macmillan, pp. 24–45.
- Jha, R. (2006), 'Pro-poor Fiscal Policy in the Globalized Economy', in G.A. Cornia (ed.), *Pro-Poor Macroeconomics: Potential and Limitations*, Basingstoke: Palgrave Macmillan, pp. 27–48.
- Lao-Araya, K. (2003), 'Emerging Tax Issues: Implications of Globalization and Technology', ERD Policy Brief No. 16, Asian Development Bank, Manila.
- Loayza, N., K. Schmidt-Hebbel and L. Serven (2000a), 'What Drives Private Saving Around the World?', *Review of Economics and Statistics*, 82: 165–81.
- Loayza, H., K. Schmidt-Hebbel and L. Serven (2000b), 'How Effective is Fiscal Policy in Raising National Saving?', *Review of Economics and Statistics*, 82: 226–38.
- Mendoza, E. and J. Ostry (2007), 'International Evidence on Fiscal Solvency: Is Fiscal Policy "responsible"?' IMF Working Paper 0756.
- Oates, W. (1972), *Fiscal Federalism*, New York: Harcourt Brace Jovanovich.
- OECD (2000), <http://www.OECD.org> (OECD fiscal information).
- Piggott, J. and J. Whalley (2001), 'VAT Base Broadening, Self Supply, and the Informal Sector', *American Economic Review*, 91: 1084–94.
- Rudra, N. (2004), 'Openness, Welfare Spending and Inequality in the Developing World', *International Studies Quarterly*, 48: 683–709.
- Sah, R. (1983), 'How Much Redistribution is Possible through Commodity Taxes?' *Journal of Public Economics*, 20: 89–101.
- Talvi, E. and C. Vegh (2005), 'Tax Base Variability and Procyclical Fiscal Policy in Developing Countries', *Journal of Development Economics*, 78 (1): 156–90.
- UNCTAD (2000), *The Least Developed Countries 2000 Report*, Geneva: UNCTAD.
- United Nations Network in Public Administration and Finance (UNPAN), http://www.unpan.org/statistical_database_publicsector.asp, accessed 8 February 2007.
- van de Walle, D. and K. Nead (eds) (1995), *Public Spending and the Poor: Theory and Evidence*, Baltimore, MD: Johns Hopkins University Press for the World Bank.
- World Bank (2006), *World Development Indicators*, Washington, DC: World Bank.

56 Stabilization policy and structural adjustment

Valpy FitzGerald

Introduction¹

Economic development strategy in open industrializing economies is dominated by stabilization policy and structural adjustment. Developing countries are faced by an exogenous and changing set of world prices and export demand schedules on the one hand, and rationed global credit markets whose dynamic is determined by financial cycles in the core economies on the other. In consequence, domestic demand stabilization in response to unexpected temporary shocks and supply adjustment to permanent shifts in global markets determine the growth path, rather than a process of intertemporal optimization in the stable and foreseeable world of textbook economic theory.

Global economic shocks are exacerbated by armed conflicts and natural disasters, which often affect entire regions, while accumulated debt positions affect not only fiscal and current account balances but also the future expectations (and thus current behaviour) of the private sector. The political economy issues arising from accompanying changes in employment levels, wage rates and sectoral output are complicated by the key role played by international institutions – particularly the International Monetary Fund (IMF) and the World Bank – both as providers of financial resources to governments and as arbiters of ‘sound’ economic management.

These orthodox macroeconomic management criteria are based on specific models of economic behaviour for small open economies as well as a bias against active intervention in markets. In consequence the disappointing record of orthodox stabilization and adjustment policies in Eastern Europe, Latin America and Asia in the 1990s has led to a renewed interest in ‘heterodox’ methods of macroeconomic management that correspond more closely to the structure and behaviour of the emerging market economies with domestic capital markets integrated to the international financial system. However, an alternative set of underlying macroeconomic models with the generality and power of the orthodox models has yet to be constructed.

This chapter opens with a brief survey of the analytical literature that reveals an excessive emphasis on inflation targeting in stabilization policy

and a lack of attention to issues of investment and distribution in structural adjustment. It then outlines an alternative new-Keynesian approach to stabilization policy that explicitly contrasts with the standard IMF monetary programming framework. When capacity utilization and open capital account are included, inflation targeting is shown to have destabilizing consequences, requiring a return to active fiscal and monetary policy. The chapter then presents a new-Keynesian view of structural adjustment in explicit contrast to the standard '1-2-3' open economy model that underpins World Bank analysis. It is shown that resource reallocation in the medium term can only take place through new investment, with significant distributional effects via employment and real wages. The chapter concludes by suggesting that emerging market authorities should and can engage in active macroeconomic intervention based on a more realistic analysis of the structure and behaviour of their economies.

Stabilization policy and structural adjustment: the analytical debates

I start by examining the macroeconomic programming models used by the Fund and the Bank in designing stabilization policy and structural adjustment, respectively.² These two models are not entirely consistent as they are built on different assumptions as well as referring to different time horizons, despite earlier attempts to reconcile them (Khan et al., 1990), and thus must be treated separately.

The IMF Basic Financial Programming Framework (BFPF) is the standard model used by the Fund in designing stabilization programmes, the object of which is to reduce inflation to as near zero as possible and ensure debt service payments (IMF, 1987; Mussa and Savastano, 1999). The intellectual origins of the BFPF are Hicks's interpretation of Keynes, expressed as the 'absorption approach' where excess domestic demand creates current account deficits and/or domestic inflation when imports are constrained by lack of external finance (Polak, 1957). The model subsequently metamorphosed into a version of the Chicago 'monetary approach to the balance of payments'. It is concerned with the short run, where both real output and exports are taken as given, so the focus is on aggregate demand management. The budgetary balance plays a key role in this (particularly when financed by money supply) as does the nominal exchange rate because it sets import purchasing power.

Apart from the usual set of national accounting identities linking the macroeconomic and monetary variables, there are only two behavioural relationships in the BFPF: the demand for money and the demand for imports. A budget deficit beyond that warranted by output growth and the inflation target (that is, 'seignorage') is then reflected fully in the current account deficit if the exchange rate is fixed and foreign finance is available;

or fully in domestic inflation if the exchange rate is flexible and foreign finance constrained. Stabilization policy design then consists in the required fiscal adjustment³ in order to bring inflation down to target and restore external reserve levels to a prudent proportion of imports. This is supported by last-resort short-term lending from the Fund, which alleviates the fiscal adjustment required to meet these targets; and also provides powerful external leverage ('conditionality') to ensure compliance with this form of stabilization policy (Collier and Gunning, 1999). However, the behavioural relationships are clearly far too simplistic in the BFPP: for instance, interest rates and wealth (especially debt) stocks do not enter the model, while capital flows are exogenous, despite the central place of both in modern monetary theory.

The World Bank Revised Minimum Standard Model (RMSM) is used in designing structural adjustment programmes, the object of which is to restore current account stability and raise output growth (Addison, 1989 [1999]). The RMSM has Keynesian roots too: in the Harrod–Domar theory of growth constrained by savings, modified to include an external constraint reflecting the dependence of developing countries on imports of producer goods, leading to the 'two-gap' model (Chenery and Strout, 1966). However, it has since involved into the more neoclassical framework of a 'computable general equilibrium' model discussed below. The RMSM is concerned with the medium term, so aggregate supply is endogenous. Exports respond to the real exchange rate, which acts so as to allocate production factors between the traded and non-traded sectors: in other words, a relative price effect on supply instead of the income effect of the nominal exchange rate on demand in the Fund model. Investment is simply driven by the availability of savings: private saving (a constant proportion of disposable income) less the budget deficit plus external finance ('foreign saving').

In addition to the usual national accounting identities, the RMSM contains five behavioural relationships for the investment–growth linkage, import demand and export supply, fiscal income and private saving. Structural adjustment design seeks to relax the current account constraint on growth by raising exports through real exchange rate devaluation; and to raise the growth rate itself by reducing government expenditure and thus reversing the 'crowding-out' of private investment. Regulatory reforms follow the same logic, emphasizing trade and financial liberalization combined with extensive privatization to reduce the size and scope of the public sector. External finance in the RMSM plays three roles therefore: directly increasing public investment (for example in infrastructure) and output growth; reducing domestic borrowing to fund the budget deficit, and thus allowing private investment to rise; and funding more imports and thus

output.⁴ As in the case of the Fund, the Bank's role as a leading provider of long-term official loans to poor countries, and its influence on other aid donors, ensures the adoption of this 'sound' approach to structural adjustment (Mosley et al., 1995).

As the RMSM is a one-sector model it is useful for macroeconomic programming, but not entirely appropriate for the analysis of structural adjustment, so World Bank policy design also has analytical foundations derived from the 'dependent'⁵ economy' model set out by Dornbusch (1986) and Buiter (1988). This is disaggregated to generate what has now become the 'industry standard'⁶ with three products – exportables, importables, and non-tradable or 'home' goods and services – which we use as a framework in the fourth section below. This '1-2-3' model⁷ has generated a wide range of applied computable general equilibrium models (Devarajan and Robinson, 1993) used by the World Bank to inform structural adjustment programmes and to link macroeconomic policy to poverty reduction strategies (Bourgignon and Morrison, 1992).

A number of significant lessons are drawn from this simple yet powerful model. One of these is the well-known 'Dutch disease'⁸ interpretation of the effect of an unexpected increase in world primary commodity prices or a rise in external aid flows: the real exchange rate appreciates, the non-traded sector expands, other traded sectors contract, so imports rise and exports fall, which is unsustainable in the long run. Another lesson is the effect of fiscal expansion: as government expenditure is intensive in non-traded goods and services, the real exchange rate appreciates and traded export production falls as non-traded output rises, leading to unsustainable debt problems.

However, neither the RMSM nor the 1-2-3 model are dynamic and thus do not allow for intertemporal optimization by economic agents: that is, the fact that households, firms and governments take investment, saving and borrowing decisions looking forward over many years. This is the basis of modern neoclassical macroeconomics and allows resource allocation behaviour to be endogenized.⁹ Further, they fail to reflect the elements of modern growth theory in general and the role of public expenditure in physical and human capital formation in particular.¹⁰ Last but not least, the simplistic view of the negative effects of budget deficits (on inflation for the Fund and on private investment for the Bank) ignores the modern macroeconomic theory of intertemporal budgetary and financial policy.¹¹ Indeed from a strictly neoclassical viewpoint this persistence of the 'financing gap' tradition can be seen as invalidating the proposals from the Bank and the Fund on additional lending and debt forgiveness (Easterly, 1999).

These orthodox models have also been subjected to a much broader critique from non-neoclassical standpoints. Four such lines of argument are:

(1) the Keynesian critique of the failure to understand the exogenous nature of cycles in open economies; (2) the structuralist critique of the neglect of supply constraints in developing countries; (3) the Kaleckian critique of implausible assumptions on investment and savings; and (4) the Fabian critique of the exclusion of poverty reduction from macroeconomic strategy. These critical theoretical views have been fuelled by the evident failure in most cases of stabilization policy to get beyond inflation reduction, and of structural adjustment to achieve sustained growth (Williamson, 1997).

The underlying assumption in the IMF model that output is unaffected by demand and indeed that the economy operates at full factor employment is clearly implausible. Excess capacity in the Keynesian sense is often present in practice, as well as chronic underemployment in the Lewis sense. Moreover, the central issue in monetary policy for most developing countries today¹² is not inflation as such but rather countering the effects of externally generated cycles exacerbated by inherited debt positions (Ocampo, 2000). In the upswing of a cycle the interest rate declines and the exchange rate appreciates, but any attempt to counter the boom attracts still more funds and the exchange rate appreciates still further. In the downswing, markets push for devaluation but this forces up interest rates and exacerbates production declines, promoting further capital flight and debt default. Moreover, the budgetary dependence on foreign borrowing makes the fiscal stance automatically procyclical. The application of the standard IMF policy model during these financial crises worsens economic recessions and further destabilizes capital flows (Stiglitz and Greenwald, 2003).

A central feature of the standard theory of structural adjustment is that any imbalance between traded and non-traded sectors in the dependent open economy is a result of distorted domestic relative prices (that is, differing from 'world' prices): thus the emphasis on real exchange rate correction and trade liberalization. However, this analysis rests on the twin assumptions of full employment of labour and capital on the one hand and the perfect substitution of existing factors between sectors in response to relative prices on the other; which is clearly unrealistic. Indeed, excess capacity and immobile factors explain much of the lack of supply response to structural adjustment (Taylor, 1993). Moreover, the assumption that domestic prices are not affected by exchange rates (and thus that devaluation is not 'passed through' into inflation) is similarly implausible for small open economies (Taylor, 1988). These supply response failures are exacerbated by the lack of business liquidity caused by restrictive monetary policies, because even under normal circumstances credit rationing prevails and output as well as prices are affected by interest rates (Blinder, 1987).

More generally, there is a clear parallel between orthodox structural adjustment theory and neoclassical trade theory because the internalization

of world prices is intended to bring about an intersectoral resource reallocation in line with comparative advantage. Trade liberalization raises the return to the abundant factor of production (assumed to be unskilled labour in developing countries); and because primary exports are taken to be more labour-intensive than tariff-protected industry there should also be a net employment creation (Obstfeld and Rogoff, 1997). However this does not often occur in practice: either because exports are based on natural resources in which case rents rise and little unskilled employment is generated; or because skilled labour is the scarce resource and export expansion opens up wage differentials (Wood, 1994). Moreover, there is in consequence no theoretical reason to believe that income distribution will necessarily improve with structural adjustment.

Central to the theoretical approach of both Bank and Fund is that private saving¹³ is a fixed proportion of disposable private income, and that private investment (and thus growth) is determined by private savings less the budget deficit plus external finance. However the large fluctuations observed in the savings rate for developing countries and the empirical evidence of the influence on private investment of other factors such as profit rates, credit conditions, public infrastructure, debt overhang, regulatory change and political stability all suggest that in developing countries at least investment is not constrained by private saving (FitzGerald, 2003). Indeed, the policy uncertainty caused by violent and unpredictable stabilization and adjustment episodes is among the most depressive influences on investment in developing countries (Rodrick, 1991). Nonetheless, successful structural adjustment and sustained growth require high rates of investment so that production capacity can change and thus the desired structural adjustment takes place.

This process cannot simply be considered as an overall proportionate expansion based on a fixed savings rate (augmented as necessary by external funds) once domestic resources have been reallocated, as the RMSM does; nor as a smooth process of reallocation of labour and capital between sectors in response to changing relative prices as the 1-2-3 model does. Both modern intertemporal macroeconomics and traditional Keynesian theory tell us that the investment process has its own dynamic based on future profitability, and this has profound implications for adjustment policy. Further, the financing gap theory used in both Bretton Woods models assumes that extra external finance always contributes to growth, by simply and directly adding to investment funds: but it is well established that capital inflows often lead to increased consumption (Jansen and Vos, 1997).

Last, but far from least, the neglect of distributional considerations in both the Bank and Fund models is not only inconsistent with their institutional commitment to poverty reduction but also leads them to

underestimate the political economy constraints on macroeconomic policy. There exists a long-standing critique of adjustment policy in this respect in terms of the negative effect on social service provision of fiscal expenditure cuts as the central macroeconomic policy tool (Cornia et al., 1987). Targeted poverty reduction programmes, while desirable in themselves, do not redress the effects of macroeconomic policy design on employment and wages, which are more significant in determining the welfare of the majority of the population. These effects in turn determine social support for economic policy, and thus its political sustainability.

Stabilization policy, inflation targeting and monetary autonomy

We have seen that the open developing macro-economy works in a different way from that which the Fund model supposes – the role of domestic credit rationing and external capital flows being crucial in the short-run context. This section sets out, therefore, a model with a formal framework similar to that in IMF (1987) except that: (1) output can be below capacity and is determined by the level and costs of credit; and (2) the interest rate and exchange rate are related through arbitrage across the capital account.

The standard inflation-targeting model can be set out as follows (IMF, 1987). As this is a short-run model exports (X) and real output (Q) are exogenous, as are the capital flows, net of debt service, that determine the net change in external liabilities (\dot{F}). The level of domestic debt (D) and foreign exchange reserves (R) are set according to fixed prudential rules. The endogenous variables are thus domestic aggregate income (Y), the level of imports (M) and the demand for money (B) and for credit (H) from the private sector. The nominal exchange rate (E) floats under the current Fund doctrine, and thus is also endogenous. The target variable is the price level (P) and the policy instrument is the interest rate (i).

I start with three accounting identities. Nominal income (Y) and inflation (p) are:

$$Y \equiv Q \cdot P$$

$$p = \frac{\dot{P}}{P} \quad (56.1)$$

The balance of payments (denominated in foreign currency) is the familiar:

$$X - M \equiv \dot{R} - \dot{F} \quad (56.2)$$

and the domestic monetary balance (Khan et al., 1990, p. 158) is:

$$B \equiv D + H + E \cdot R \quad (56.3)$$

There are three behavioural equations in this standard model, each of which reflect a key aspect of aggregate private sector macroeconomic behaviour. Import demand in nominal terms is a proportion (m) of aggregate income (Y):¹⁴

$$M \cdot E = mY \quad (56.4)$$

Deposits in the banking system (that is, 'demand for money') depend on income (Y) and the interest rate (i) for a given velocity of circulation (v) and positive interest 'elasticity' coefficient (α):

$$B = Yv + \alpha i \quad (56.5)$$

The credit (and cash) requirements of the private sector – that is, the supply of money – have a similar form because the Fund model assumes that the authorities always accommodate the monetary needs of the market (that is, passive rather than active monetary stance) and that the impact (β) of the interest rate on this demand is of course negative:

$$H = Yu - \beta i \quad (56.6)$$

The 'prudential rules' for domestic debt (D) and reserves (R) are:

$$\begin{aligned} \dot{D} &= \lambda Y \\ R &= \theta M \end{aligned} \quad (56.7)$$

The reserves rule (θ) is based on a specific degree of 'liquidity' in the form of import coverage;¹⁵ while the domestic debt rule (λ) effectively constrains the fiscal deficit as a proportion of GDP.¹⁶

This model is simple to solve because it can be distilled down to two reduced-form equations based on (56.2) and (56.3). The domestic price level (P) is determined from the domestic monetary balance by substituting (56.5), (56.6) and (56.7) into (56.3) using (56.1) and (56.4) to yield:

$$P = \frac{D_{-1} - (\alpha + \beta)i}{Q\{v - u - \lambda - m\theta\}} \quad (56.8)$$

from which it is clear not only that higher interest rates (i) reduce the price level and thus inflation, but also that a key determinant of inflationary pressure is the domestic debt overhang (D_{-1}), as indeed is the prudent fiscal deficit (λ), thus the emphasis on fiscal retrenchment in Fund stabilization programmes.

An alternative formulation of the reserves rule that has recently found favour in the Fund with the spread of full currency convertibility is that there should be maintained a constant proportion (κ) of the money supply (H). This gives a similar result:

$$ER = \kappa H \quad (56.7a)$$

$$P = \frac{D_{-1} - \{\alpha + \beta(1 + \kappa)\}i}{Q\{v - u(1 + \kappa) - \lambda\}} \quad (56.8a)$$

The external foreign currency balance found by substituting (56.1), (56.4) and (56.7) into (56.2) and using (56.8) serves to determine the nominal exchange rate (E) for a given domestic price level under the floating exchange rate regime espoused by the Fund:

$$E = \frac{mQP(1 + \theta)}{X + (\theta M_{-1} + \dot{F})} = \frac{\{D_{-1} - (\alpha + \beta)i\}}{(v - u - \lambda - m\theta)} \frac{m(1 + \theta)}{\theta M_{-1} + \dot{F}} \quad (56.9)$$

This in turn implies that the real exchange rate (e) using world prices as numeraire¹⁷ – and thus export competitiveness in the medium term – is endogenous and appreciates (that is, e falls) with positive external shocks such as capital inflows (\dot{F}) or increased commodity export income (X) because:

$$e = \frac{E}{\bar{P}} = \frac{mQ(1 + \theta)}{X + (\theta M_{-1} + \dot{F})} \quad (56.10)$$

However, note also that although from (56.9) inflation targeting will affect the nominal exchange rate (higher interest rates leading to appreciation) the real exchange rate in (56.10) remains unaltered. As a whole, therefore, the policy stance applied by the Fund model is procyclical because not only is the impact of external shocks on the economy unmitigated by active domestic fiscal or monetary policy, but also any unexpected decrease in output (Q) must be met by higher interest rates to keep P (in 56.8) stable and vice versa in the inflation targeting approach.

I now adapt the model to allow for the two characteristics noted at the beginning of this section, which are essential in order adequately to describe middle-income ‘emerging market’ economies and the larger low-income countries with a domestic capital market open to foreign investment. I retain the same basic modelling framework in order to facilitate comparison between our new-Keynesian approach and the orthodox model.

The first modification is to the private credit channel. The relationship in (56.11) appears superficially similar in form to (56.6) but in fact causality

has been reversed: in a credit-rated economy monetary policy determines the level of real output (Q) as this responds to credit supply (H) within the limit of capacity (AK).¹⁸ This response comes about from both consumer credit expansion affecting demand and working capital availability affecting supply, and is a more realistic representation of emerging market economies than the Fund model. The authorities can alter the supply of money by straightforward monetary emission as an alternative to debt issue for fiscal deficit finance, by varying reserve requirements on banks or by changing the mode of financing foreign exchange reserve holdings. Note that this relationship means that raising interest rates (i) will reduce output. Only with full capacity utilization is the demand effect felt on prices (P) rather than output, and only then will reductions in money supply or higher interest rates reduce inflation. In contrast, my bank deposit function (B) function is similar to that of the Fund model:

$$\begin{aligned}
 Y &= H(\phi - \beta i) \\
 Y &= QP \\
 Q &= \frac{H}{P}(\phi - \beta i) \text{ when } Q < AK \\
 P &= \frac{H}{K}(\phi - \beta i) \text{ when } Q = AK \\
 B &= Y(v + \alpha i)
 \end{aligned}
 \tag{56.11}$$

The second modification is to open up the capital account of the balance of payments in (56.2) by expressing capital flows – changes in external liabilities (F) – as a function of domestic interest rates (i) and changes in the nominal exchange rate (E).¹⁹ This contrasts with the Fund model where capital flows (limited to aid and foreign direct investment, FDI) are entirely exogenous:

$$\dot{F} = \gamma \left(i - \frac{\dot{E}}{E} \right)
 \tag{56.12}$$

Note also that foreign investor risk appetite or world interest rates (both reflected in the parameter γ) can shift suddenly in practice, and that a sufficient imbalance between interest rates and exchange rate changes can lead to a capital outflow ($\dot{F} < 0$).

Reflecting observed practice in emerging market economies, the ‘prudential rules’ for the management of external reserves and domestic debt in our model need to be adapted to these structural characteristics. Thus instead of the import coverage rule in (56.7), the central bank maintains a reserve level adapted to the external debt position (F) as a form of insurance against external capital account shocks.²⁰

$$R = \pi F \quad (56.13)$$

And the debt solvency rule in (56.7) is applied in an intertemporal context, changing with the primary fiscal deficit (Z) and limited to a given ratio (λ) of output at full capacity:²¹

$$\begin{aligned} D &= Z + (1 + i)D_{-1} \\ \frac{D}{PK} &\leq \lambda \end{aligned} \quad (56.14)$$

Finally, although the emerging market economy is exposed to external shocks even if the nominal exchange rate is allowed to float, the real exchange rate in our model is no longer indeterminate, unlike the Fund model (56.10), because it is affected by the domestic interest rate through the capital account. This endows the monetary authorities with a degree of freedom both to ensure that exports remain competitive and to respond adequately to external shocks. The primary deficit (Z) thus becomes the policy instrument affecting inflation, while interest rates set the real exchange rate and credit levels determine capacity utilization.

We can see how this policy framework works in practice by condensing the model into three reduced-form equations.²² The first is that real credit supply (H/P) be set so as to ensure full capacity utilization. From (56.11) we have this condition as:

$$\begin{aligned} Q &= K \\ \frac{H}{P} &= \frac{K}{\phi - \beta i} \end{aligned} \quad (56.15)$$

The second is the balance of payments identity (56.2), re-expressed in terms of the exchange rate, interest rates and output by substituting in equations (56.1), (56.4), (56.12) and (56.13). Assuming that the target of stabilizing the real exchange rate is in fact achieved (that is, $\dot{e} = 0$ and thus $p = \dot{E}/E$), inserting expressions for the real exchange rate (56.10) and inflation (56.1) yields the real exchange rate (e):

$$\begin{aligned} X - \frac{mPK}{E} &= (\pi - 1)\gamma \left(i - \frac{\dot{E}}{E} \right) \\ e = \frac{E}{P} &= \frac{mK}{X + \gamma(1 - \pi)(i - p)} \end{aligned} \quad (56.16)$$

In other words, the policy instrument that determines the real exchange rate in an emerging market economy with an open capital account is the real interest rate ($i - p$). The higher the real interest rate, the more the real

exchange rate will appreciate (that is, e falls). The desirable policy target is thus clearly to maintain the real exchange rate at a stable and competitive level, keeping real interest rates low and adjusting them actively according to world market conditions.

The third reduced-form equation is derived from the monetary balance (56.3): substituting in (56.5), (56.13), (56.14) and (56.15) yields an equation for the price level (P) in terms of the other targets and instruments:

$$P = \frac{Z + (1 + i)D_{-1}}{K \left\{ (v + \alpha i) - \frac{1}{\phi - \beta i} \right\} - e\pi \{ F_{-1} + \gamma(i - p) \}} \quad (56.17)$$

Because the interest rate (i) instrument is already employed to stabilize the real exchange rate (e), and the credit level (H) is already set so as to stabilize output ($Q = K$), the budget deficit (Z) plays the key role of price stabilization. However from (56.17) it is clear that a low level of inflation does not mean that the budget should always be in balance ($Z = 0$) or even a fixed proportion of aggregate income ($Z = \lambda Y$), but rather should compensate for exogenous shocks to international asset demand (γ) or production capacity (K) even when international price fluctuations are smoothed by the real exchange rate.

In sum, in this type of open economy integrated – albeit asymmetrically – to international capital markets, a decision by the central bank to raise the interest rate in order to curb inflation from (56.17) will actually have three undesirable effects: inflation will rise due to the effect of debt service on the budget deficit; the exchange rate will appreciate through the capital account effect; and real output will fall from the credit channel effect. My more realistic new-Keynesian model for the emerging market economy thus shows not only that this is a misguided stabilization policy, but also that a wider range of policy instruments should be used to achieve multiple stabilization targets, with particular emphasis on trade competitiveness and full employment. These instruments include low real rates of interest, a balanced budget over the cycle and above all strong prudential control of bank credit.

Structural adjustment, sectoral investment and income distribution

We have also seen that emerging market economies work in different ways from that which the World Bank model supposes – the existence of surplus labour and sector-specific installed capital being crucial. This section sets out, therefore, a medium-term model with a framework similar to the 1-2-3 model used as a formal analytical basis for the RMSM, except that: (1) although production capacity may be fully utilized, this does not involve

the full employment of the labour force; and (2) shifts in production patterns are not possible except through new investment.

In the canonical dependent economy model²³ there are two sectors producing traded (T) and non-traded (N) goods from homogeneous production functions with sector-specific labour (L) that is intersectorally mobile, so that the nominal wage (W) is equalized across sectors. Total labour supply is fixed and fully employed at the equilibrium wage. For the two sectors (j):

$$Q_j = Q_j(L_j) \quad (56.19)$$

and the real exchange rate (e) is now defined as the ratio of traded (P_T) to non-traded prices (P_N):

$$e = \frac{P_T}{P_N} \quad (56.20)$$

Each sector employs labour up to the point where the marginal product of labour is equal to the single intersectoral wage,²⁴ which is thus equalized between the two sectors. Defining the real wage (w) in terms of non-traded prices (P_N) we thus get:

$$\begin{aligned} w &= \frac{W}{P_N} \\ Q'_T(L_T) &= \frac{w}{e} \\ Q'_N &= w \end{aligned} \quad (56.21)$$

It follows that the ratio of the marginal products of labour in the two sectors is equal to the real exchange rate (e) at equilibrium:²⁵

$$\frac{Q'_N(L_N)}{Q'_T(L_T)} = e \quad (56.22)$$

Demand for labour is the inverse function of the product wage and there is full employment of the total labour force (\bar{L}):

$$L_T(ew) + L_N(w) = \bar{L} \quad (56.23)$$

The real exchange rate thus drives labour allocation across the two sectors, and hence output. The sectoral supply functions become:

$$\begin{aligned} Q_j &= Q_j(e/w) \\ Q'_T &> 0, Q'_N < 0 \end{aligned} \quad (56.24)$$

And generally an inverse relationship between the real wage (w) and the real exchange rate (e) is also implied by this result, because with full employment in (56.23) and exploiting the implicit function rule:²⁶

$$w' = -\frac{wL'_N}{L'_T + eL'_N} < 0 \quad (56.25)$$

Disaggregating traded goods into exportables (x), importables (m) and non-tradable or 'home' goods and services (h) generates the 1-2-3 model (Devarajan and Robinson, 1993), which we use as a framework in order to facilitate comparison. The prices for exportables (P_x) and importables (P_m) are determined by world prices multiplied by the nominal exchange rate (E), while home goods prices (P_h) depend upon supply conditions, as the domestic market must clear to ensure full capacity utilization.

However, in addition to the unwarranted assumption of full employment of labour, the notion in the 1-2-3 model that capital can simply be moved between sectors with a constant elasticity of transformation (CET) production function is obviously implausible. In fact the key issue in structural adjustment is investment behaviour because installed capital cannot be shifted between traded and non-traded sectors, while labour cannot easily be substituted for capital when technology is largely imported. The sectoral production functions are thus better seen as separate, limited by the installed capital stock in each.²⁷

World prices for exports and imports (P_j, P_m) and thus terms of trade (τ) are exogenous, while the unit import (m_j) and labour (l_j) input coefficients are technologically fixed. As before there is a single nominal wage rate (W), but in our model it is set either institutionally or by the reserve price of labour from the household sector, rather than by labour market clearing, and unemployment persists. Sectoral employment (L_j) is thus determined by output and there is excess labour supply, due to the Leontief fixed technical coefficients:

$$\begin{aligned} Q_j &\leq A_j K_j \\ G_j &= Q_j(P_j - Wl_j - m_j E P_m) \\ L_j &= l_j Q_j, \quad \sum_j L_j < \bar{L} \\ \tau &= P_j / P_m \end{aligned} \quad (56.26)$$

Traded exportable prices (P_x) are as before, but home goods prices (P_h) are formed by a mark-up (g) on production costs, where the nominal wage as well as the exchange rate plays a central role. I use home goods prices (P_h) as the numeraire in order to define the real exchange rate (e) and the real wage rate (w):

$$\begin{aligned}
 P_x &= EP_f \\
 P_h &= (1 + g)(Wl_h + Em_h P_m) \\
 e &= \frac{EP_f}{P_h} \quad w = \frac{W}{P_h}
 \end{aligned}
 \tag{56.27}$$

I assume as before that the export sector produces at capacity because world demand is infinitely elastic to supply by our country. The demand for home goods is a proportion (a) of the aggregate factor income,²⁸ but the domestic market clearing condition will not now determine price as in the standard model, but will rather determine output within the capacity constraint. Substituting the income (56.26) and price (56.27) equations into the demand function for home goods gives the output level, therefore:

$$\begin{aligned}
 Q_x &= A_x K_x \\
 Q_h P_h &= a \sum_j (G_j + WL_j) \\
 Q_h &= \frac{a}{P_h} \left[Q_h \left\{ 1 - \frac{m_h e}{\tau} \right\} + Q_x \left\{ e - \frac{m_x e}{\tau} \right\} \right] \\
 &= Q_x \frac{\tau - m_x}{\frac{ae}{1-a}\tau + m_h} \leq A_h K_h
 \end{aligned}
 \tag{56.28}$$

Note the so-called ‘Dutch disease’ effect of an improvement in the terms of trade (τ) leading to an expansion of the home goods sector as incomes rise.

Taking the home good price (P_h) as the numeraire and rearranging (56.27), we now have real sectoral profit levels (\bar{R}_j) in terms of the real exchange rate (e_j), the real wage rate (w_j) and sectoral output (Q_j):

$$\begin{aligned}
 \bar{G}_x &= Q_x \left[e \left(1 - \frac{m_x}{\tau} \right) - wl_x \right] \\
 \bar{G}_h &= Q_h \left[1 - \frac{em_h}{\tau} - wl_h \right] \\
 \bar{G}_x(e) &> 0, \quad \bar{G}'_h(e), < 0 \quad \bar{G}'_j(w) < 0
 \end{aligned}
 \tag{56.29}$$

The channel through which the real exchange rate and the real wage rate affect profits is now evident and the implications for investment decisions and income distribution can be explored properly – something that is not done in the standard model.

I now examine in detail the investment decision for the case when e rises (that is, real depreciation) but the obverse is simple to work out; as are the consequences of other shocks such as changes in labour productivity (l). I assume for convenience that in previous periods firms have been able to adjust their capital stocks to the desired level (that is, $Q_j = A_j K_j$). From (56.29) real depreciation will raise real profits in the traded export sector and reduce them in the non-traded home goods sector; but any shift in output depends on investment, which is irreversible. The two-period ($0, 1$) problem for firms in each sector is whether to invest or not. If they do not invest, then capacity (and thus production) falls by the amortization rate (δ). If they do invest then they must do so at the level that maximizes its present value (V) discounted at the interest rate (i) where the installation cost (J) expressed in home goods prices is an increasing function²⁹ of real investment (I) using imported equipment at the world price (P_m) and the corresponding import coefficient (m_k):

$$\begin{aligned} V_{j,0} &= \bar{G}_{j,0} + \frac{\bar{G}_{j,1}}{1+i} - J_j \\ J_j &= I_j(em_k P_m + \frac{1}{2}\sigma I_j) \\ I_j &\geq 0 \\ K_{j,1} &= K_{j,0}(1 - \delta) + I_j \end{aligned} \tag{56.30}$$

The optimal investment level (\bar{I}) is simply found by differentiating V with respect to I in order to maximize the present value of the firm:

$$\begin{aligned} \frac{dV}{dI} &= \frac{1}{1+i} \frac{d\bar{G}}{dI} - e \frac{m_k}{\tau} - \sigma I = 0 \\ \bar{I} &= \frac{1}{\sigma} \left\{ \frac{1}{1+i} \frac{d\bar{G}}{dK} - e \frac{m_k}{\tau} \right\} \end{aligned} \tag{56.31}$$

Because the capital stock in the home goods sector (K_h) was adjusted to the previous real exchange rate so as to maximize profits, from (56.29) no investment takes place in the sector (that is, $I_h = 0$) and capacity declines by the amortization rate:

$$K_{h,1} = K_{h,0}(1 - \delta) \tag{56.32}$$

In contrast, the export firms do invest as real profits have risen and:

$$\begin{aligned} \tilde{I}_x &= \frac{1}{\sigma} \left\{ \frac{A_x}{1+i} \left[e \left(1 - \frac{m_x}{\tau} \right) - w l_x \right] - e \frac{m_k}{\tau} \right\} \\ K_{x,1} &= K_{x,0}(1 - \delta) + \tilde{I}_x \\ \tilde{I}'_x(e) &> 0, \tilde{I}'_x(\tau) > 0, \tilde{I}'_x(w) < 0, \tilde{I}'_x(i) < 0 \end{aligned} \quad (56.33)$$

Note that improved terms of trade or real depreciation raise export sector investment, while higher wages or interest rates reduce it.

Traded output capacity thus rises while non-traded output capacity falls, due to the changes in their respective capital stocks. Aggregate real output (Y) only rises if the net output capacity shift is large enough:

$$\begin{aligned} Y &= eQ_x + Q_h \\ Y_1 &> Y_0 \text{ if } eA_x(K_{x,1} - K_{x,0}) > \delta A_h K_{h,0} \end{aligned} \quad (56.34)$$

Moreover, even if the adjustment is sufficient to cause aggregate real output to rise, there is no necessary reason why net employment should do so as well. The general condition for this to happen is found by substituting the employment functions from (56.27) into (56.33) to yield:

$$L_1 > L_0 \text{ when } Y_1 > Y_0 \text{ if } l_x > e l_h \quad (56.35)$$

This result has three interesting characteristics. First, there is no guarantee that total employment will increase with real devaluation because this depends on the labour intensities (l) of each sector as well as the investment outcome: clearly only if the export sector is the more labour-intensive ($l_x > l_h$) is this likely to occur. Second, the ratio of the real wage rate to the real exchange rate (w/e) is clearly critical to the outcome. If real wages fall then employment will rise, although this trade-off is not the result of factor substitution along the constant elasticity of transformation (CET) curve, but rather of investment incentives. Third, an increase in the interest rate (i) will reduce the employment gain due to the investment disincentive (56.33): this is the reverse of the factor substitution effect textbook theory would predict. Moreover, for investment to take place, private investors must have confidence in future profits and be provided with sufficient credit and infrastructure.

As we have seen in (56.25), the dependent economy model implies that depreciation of the real exchange rate will reduce the real wage rate, which with full employment implies a deterioration of the overall distribution of income. The new-Keynesian model reveals a more complex relationship. Rearranging (56.27) we have:

$$w = \frac{W}{P_h} = \frac{1}{l_h} \left\{ \frac{1}{1+g} - e^{-\frac{m_h}{\tau}} \right\} \quad (56.36)$$

which has interesting characteristics: there is again a negative relationship between real wages and the real exchange rate; but the home goods profit mark-up also plays an important part in income distribution, and productivity in the home goods sector (the inverse of the labour input coefficient l_h) is also a key determinant of real wages. This last point should remind us that while traded investment is the central focus of adjustment policy, improved living standards require a greater supply of wage-goods.

Further, if I define workers' living standards (ω) as the nominal wage deflated by the cost of living (P_c) defined by the mean domestic product prices weighted by home goods consumption propensity (a) in (56.28):

$$\omega = \frac{W}{P_c}$$

$$P_c = aP_h + (1-a)P_x \quad (56.37)$$

Substituting the relevant definitions from (56.29) into (56.37) then gives:

$$\omega = \frac{w}{a + (1-a)e} \quad (56.38)$$

Here the welfare consequence of the fall in the w/e ratio in (56.33) required to raise investment and permit structural adjustment is clearly revealed as a deterioration in workers' living standards.

In other words, the targeting of the real exchange rate necessary in order to maintain export competitiveness is in fact an 'incomes policy'. If employment expands then overall income distribution can improve, but this requires active intervention in order to raise investment rates as we have shown. Monetary policy should be geared to low real interest rates and producer credit provision on the one hand, and an active fiscal stance geared to damp exogenous macroeconomic shocks on the other. This desirable outcome can be reinforced by a system of dividend taxation (designed to stimulate investment) and social spending specifically aimed at raising the living standards of employees' families.³⁰

Conclusions

In this chapter we have seen how the macroeconomic models conventionally used to analyse stabilization policy and structural adjustment are open to a wide range of criticism from both neoclassical and Keynesian standpoints. Specifically, the existence of excess productive capacity, mark-up pricing by firms, credit rationing by banks and open capital accounts on the

one hand; and the central role of investment in determining sectoral output and the impact of the real exchange rate on both export levels and real wages on the other; mean that the standard IMF and World Bank models can become seriously misleading as the basis for macroeconomic policy.

By including more realistic formulations of credit supply and external capital flows for emerging markets in the standard stabilization model, I have shown not only that inflation targeting using the interest rate can have serious procyclical consequences, but also that a more active monetary policy based on fiscal and credit instruments can make for higher and more stable output solutions – that is, sustainable stabilization. Similarly, by including the pricing and investment behaviour of firms in the standard adjustment model, I have shown that effective structural adjustment will not take place unless active exchange rate and monetary policies are implemented in support of traded production.

Finally, this new-Keynesian approach also differs from that of the Bretton Woods institutions in its attitude to macroeconomic intervention. The Bank and the Fund insist that inflation-reduced public expenditure and balanced budgets are essential in order to promote growth. This passive ‘rules-bound’ approach is central to their lending conditionality, and is built into significant institutional reforms such as central bank independence. In marked contrast, this chapter has shown how emerging market authorities can combine active fiscal and credit management with real exchange rate targeting in order to cope with exogenous shocks and promote longer-term export-led growth in a more purposive stabilization policy. Finally, it has demonstrated that active intervention to maintain both a competitive real exchange rate and a low real interest rate is necessary in order to promote sufficient investment in the traded sector and thus ensure not only export growth but also employment expansion so that real wage constraints do not lead to a worsening income distribution as a consequence of structural adjustment.

Notes

1. This chapter does not address issues such as privatization of public enterprise or government expenditure reform that, while central to structural adjustment and stabilization policy in practice, are covered in other chapters of this *Handbook*.
2. These two well-known models can be found in the official sources referenced below and are summarized clearly in Khan et al. (1990) and Agénor (2000).
3. Usually government expenditure cuts rather than increases direct tax pressure – which would depress saving, assumed to be a constant proportion of private disposable income in this model – with obvious distributional consequences.
4. In consequence, it might better be described as a ‘three gap model’ (Bacha, 1990).
5. In the sense a small open economy that is a price-taker in world markets, popularized by Dornbusch (1986), rather than the wider Latin American notion of *dependencia*.
6. This model is lucidly set out in Montiel (2003, Part V).
7. Because there are one country, two sectors and three products in the model.

8. See Corden (1984). In fact 'British disease' would be more appropriate. The Netherlands did indeed experience real exchange rate appreciation in the 1980s due to North Sea gas finds, which rendered manufactured exports uncompetitive; but the fiscal resources were reinvested in infrastructure and skills, with subsequent growth based on advanced services exports. The Thatcher administration used the North Sea royalties to reduce UK profit taxes, stimulating consumption and depreciating the real exchange rate; but the consequences are now visible in deteriorating public transport and education.
9. See Obstfeld and Rogoff (1997).
10. See Aghion and Howitt (1998).
11. See Buiter (1990).
12. As opposed to the justifiable concern with hyperinflation in the 1980s.
13. Strictly speaking, bank deposits in the Fund's BFPF model.
14. This implies unitary price and income elasticities of import demand, which are much higher than those found empirically and means that import contraction through reductions in real demand (Y/P) become the main channel for stabilizing the current account of the balance of payments.
15. The conventional rule of thumb is three months' cover (that is, $\theta = 0.25$).
16. As in the 'Maastricht Criteria', where $\lambda = 0.03$.
17. This corresponds to the IMF definition of the 'effective real exchange rate' as the nominal rate divided by the ratio of the domestic price level (P) to the weighted mean of the price levels in trading partners (unity in our case). The alternative definition of the real exchange rate in terms of the ratio of traded to non-traded prices is discussed below. On both definitions, see Dornbusch and Helmers (1988) and Montiel (2003).
18. Which in turn is given in the short run but depends on investment in the medium term, a point taken up below.
19. This is a simplified form of the full international demand function for emerging-market assets, which itself can be derived from standard portfolio theory (FitzGerald, 2006).
20. The most cautious position would be to maintain reserves equal to short-term external liabilities, commonly known as the 'Greenspan rule'. The value of π will then depend on the maturity structure of external liabilities (F).
21. See Missale (1999) for a discussion of optimal debt models in a full intertemporal context, from which this familiar rule is derived.
22. Formally, Walras' Condition is satisfied because the model has 11 equations and 11 variables (eight endogenous and three targets); while Tinbergen's criterion is met by having three instruments (H , i and Z) with which to hit the three targets (Q , e and P).
23. The model is very well explained in Chapter 2 of Agénor and Montiel (1999).
24. That is, $P_j Q_j(L_j) = W$
25. Note that this is the formulation used to derive the Harrod–Belassa–Samuelson model of real exchange rate trends in the long run.
26. If $F(x,y) = \text{constant}$, then $dx/dy = -F'_x/F'_y$
27. In the standard 1-2-3 model set out above, there is a domestic production function with a constant elasticity of transformation (σ)

$$Q = A[\gamma Q_x^\rho + (1 - \gamma) Q_h^\rho]^{\frac{1}{\rho}} \quad \sigma = \frac{1}{\rho - 1}, \quad 0 < \sigma < \infty$$

where the convention is to adopt a Cobb–Douglas unitary elasticity ($\sigma = 1$). Below I adopt the more realistic Leontief form ($\sigma = 0$) to reflect the fact that once installed, capital is entirely immobile.

28. This form is in fact the constant elasticity of substitution (CES) consumption function used in the canonical 1-2-3 model discussed above, but with unitary own-price elasticity and zero (Leontief) substitution between home and export goods.
29. This is a standard representation of the intertemporal optimisation process for the firm, which as part of the 'AK' model underpins endogenous growth theory – see Heijdra and van der Ploeg (2000, Chapter 2). In developing countries this can also be seen as reflecting limited local project implementation capacity.

30. This topic lies beyond the scope of this chapter, but see FitzGerald (1993) for a further discussion of such a policy based on dividend taxation and social expenditure; and FitzGerald (2002) for the derivation of an optimal profits tax to fund infrastructure provision.

References

- Addison, D. (1989), 'The World Bank Revised Minimum Standard Model: Concepts and Issues', Working Paper Series 231, Washington, DC: World Bank; updated (1999) version at <http://www.worldbank.org/data/rmsm/index.htm>.
- Agénor, P.-R. (2000), *The Economics of Adjustment and Growth*, San Diego, CA: Academic Press.
- Agénor, P.-R. and P.J. Montiel (1999), *Development Macroeconomics*, 2nd edn, Princeton, NJ: Princeton University Press.
- Aghion, P. and P. Howitt (1998), *Endogenous Growth Theory*, Cambridge, MA: MIT Press.
- Bacha, E.L. (1990), 'A Three-Gap Model of Foreign Transfers and the GDP Growth Rate in Developing Countries', *Journal of Development Economics*, 32 (2): 279–96.
- Blinder, A.S. (1987), 'Credit Rationing and Effective Supply Failures', *Economic Journal*, 97: 327–52.
- Bourgignon, F. and C. Morrison (1992), *Adjustment with Equity in Developing Countries*, Paris: OECD.
- Buiter, W. (1988), 'Structural and Stabilization Aspects of Fiscal and Financial Policy in the Dependent Economy', *Oxford Economic Papers*, 40: 220–45.
- Buiter, W.H. (1990), *Principles of Budgetary and Financial Policy*, New York: Harvester Wheatsheaf.
- Chenery, H.B. and A.M. Strout (1966), 'Foreign Assistance and Economic Development', *American Economic Review*, 66: 679–733.
- Collier, P. and J.W. Gunning (1999), 'The IMF's Role in External Adjustment', *Economic Journal*, 109: 634–51.
- Corden, M. (1984), 'Booming Sector and Dutch Disease Economics: Survey and Consolidation', *Oxford Economic Papers*, 36: 359–80.
- Cornia, G.A., R. Jolly and F. Stewart (eds) (1987), *Adjustment with a Human Face*, Oxford: Oxford University Press.
- Devarajan, D. and S. Robinson (eds) (1993), *Applied General Equilibrium Modeling*, New York: Elsevier Science.
- Dornbusch, R. (1986), *Open Economy Macroeconomics*, New York: Basic Books.
- Dornbusch, R. and F. Helmers (1988), *The Open Economy: Tools for Policymakers in Developing Countries*, New York: Oxford University Press.
- Easterly, W. (1999), 'The Ghost of the Financing Gap: The Growth Model Used in the International Financial Institutions', *Journal of Development Economics*, 60: 423–38.
- FitzGerald, V. (1993), *The Macroeconomics of Development Finance*, Basingstoke: Macmillan.
- FitzGerald, V. (2002), 'International Tax Cooperation and Capital Mobility', *Oxford Development Studies*, 30 (3): 251–66.
- FitzGerald, V. (2003), *Global Markets and the Developing Economy*, Basingstoke: Palgrave.
- FitzGerald, V. (2006), 'International Risk Tolerance, Capital Market Failure and Capital Flows to Developing Countries', in A. Shorrocks (ed.), *Advancing Development: Core Themes in Global Development*, Basingstoke: Palgrave Macmillan for WIDER, pp. 299–318.
- Heijdra, B.J. and F. van der Ploeg (2000), *The Foundations of Modern Macroeconomics*, Oxford: Oxford University Press.
- IMF (1987), 'Theoretical Aspects of the Design of Fund-Supported Adjustment Programs', IMF Occasional Paper 55, Washington, DC, International Monetary Fund.
- Jansen, K. and R. Vos (eds) (1997), *External Finance and Adjustment: Failure and Success in the Developing World*, Basingstoke: Macmillan and New York: St Martin's Press.
- Khan, M.S., P.J. Montiel and N.U. Haque (1990), 'Adjustment with Growth: Relating the Analytical Approaches of the World Bank and the IMF', *Journal of Development Economics*, 32: 155–79.

- Missale, A. (1999), *Public Debt Management*, Oxford: Oxford University Press.
- Montiel, P.J. (2003), *Macroeconomics in Emerging Markets*, Cambridge: Cambridge University Press.
- Mosley, P., J. Harrigan and J. Toye (1995), *Aid and Power: The World Bank and Policy-Based Lending*, 2nd edn, 2 vols, London: Routledge.
- Mussa, M. and M. Savastano (1999), 'The IMF Approach to Economic Stabilization', in B.S. Bernanke and J.L. Rotemberg (eds), *NBER Macroeconomics Annual 1999*, pp. 79–128.
- Obstfeldt, M. and K. Rogoff (1997), *Foundations of International Macroeconomics*, Cambridge, MA: MIT Press.
- Ocampo, J.A. (2000), 'Developing Countries' Anti-cyclical Policies in a Globalized World', *Temas de Coyuntura 13*, Santiago de Chile: Economic Commission for Latin America and the Caribbean.
- Polak, J.J. (1957), 'Monetary Analysis of Income Formation and Payments Problems', *IMF Staff Papers*, 6 (1): 1–50.
- Rodrik, D. (1991), 'Policy Uncertainty and Private Investment in Developing Countries', *Journal of Development Economics*, 36: 229–42.
- Stiglitz, J.E. and B. Greenwald (2003), *Towards a New Paradigm in Monetary Economics*, Cambridge: Cambridge University Press.
- Taylor, L. (1988), *Varieties of Stabilization Experience*, New York: Basic Books.
- Taylor, L. (ed.) (1993), *The Rocky Road to Reform: Adjustment, Income Distribution, and Growth in the Developing World*, Boston, MA: MIT Press for UNU/WIDER.
- Williamson, J. (1997), 'The Washington Consensus Revisited', in L. Emmerij (ed.), *Economic and Social Development into the XXI Century*, Washington, DC: Inter-American Development Bank.
- Wood, A. (1994), *North–South Trade, Employment and Inequality: Changing Fortunes in a Skill-Driven World*, Oxford: Clarendon Press.

57 Economic planning in developing economies

*Bill Gibson*¹

Introduction

'Planning' is a term that generally has fallen into disuse. Todaro defines development planning as 'the conscious effort of a central organization to influence, direct and in some cases even control changes in the principal economic variables (such as GDP, consumption, investment, savings, etc) of a certain country or region, over the course of time in accordance with a predetermined set of objectives' (Todaro, 1971, p. 1). Planning connotes, but does not logically imply, command-and-control mechanisms by which authorities issue directives for which compliance becomes a matter of administrative law.

Development planning was attempted in the Soviet Union and Eastern Europe and to a degree in India, Cuba, Egypt and Tanzania and other countries in the immediate post-World War II period. Indeed, it was largely the success of the Soviet Union in raising per capita incomes in the first half of the twentieth century that demonstrated the existence of a practical alternative to market allocation. Soviet performance impressed policy-makers in developing economies who had come to see the market as inadequate to the task of industrialization. In non-communist countries, planning without enforceable command-and-control mechanisms was widespread in the immediate post-World War II period. The United Nations and other sources even withheld development aid unless a plan was in place and as a result, planning ministries became commonplace throughout the developing world. Planning models that demonstrated how foreign aid could be coordinated to achieve maximum impact on growth and development were especially popular. Despite its increasing technical sophistication and theoretical appeal, planning in the post-World War II period led to widespread disillusion and rejection by even formerly ardent supporters. By the end of the 1970s, Chowdhury and Kirkpatrick noted that many economists were talking openly about the failure of planning, and as early as 1965, Waterson had concluded on the basis of a study of 55 country experiences that 'the majority of countries have failed to realize even modest income and output targets' (Chowdhury and Kirkpatrick, 1994, p. 2).

Since the 1970s explicit plans in developing countries have largely been abandoned. Many of the problems planning was designed to confront are

still present, of course, and the need for some kinds of planning persists. As a result, planning has re-emerged in a more market-friendly variant, development policy management, with emphasis on the price mechanism, incentives and schemes such as 'cap-and-trade', that rely heavily on decentralized implementation.

This chapter critically reviews planning as applied to developing countries. The next section addresses the general question of the relationship of planning to the market. Economy-wide planning models and techniques are discussed in the following section, while the subsequent section turns to microeconomic planning and cost-benefit analysis. A concluding section discusses the uses of surviving planning models in the current context.

Plan versus market

Plans may either be economy-wide or partial. Heal reviews the theory underlying the economy-wide planning procedures (Heal, 1973). He notes that much of the early writing on planning, by distinguished economists such as Lange, Lerner, Arrow and Hurwicz, sought to establish that an efficient centrally planned system would employ the same marginal equalities as in the Walrasian system, with the central planning board playing the role of the auctioneer. Plans in which individual preferences are constitutive of the social objective function, therefore, yield the same pattern of resource allocation as would a competitive market. In other words, there is nothing inherently inefficient about planning. This conclusion is today widely accepted inasmuch as planners' preferences often proxy a social welfare function under the assumption that a freely functioning competitive market mechanism would produce an identical allocation of scarce resources.

In partially planned economies, planning is generally conceived as a response to market failure, including externalities, informational asymmetries and public goods. If market failure is widespread then it follows that central planning can serve as a substitute for the market; if not, then planning can, in principle, resolve allocational issues related to market failure.

In addition to concerns about market inefficiencies, equity was also considered a legitimate objective. The Coase theorem holds that efficiency and equity are separable, but the distinction in the early days of planning was not clearly recognized. Force-draft industrialization had achieved rapid modernization in the Soviet Union, but at the great expense of a debilitated agricultural sector. The First Five Year Plan under Nehru in India in the early 1950s explicitly prioritized reducing unemployment and poverty over maximizing the rate of economic growth. The principle that income could be redistributed without disturbing the price-guided marginal equality of social costs and benefits was ignored.

There is significant disagreement as to the extent to which government can improve outcomes by realigning social and private costs. In standard theory, a properly tuned set of taxes and subsidies could repair markets that failed and public sector institutions could fill in when markets were missing altogether. In developing-country practice, however, public policy often did not improve outcomes and the term 'government failure' gained currency to describe counterproductive intervention by states. The necessity of a one-to-one relationship between policy objectives and policy instruments, originally due to Tinbergen, shows how precarious is the entire mission. The collapse of earlier planning initiatives was in part due to a mismatch in this relationship, with goals grossly exceeding the number of instruments, other than command-and-control, available for implementation.

Killick (1976) provides a comprehensive discussion of government failure in development planning. He argues that the plans failed because their creators assumed that politicians see the planning problem 'essentially as economists do'. The assumption that governments are composed of 'public spirited, knowledgeable and goal oriented politicians . . . clear and united in their objectives, choosing policies which will achieve the optimal results for the national interests', is unwarranted. Anticipating much of the subsequent public choice literature, Killick argues that politicians should be seen as rational, self-interested, acting to maximize the short-term probability that they will be re-elected. The implicit assumption of the existence of a 'benevolent despot' was at variance with both reality and the 'liberal-individualist' tradition of Western civilization.

Modern public choice theory does indeed suggest that planning will be undertaken for the benefit of the planners themselves or their clients, and that command-and-control directives will give rise to rent-seeking behavior and other principal-agent problems that deprive a country of needed resources and talents. In democratically organized societies, a major problem arises when costs of a directive are widely distributed, while benefits accrue to a smaller set of individuals. Significant pressure to change course can develop as a result, with powerful groups lobbying effectively to push the economy away from its socially optimal path.

Planning, *ipso facto*, could never have resolved these deeper issues of policy-making. Problems of coordination, incentives and the trade-off between efficiency and equity are at the root of the problem of underdevelopment and were beyond the reach of technocratic planners and their tools. Planning was abandoned within a broader current of change that involved rethinking the role of the state generally. Much of the planning literature reviewed above is now seen as archaic, and to say that planning is 'out of fashion' is an understatement. As demand management, automatic stabilizers, incomes policy and the Phillips curve gave way to rational expectations,

new public choice theory and inflation targeting, planning became caught up in a generalized retreat from *dirigisme* and ascendancy of the market mechanism. Killick, who had so thoroughly excoriated the planning process earlier on, came later to wonder if there had not been ‘a reaction too far’ in moving so decisively away from planning toward the market. We shall return to this issue in the concluding section.

Economy-wide planning models

Planning models can be classified in several different categories: aggregate, main sector, multisectoral, regional and project-specific models. Economy-wide models include the first three categories, but not the last two, and may be static or dynamic. They typically reflect the accounting regularities and conventions of national income and product accounts, balance of payments, and income and expenditure balances of the public sector (Taylor, 1979). These can be simulation models or more traditional econometric constructs. The former employ informal calibration procedures, while the latter are estimated formally, using statistical theory under the usual assumptions. The simulation approach does not rely on statistical theory, but rather on whether the model captures salient features of the economy (Gibson, 2003).

Input–output models

Economy-wide planning models have their roots in the model first described by the young Harvard economist W. Leontief in the 1930s and 1940s. The inter-industry or input–output approach pioneered by Leontief, and first implemented with his help in the Soviet Union, served a means by which consistent intersectoral plans could be drawn up. Input–output models have their roots in Quesnay’s *Tableau Economique*, a Physiocratic device that was the first effectively to separate real from nominal resources flows.

Input–output models are used to analyze the impact of a change in final demand on the levels of production. Let $A = \{a_{ij}\}$ be the coefficient matrix such that each a_{ij} describes the use of input i for the production of one unit of output j , and $X = \{X_j\}$ be a column vector of gross outputs, including intermediate goods. Factors of production, labor, L , and capital, K , are treated separately, usually with fixed coefficients under the assumption that factor prices remain unchanged. Together with the labor and capital coefficients, the A matrix represents the technique by which goods are produced.

Final demand is denoted by $F = \{F_j\}$, a column vector of outputs, and may be disaggregated into consumption, government spending, exports and imports as needed. The essential equation of input–output analysis, known as the material balance, is then:

$$X = AX + F. \tag{57.1}$$

In a one-commodity world, say corn, let the output of six ears require the input in the form of seed corn of one ear. In order to consume 500 ears of corn, we must then produce $X = (1/6)X + 500$ or $X = 600$ ears to make sure that there is enough for both final, 500 and intermediate, or seed, of 100. Since we could just as easily produce 1200 ears with 200 ears of seed, the model evidently assumes constant returns to scale.

So-called dual variables can also be defined and interpreted as prices, denoted here by row vector $P = \{p_j\}$. The equation dual to the material balance is then:

$$P = PA + V_A \tag{57.2}$$

where V_A is a row vector of value added, and may be disaggregated into wages, profits, imports, taxes and rents as needed.

As a result of a linear production technology, input-output models are relatively inexpensive and easy to formulate and run. Since prices were often administered and incentives less relevant, the absence of a functioning price mechanism in the model was unimportant. They were, consequently, enormously popular in early development planning.

The model can be made dynamic if investment, I , is first disaggregated from final demand, F , and then used to determine the time path of capital stock, K . This is done by way of the ‘stock-flow equation’:

$$K_t = K_{t-1}(1 - \delta) + I$$

where δ is the depreciation rate. Consistent forecasts of intermediate demand, labor and foreign exchange requirements, for example, could then be made, contingent on a forecast for investment.

The framework just presented is the ‘open Leontief model’, but a closed version is available in which all elements of final demand and value added are made dependent on X . It was left to von Neumann to show that a maximum rate of sustainable growth is well defined by the model. Despite the elegance of the von Neumann model itself, there were limited direct practical implications of the closed Leontief model for planning. One reason was that prices played virtually no role whatsoever; the technical coefficients, whether for capital or feeding labor, determined the entire balanced growth path. ‘Turnpike optimality’, as exhibited by the von Neumann model, was intellectually appealing, but offered little insight into the nature of the far bumpier road on which developing countries were traveling.

Linear programming models

The chief limitation of all linear models is that they do not allow for substitution in response to changing prices of goods and factors. A partial solution to this problem is provided by the linear programming approach. Introduced by Dantzig for the US Air Force in 1947 and popularized in a classic text by Dorfman, Samuelson and Solow, linear programming models allowed for prices to have a limited impact on the allocation of resources (Dorfman et al., 1958). Unlike their more rigid input–output counterpart, linear programming models could be set, for example, to maximize employment by choosing a sectoral pattern of output consistent with a foreign exchange constraint or some other supply-side limitation (Blitzer et al., 1975).

In a typical linear programming model, there is usually more than one feasible solution. The feasible solutions are then ranked according to an explicit objective function that depends on prices of goods and factors, or some other methods of valuation. An optimal primal solution satisfies all constraints and provides a maximum of the objective function. Like input–output models, there is a dual solution which minimizes the value of the dual objective function. A powerful and fundamental duality theorem of linear programming establishes complementary slackness, which holds that if a constraint in the primal solution of a linear program does not bind, that is, it is satisfied only as an inequality, then the corresponding dual variable is zero. In less formal language, an additional unit of a resource that was already in excess supply could have no effect on social welfare.

The impact of complementary slackness on development planning cannot be overestimated, since at once the notion clarified the relationship of a ‘social optimum’, however planners wished to define it, to factor abundance and the related production technology. From linear programming and complementary slackness, practitioners derived the idea of a ‘shadow price’, or the change in the value of the social objective function with respect to a change in the quantity of a specific binding resource. Now the size of the wedge between the social and private cost of resources could be computed. The application was immediate: in economies with surplus labor, the shadow value of unskilled labor was effectively zero and thus planners would be justified in substituting a lower than market wage when computing the social cost of any particular project or policy intervention.

While linear programming models allow for choice of technique, they do not allow for smooth substitution and infinite divisibility between discrete techniques. This may in practice be more realistic but does give rise to jumps in the values of the solution variables. Data permitting, smooth substitution can always be approximated to any degree of accuracy by

increasing the number of available techniques of production. Moreover, since linear programming models are special cases of non-linear programming models, computer software available for the solution of the latter, for example, General Algebraic Modeling System (GAMS), Matlab, Mathematica, and so on, also compute solutions for the former. Specialized packages exist for linear programming problems, such as Lindo, that are fast, efficient and give highly detailed computational results.

That linear programming would show only how one analyzes resource constraints given the objective function, rather than the deeper problem of how social objectives are themselves to be defined, would ultimately lead to its undoing. But for a while, the technique enjoyed immense popularity, and still does in many specialized applications. Moreover, that it could neatly separate the role of the policy-maker, who determined the coefficients in the objective function, from that of the economist or planner, who designed, built and ran the model, only enhanced its scientific patina.

SAMs and CGE models

Social accounting matrices (SAMs) extend the usual conceptual categories of input–output frameworks to account for more detailed expenditure and distributional categories. The constructs are not properly referred to as ‘models’, but rather serve as a database to which behavior equations can be calibrated.

Just as linear programming generalized input–output analysis, so computable general equilibrium (CGE) models take the next step in integrating price signals in more fundamental ways (Gunning and Keyzer, 1995). They are usually multisectoral, economy-wide models, calibrated to SAMs. They may be static or dynamic with short-run coverage of one to three years, three to seven for the medium run, and long-term models that extends beyond a decade. Static models compare two points in time without explicit attention to the path connecting these points, while dynamic models trace out a locus of points with explicit stock-flow adjustment processes. The models may exhibit a wide range of adjustment mechanisms, from closed, purely competitive, Walrasian models to macro structuralist models in which foreign exchange availability determines the level of output in some key sectors.

The structure of a typical CGE model can be briefly sketched as follows. Beginning with the material balance, in equation (57.1), the model links the various elements of final demand to goods prices and incomes. Factor demand equations determine factor prices when supply is binding, but this need not be the case and some other mechanism might be introduced to determine nominal factor prices. CGE models (CGEs) can be constructed

in real or nominal terms, but it is a characteristic feature of structuralist CGEs that equations are given for nominal quantities, which are then converted into real terms by the price vector that results from general equilibrium. This implies that money or some other nominal quantity be fixed exogenously and thus, inflation can be modeled in dynamic systems.

There is, of course, no need to specify supply and demand equations in CGEs since the underlying determinants are modeled directly. Production functions combine labor, L , and capital, K , so that equation (57.1) can be expressed as:

$$X(K,L) = AX + F(P, Y)$$

with final demand written as function of income, Y , given by:

$$Y = V_A X.$$

Since value added depends on factor supplies, equation (57.2) should be re-expressed as:

$$P = PA + V_A(K,L).$$

Unlike the input–output or linear-programming models, both X and P must be solved for simultaneously. Prices appear throughout the model in a more integral way, causing substitution of both goods and factors and determining incomes.

It follows that the elasticities of substitution must be carefully calibrated for each application. Overestimating these elasticities implies a failure to recognize structural rigidities that may be present in the actual economy. Models in which response elasticities are too high underestimate the effect of policies, since the model allows adjustment in both production and consumption to be smoothly and easily accomplished. In the real economy, there may be significant transactions costs associated with substitution, and thus policies may be more effective in the real economy than in the model.

Dynamic CGEs are more cumbersome, and to the extent that they are designed to reflect Walrasian dynamic adjustment mechanisms, with perfect foresight, are less realistic than models which depend on an explicit investment function. The latter can employ parameters that are econometrically estimated to enhance realism. Dynamic CGEs can be calibrated to historical time series in the same way as large econometric models can be, and provide much more detailed and consistent information than typical time-series models.

Environmental planning models

Computable general equilibrium models in theory can be extended to address a range of related policy problems, such as environmental components. So long as stable contaminant coefficients can be found and linked to production and consumption levels, the models can generate an endogenously determined estimate of environmental quality along with its forecasts for production, consumption, investment and international trade. There are several important problems of implementation, however, the first of which is that contaminant levels can vary significantly between two industries that have been aggregated into a larger category, and even within an industry pollution levels can vary between two firms. Moreover, the coefficients presently in use are derived from studies of US manufacturing firms and one can only guess how these coefficients would need to be adjusted to conform to conditions in developing countries.

Environmental policy analysis thus requires considerable sophistication. Without some detailed microeconomic analysis built into the model, it might become difficult or impossible to judge how firms would react to the introduction of tradable emissions permits, that is, pollution rights that can be bought or sold in a specified market. Earlier planning models could adequately capture a command-and-control system that targeted output levels, but would fail to capture more nuanced response to cap-and-trade policies, such as the time-phasing of investment in compliant technologies. Moreover, models that do not include a feedback loop from the toxic contaminants to price or output levels would also fail to capture reality. While environmentally augmented CGEs have been employed in a small subset of developing countries, they are in their infancy.

Growth and long-term planning models

Even if resources are efficiently allocated statically, a sequence of Pareto-optimal states need not be Pareto optimal when viewed as a sequence (Dorfman et al., 1958). Hence markets may function well to allocate resources over space, yet do a poor job over time. This is especially difficult when the allocation problem stretches over generations, some of which are not yet born. Heal has argued recently that markets systematically err in valuing the future. Thus, inadequate capital accumulation due to uninsurable risk, credit rationing, asymmetric information and other imperfections is related to, though not the same thing as, imperfections that block trades between agents who happen to be alive at the same time. In this limited but important regard, the coefficients in the planner's objective function may be more accurate than market-determined weights.

Growth models have a distinguished history in planning, stretching back to the 1920s when issues of capital accumulation were first addressed in the

Soviet Union. Following Fel'dman's work in the USSR, Indian statistician P.C. Mahalanobis in 1953 developed a two-sector model that examined the allocation of investment between capital-good-producing and consumption-good-producing sectors and implied that investment needs to be allocated to the former sector to increase the rate of growth of the economy. By implicitly ignoring the agricultural sector (under the assumption that cheap food could be extracted from the agricultural sector through favorable terms of trade), it was a real failure of planning inasmuch as agriculture stagnated and 'cheap food' became expensive, sometimes prohibitively so. The models eventually fell out of fashion.

Following the emergence of the one-sector Solow model in the 1950s, gap models, essentially aggregate growth models with both savings and trade constraints, became popular planning tools. Gap models continue to be used to resolve issues of whether faster growth will be self-canceling by stimulating imports to the point that a balance-of-payments crisis develops.

Under very restrictive conditions, dynamic planning models can be used to determine optimal accumulation paths far into the future. One of the most well-known early models in economics, due to Ramsey, employs the calculus of variations to find the optimal savings rate, the one that maximizes the discounted value of future consumption. Despite their technical sophistication, these optimal growth models, like the von Neumann model, never guided real planning exercises in any important or practical way. Similarly, endogenous growth models have been current since the 1990s, but neither have they gained much traction for development planning.

Regional models

Regional models comprise a final subcategory of planning models. Since data requirements are hefty and data availability is sometimes scanty, regional models have lagged in application. The exception was in Eastern Europe, where data were more abundantly available, even if fabricated out of legal necessity. It is clear, however, that in the case of India and China, which together comprise almost half the developing world, regional models are not merely desirable, but unquestionably necessary. Combining regions in China could be as misleading as aggregating North and South America, and therefore aggregate models could grossly distort the true state of economic activity.

Micro-level planning

In the 1970s there was an explicit attempt to integrate micro planning into comprehensive models that were used to check consistency and direct and indirect effects of policies. The two best known were the Organisation for Economic Co-operation and Development (OECD) manual written

by Little and Mirrlees (1974) and the *UNIDO Guidelines for Project Evaluation* by Dasgupta et al. (1972).

Cost-benefit analysis

Public sector projects for electrification, hydrological development or transportation and communications infrastructure are key components of any development plan. Costs and benefits of projects are optimally evaluated using an hierarchical methodology in which the project is sequentially evaluated at ever higher levels of aggregation. Eventually, of course, the model may not 'see' the project, simply because the project is too small to matter at the aggregate level.

The private sector criterion for project acceptance is either that the present discounted value of costs and benefits as they are distributed over time should be positive, or that the internal rate of return of these same costs and benefits exceeds the cost of capital to the firm. Because externalities are so prominent in developing countries, however, the private project selection procedure has long been considered inadequate for use by development planners. While the present value template itself is appropriate, it is social – rather than private – costs and benefits that must be reconciled. Shadow, rather than market, prices are then used to evaluate project costs and benefits.

As discussed above, shadow prices are intended to reflect the marginal social benefit of available resources. Computation of these shadow prices, however, is fraught with controversy due to the large number of assumptions required for their determination. Projects that would utterly fail a private screening can, perhaps, be accepted using one method of computing shadow prices, but not another. Since shadow prices purportedly measure the marginal impact of aggregate welfare, the whole procedure had now to be vetted by the political process. This dulled the technocratic gloss that project evaluation had acquired under the direction of the authors cited above.

As noted, the linear programming approach imputes a shadow value of zero to factors of production in excess supply. Much of the early literature was devoted to calculating shadow prices in specific markets: labor, both skilled and unskilled, foreign exchange, and capital markets. The United Nations Industrial Development Organization (UNIDO) guidelines developed an extensive analysis consistent with optimal accumulation paths in surplus labor economies, all done in an analytically rigorous fashion (Dasgupta et al., 1972). Recently, more elaborate economy-wide simulation models have been used to calculate shadow values, but have not escaped intense methodological criticism since so much depends on the objectivity of the price scheme.

In retrospect, it is hardly surprising that the less analytically demanding scheme of Little–Mirrlees became dominant. It is the approach to shadow pricing most widely accepted today (Little and Mirrlees, 1974). The economy is divided into traded and non-traded goods markets and there is a competitive primary factor market as well. The shadow price of traded goods is simply the border price, since the import border price is the clearest measure of what the country is willing to give up in order to secure an additional unit of a good. Similarly, if foreigners are willing to pay the border price for our exports, that stands as the next-best alternative to any domestic use. It is a straightforward application of the basic principle of opportunity cost and requires no political justification, defense or intervention.

Non-traded goods are still difficult to shadow price. If there happens to be a separate factor of production for every traded good, and input–output relationships are known, it would be possible to solve for the shadow prices of non-traded goods and factors as a function of the known traded goods prices. If the number of factors is greater than the number of tradables, then the indeterminacy must be removed by additional information. If, for example, it is possible to deduce the foregone output of a traded good upon removing a unit of unskilled labor, then we would have a measure of the shadow value of unskilled labor that could be used to reduce the number of unknowns. If the number of factors were less than the number of tradables, the system would be overdetermined and there would exist two shadow prices for the same good.

Eventually shadow prices would be calculated directly from computable general equilibrium models, but this did not fully resolve the problem either. Model structure clearly matters and moreover, shadow prices are sensitive in general equilibrium models to how projects are financed. If a project is offset by an increase in lump-sum taxes, then the effect on aggregate welfare is the simplest to calculate. But since these tax vehicles are not usually available in developing countries, one immediately has to contend with distortionary mechanisms like income or sales taxes, which add another assumption-laden level of complexity to the analysis. Other complications include economies with segmented goods (traded and non-traded) and labor markets (which may also be regulated), large informal sectors, credit rationing, an inadequately developed or captured regulatory apparatus, and the like (Squire, 1996).

Projects that do not represent Pareto improvements, since they may easily imply a loss of welfare to some members of society while others gain, can be accepted. Income distribution need not, however, be taken into account in project appraisal if an appropriate scheme of taxes and subsidies is available to compensate losers. This is a big “if” however, and some

authors have tried to incorporate distributional concerns directly into the procedures for project evaluation. Government policy-makers may choose to redistribute income from current to future generations or within the current generation from one class of households to another. As Chowdhury and Kirkpatrick note, distributional weights applied to utility representations of individual households is an explicitly subjective exercise, which varies across both time and space (Chowdhury and Kirkpatrick, 1994, p. 2). Efficiency calculations are rarely of such magnitude that they cannot be reversed by small changes in weights in the aggregate welfare function. For this reason, planners have been reluctant to mix concerns of equity and efficiency.

Public investment in infrastructure projects including electrification, telecommunications, transportation and marketing facilities would seem to address problems of static and dynamic market failure. Oddly, it has been argued that there in fact has been too much investment in infrastructure. Project evaluation techniques – even when undertaken by competent economists, such as the staff of the World Bank – fail to account properly for the welfare loss in cost recovery. On the other hand, welfare losses per dollar of public revenue raised are typically calculated using static computable general equilibrium models and therefore cannot account for the dynamic market failure of the underproduction of public goods. Getting prices wrong ultimately means they will not be used for any politically sensitive decision. Planning succumbed in large measure because, in democratically organized societies, only the market has been able to claim objectivity in determining shadow values.

Current uses of planning models

Planning and planning models may be out of fashion, but they can still serve a useful purpose. The most obvious use is that they allow policy-makers to form quantitative estimates of the various trade-offs in preparing development policies. They can be used to comb out inconsistencies in the ways in which policy-makers believe the economy is working. The models also enhance internal communication, adding clarity to discussions within the policy establishment as well as between these individuals and politicians, the public and other interested parties, such as non-governmental organizations (NGOs). Planning models also serve as a means of external communication. The models communicate the thinking about how resources are employed and the explicit assumptions (behavioral parameters, elasticities and the like) underlying the model can be reviewed and evaluated by outsiders. Models can signal to donors that contributed resources will be used wisely and in ways consistent with broad development objectives. Finally, planning models with sufficient structural detail

also can be used to counterbalance any undue influence of generic, one-size-fits-all models.

Proper incentives were often ignored in early planning and this was reflected in the models themselves. More recently, CGE models explicitly incorporate the incentive structure. They derive their strength from the comprehensive picture they paint of the economy and can account for the combined effects of numerous simultaneous policies, from labor markets to exchange rates, taxes and transfers. Planners can conduct realistic 'what if' experiments, refining their understanding of the various channels by which adjustment processes unfold. Some, although not all, unintended consequences are likely to be anticipated, allowing for corrective policies to be put in place.

Planning, as an institution throughout the developing world, has not entirely disappeared but rather has changed forms in significant ways. Policies often have unintended consequences, most often when they are blind to the implicit incentive structures they erect. Consequently, planning ministries have given way to development policy management offices. The latter explicitly strive to enhance market outcomes. Rather than having to anticipate the various ways in which the private sector may try to evade the planners' directives, modern theory suggests that a market-driven approach can yield more satisfactory results. Planners set broad overall planning objectives and then encourage the private sector to maximize their own interests subject to these imposed constraints. Decision-making is decentralized and the social cost of compliance is minimized.

This enlightened approach takes much of the conflict out of planning and the negative connotation associated with command-and-control is thereby lessened. As states abandoned coercive methods, fewer trades were blocked, and economic efficiency automatically increased. This is not planning 'lite', but rather a different approach that tries to exploit fully the informational content of prices rather than issue legally binding directives.

Note

1. Thanks to Diane Flaherty and the editors of the *Handbook* for many constructive comments and criticisms.

References

- Blitzer, C., P. Clark and L. Taylor (1975), *Economy-Wide Models and Development Planning*, New York and London: Oxford University Press.
- Chowdhury, A. and C. Kirkpatrick (1994), *Development Policy and Planning: An Introduction to Models and Techniques*, London: Routledge.
- Dasgupta, P., A. Sen and S. Marglin (1972), *UNIDO Guidelines for Project Evaluation*, New York: United Nations.
- Dorfman, R., P. Samuelson and R. Solow (1958), *Linear Programming and Economic Analysis*, Rand Corporation, New York: McGraw-Hill.

- Gibson, B. (2003), 'An Essay on Late Structuralism', in A. Dutt and J. Ros (eds), *Development Economics and Structuralist Macroeconomics: Essays in Honor of Lance Taylor*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 52–76.
- Gunning, J.W. and M.A. Keyzer (1995), 'Applied General Equilibrium Models for Policy Analysis', in J. Behrman and T.N. Srinivasan (eds), *Handbook of Development Economics*, Vol. 3A, Amsterdam: Elsevier, pp. 2025–2107.
- Heal, G. (1973), *Theory of Economic Planning*, Amsterdam, Oxford and New York: North-Holland.
- Killick, T. (1976), 'The Possibilities of Development Planning', *Oxford Economic Papers*, 28 (2): 161–84.
- Little, I. and J. Mirrlees (1974), *Project Appraisal and Planning for Developing Countries*, London: Heinemann.
- Squire, L. (1996), 'Project Evaluation in Theory and Practice', in H. Chenery and T.N. Srinivasan (eds), *Handbook of Development Economics*, Vol. 2, Amsterdam: Elsevier, pp. 1093–1137.
- Taylor, L. (1979), *Macromodels for Developing Countries*, New York: McGraw-Hill.
- Todaro, M. (1971), *Development Planning*, Nairobi: Oxford University Press.

58 State-owned enterprises and privatization

Anindya Sen

Introduction

The 1960s and 1970s were characterized by the rapid expansion of the public sector in both developed and developing economies. But from the 1970s there was a breakdown of the social-democratic Keynesian consensus of the preceding 30 years about the strong social and economic role of the state. Ideas of market failure came to be replaced by ideas of government failure. The 1980s therefore witnessed widespread attempts by policy-makers to curb the state's economic role. Privatization was a significant element of these attempts. Moreover, the fiscal crises faced by many developing countries in the 1980s considerably curtailed the capacity of the state to invest in state-owned enterprises (SOEs). The consequent attempts at reforms were sometimes part and parcel of structural adjustment programmes that emphasized speedy privatization. In particular, countries facing foreign exchange constraints which had to approach the international financial institutions – the IMF and the World Bank – for assistance were given such assistance subject to certain conditionalities, including privatization of SOEs.

It is interesting to note that the term 'privatization' was originally coined by Peter Drucker and replaced the term 'denationalization'. The first large-scale 'denationalization' programme of the post-World War II era was launched by the Adenauer government in the Federal Republic of Germany. In 1961, the German government sold a majority stake in Volkswagen in a public share offering. The offering was heavily weighed in favour of small investors. Four years later another larger offering took place. Both offerings were initially received favourably. However, a later cyclical downturn in share prices forced the government to bail out many small shareholders.

The next major European initiative came 20 years later with the successful British Telecom initial public offering in November 1984. This was followed by increasingly massive share issue privatizations in the late 1980s to early 1990s. As a result, the share of SOEs in British gross domestic product (GDP) reduced from around 10 per cent to almost 0 per cent in the space of 18 years. France, Italy, Germany and Spain followed with their own programmes. Typically these were public share offerings, often launched by socialist governments.

Among Asian countries, Japan has sold only a relatively few SOEs, mostly via share issue privatization. Some of these have been truly enormous. The \$40 billion Nippon Telegraph and Telephone offer in November 1987 remains the single largest security offering in history. In China, numerous small privatizations have taken place, but relatively few outright sales of SOEs. The reason may be that most Chinese SOEs are burdened with social welfare responsibilities. It will be difficult to implement a very large privatization programme since that will seriously undermine the state's economic role.

India has undertaken a 'disinvestment programme' since 1991, but the targets have consistently exceeded the actual proceeds realized, and major controversies have erupted over some of the SOEs privatized. In Latin America, many countries have undertaken large-scale privatization – Chile, Mexico and Brazil being the most prominent. In sub-Saharan Africa, privatization can be characterized as a 'stealth economic policy' (Megginson and Netter, 2001) since few governments have openly adopted an explicit divestment strategy. But there has been more privatization than is commonly believed. For example, in South Africa the African National Congress came to power on the planks of nationalization and redistribution of wealth, but the Mandela and Mbeki governments have almost totally refrained from nationalization and have even sold off several SOEs.

In Central and Eastern Europe, privatization efforts are part of a broader effort to transform from command to market economies. Two imperatives facing these countries are the lack of financial savings and the infeasibility of foreign purchases of divested assets because of political considerations. These two imperatives compelled the launch of 'mass privatization' programmes that involved the distribution of vouchers to the population. The vouchers could be used to bid for shares in companies being privatized. Such mass privatizations resulted in massive reduction in state ownership. Later this technique became very unpopular, especially in Russia where a perception grew that privatization had led to robbery by the old elites and new oligarchs.

It has been estimated that cumulative value of proceeds raised from privatization exceeded \$1 trillion in mid-1990. Annual proceeds peaked at \$160 billion in 1997.

Definition and techniques

Privatization involves a transfer of ownership and control from the public to the private sector. Privatization can be accomplished in two ways. The government can sell its assets to private buyers. Or the government can stop providing a service directly and rely on the private sector to deliver the

service. Given that a government often does not unload 100 per cent of its holdings in a company, the distinction between public and private entities can become nebulous. The problem is compounded by our inability to identify precisely the point at which control passes to private agents.

The public enterprises relevant here are revenue-generating entities originally owned or controlled by the state. An example of a revenue-generating entity is a municipal corporation that gets property taxes.

There are a number of techniques of privatization (Lopez-Calva, 1998). The most well known among these is the sale of equity to the general public. This is called 'divestiture' (divestment or disinvestment), and may be complete or partial. Divestiture can be done through both direct sales and equity offerings. While the developed countries have often utilized equity offerings as a technique for privatization, developing countries have found the process to be difficult for a number of reasons. The inadequacy of national stock markets and the lack of domestic capital in these countries have sometimes led to a shortage of local buyers, while foreign investors, unable to obtain sound information on the enterprises offered, often lacked sufficient interest. Even in developed countries, the direct sales approach may be costly and slow, owing to the complexity of preparing each state asset for sale individually, and then ensuring that buyers observed all contract provisions.

One special type of direct sales is a management–employee buyout. Shares of an enterprise are sold or given to some combination of managers and other employees. Well-structured management–employee buyouts can sometimes lead to efficient results, since the people who know best about an enterprise, that is, the employees and managers, become the owners. It is also rapid and easy to implement. Nevertheless, experience shows that these buyouts suffer serious disadvantages. Yielding to insider interests often entails large costs in inefficiency and poor management. Insiders may also lack many of the skills necessary to function in a market-oriented economy. Further, the process is seen to be inequitable, handing employees, rather than the population at large, most of the benefits.

In a number of former socialist countries, restitution has been employed to privatize SOEs. Restitution refers to the return of state assets to their former private owners in situations where the government's original acquisition is seen as unjust, such as uncompensated seizure. Restitution, in such cases, it is argued, is essential on moral grounds. Opponents of restitution counter that the process is necessarily selective, and therefore an unsatisfactory way of achieving justice retroactively. Moreover, private claims can often be complicated and drawn out, bogging down privatization unnecessarily. In practice, the transition countries have seldom used restitution, except for Estonia and, to a lesser extent, the Czech Republic.

In such economies, another important technique has been mass privatization. In mass, or equal-access, voucher privatization, the government generally gives away, or sells for a nominal fee, vouchers that can be used to purchase shares in enterprises. This technique has proved to be popular particularly in the Czech Republic. Voucher privatizations can not only help to overcome the shortage of domestic capital, but they are also politically popular because they address the perceived unfairness of other approaches and avoid the charges of a sell-out of national assets to foreigners. The main risk is that a dispersed ownership structure will lack the focus and power to direct effective corporate management. These problems have been partly addressed by pooling ownership interests in investment or mutual funds. The funds, however, do not always have adequate management, control and supervisory powers. In such cases, voucher privatization becomes merely ineffective absentee ownership.

Contracting out or leasing out of government services can be another technique of privatization. For example, a municipal corporation can contract out the task of garbage collection to a private party.

Associated with privatization usually are processes of liberalization and deregulation. Liberalization refers to the introduction or promotion of competition in a traditionally monopolized industry. Deregulation refers to the abolition of statutory barriers to the operation of market forces. For example, the government of India controls the prices of many commodities through the administered pricing mechanism. If some commodity is taken out of the purview of this mechanism, then this is a deregulatory measure, because the price will now be determined by market forces.

Reasons for the establishment of SOEs

State owned enterprises (SOEs) were created for a number of reasons. It was believed that nationalization of private sector enterprises and establishment of SOEs would provide governments with access to much-needed revenues. These profits or surpluses could then be channelized to develop the priority sectors of the economy. Implicit in this line of thinking was the assumption that the private sector would not help in the rapid and sustained development of the economy if left to itself. There was a need to control the 'commanding heights' of the economy, that is, the strategic industries. If the government controlled these industries, it would be able to steer the economy in the right direction and overcome critical bottlenecks. National security reasons were sometimes added as justifications to the above, particularly in the context of heavy industries.

In many developing countries, lack of private entrepreneurs was also a major concern and forced the state to take an active role in the process of industrialization. Local private entrepreneurs often were in short supply.

Even if they did exist, they might not have access to adequate capital, partly because stock markets were not well developed. In some countries, private entrepreneurs came from unpopular minorities or were linked to foreign powers.

Distributional considerations also played a role. In India, SOEs were set up in backward regions to lessen regional inequalities. SOEs were also used to increase employment generation.

In political terms, SOEs constituted important resources for state elites – politicians and bureaucrats. They could be used to provide jobs to potential voters and service constituencies (for example, a railway minister could order railway officials to provide better links to his or her constituency).

Why privatize?

Over time, in many countries, the performance of SOEs turned out to be, by and large, unsatisfactory. They incurred losses, or did not make as much profit as they should have, given that they had privileged access to capital, various subsidies and protection from domestic and foreign competition. The main reason for this failure seems to have been the problems of gathering appropriate information and devising appropriate incentives for the pursuit of public interest. The incentives for serving customer interests and controlling costs were usually weak. Management was given ill-defined objectives and these frequently conflicted with political objectives. 'The accounting systems were not appropriate to the information needs required for efficient pricing, i.e. setting prices equal to long-run marginal costs and investment projects were often appraised against technical criteria rather than economic hurdle rates of return' (Jackson and Price, 1994, p. 2) As a result, many governments started considering privatization of SOEs as a solution to these problems because it was felt that markets provide better incentives to participants and use information more efficiently. Privatization would provide greater incentives for cost minimization, encourage more effective managerial supervision and stimulate greater employee effort.

We next examine the possible impacts of privatization in economic terms. In addition to efficiency, distribution and stability factors also need to be taken into consideration.

Fiscal impact

When a SOE is sold off to the private sector, the government gets the sales proceeds. Further, if the SOE had been making losses and was being subsidized, then these subsidies come to an end, which further helps the government. Thus the immediate generation of revenues is supplemented by reduction in recurrent expenditures.

But does the government really gain? In the simplest case, the buyer will be willing to pay only so much as the SOE is expected to bring in the future. The discounted sum of the future stream of returns from the SOEs is what a buyer will pay (Van de Walle, 1989). The government would have got the same revenue had it not sold the SOE. Therefore, it would seem that privatization does not have any real impact on the government's finances.

There are two reasons why privatization might still make a difference. First, a privatized firm might be expected to be more efficient than a SOE. Hence, the sum of discounted returns will be higher than that under government ownership. Secondly, the government, when it privatizes, is getting funds immediately. This added liquidity might be desirable for a number of reasons: for example, because the government might want to spend on education or infrastructure.

It is interesting to note that in theory, for a loss-making SOE the price might be negative. This is not very far-fetched. Governments have sometimes given so many concessions to the buyer to induce them to buy loss-making concerns that in effect the price has turned out to be negative.

One pertinent question here is whether selling bonds is a better means of raising revenues than equity sales. Suppose that privatization via equity sales does not change the earnings prospects of the firm in question. In countries with liquid bond markets, selling bonds might involve lower transaction costs than privatization. Moreover to garner popular support for privatization, equity is often underpriced. However, countries constrained in their ability to sell bonds, for example those facing debt crises, or those that are trying to limit borrowings to commit to an anti-inflationary policy, might be compelled to sell equity. In addition, the perceived risk of default on bonds might be higher than equity risk.

Efficiency gains

The argument for privatization often rests on the supposed superiority of the private sector in attaining the goal of economic efficiency – both allocative efficiency and productive efficiency. Proponents of privatization have argued that a change in ownership can have an important effect on economic efficiency. In SOEs, prices sometimes did not reflect scarcities properly. For example, if the government gives a subsidy for an input used by an SOE, the SOE would tend to overuse that resource. Or, if a SOE is a monopoly, then it can set its own price. SOEs therefore would not attain allocative efficiency.

It has also been argued that SOEs are likely to exhibit greater internal inefficiencies than private firms for various reasons. Public managers are given numerous and inconsistent objectives. Instead of control by

shareholders who are interested in profit-maximization, there is bureaucratic control which puts more emphasis on 'playing it safe'. Suppose that productive efficiency requires use of an input that is not available in a competitive market. The manager in an SOE is required to obtain competitive quotations for almost everything, and hence may have no way of using this particular input because there is only one seller.

Whether such efficiency will indeed be achieved depends in turn on the goal(s) that managers pursue in private enterprises *vis-à-vis* public enterprises. In theoretical literature, it is often assumed that a manager of a private enterprise maximizes economic profit, while the manager of a public sector has the liberty or is forced to pursue a more diffused agenda. If this assumption is correct, then of course ownership per se becomes a determinant of efficiency, and there is no dearth of empirical research looking into the ownership issue. However, there are reasons to believe that the monitoring system and incentive system in private enterprises may not always work perfectly, and to a large extent, the efficacy of these systems depends on the market structure and the regulatory policy. Then not only is it necessary to turn our attention to a comparative evaluation of the incentive structure in these two types of enterprises, but we are forced to consider other alternatives for achieving efficiency.

In many economies today, the incidence of owner-managers is declining and most large firms are coming to be characterized by the separation of ownership from control. Even if the shareholders can be assumed to have profit-maximization as the overriding objective, managers can pursue their own goals because of the existence of asymmetric information: in general the managers have much better knowledge about market conditions, technology and their own effort levels than shareholders. Since shareholders cannot control managers, various instruments for curbing managerial discretion have been suggested. These include: product market competition which imposes a Darwinian survival requirement of profit maximization; labour market signalling which leads to diminution of market value of non-performing managers; the threat of hostile takeover and use of incentives like employee stock option plans to align the interests of managers with those of shareholders.

All these instruments have their own requirements to be successful. In many markets, competition is muted. The Old Boy network bypasses the signalling effect of non-performance. In response to the threat of hostile takeovers, managers have developed a host of takeover defences to discourage or foil such bids. Moreover, as Grossman and Hart (1980) have pointed out, a free-riding consideration which makes shareholders refuse to part with their shares in the anticipation of an imminent rise in share prices can abort attempted takeovers. The experiences of offering stock

option plans have been mixed, mainly because the design of these plans often insures managers against the downside risk.

Of course, in addition, market failures from externalities and public goods nature of certain commodities may mean that profit-maximization does not lead to efficiency.

For public sector enterprises, on the other hand, there is no market for shares and hence no market for corporate control. How does government monitoring compare with monitoring in private enterprises? In theory, the government has the ability to correct for all types of deviations between social and private returns in goods and factor markets (Yarrow, 1986). The dissatisfaction with government control stems from the fact that the 'market' for political control is highly imperfect. The voting public controls the government, but its knowledge about specific enterprises may be very weak, and normally it cannot vote separately on the issue of running public sector enterprises (though when the public sector constitutes a very large part of the economy, poor performance by such enterprises can become a significant political issue in elections). The resources at the government's command can also lead to the soft budget constraint – continuing support of even non-viable and inefficient enterprises to keep them going. If it is argued that it is necessary to keep these enterprises alive to attain distributional goals (for example, provision of cheap products to poorer sections), then one is faced with the question whether there are alternative and better ways of attaining these goals. Again, sometimes, a fraction of the shares of an SOE are sold to the public, and then outsiders can monitor the enterprise's performance to some extent. However, it still remains unclear how much divestment must take place before the outsiders can exert a significant oversight on the running of such enterprises.

As already mentioned, product market competition can be one instrument for checking managerial discretion. Product market competition forces firms to minimize costs and maximize profit for long-term survival. Moreover the observance of the performance of competing firms provides shareholders and governments with additional information about managerial inputs and the firm's true opportunities. Thus in any year, profitability will depend on the levels and the quality of managerial inputs as well as a host of other factors outside of the manager's control. It then becomes difficult to employ profitability as a correct indicator of managerial inputs and base rewards on such a measure. However, when the profitability of other firms can be observed, this to some extent aids shareholders in disentangling managerial contribution from random factors in a firm's performance.

It is clear that for SOEs operating in competitive markets, prices would better reflect scarcities and therefore allocative inefficiency would be less. Then the gains from privatization would also be less. On the other hand,

transforming a public sector monopoly into a private sector monopoly would also not lead to increases in allocative efficiency. We can expect large increases in allocative efficiency to be achieved when a public sector monopoly is privatized and the market opened up to other players.

Distributional impact

Privatization can have rather significant effects on income distribution. One can discuss distributional issues either by looking at the sources of distributional changes or by identifying the major groups who might be affected by privatization. Perhaps the most important sources of redistribution effects are the changes in the prices of privatized assets and the pricing of commodities after privatization. If, for example, water and electricity supplies are privatized and water and electricity charges go up after privatization, this can affect large segments of the population adversely. If privatized assets are sold at discounted prices, there is a transfer of wealth to the new owners from the wider public and taxpayers in particular (Vickers and Yarrow, 1988). Such sales at discounts are sometimes politically attractive because the risk of shares being unsold is minimized, and 'because the beneficiaries tend to be more aware of their gains than the losers feel the losses' (Yarrow, 1986). In all these instances, it is clear that different groups are affected differently. If privatization is followed by layoffs, employees are affected. If privatization leads to higher profits, shareholders gain.

Sometimes the goods and services made available by the SOE to the poor may become less accessible after privatization. For example, a privatized airline may choose not to fly on unprofitable routes. On the other hand, if privatization is accompanied by a more competitive environment, then greater varieties of commodities may be available at lower prices. Especially in the telecommunications sector, privatization has given access to new and cheaper services to the population at large.

While privatization can have some adverse distributional implications, one can also argue that the SOEs have not had a very successful record of reaching the poor and the disadvantaged sections of the population. For example, in India, it has been repeatedly shown that the Public Distribution System does not do a good job of enabling essential commodities to reach the poorest sections; it is more useful to the richer sections.

Some key issues

Implementation issues

Worldwide experience shows that implementation of privatization programmes has lagged well behind stated intentions. Barring a few countries,

privatization has been limited to small SOEs of the manufacturing and the services sector. There are quite a few problems faced by countries trying to privatize SOEs. Firstly, in some developing countries, there is a lack of well-established, competent management consulting groups, accounting firms and investment bankers. These are needed to provide technical advice and valuation of SOEs. As a result, in some instances, foreign experts have been brought in.

Secondly, a valuation of the SOE has to be carried out before it can be offered for sale and the valuation exercise has faced severe problems. Valuation is a sensitive subject politically, because governments want to get high sales prices and at the same time the valuation process might raise questions about past public management and investment decisions. There have been inordinate delays in valuation. The problem is aggravated when poor records are maintained by SOEs.

Thirdly, once the valuation has taken place, administrative capacity is needed to assess buyers' bids, arrange finance and insurance, and deal with a host of complex legal issues. Sometimes, a comprehensive rehabilitation plan for the SOE has to be designed, evaluated and financed before privatization is possible. Moreover, appropriate regulatory structures may not exist and may have to be set up, particularly when privatization leads to the creation of a monopoly.

Fourthly, capital markets in many developing countries are typically weak and poorly regulated. Large investments in equity are quite unusual. SOEs are some of the largest firms in the country and the private sector may not be in a position to fund the purchase of large assets. The private sector may also be suspicious about the government's intentions, given the record of nationalization in the past. On the other hand, the government may not be willing to sell assets to foreign investors.

Political constraints

Generally, the costs of privatization are borne by a small group of people, for example the workers of the enterprise who may lose their jobs, or the suppliers who may lose favoured contracts. The benefits, however, are spread out over a large number of people, sometimes a very large section of the population. Public choice theory suggests that in such situations, it will be easier to organize opposition to the privatization programme than support. Experience tells us that in many countries, privatization programmes fail to mobilize popular support and in fact give rise to strong opposition.

Trade unions, in particular, tend to react strongly against privatization. Trade union power is often concentrated in the public sector and the public sector provides a base for such power. Unions oppose privatization, not

only because of the direct effect on employment, but also because of a fear that trade union power will be reduced in the private sector. The restructuring process generally involves laying off part of the workforce. Usually, forced dismissals are politically infeasible and only generate more opposition to privatization. Governments therefore try to adopt some kind of voluntary approach. Components of voluntary approaches that have been tried out include monetary compensation (for example through voluntary retirement schemes), retraining and redeployment. Cash-strapped governments may find it difficult to cover the cost of laying off workers. Sometimes the government agrees to accept a lower price for the enterprise in return for an assurance from the new owner that employees will be retained even after privatization. In the East German privatization programme, there is an instance where an enterprise was sold for 1 Deutschmark, because the bidder promised to retain all the workers.

Should restructuring occur before or after sale of the unit?

Most SOEs will not fetch a good price if they are sold in their current condition. For historical reasons, many have excess workers, are burdened with obsolete machinery and technology, and often are run bureaucratically. One option before the government is to restructure these enterprises before placing them on the market, for example, by laying off excess workers, by inducting new workers with appropriate skills, by selling off non-strategic parts of the business, computerization of operations, and so on. These enterprises will then become attractive to private investors who would be willing to pay high prices for them. On the other hand, proponents of speedy privatization (the 'big bang approach') argue that the attempt to restructure these enterprises before sales will inevitably lead to delays and the entire momentum for privatization will be lost. Further, it is doubtful whether the governments are at all adept at restructuring.

The evidence

Historically, it appears that SOEs have contributed quite significantly to the gross fixed capital formation of many economies. They have played an important role even in the highly successful East Asian newly industrializing countries. Moreover, there is no clear evidence of a negative correlation between the size of the SOE sector in an economy and its economic performance in terms of the rate of growth. There are also acknowledged outstanding cases of efficient SOEs, for example the giant Korean state-owned steel enterprise POSCO.

There are two separate empirical issues that must be kept in mind. One is the question of whether SOEs are necessarily less efficient than comparable private sector enterprises. The other is the question of the success of

privatization programmes. The first question can be summarized as 'ownership matters', that is, regardless of all other factors, the mere act of transferring a SOE to the private sector will increase its efficiency. The discussion in the section on the reasons for privatization has demonstrated that there is no a priori theoretical reason to believe that ownership alone matters.

Empirical investigations of these two questions run into several problems. For one, using profitability as a performance indicator for comparing SOEs with private enterprises misses out the point that SOEs are often established for reasons other than making profits. Therefore, if this measure is used it will tend to flatter privatization if under state ownership non-profit goals had been pursued. Even studies using profitability do not establish that SOEs are invariably inefficient. Moreover it is difficult to control for the effects of factors other than ownership which might affect a firm's performance. Chang (2003) notes that there might be country-specific, industry-specific and firm-specific factors that are the determinants of performance differentials rather than ownership per se. Thus, countries with successful records of privatization sometimes underwent substantial macroeconomic changes that created a climate conducive for realizing microeconomic efficiency gains. So far as industry-specific results are concerned, the evidence of successful privatization in the telecommunications sector, to take one example, cannot be generalized to other sectors. Rapid technological changes in this sector have increased competition and reduced regulatory problems.

Parker and Kirkpatrick (2005) note that to assess the impact of privatization, in addition to using the correct performance measure, there is the problem of taking into account relative price changes with spillovers into other sectors of the economy, and redistribution effects on different socio-economic groups. Their review of the empirical literature leads them to conclude that: 'The studies vary in terms of the financial and economic performance measures and show that privatization measures can lead to widely differing results.'

Conclusion

Privatization in essence represents a reduction in the role of the government in the economic activities of the nation. The process of privatization has been impelled by ideology as much as by necessity. Even if we restrict ourselves to efficiency concerns, there seem to be no convincing theoretical reasons for arguing that privatization per se will improve efficiency. This is especially important in developing countries where a number of preconditions needed to carry out successful privatization programmes are often missing. It is therefore no wonder that the empirical evidence is also mixed, particularly in the context of developing countries.

References

- Chang, Ha-Joon (2003), *Globalisation, Economic Development, and the Role of the State*, London and New York: Zed Books.
- Grossman, S. and O.D. Hart (1980), 'Takeover Bids, the Free-Rider Problem and the Theory of the Corporation', *The Bell Journal of Economics*, 11 (1): 42–64.
- Jackson, Peter M. and Catherine M. Price (eds) (1994), *Privatisation and Regulation: A Review of the Issues*, London and New York: Longman.
- López-Calva, Luis Felipe (1998), 'On Privatization Methods', Development Discussion Paper No. 665, Harvard Institute for International Development.
- Meggison, William L. and Jeffrey M. Netter (2001), 'From State to Market: A Survey of Empirical Studies on Privatization', *Journal of Economic Literature*, 39 (2): 321–89.
- Parker, David and Colin Kirkpatrick (2005), 'Privatisation in Developing Countries: A Review of the Evidence and the Policy Lessons', *Journal of Development Studies*, 41 (4): 513–41.
- Van de Walle, Nicholas (1989), 'Privatization in Developing Countries: A Review of the Issues', *World Development*, 17 (5): 601–15.
- Vickers, J. and G. Yarrow (1988), *Privatization: An Economic Analysis*, Cambridge, MA: MIT Press.
- Yarrow, George (1986), 'Privatization in Theory and Practice', *Economic Policy*, 323–64.

59 Corruption and development

Cheryl W. Gray

The attention given to corruption and its causes and consequences for economic development has expanded enormously in recent years. While in the 1970s and 1980s corruption was often treated in the economics literature as a benign redistribution of economic rents, since 1990 there has been increasing recognition in both academic and policy circles that high levels of corruption can have a large detrimental impact on economic growth and development. The change in thinking reflects the increasing focus of the economics profession on the role of institutions in economic development, as a rapidly growing body of economic research in the 1990s examined the implications of different institutional arrangements on the functioning of markets and the supply of public goods.¹ This change also partly reflects political developments – most notably the end of the Cold War, the transition from communism in Central and Eastern Europe, and the break-up of the Soviet Union – as the softening of superpower competition created conditions in which international organizations and donor agencies could be more forthright in challenging corrupt regimes in developing countries. The opening up of political regimes in other parts of the world, including increasing democratization in Africa, Asia and Latin America, also provided a window for citizens to organize and put increasing pressure on their governments to fight corruption.

Corruption can be defined as ‘the use of public office for private gain’, or it can be defined more broadly also to encompass the abuse of positions of trust in the private sector for personal gain. For the purposes of this chapter the term is limited to the narrower definition pertaining to the public sector, around which most research and policy work has focused. Fraud and abuse in the private sector is most often addressed by literature on corporate governance.

Measuring corruption

Before the mid-1990s it was generally assumed that corruption could not be measured – both because it was difficult to define and because perpetrators wanted to keep it secret – and thus it was a phenomenon that could only be discussed in generalities. Since the mid-1990s, however, survey tools have been developed to measure levels and patterns of corruption. While significant inaccuracies no doubt remain, these survey tools have added

immensely to economists' ability to analyze the phenomenon and its impact on other economic variables.

There are two general types of survey tools: polls of 'experts' and surveys of economic actors. In each case the questions can be focused either on opinions and perceptions or on actual experiences. The first type of tool asks questions of people who are considered to have expert knowledge on a particular country, and tends to focus primarily on their opinions and perceptions on various issues, including the incidence and severity of corruption.² The second type of tool surveys economic actors – whether enterprise managers, citizens or public officials. While such surveys may also contain questions about opinions and perceptions, surveys of economic actors often try to gather information about actual experiences in dealing with government – for example, the amount of bribes paid in obtaining a business license, getting access to medical care or interacting with tax officials.³ A third type of tool, pioneered in Transparency International's Corruption Perceptions Index, is an attempt to combine all surveys on an individual country into one composite indicator that facilitates rankings and comparisons among countries.⁴

Economics literature and applied policy analyses have utilized these various survey tools extensively since the mid-1990s to understand the extent and patterns of corruption and their implications for economic development. Aid donors have also relied increasingly on these tools to help focus aid in countries with lower levels of corruption.⁵ As the number of types of measurement tools have increased and begun to have greater practical impact in recent years, it has also become increasingly important to understand the pros and cons of different approaches and how they compare to each other.⁶

Levels and types of corruption

The surveys undertaken since the mid-1990s have proven what casual observers already know – that levels and patterns of corruption vary widely among countries. It is indeed not really possible to talk about 'corruption' as one phenomenon, as public office can be abused for private gain in a myriad of different ways. Some literature distinguishes between 'petty' and 'grand' corruption, with the difference between the two being largely a function of the size of the bribe and the status of the briber. Large bribes paid to senior officials for major public contracts are seen as 'grand' corruption, while small bribes paid to the traffic police, for example, are classified as 'petty'. More recent literature distinguishes between 'state capture' and 'administrative corruption', with the former referring to corruption in the formation of laws and regulations (for example bribes to legislators or regulators to tilt lawmaking in certain directions) and the latter

referring to corruption in their implementation (for example bribes to get goods through Customs or enroll children in favored schools)⁷. State capture is often seen as the more pernicious, because it affects the rules by which markets function. Both state capture and administrative corruption come in various forms, depending on who the parties to the corruption transaction are and what is being bought, and both can occur on a large or small scale. ‘Unbundling’ corruption into its various types can give analysts and policy makers a more nuanced sense of its economic and political causes and consequences.

The costs of corruption

High levels of corruption can have devastating impacts on an economy and a society. Among the most pernicious are bribes that allow people to evade laws that protect public safety. Importers, for example, may bribe customs officials to allow dangerous or ineffective drugs into the market, or builders may bribe regulatory agencies to erect buildings that do not meet safety codes. Also costly to public welfare is corruption that affects access to public services, as when parents must bribe doctors or teachers for medical care or education for their children. In all cases, widespread corruption undermines citizen trust and rule of law, and thereby impedes the arm’s-length transactions among strangers that are so fundamental to a market economy.

A large body of academic analysis has been undertaken since the mid-1990s to try to measure the economic impact of corruption more precisely, with various dimensions of economic impact being studied.⁸ One of the first studies was the 1995 analysis by Mauro of investment in a cross-section of 67 countries, which found that corruption has a significant negative impact on the level of investment in relation to GDP.⁹ Further studies in the late 1990s using a variety of corruption indicators reinforced this overall finding. One such study differentiated among types of corruption and found that in settings in which bribes had less predictable outcomes – that is, where bribers were less confident about getting what was paid for – corruption had a stronger negative impact on investment.¹⁰ Another study differentiated between ‘centralized’ (or coordinated) and ‘decentralized’ (or uncoordinated) bribe-taking, arguing that the economic impact of the latter is likely to be more severe, and using the model to illustrate the increased cost of corruption in post-communist Russia as compared with the centralized monopolistic corruption imposed by the Communist Party.¹¹ Analytic work has also focused on the negative impact of corruption on foreign direct investment (FDI). Wei found that corruption at the level found in Mexico was equivalent to a 20 percent tax as compared to that found in Singapore.¹²

While corruption and gross domestic product (GDP) per capita are highly correlated, the direction of causation is more difficult to untangle. On the one hand, corruption can hamper growth by reducing the efficiency of public spending and the effectiveness of public service delivery. On the other hand, poorer countries have a more difficult time tackling corruption, both because bribes may be more tempting when public sector salaries are low, and because it takes resources to fund ‘watchdog’ groups needed to prevent corruption, such as the press, accounting and auditing services, and police and other investigative and law enforcement agencies. A large body of recent literature attempts to unravel the effects of corruption on either the level or the rate of economic growth.¹³ While many methodological difficulties make this line of research difficult and skepticism remains about the findings of individual studies, the broad consensus supports the view that corruption – rather than ‘greasing the wheels’ of commerce – has a substantial negative impact on economic growth.

Other economic costs and consequences have also been identified through research in recent years. Corruption has been found to increase inequality, and the reverse – that higher inequality leads to more corruption – has also been shown.¹⁴ Corruption reduces public revenues,¹⁵ leads to lower quality in public investments and public services,¹⁶ and tends to skew public spending away from education to other types of spending – such as large infrastructure projects or military procurement – where bribes are more lucrative.¹⁷ Corruption has detrimental effects on the environment by reducing the effectiveness of environmental regulation,¹⁸ and it is positively correlated with a country’s rates of inflation¹⁹ and crime.²⁰

Most fundamentally, corruption reduces citizens’ trust in government and the political system, which undermines the ability of governments to cooperate with the citizenry in the formulation of policies and enforcement of laws.²¹ A dysfunctional and mutually reinforcing equilibrium of citizen distrust, poor government performance and corruption can result, undermining economic growth and development in the many ways outlined above.

Tackling corruption

In its simplest form, corruption is facilitated by the ability of a public official to exercise discretion in the carrying out of his or her duties without having to answer to formal organs of accountability. Hence the formulation that $C = M + D - A$ (corruption equals monopoly plus discretion minus accountability).²² This simple formulation points to several variables that can affect the level of corruption in an economy. A large role for government and greater discretion for individual public officials opens avenues for potential corruption, and such corruption is more likely to materialize

if formal institutions of government accountability are weak. Corruption is also more likely when valuable and sought-after assets are under government control, as, for example, in countries with abundant state-owned natural resources (for example, oil- and gas-producers) or in transition countries with large-scale privatization programs. The simple lesson would be to minimize government's role and public officials' discretion while strengthening oversight institutions in an economy.

Reality is much more complex, however. First of all, although people may disagree at the margin about the appropriate roles for government, such roles will never be eliminated entirely, as there are certain public goods (for example, defense, law and order, basic infrastructure and education, environmental protection) that governments have an important role in supplying. Furthermore, giving discretion to public officials is often critical to getting good results, particularly in more complex areas of public policy. Building in mechanisms to ensure transparency and accountability in public decision-making – for example through public hearings on draft laws, 'freedom of information' and publication of government decisions, or internal or external audit procedures – is always desirable, but these can be costly or difficult to implement, particularly in poor countries with few public resources and a severe shortage of skills. The challenge in any particular case is to understand these trade-offs and try to tailor the role of government, the discretion granted to officials, and efforts to strengthen accountability to the particular needs and characteristics of each country situation.

Controlling corruption also requires the active commitment and involvement of the citizenry. Formal public law enforcement is expensive and necessarily operates only at the margin in any country; most laws are enforceable in practice because citizens willingly obey them and are willing to take action when someone else breaks them. If, in contrast, people lack confidence that laws can be enforced, this can become a self-fulfilling prophecy, as they will neither respect and follow the laws themselves, nor bother to report others who break them. One can envision two situations of equilibrium: one where most people follow the law and expect others to follow it, and thus where transgressions are few and can be managed by formal law enforcement; and the other where most people do not follow the law nor expect others to, and transgressions are too numerous for formal law enforcement to handle. Both equilibrium conditions are common in today's world: corruption is the rare exception in some countries, but it is systemic and widespread in others. The challenge that many systemically corrupt countries face is how to move from one equilibrium state to the other.

In practice, governments that are strongly motivated to reduce corruption have many policy and institutional levers they can use. A multi-pronged

approach to tackling corruption would address many institutional dimensions, as shown in Figure 59.1.²³

State capture tends to thrive when the private sector is monopolized, and economic conditions and policies that enhance private sector competition and transparency will tend to reduce incentives and opportunities for corruption. These include, for example, policies that promote international trade,²⁴ low entry barriers for new firms, a stable macroeconomic framework with low inflation, and well-designed regulatory, corporate governance and anti-monopoly laws.²⁵ Competition and transparency in the political realm can also help, including clear and binding rules for political party financing or asset declaration and conflict of interest rules for senior government officials.²⁶ More fundamentally, government structures that limit power and create horizontal or vertical checks and balances – for example, through legislative or judicial²⁷ oversight of government decisions or through multi-tiered layers of government²⁸ – can greatly enhance political competition and accountability. On a day-to-day basis governments can take steps to improve public sector management by adopting rules and enhancing incentives and skills for meritocratic civil service staffing,²⁹ honest tax and customs administration, and transparent public procurement and budgeting procedures. As there are almost always losers as well as winners in these reforms, governments must make intensive efforts to build public support and publicize early ‘wins’ to overcome countervailing pressures. Governments can also restrict opportunities for corruption by taking steps to reduce their role in an economy – for example, by privatizing commercial firms and reducing regulations on business activity. While it is true that cross-country regressions do not show a statistically significant correlation between the overall size of government and the level of corruption,³⁰ it is also true that opportunities for corruption can be reduced in any particular country by shedding unnecessary activities in the public sector. Finally, governments can help create the legal basis and political openness for a vibrant and independent media³¹ and a strong civil society to function. Both are critical counterweights to government power and are essential to the control of corruption.

Reforming the political system to increase transparency and accountability is part of the challenge, as noted above. However, the presence or absence of democracy per se is not necessarily the determining factor. Young democratic systems with limited transparency and accountability can be highly corrupt, and pay-offs often increase during election years as politicians trade bribes for votes or private parties buy jobs or favors from new governments. However, political reformers and the international community need to persist in promoting democratic values, because there is clear evidence that a longer exposure to democracy lowers corruption.³²

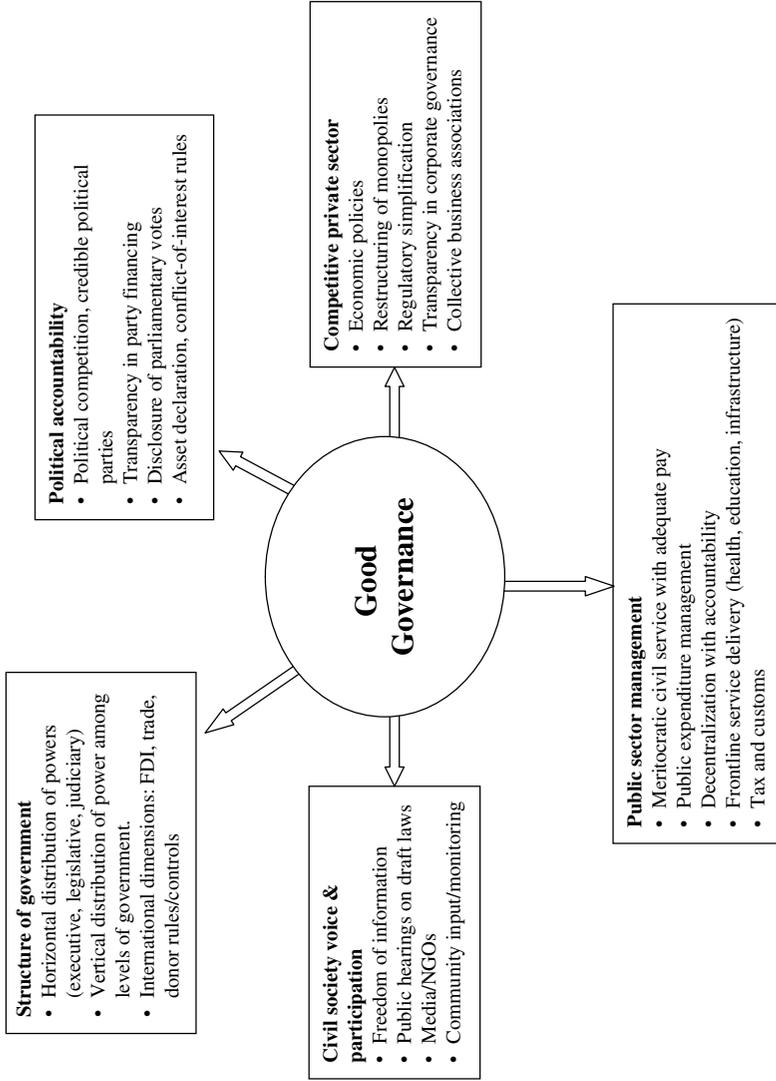


Figure 59.1 Addressing corruption requires action across many fronts

And when there is a window of opportunity to change the design of the democratic system, reformers should note that parliamentary systems appear to do a better job than presidential systems in controlling corruption,³³ at least as long as political parties are generally oriented toward the public interest. The Westminster system of democracy found in the UK – with a long and unbroken history, direct election of candidates in lieu of proportional representation, strong and accountable political parties, high electoral participation, and voting districts that are not unduly small – is considered by many to be the model of democracy most likely to control corruption.

The question of whether decentralization of government power reduces corruption is a subject of active debate, and many aid donors are pressing for further decentralization and community participation in countries with weak governance. Decentralization advocates argue that moving government closer to the citizenry will force greater transparency and accountability, while more cautious observers question the capacity of local governments (particularly in poor countries) and the likelihood of ‘capture’ of decentralized power by local elites. On the one hand, it is clear that some of the best-governed countries in the world (Denmark, for example) are also among the most decentralized. On the other hand, it is also clear that state and local governments in many settings (the United States, for example) have been more prone to corruption scandals than national governments. The extensive research on this issue is inconclusive and suggests that the details of each country’s historical, cultural,³⁴ political and economic situation are likely to be more important than the extent of decentralization per se in determining how decentralization will affect governance and corruption.³⁵

Progress is possible

Experience around the world shows that strong and committed leaders can indeed reduce corruption in government. Among the more advanced countries, for example, both the United States and the United Kingdom undertook far-reaching civil service reforms in the nineteenth century that helped to reduce corruption.³⁶ More recently, corruption in state governments in the USA is widely thought to have been far worse 50 years ago than it is today. And, more recently still, many of the former communist countries of Central and Eastern Europe appear to have made significant strides in reducing corruption since the mid-1990s.³⁷ In-depth surveys of businesses undertaken in 1999, 2002 and 2005 indicate that the frequency of bribery in many types of public services and the constraint that corruption imposes on doing business declined significantly in many countries over this period. A combination of better economic policies, stronger

economic and political institutions, faster economic growth and a more vibrant civil society combined to put pressure on governments for greater honesty and accountability.

Eliminating all corruption is not feasible. It exists and will continue to exist in every government in every country in the world. However, a much more feasible – and in the end more important – goal in every country should be to reduce the level of corruption and its negative social and economic impacts. For developing countries with widespread and systemic corruption, the goal is to move to an equilibrium where corruption is the exception rather than the norm. There is little doubt that declining corruption, enhanced government effectiveness, greater citizen trust and increased private investment are mutually reinforcing phenomena that together spur economic growth and development.

Notes

1. See, for example, North (1990), Bardhan (1997a) and World Bank (2002).
2. Some examples include the Nations in Transit indicators from Freedom House (www.freedomhouse.org), the Political Risk Service's International Country Risk Guide (www.prsgroup.com), and the World Bank's Country Policy and Institutional Assessment (www.worldbank.org/ida).
3. Examples including the EBRD–World Bank Business Environment and Enterprise Performance Survey (BEEPS) (worldbank.org/eca/governance) and the World Economic Forum's Global Competitiveness Survey (www.weforum.org).
4. Examples include the Transparency International (TI) indicators (www.transparency.org) and indicators put together by the World Bank Institute (www.worldbank.org/wbi/governance).
5. For example, both the level of resources made available to poor countries by the World Bank's International Development Association (IDA) and the eligibility of countries to access grants from the US Millennium Challenge Account (MCA) are dependent in part on countries' scores on various governance and corruption indicators.
6. Knack (2006) and Arndt and Oman (2006).
7. For further explanation of the concept of state capture, see World Bank (2000) and Hellman et al. (2000).
8. For a synopsis of some of the vast literature on corruption and development, see Bardhan (1997b) and Lambsdorff (2006).
9. Mauro (1995, 1997).
10. Campos et al. (1999).
11. Shleifer and Vishny (1993).
12. Wei (2000b).
13. Some recent research that attempts to use instrumental variables to address these simultaneity problems finds that corruption leads to lower GDP per capita. See, for example, Hall and Jones (1999) and Kaufmann et al. (1999).
14. Gupta et al. (2002), Husted (1999).
15. Tanzi and Davoodi (1997).
16. Gupta et al. (2001).
17. Tanzi and Davoodi (1997) and Gupta et al. (2002).
18. Esty and Porter (2002).
19. Al-Marhubi (2000) and Gerring and Thacker (2004).
20. Azfar (2004).
21. La Porta et al. (1997).
22. Klitgaard (1998).

23. World Bank (2000).
24. Research points to the value both of low international trade barriers and of a uniform (rather than highly diversified) tariff structure in reducing opportunities for corruption: Gatti (1999) and Ades and Di Tella (1999). Wei (2000a) argues that a country's 'natural openness' (that is, small size and non-remote location) is more important than trade policy per se in increasing competition from trade and thereby reducing corruption.
25. Ades and Di Tella (1995).
26. With regard to electoral rules, there is some support for the view that corruption is less likely in countries where voting districts are larger (allowing greater political competition) and where citizens vote for individual candidates rather than party lists (which increases individual accountability of politicians to the electorate). Persson et al. (2003).
27. For an analysis of the importance of an independent judiciary in controlling corruption, see World Bank (1997) and Ades and Di Tella (1997).
28. See fuller discussion on decentralization below.
29. Evans and Rauch (2000). Raising civil service salaries can also help to reduce corruption (van Rijckeghem and Weder, 2001), but, above a certain reasonable minimum, raising salaries is not as important as promoting merit-based hiring and promotion.
30. It is not surprising that cross-country regressions show no significant correlation between government size and the level of corruption, given the two-way nature of causation. While an overextended public sector may create more opportunities for corruption, corruption (and the low level of trust and public sector effectiveness it causes) may also make it more difficult for governments to collect public revenue. Thus only the more effective governments may have the capacity and citizen trust to grow and remain large. Indeed, many of the world's least corrupt countries – most notably in Scandinavia – are also among the countries with the largest public sectors. When the Scandinavian countries were omitted in one study, size of government was found to be positively correlated with the level of corruption. LaPalombara (1994).
31. Brunetti and Weder (2003).
32. Treisman (2000), Sung (2004) and Anderson and Gray (2006).
33. Gerring and Thacker (2004) and Kunicova and Rose-Ackerman (2005).
34. With regard to culture, Husted (1999), drawing on Hofstede (1997), finds that more hierarchical, more materialist, and more risk-averse cultures are likely to be more corrupt.
35. A large number of studies have been undertaken in recent years to try to assess the links between decentralization and corruption, with some supporting the view that greater decentralization is correlated with lower corruption and some refuting that view (generally finding that the correlations disappear when the sample changes or when decentralization is measured in a different way). Another strand of this literature looks at federal and unitary states, with some studies finding that federalism is correlated with higher levels of corruption and others showing no effect. See Lambsdorff (2006) for further discussion.
36. Delay and Moran (2003) and Johnson and Libecap (1994).
37. Gray et al. (2004) and Anderson and Gray (2006).

References

- Ades, A. and R. Di Tella (1995), 'Competition and Corruption', unpublished paper, Keble College, Oxford University.
- Ades, A. and R. Di Tella (1997), 'The New Economics of Corruption: A Survey and Some New Results', *Political Studies*, 45 (3): 496–515.
- Ades, A. and R. Di Tella (1999), 'Rents, Competition, and Corruption', *American Economic Review*, 89: 982–94.
- Al-Marhubi, F.A. (2000), 'Corruption and Inflation', *Economics Letters*, 66 (2): 199–202.
- Anderson, J. and C. Gray (2006), *Anticorruption in Transition 3: Who is Succeeding . . . and Why?* Washington, DC: World Bank.
- Arndt, C. and C. Oman (2006), 'Uses and Abuses of Government Indicators', Paris: OECD Development Centre.
- Azfar, O. (2004), 'The Rule of Law, Corruption and Homicides', mimeo, IRIS.

- Bardhan, P. (1997a), *The Role of Governance in Economic Development: A Political Economy Approach*, Paris: OECD Development Centre.
- Bardhan, P. (1997b), 'Corruption and Development: A Review of Issues', *Journal of Economic Literature*, 35: 1320–46.
- Brunetti, A. and B. Weder (2003), 'A Free Press is Bad News for Corruption', *Journal of Public Economics*, 87 (7–8): 1801–24.
- Campos, J.E., D. Lien and S. Pradhan (1999), 'The Impact of Corruption on Investment Predictability Matters', *World Development*, 27 (6): 1059–67.
- Delay, S. and D. Moran (2003), 'UK Civil Service Reform in the 19th Century and the Lessons for Future Reforms in Developing and Transitional Countries', unpublished paper, School of Public Policy, University of Birmingham.
- Esty, D. and M. Porter (2002), 'National Environmental Performance Measurement and Determinants', in D. Esty and P. Cornelius (eds), *Environmental Performance Measurement: The Global Report 2001–2002*, New York: Oxford University Press.
- Evans, P.B. and J.E. Rauch (2000), 'Bureaucratic Structures and Economic Performance in Less Developed Countries', *Journal of Public Economics*, 75: 49–71.
- Gatti, R. (1999), 'Corruption and Trade Tariffs, or a Case for Uniform Tariffs', World Bank Policy Research Working Paper No. 2216, Washington, DC: World Bank.
- Gerring, J. and S. Thacker (2004), 'Political Institutions and Corruption: The Role of Unitarism and Parliamentarism', *British Journal of Political Science*, 34 (2): 295–330.
- Gray, C., J. Hellman and R. Ryterman (2004), *Anticorruption in Transition 2: Corruption in Enterprise-State Interactions in Europe and Central Asia 1999–2002*, Washington, DC: World Bank.
- Gupta, S., H. Davoodi and R. Alonso-Terme (2002), 'Does Corruption Affect Income Inequality and Poverty?' *Economics of Governance*, 3: 23–45.
- Gupta, S., H. Davoodi and E.R. Tiongson (2001), 'Corruption and the Provision of Health Care and Education Services', in A.K. Jain (ed.), *The Political Economy of Corruption*, London: Routledge, pp. 111–41.
- Hall, R. and C. Jones (1999), 'Why do Some Countries Produce so much more Output per Worker than Others?', *Quarterly Journal of Economics*, 114: 83–116.
- Hellman, J.S., G. Jones and D. Kaufmann (2000), 'Seize the State, Seize the Day: State Capture, Corruption and Influence in Transition', World Bank Policy Research Working Paper No. 2444.
- Hofstede, G. (1997), *Cultures and Organizations: Software of the Mind*, New York: McGraw Hill.
- Husted, B. (1999), 'Wealth, Culture, and Corruption', *Journal of International Business Studies*, 30 (2): 339–60.
- Johnson, R. and G. Libecap (1994), *The Federal Civil Service System and the Problem of Bureaucracy: The Economics and Politics of Institutional Change*, Chicago, IL: University of Chicago Press.
- Kaufmann, D., A. Kraay and P. Zoido-Lobaton (1999), 'Governance Matters', World Bank Policy Research Working Paper No. 2196, Washington, DC: World Bank.
- Knack, S. (2006), 'Measuring Corruption: A Critique of Indicators', in Eastern Europe and Central Asia', *Journal of Public Policy*, 27 (3): 255–91.
- Kunicova, J. and S. Rose-Ackerman (2005), 'Electoral Rules and Constitutional Structures as Constraints on Corruption', *British Journal of Political Science*, 35 (4): 573–606.
- Klitgaard, R. (1998), 'International Cooperation Against Corruption', *Finance and Development*, March: 3–6.
- Lambsdorff, Johann Graff (2006), 'Consequences and Causes of Corruption: What do We Know from a Cross-Section of Countries?', in S. Rose-Ackerman (ed.), *International Handbook on the Economics of Corruption*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 3–51.
- LaPalombara, J. (1994), 'Structural and Institutional Aspects of Corruption', *Social Research*, 61: 325–50.
- La Porta, R., F. Lopez-De-Silanes, A. Shleifer and R.W. Vishny (1997), 'Trust in Large Organizations', *American Economic Review, Papers, and Proceedings*, 137 (2): 333–8.

- Mauro, P. (1995), 'Corruption and Growth', *Quarterly Journal of Economics*, 110: 681–712.
- Mauro, P. (1997), 'The Effects of Corruption on Growth, Investment, and Government Expenditure: A Cross-Country Analysis', *Corruption and the Global Economy*, Washington, DC: Institute for International Economics.
- North, D. (1990), *Institutions, Institutional Change and Economic Performance*, New York and Melbourne: Cambridge University Press.
- Persson, T., G. Tabellini and F. Trebbi (2003), 'Electoral Rules and Corruption', *Journal of the European Economic Association*, 1 (4): 958–89.
- Shleifer, A. and R. Vishny (1993), 'Corruption', *Quarterly Journal of Economics*, 108 (3): 599–617.
- Sung, H.-E. (2004), 'Democracy and Political Corruption: A Cross-National Comparison', *Crime, Law and Social Change*, 41 (2): 179–93.
- Tanzi, V. and H. Davoodi (1997), 'Corruption, Public Investment, and Growth', International Monetary Fund Working Paper 97/139.
- Treisman, D. (2000), 'The Causes of Corruption: A Cross-National Study', *Journal of Public Economics*, 76: 399–457.
- van Rijckeghem, C. and B. Weder (2001), 'Bureaucratic Corruption and the Rate of Temptation: Do Wages in the Civil Service affect Corruption, and by How Much?' *Journal of Development Economics*, 65 (2): 307–31.
- Wei, S.-J. (2000a), 'Natural Openness and Good Government', World Bank Policy Research Working Paper No. 2411 and NBER Working Paper 7765.
- Wei, S.-J. (2000b), 'How Taxing is Corruption on International Investors', *Review of Economics and Statistics*, 82 (1): 1–11.
- World Bank (1997), *World Development Report 1997: The State in a Changing World*, Washington, DC: World Bank.
- World Bank (2000), *Anticorruption in Transition: A Contribution to the Policy Debate*, Washington, DC: World Bank.
- World Bank (2002), *World Development Report 2002: Building Institutions for Markets*, Washington, DC: World Bank.

60 Law and development

Pranab Bardhan

Rule of law

In his manuscript *Lectures on Jurisprudence*, based on his lectures at Glasgow University in the early 1760s, Adam Smith stated that a factor that ‘greatly retarded commerce was the imperfection of the law and the uncertainty in its application’. This is still one of the main messages of the Law and Economics literature as it pertains to development. Law and Economics is a thriving subject in the USA, and it is now being widely adopted in other countries, including in law schools of developing countries. But its Chicago origins and the general American mold may have given a particular slant to the development of the subject, which is not always quite appropriate for these countries. In this chapter I shall focus very generally on some of the special issues that arise in the context of developing countries that the literature on Law and Economics needs to address if it is to be applicable there. These special issues arise primarily because the institutional, political and behavioral context in these countries is different from the usual context of the literature.

In this literature as well as that of recent Institutional Economics the major emphasis is on contract law and security of property rights. In the pervasive context of incomplete contracts the emphasis is rightly on the residual rights of control, and the security and predictability of these property rights are crucial for economic performance and long-term investment. Throughout history in any time-separated activity – for example, if the seed planter cannot be secure in reaping the harvest, if a trap-setter cannot claim the trapped game, or a lender is uncertain of being repaid – economic life is hampered by insecurity of property rights. North and Weingast (1989) trace the success story of development in English history to the King giving up royal prerogatives and increasing the powers of the Parliament in 1688, thus securing private property rights against state predation and allowing private enterprise and capital markets to flourish. The more recent empirical literature has tried to quantify the effect of these property rights institutions – or what they call in this literature the ‘rule of law’ variable (one standard measure combines indices of effectiveness and predictability of judiciary, enforceability of contracts and incidence of crime) – on economic performance from cross-country aggregative data. Since these institutions may be endogenous (that is, economically better-off countries may have

more of those institutions, rather than the other way round), the literature tries to resolve the identification problem by finding exogenous sources of variations in those institutions. See Acemoglu et al. (2001, 2002).¹ Rodrik et al. (2002) use similar data to show that once the property rights institutions are accounted for, the role of other factors like geography or openness to trade in explaining cross-country variations in per capita income is minimal.

What is often ignored in this literature is that the 'rule of law' actually involves a whole bundle of rights, and we need to 'unbundle' it. Even for security of property rights, different social groups may be interested in different aspects of these rights. For example, the poor may be interested primarily in very simple rights like land titles, and also, to a very important extent, in protection against venal government inspectors or local mafia; to them that is the most salient aspect of security of property rights. For the richer investors, however, a whole range of other issues like protection of the minority shareholders in corporations, oversight of capital markets against insider abuse, bankruptcy laws, and so on, loom large; these are what investors emphasize when they talk about security of property rights. As different groups are thus interested in different aspects of security of property rights, these rights may have differential political sustainability, depending on how politically influential the corresponding groups are in a given polity.

'Rule of law' should also include other rights, some quite different from mere security of property rights. For example, one part may involve various democratic rights of political participation, association, mobilization and expression of 'voice'. An analysis of cross-country variations in human development indicators (which includes education or health variables like mass literacy or life expectation) shows that an institutional variable measuring 'voice' or participation rights is just as important as that measuring security of property rights as an explanatory variable; – see Bardhan (2005), Chapter 1. In other words, the part of 'rule of law' that refers to democratic participation rights explains a significant amount of variation in human development indices across countries. Those who emphasize property rights often ignore the effects of participatory rights, and there is some obvious tension between these two types of rights included in the standard package of 'rule of law'.

The idea of security of property rights has been extended to the case of intellectual property rights for the preservation of incentives for innovation. Since innovations are the main source of economic growth, laxity in the enforcement of international patents and copyrights in developing countries for products that are knowledge-intensive or require expensive investments in research and development is often regarded as harmful for

long-term economic growth. This has been the rationale for the incorporation of TRIPS (Trade-Related Intellectual Property Rights) in World Trade Organization (WTO) rules, when developing countries accepted these rules under some pressure from rich countries. While keeping incentives alive for new research and innovations is extremely important, the question from the point of view of a developing country is usually whether the enormous costs (including the often exorbitant monopoly prices charged by the patent holder for a prolonged period)² are always worth the benefits, and whether there are better alternative ways of encouraging research. It is recognized now by many scientific researchers that existing patents often act as an obstacle to further research that tries to build on earlier findings (in developing countries this includes research for adapting new technology to the special conditions there). This is linked with the question of the optimal patent breadth, which is about how broadly the protection of existing innovations ought to extend to related innovations in the future.³ The alternative method of subsidizing research inputs (rather than rewarding research output with temporary monopoly) has the advantage of encouraging information-sharing and collaborative research. Of course, upfront funding carries with it the moral hazard problem that researchers, once having secured funding, may be tempted to pursue activities or lines of research other than those most desired by the public sponsor. This problem may be mitigated if researchers expect to apply for public funding in future.

The problem of international patents in life-saving drugs in poor countries recently caught public attention in connection with the controversies about the prices of anti-retroviral drugs for AIDS patients in Africa. The major problem in corporate drug research is that only a tiny fraction of what the companies spend on finding new diet pills or anti-wrinkle creams is spent on drugs or vaccines against major killer diseases of the world's poor, like malaria or TB, and the situation has not changed with the onset of TRIPS and is not expected to change even with a more stringent enforcement of TRIPS in poor countries. So alternative avenues of encouraging such research have to be sought. There are now the beginnings of some international attempts to make credible arrangements on the part of international organizations like the World Health Organization (WHO) in collaboration with non-governmental organizations (NGOs) like Médecins sans Frontières, private foundations (like the Gates Foundation) and donor agencies and governments to a commitment to purchase vaccines to be developed by pharmaceutical companies against some of these diseases. For a discussion of the incentive issues in vaccine purchase commitments, see Kremer (2001). For other diseases (like diabetes or cancer) which kill large numbers of people in both rich and poor countries, the incentive argument for enforcing patents in poor countries is weak, since that research will

be carried out by the transnational drug companies in any case as the market in rich countries is large enough (provided resale can be limited).

We have earlier commented upon the different kinds of security of property rights being relevant for different social groups. In the case of intellectual property rights as well, the transaction costs may limit the symmetry of access of different groups to those rights. Khan and Sokoloff (1998), in a historical comparison of the patent systems in the USA and Britain in the first half of the nineteenth century, show that while the British system used to effectively limit access to intellectual property rights to the relatively wealthy and well-connected, access in the American system was much more broad-based, and this contributed to a much more vigorous and wider spread of patenting activity in the USA in that period.

Beyond formal legal institutions

While nobody will deny the importance of innovations in the process of economic growth, in the case of manufacturing technology in most developing countries the problem is really in adaptation of technology theoretically available elsewhere. Much of the effective use of that technology particularly in these alien circumstances is not codified, but implicit or tacit, and cannot be just transplanted from abroad. Learning by doing and domestic efforts to adapt and assimilate are critical, costly and time-consuming, and in this, government investment in market-supporting infrastructure and in research and training and extension are quite important. Just putting in place a legal system facilitating private efforts may not be enough. As Pack (2003) points out, in recent years many developing countries have liberalized domestic and international trade regulations but have not realized high total factor productivity, in the absence of a set of institutions constituting a national innovation system and extension services that facilitate appropriate training and technology absorption.

There are also corresponding implications for the inadequacy of just a legal framework in developing credit and equity markets or the requisite financial infrastructure in general. Investment in learning by doing is not easily collateralizable and is therefore particularly subject to the high costs of 'imperfect information'. At an early stage (which can be prolonged in poor countries) when firms are not yet ready for the securities market (with its demands for codifiable and court-verifiable information), there is often a need for some support and underwriting of risks by some centralized authority (with, of course, its attendant dangers of political abuse). There is also the problem of interdependence of investment decisions with externalities of information and the need for a network of proximate suppliers of components, services and infrastructural facilities with large economies of scale. Private financiers willing and able to internalize the externalities

of complementary projects and raise large enough capital from the market for a critical mass of firms are often absent in the early stage of industrialization. Historically, the state has played an important role in resolving this kind of coordination failure by facilitating and complementing private sector coordination – as the examples of state-supported development banks in nineteenth-century France, Belgium and Germany, and more recently in Japan, Korea, Taiwan and China, suggest. There are, of course, many examples of state failures in this respect and politicization of financial markets in other developing countries. In much of the literature on Law and Economics, as in Institutional Economics, the importance of the state is recognized only in the narrow context of how to use its power in the enforcement of contracts and property rights, and at the same time how to establish its credibility in not making confiscatory demands on the private owner of those rights. The history of the successful as well as failed cases of the state as a coordinator of technology assimilation and financial market development has lessons which should be analyzed in a framework that goes beyond this narrow context.

Why does a society not always adapt its legal and institutional set-up to facilitate productivity-enhancing innovations? Such innovations have gainers and losers, but in most cases the gainers could potentially compensate the losers. The problem is that it is politically difficult for the gainers from a change to commit credibly to compensate the losers *ex post*.⁴ As Acemoglu (2003a) puts it, there may not be any political Coase Theorem, whereby politicians and powerful social groups could make a deal with the rest of society, give up some of their control on existing rules and institutions that are inefficient, allow others to choose policies and institutions that bring about improvements in productivity, and then redistribute part of the gains to those politicians and groups. Such deals have severe commitment problems; those in power cannot credibly commit to not using this power in the process, and others cannot credibly commit to redistribute once the formerly powerful really give up their power for the sake of bringing about new rules and institutions.

A central issue of development economics is thus the persistence of dysfunctional regulations and institutions over long periods of time, as we discuss in Bardhan (2005), Chapter 2. In particular, the history of underdevelopment is littered with cases of formidable institutional impediments appearing as strategic outcomes of distributive conflicts. Acemoglu and Robinson (2002) develop a theory where incumbent elites may want to block the introduction of new and efficient technologies because this will reduce their future political power; they give the example from nineteenth-century history when in Russia and Austria-Hungary the monarchy and aristocracy controlled the political system but feared replacement and so

blocked the establishment of rules and institutions that would have facilitated industrialization. These replacement threats are, of course, often driven by extreme inequality in society.

In explaining the divergent development paths in North and South America since the early colonial times, Engerman and Sokoloff (2002) have provided a great deal of evidence of how in societies with high inequality at the outset of colonization rules and institutions evolved in ways that restricted to a narrow elite access to political power and opportunities for economic advancement. Initial unequal conditions had long-lingering effects, and through their influence on public policies (in distribution of public land and other natural resources, the right to vote and to vote in secret, primary education, patent law, corporate and banking law, and so on) tended to perpetuate those institutions and policies that atrophied development. Even in countries where initially some oligarchic entrepreneurs are successful in creating conditions (including securing their own property rights) for their own economic performance, as long as that oligarchy remains powerful, they usually get away with regulations that raise entry barriers for new or future entrepreneurs, and this blocks challenges to their incumbency and thus sometimes new technological breakthroughs. See Acemoglu (2003b) for a theoretical analysis of this kind of dynamic distortion in oligarchic societies even when property rights are protected for the initial producers. The classic example of inefficient rules and institutions persisting as the lopsided outcome of distributive struggles relates to the historical evolution of land rights in developing countries. In most of these countries the empirical evidence suggests that economies of scale in farm production are insignificant (except in some plantation crops), and the small family farm is often the most efficient unit of production. Yet the violent and tortuous history of land reform in many countries suggests that there are numerous roadblocks on the way to a more efficient reallocation of land rights, put up by vested interests for generations.

Inequality in power distribution in society also influences the social legitimacy of laws enacted or decreed by the powerful, and the degree of commitment of the general population to the rule of law. When the state is captured by a narrow clique, or when the state is weak so that there is an 'oligopoly' of coercion and authority (as opposed to the 'monopoly of violence' that Max Weber attributed to the state) shared by various protection rackets and corrupt officials (police, judges, bureaucrats), there is usually a big gulf between laws that are in the statute books and their enforcement, and, most importantly, a deficiency in every citizen's expectations about others' compliance, which form the foundation of the rule of law. Along with the underlying power distribution and enforcement mechanisms in society, some overarching social norms and political commitments provide

the main structure within the confines of which the formal legal system operates, and compared to the former the latter – which is the focus of much of the Law and Economics literature – is often in a secondary role.

These important elements of the institutional, political and social framework are ignored in a recent burgeoning of empirical literature on the effects of legal origins of a system. La Porta et al. (1997, 1999) have called attention to the superior effects, across countries, of the Anglo-Saxon common law system based on judicial precedents over the civil law system based on formal codes, on the corporate business environment both in terms of more flexibility with changing needs of business and in terms of better protection for external suppliers of finance to a company (whether shareholders or creditors). Apart from some doubts about the establishment of causality in these cross-national studies,⁵ one can also question the historical evidence even in the rich countries themselves. Lamoreaux and Rosenthal (2005) have done a comparative study of the constraints imposed by their respective legal systems on organizational choices of business in the USA (with its common law system) and France (with its civil law codes) during the middle of the nineteenth century around the time when both countries were beginning to industrialize. They conclude that there was nothing inherent in the French legal regime that created either a lack of flexibility or a lack of attention to the rights of creditors or small stakeholders. Many of the rules in the USA for minority shareholder rights actually came after the insider scandals of the Great Depression period. Franks et al. (2003) point out that in the UK it was not until as late as 1948 that the Parliament began to enact limited legislation to protect minority shareholder rights. Rosenthal and Berglof (2003) also question the primacy of legal origin in explaining institutions of investor protection; drawing upon the legislative history of US bankruptcy law they show how the USA, with an English common law legal origin, ended up with a bankruptcy regime quite different from that in the UK, and how political and ideological forces shaped financial development. Several legal scholars – see, for example, Roe (2003) – have pointed out how the nature of corporate governance even in American large firms depends more on socio-political factors than on the form of corporate laws.

In any case, as we have indicated earlier, the importance of the legacy of the formal legal system is rather moot where much too frequently in developing countries the enforcement of whatever the laws are in the statute books is quite weak, and the courts are hopelessly clogged and corrupt. Take the two largest developing countries, China and India. India has inherited the English common law system, and being a democracy, legal rights there are more well defined and the legal system is less subject to political discretion than in China under the monopoly control by a Communist Party. And yet, according to the World Bank Report on *Doing*

Business in 2005 (World Bank, 2005), it is China which seems less disadvantaged in most indicators of regulatory and judicial effectiveness in business matters. For example, registering property requires 67 days and costs about 14 per cent of property value in India, whereas in China it is 32 days and 3 per cent of property value. In enforcing debt contracts it requires 425 days and costs about 43 per cent of debt value in India, whereas in China it is 241 days and 26 per cent of debt value. On closing an insolvent business it takes about ten years in India, in China 2.4 years.

In many developing countries the efficiency of courts as mechanisms of resolving disputes or enforcing contracts is shaped by a rather warped system of incentives: judges, even when they are not corrupt, do not care about delays, lawyers earn more when court proceedings are prolonged, appeals are too easy and some defendants deliberately seek continual delay in judgment. Courts are congested because of too-lengthy procedures and built-in incentives for over-litigation, apart from administrative delays in appointments of judges. Such low judicial effectiveness in commercial law, apart from raising transaction costs all around, has important effects on the size and structure of firms. This is because the more effective the judicial process, the more you can have relatively complex contracts, larger firms can thrive and more complex goods be produced.

Social and behavioral presumptions

Finally, I am going to comment on some of the broad presumptions of the Law and Economics literature which may need to be changed or made more flexible if it is to be applied to developing countries. One relates to the scale of economic activity. In small peasant communities where the scale of economic activity is not large, informal relational contracts may be more efficient than rule-based contracts supported by elaborate legal-judicial procedures. Breaches of relational contracts are often observable by other community members even when not verifiable by courts, and punishment is usually through social sanctions and reputation mechanisms. Another advantage is flexibility and ease of renegotiation. But as the scale of economic activity expands, as the need for external finance becomes imperative, and as large sunk investments increase the temptation of one party to renege (and as increased mobility and integration with the outside world improve exit options), relational contracts and reputational incentives become weaker.⁶ As Li (2003) points out, relation-based systems of governance may have low fixed costs (given the pre-existing social relationships among the parties and the avoidance of legal-judicial and public information and verification costs of rule-based systems), but high and rising marginal costs (particularly of private monitoring) as business expansion involves successively weaker relational links.⁷

Of course the transaction costs of legal-judicial systems are asymmetric in their incidence on the rich and the poor as they try to get legal remedies, and it is not surprising that the legally handicapped poor often feel that the law is just another 'stick' with which the resourceful rich can beat them. In small face-to-face communities what anthropologists call the 'politics of reputation' may provide some modest measure of protection for the weak against the strong; as long as all parties belong to what is perceived to be the same 'moral community' in terms of which reputation is defined, there are some accepted limits and symbolic sanctions against the kind of ruthless exercises of power that sometimes accompany the cut-throat impersonality of the legal system enforced by the gendarmerie of the state.

It also needs to be recognized that in a world of highly imperfect information and the interlinked and multiplex nature of traditional informal contracts, the establishment of market relations enforced by the legal system in one market can crowd out implicit contracts in other related markets. Kranton and Swamy (1999) show in a study of the impact of the introduction of civil courts in British India on the agricultural credit markets of the Bombay Deccan that while it led to increased competition in the credit market, it reduced lenders' incentives to subsidize farmers' investments in times of crisis, leaving them more vulnerable in bad times, with insurance markets largely absent. In the context of environmental management of the village commons, Seabright (1993) has pointed out that as contracts are necessarily incomplete, attempts to enforce private property rights may weaken the mechanisms of cooperation that previously existed among the resource users, who may have shared implicit non-contractual rights in the common property resource.

The Law and Economics literature has inherited from mainstream economics the latter's behavioral postulate of rational self-interested individuals. This postulate is being increasingly questioned in the branch of economics that is now called 'behavioral economics', but there may be special reasons for questioning it in the context of poor countries. In traditional communities where your conformity to community norms is at a special premium, we may have to pay particular attention to social preferences ('other-regarding' as opposed to self-centered, or 'process-regarding' as opposed to simply outcome-oriented) which may go beyond the narrow interpretation of self-interested behavior. For example, social reciprocity (individuals going out of their way to reward helpful actions by other members of the community, or taking revenge for perceived unfair or nasty behavior on the part of others at some considerable cost to the revenge-taker – 'honor killings' in many traditional societies being the extreme but not uncommon case) is often a foundation stone of community norms,

which define the informal institutional framework within which particular legal rules can be implemented.

It is also a questionable presumption of the Law and Economics literature that individuals always behave in their best interests. Common observations of myopic, weak-willed, procrastinating and time-inconsistent behavior fly in the face of the inexorably rational economic man of our textbooks.⁸ This may be a special problem in poor countries where public information media are weak, many people are uneducated and superstitious, and there is a surfeit of touts, middlemen and operators trying to manipulate people to make hasty uninformed decisions. The innate psychological characteristics of people may not be different in poor countries, but their circumstances and information sources are often quite different, and capacity for complex calculations is an acquired trait, honed only as transactions become more complex. Also, people often internalize their constraints and by all accounts the constraints are much more severe in the case of poor people. All this may sometimes call for more paternalistic regulations than are admitted in the rational-choice framework of Law and Economics. For example, consumer protection regulations in food labeling and health warnings, publicizing of information about often the exorbitant implicit interest rates charged in instalment purchases of durables from retailers and pawnbrokers, publicizing the odds of winning lotteries (which are very popular, as most people systematically overestimate their chance), are all instances of paternalistic regulations that are particularly important in poor countries. One, of course, has to be wary of the slippery slope here that may easily end up in heavy-handed regulations or regulatory capture, but one cannot deny that the sovereignty of the rational consumer is a particularly egregious myth in such contexts.

Furthermore, the Law and Economics literature, particularly through its Chicago origins, has inherited a presumption about voluntary contracts that one may have to be careful about. Milton Friedman and others have repeatedly asserted that if parties enter into a transaction voluntarily (without adverse effects on third parties), legal rules should not interfere; they should play only an enabling or facilitating role in that transaction. There are, however, many cases, particularly in poor countries, where it is possible to show that one party in this transaction would have been actually better off if the law intervened to take out certain options from the choice set. Take the case of 'bonded labor'. Genicot (2002), in describing what she calls 'the paradox of voluntary choice', constructs a case where the strategic interaction between the landlord and the local credit institutions can constrain the poor peasant to 'choose' a bonded labor contract, whereas if bonded labour were banned it would have resulted in welfare-enhancing credit opportunities for the peasant. Basu (2000) models a

somewhat similar case of a woman choosing a 'sexual harassment contract' where she would have otherwise been better off if such contracts were disallowed. Similar cases can be argued for legally taking out the option for a poor worker to work in unsafe or hazardous conditions. These are all cases for interventionist regulations in the context of extremely unequal but 'voluntary' contracts.

Let us end with a comment on a fashionable attitude to the rule of law in the context of development that is sometimes expressed at the opposite end of the political spectrum. We have indicated earlier in this section as well at the beginning of this chapter that the rule of law is often an instrument in the hands of the propertied, ruling over and restricting the activities of the propertyless. This undoubted fact sometimes leads commentators to dismiss the rule of law merely as an instrument of class oppression or as part of a modernizing elitist project that rides roughshod over the 'subaltern'. In the face of such tendentious simplifications we can do no better than to quote here from the far more nuanced historical analysis of E.P. Thompson. At the conclusion of his 1975 book, *Whigs and Hunters* (which shows how a political oligarchy in eighteenth-century England invented callous and oppressive laws to serve its own interests) Thompson writes:

We reach, then, not a simple conclusion (law = class power) but a complex and contradictory one. On the one hand, it is true that the law did mediate existent class relations to the advantage of the rulers . . . On the other hand, the law mediated these class relations through legal forms, which imposed, again and again, inhibitions upon the actions of the rulers. . . . In a context of gross class inequalities, the equity of the law must always be in some part sham . . . We ought to expose the shams and inequities which may be concealed beneath this law. But the rule of law itself, the imposing of effective inhibitions upon power and the defence of the citizen from power's all-intrusive claims, seems to me to be an unqualified human good. To deny or belittle this good is . . . a desperate error of intellectual abstraction. (pp. 264–6)

Notes

1. For a discussion of the limitations of such exercises see Bardhan (2005, Chapter 1).
2. Even when the original patent is about to run out, the transnational company holding the patent often has various ways of effectively extending it: by slightly changing the composition of ingredients in the product and then taking out a new patent, bribing or intimidating the potential producers of the generic substitute, and through high-pressure advertisement keeping many of the customers hooked on to the original brand.
3. For a discussion of some of these issues see the papers by Gallini and Scotchmer (2001) and Kremer (2001).
4. For a review of the theoretical political economy literature on credibility of commitment see Bardhan (2005, Chapter 4).
5. For example, among developing countries many French legal origin countries are in Africa or Latin America and it may be standing as a proxy for other (unmeasured) deficiencies in state capacity in several of these countries.

6. Some of the pros and cons of relational contracting are empirically studied in the case of Vietnam's emerging private sector by McMillan and Woodruff (1999).
7. For a formal treatment of the subject see Dixit (2003).
8. The standard argument that 'irrational' behavior is weeded out in the evolutionary process is much too limited. Other-regarding cooperative behavior may be more successful in many cases. Evolutionary success in replication and the economist's narrow conception of efficiency may not go together if pay-offs to adherence to particular behavioral rules depend on adherence by others, or if there are positive and negative interactions of different behavioral rules.

References

- Acemoglu, D. (2003a), 'Why Not A Political Coase Theorem? Social Conflict, Commitment and Politics', *Journal of Comparative Economics*, 31: 620–52.
- Acemoglu, D. (2003b), 'The Form of Property Rights: Oligarchic vs. Democratic Societies', NBER Working Paper No. 10037, 2003.
- Acemoglu, D. and J.A. Robinson (2002) 'Economic Backwardness in Political Perspective', NBER Working Paper.
- Acemoglu, D., S. Johnson and J.A. Robinson (2001), 'The Colonial Origins of Comparative Development: An Empirical Investigation', *American Economic Review*, 91 (5): 1369–401.
- Acemoglu, D., S. Johnson and J.A. Robinson (2002), 'Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distribution', *Quarterly Journal of Economics*, 117 (4): 1231–94.
- Bardhan, P. (2005), *Scarcity, Conflicts and Cooperation: Essays in Political and Institutional Economics of Development*, Cambridge, MA: MIT Press.
- Basu, K. (2000), *Prelude to Political Economy: A Study of the Social and Political Foundations of Economics*, Oxford: Oxford University Press.
- Dixit, A. (2003), 'Trade Expansion and Contract Enforcement', *Journal of Political Economy*, 111 (6): 1293–317.
- Engerman, S.L. and K.L. Sokoloff (2002), 'Factor Endowments, Inequality, and Paths of Development among New World Economies', NBER Working Paper.
- Franks, J., C. Mayer and S. Rossi (2003), 'The Origination and Evolution of Ownership and Control', Oxford Financial Research Centre Working Paper No. 1003-FE-01.
- Gallini, N. and S. Scotchmer (2001), 'Intellectual Property: When is it the Best Incentive System?' NBER Working Paper.
- Genicot, G. (2002), 'Bonded Labor and Serfdom: A Paradox of Voluntary Choice', *Journal of Development Economics*, 67 (1): 101–27.
- Khan, B.Z. and K.L. Sokoloff (1998), 'Patent Institutions, Industrial Organization and Early Technological Change: Britain and the United States, 1790–1850', in M. Berg and K. Bruland (eds), *Technological Revolutions in Europe*, Cheltenham, UK and Lyme, USA: Edward Elgar, pp. 292–313.
- Kranton, R.W. and A.V. Swamy (1999), 'The Hazards of Piecemeal Reform: British Civil Courts and the Credit Market in Colonial India', *Journal of Development Economics*, 58 (1): 1–24.
- Kremer, M. (2001), 'Creating Markets for New Vaccines', *Innovation Policy and the Economy*, Cambridge, MA: MIT Press.
- Lamoreaux, N.R. and J.L. Rosenthal (2005), 'Legal Regimes and Business's Organizational Choice: A Comparison of France and the United States during the Mid-Nineteenth Century', *American Law and Economics Review*, 7 (1): 28–61.
- La Porta, R., F. López-de-Silanes, A. Shleifer and R. Vishny (1997), 'Legal Determinants of External Finance', *Journal of Finance*, 52 (3): 1131–50.
- La Porta, R., F. López-de-Silanes, A. Shleifer and R. Vishny (1999), 'The Quality Of Government', *Journal of Law, Economics and Organization*, 15: 222–79.
- Li, J.S. (2003), 'Relation-based versus Rule-based Governance: An Explanation of the East Asian Miracle and Asian Crisis', *Review of International Economics*, 11 (4): 651–73.

- McMillan, J. and C. Woodruff (1999), 'Interfirm Relationships and Informal Credit in Vietnam', *Quarterly Journal of Economics*, 114 (4): 651–73.
- North, D.C. and B. Weingast (1989), 'Constitutions and Commitment: Evolution of Institutions Governing Public Choice', *Journal of Economic History*, 49 (4): 803–32.
- Pack, H. (2003), 'Review of Parente and Prescott', *Journal of Development Economics*, 70 (1): 243–7.
- Rodrik, D., A. Subramanian and F. Trebbi (2002), 'Institutions Rule: The Primacy of Institutions over Geography and Integration in Economic Development', NBER Working Paper.
- Roe, M.J. (2003), *Political Determinants of Corporate Governance*, New York: Oxford University Press.
- Rosenthal, H. and E. Berglof (2003), 'The Political Origin of Finance: The Case of Federal Bankruptcy Law in the United States', Cambridge, MA: Harvard University.
- Seabright, P. (1993), 'Managing Local Commons: Theoretical Issues in Incentive Design', *Journal of Economic Perspectives*, 7 (4): 113–34.
- Thompson, E.P. (1975), *Whigs and Hunters: The Origin of the Black Act*, New York: Pantheon Books.
- World Bank (2005), *Doing Business in 2005*, Washington, DC: World Bank.

61 Institutions, property rights and development

Jean-Philippe Platteau and Jean-Marie Baland

Introduction

The New Institutional Economics (NIE), which has expanded rapidly since the mid-nineteenth century, has placed the issue of property rights at the forefront of its intellectual priorities. This is not surprising since, in the absence of well-established and well-enforced property rights, trade transactions may involve considerable transaction costs which have the effect of slowing market development. Perhaps to the dismay of those who believe in the absolute superiority of private ownership, a transaction-cost analysis aimed at assessing the relative efficiency of different property regimes does not necessarily point to the desirability of private property. In this chapter, we elucidate the kinds of circumstances under which private property is likely to dominate alternative property regimes, and this is done by limiting our attention to land and other natural resources (forests, lakes, pasturages, and so on). The demonstration proceeds in two steps. First, we compare a regime in which there are no property rights (the so-called open access regime) to alternative regimes in which such rights exist. Thereafter, we discuss the advantages and disadvantages of a regime of collective or community ownership *vis-à-vis* a regime of private ownership. The final section offers concluding remarks.

Open access versus property rights

A simple but basic principle taught by economics since Adam Smith is that a commodity, whether a good, service or asset, carries a low value if its supply is abundant relative to the demand expressed for it. The value can even be nil if the commodity is so plentiful that there is actually no competition between potential users. The next step is to recognize that, if a resource has no value, users have no incentive to seek to establish and acquire property rights which would assure them exclusive and free disposal of a portion of that abundant resource. The above proposition applies very well to land and other natural resources. Thus, in an environment characterized by low population density, access to land and natural resources is so easy that a portion of them can be occupied and exploited by individuals or collective entities without there being any need for them to defend against possible encroachments from outside.

Problems arise when an increasing number of individuals want to use a resource that is protected by no exclusive property rights acting as barriers to entry. More exactly, the regime under which no property rights exist (known as *res nullius*) begins to be problematic as soon as the entry of an additional user in the resource domain causes a damage to those users who were exploiting it before, and there is no penalty that this additional entrant is required to pay for such negative externalities. There are in fact two different sorts of externalities involved here.

The first type is known as congestion externalities. They manifest themselves in all the cases where the use of a natural resource necessitates that it is divided into distinct portions individually exploited by the users. When pressure on this resource increases, the activity of an individual user then unavoidably entails ecological costs for the other users, especially those located in the immediate neighbourhood. For example, a farmer who has cut off trees in the upper part of a hill in order to open a new field can cause erosion that will result in the silting of the lower parts.

The second type of externalities are rent-dissipating. They are observed in conditions where a natural resource is jointly, rather than individually exploited. And the problem arises only insofar as the resource is subject to decreasing returns, which implies that there is some degree of population pressure. In these conditions, through his additional harvesting effort, a new entrant causes a fall in the average productivity that is felt by all the previous users. A rational individual will enter into a resource domain, or continue to increase its activity therein, as long as the benefit he obtains exceeds the cost. What needs to be emphasized is that he will so act even though his income, or part of his income, is earned at the expense of previous users of the resource whose incomes are diminished as a result of his additional efforts. It is even possible that the additional effort of the new entrant does not give rise to any increase of output (the marginal productivity of effort is nil), yet the new entrant finds it individually profitable to apply this effort.

From the social viewpoint, such a decision is evidently undesirable since the costs thereby incurred are borne in vain, that is, they are entirely unproductive. A social waste of scarce production factors (labour, capital, fuel, and so on) occurs to the extent that the same aggregate output could have been obtained without applying additional amounts of these factors. The natural resource can then be said to be inefficiently managed. If the unproductive factors are used, it is because they enable the individual who owns them to participate in the exploitation of the resource and to obtain the average product which decreases as the number of harvesters rises. In the open access equilibrium, the resource rent is entirely dissipated (the costs eat the product), an absurd outcome since the natural resource is scarce

(subject to decreasing returns), and should therefore have a positive value reflected in the rent.

In fact, the root cause of this inefficient use of a natural resource lies in the absence of property rights. Indeed, had property rights prevailed, the addition of an effort unit to be applied to the resource would have been decided by the owner only if the expected additional product exceeded the cost. In other words, it is the marginal product instead of the average product that would have been compared to the cost in order to decide whether an additional unit of effort must be applied. The marginalist rule guarantees the efficient use of a resource and, which amounts to the same thing, the maximization of the rent associated with it. It has this property because it allows a perfect internalization of externalities (see Demsetz, 1967; Alchian and Demsetz, 1973, for a first formulation of this well-known principle).

Dynamic losses must also be taken into consideration when assessing the pros and contras of a given property regime. More precisely, in the same way that he has no reason to be concerned with the damage caused to other users by causing a fall in their incomes, an individual harvester is not induced to forsake present benefits in order to ensure better the long-term conservation of the resource. For example, a fisherman has no incentive to throw back juveniles into the sea so that they can grow to mature size and be caught in their adult state at a later time. His reasoning is, indeed, that if he does not keep those juveniles when they are caught in his net, other fishermen will not hesitate to kill them with the result that they will anyway be unable to reach mature size. Had private property rights existed, the owner would have seen to it that conservation measures are adopted, since the future benefits of present sacrifices would have accrued to him rather than to other users.

Private property versus communal property

General considerations

In the absence of transaction costs, communal property allows rightsholders to internalize externalities as effectively as private property. Under this condition of zero transaction costs, the two property regimes are thus strictly equivalent (Platteau, 2000, Chapter 3). As a matter of fact, a group of people who own a resource will make exactly the same decisions as an individual owner regarding the amount of effort to be applied. Such an outcome results from the fact that, like an individual, a group seeks to maximize the rent or the surplus, that is, the difference between the value of the flow of produce extracted from the resource and the capital and labour costs. In both cases, the marginalist decision rule is applied.

But the assumption of zero transaction costs is totally unrealistic: it is useful to set a theoretical reference point, but not to describe reality. Once transaction costs are taken into account, the members of the property rights school contend, private property appears superior to any system of collective or communal property. The underlying argument can be summarized as follows: while an individual owner forms an organically integrated decision unit (he need not discuss with anyone else than himself in order to reach decisions regarding the use of the resource), so as to create a one-to-one relationship between individual actions and their effects,¹ a collective owner must achieve an agreement between its members to decide how to use it. An agreement will not necessarily emerge from intra-group discussions. And, even if an agreement can be eventually reached, the negotiation process will necessarily entail non-trivial costs, such as the opportunity cost of the time spent in meetings, transport expenses, the costs of communicating the time and place of the meeting, and so on (Baland and Platteau, 1998b; Platteau, 2000, Chapter 3). Lastly, assuming that an agreement has been found and that the details of its *modus operandi* have been worked out, there remains the delicate question as to how it will be effectively enforced. At the very least, one may fear that its implementation will not remove all the possible sources of inefficiency, something which single private ownership is apparently capable of achieving. In the following, all the costs plaguing group ownership will be referred to as governance costs.

The determinants of governance costs

The feasibility of an agreement regulating the use of a natural resource at village level is itself dependent on a certain number of factors, among which the size of the user group and its degree of heterogeneity stand foremost. Regarding the first factor, it is evident enough that the smaller the number of rightsholders the lower the negotiation costs involved in the process of devising the regulatory agreement. If their number is too high, on the contrary, one may fear that no regulatory mechanism will be put into place for lack of an internal agreement.

The impact of heterogeneity is also evident. Thus, it is easy to understand that divergences between group members regarding the intended uses of a natural resource will make an agreement more difficult to reach. Illustrations are numerous, as attested by the prolonged conflicts between farmers and herdsmen around land areas claimed by each category for their own specific purpose. Herdsmen want to maintain their customary rights to large grazing areas at least during a part of the year (so that animals can feed themselves on crop residues after harvest time), while farmers are increasingly eager to win exclusive rights over well-delineated zones so as

to be able to practice more intensive forms of agriculture requiring continuous cultivation and long-term land improvements.

Income or wealth inequality constitutes another form of heterogeneity that tends to make regulation of the use of natural resources difficult to achieve. It has indeed been shown that the more unequal the distribution of income between members of a village community or any user group, the harder it is to find a regulatory scheme that satisfies all the people concerned. Moreover, if such a scheme exists, the efficiency gains that it will yield compared to a situation with no regulation diminish as income distribution becomes more unequal (Baland and Platteau, 1998a; 2003). The underlying intuition is simple: when there is a need to regulate the use of a resource, the group involved must not only determine the extent to which the intensity of use must be reduced to approach efficiency, but also the manner in which the effort reduction will be shared among the various users. If users are relatively identical, the latter problem is unlikely to be serious: a uniform reduction of individual effort levels appears as the natural solution. If, on the contrary, the users are different in terms of wealth or income and these differences are reflected in different rates of resource use, the problem of sharing the burden of effort reduction obviously becomes more complicated. In particular, the efficient solution might well imply that the larger part of this burden be borne by the less productive or the more impatient (those more preoccupied with subsistence constraints) users who often are the poorer members of the community.

Let us nonetheless assume that an agreement can be found that entails efficiency gains and allows each resource user to improve their situation compared to the status quo state of no regulation (only access rights exist). There remains the tricky issue of the enforcement of the regulatory scheme: once the rules are decided and agreed upon, each user has an incentive to violate them. The problem is especially serious because, by rendering the resource more valuable than before, effort restriction has the effect of increasing the benefits which can be obtained by exceeding one's allowed quota while other users follow the rule. To put it in another way, the outcome of the agreement is to enhance the temptation for individual users to free-ride on the sacrifices incurred by fellow users.

In addition, again assuming that an agreement is feasible, there is the question as to who will bear the costs of formation of collective action, understood as the costs involved by the very process of creating collective mechanisms for both decision-making and enforcement. Again, the characteristics of the user group – its size and heterogeneity, in particular – influence the extent to which this problem can be surmounted (Baland and Platteau 1997; 2003). To begin with, the impact of group size is identical to

the one observed when the problem consists of devising a regulatory scheme: the smaller the size of the community, the more likely the costs of formation of collective action will be actually incurred and, therefore, the more likely the agreed rules will be designed and applied. The main argument here is known as the incentive dilution argument (Olson, 1965). Each individual makes a personal calculation when he decides whether or not to contribute to the production of a local public good (or, in our context, to the creation of a collective mechanism, on the one hand, and to the implementation of a regulatory scheme, on the other hand). The individual compares the gains from abstaining from contributing to the collective effort with the cost. When a group is smaller, the cost of withdrawing participation obviously rises relative to the benefit.

In addition, when a group is smaller, members tend to know each other better and, therefore, reputation effects are more important. More attention will be paid to the future consequences of opportunistic behaviour in order to avoid punishment in the form of exclusion from the group or denial of the right of access to the resource. Furthermore, not only is communication facilitated within a smaller group, but the formation of collective identity feelings is also easier and, as a consequence, individuals are more induced to take into account the effects of their decisions on the other members (Baland and Platteau, 1996, pp. 75–8).

The impact of heterogeneity is more ambiguous than the impact of group size. In fact, it cannot easily be predicted a priori. Let us examine, in particular, the impact of inequality of income or wealth on efficiency in the production of a public good, such as the formation of a regulatory body or framework. Two effects are at work which run into opposite directions. On the one hand, a great inequality creates a situation in which the rich guy internalizes a large part of the externalities created by his particular contribution to the public good, thereby inducing him to apply the required effort. Yet, on the other hand, those who are at the lower tail of the income distribution find themselves in exactly the opposite situation: they will draw only minor benefits from the collective good and hence they will have weak incentives to contribute.

It is thus impossible to predict in a general manner whether a higher degree of income inequality will actually result in an increase or a decrease of the aggregate amount of contributions to the local public good. The aggregate amount will rise only if the increased contributions of the village elite (who better internalize the externalities) exceed the reduced contributions of the common people. What is certain, however, is that an extreme inequality corresponding to a total concentration of all the wealth in the hands of a single villager will lead to an efficient provision of the public good (see Baland and Platteau, 2007).

The evolution of governance costs

As is evident from the above discussion, communal property gives rise to serious incentive problems, especially when it involves the regulation of the use of natural resources besides the setting of access rules. A reasonable position therefore consists of admitting that inefficiencies are bound to persist under this ownership regime, whether in a static form (the dissipation of part of the resource rent) or in a dynamic form (the lack of investment to conserve the resource, including actions to fight against predators who threaten its long-term stock).

A straightforward consequence of such a situation is the following: the people's ability to cooperate in the management of common access resources determines the profitability of jointly held resources compared to their profitability when they have been individualized, and it also influences the allocation of resources between various uses. Inasmuch as this cooperation ability varies from one area to another, one must expect to observe geographic variations in the uses and rates of profitability of local-level natural resources. For example, it has been shown that in Mexico, when cooperation fails in the management of collectively grazed pastures, more land is allocated to crops than under successful cooperation and less to pastures, while the stocking rate on pastures is increased. This results in too much land in extensive crops and too many animals per hectare of pasture (McCarthy et al., 1998).

The point that we want to make now is that efficiency losses caused by externalities are likely to grow with the value of the resource, hence the frequent emphasis in the literature on the unit value of natural resources as one of the main determinants of its privatization (division). To illustrate, in his classical study of the Swiss Alps, Netting contrasts the lowlands of the valley which are fertile and therefore tend to be privately appropriated with the more arid highlands which are used as communal (summer) pastures under the authority of the village council (Netting, 1976; 1981).

Population pressure bears upon the efficiency gains of division in two different ways. For one thing, by increasing the number of users per unit area, it creates more room for external effects and, thereby, the governance costs and the inefficiencies involved in the joint exploitation of the resource rise. For another thing, by making the resource increasingly scarce, population pressure enhances its value and therefore makes for increased aggregate losses from collective exploitation. In other words, the amount of the rents foregone by not dividing the resource tends to increase with population. This is especially true when population pressure involves a transformation of the pattern of resource use, such as a shift from extensive to intensive agricultural or grazing practices, since intensive practices have the

effect of increasing the potential value of the resource per unit area and thus enhance the gains of private property.

Market penetration and the ensuing commercialization of products from primary activities is another critical determinant of the relative profitability of private property. It is by enhancing the realizable value of natural resources that growing integration of rural communities into developing market networks increases the benefits of resource division. Thus, in many developing countries, dramatic increases in prices for fuelwood or fresh fish as a result of a rapid expansion of urban markets have prompted rural inhabitants to intensify the exploitation of many forest and fish resources during the post-independence period. Significant efficiency losses have resulted from the growing pressure on these resources and the rising incidence of negative external effects that have accompanied it (see, for example, Baland and Platteau, 1996, pp. 262–70).

At this stage of our analysis, it appears that two series of factors bear upon the efficiency of regulated communal property compared to that of private property. On the one hand, there is the cooperation ability of the resource users, which is itself determined by the size and the degree of heterogeneity of the group or community to which they belong. On the other hand, there is the degree of scarcity of the resource as reflected in its value, which is determined by: (1) the intensity of population pressure in the area; and (2) the extent of market integration. A third series of factors influences the extent of efficiency losses resulting from the collective exploitation of a resource or the extent of the potential efficiency gains of its privatization. This last set of factors bears upon the costs of privatization, understood as a process of division of a jointly held resource accompanied by its partitioning into individually held portions. In the remainder of this chapter, we focus our attention on the two main costs of privatization, namely direct transaction costs and opportunity costs.

The role of direct transaction costs

When the problem of choice of ownership regime is considered from the exclusive standpoint of governance costs, the balance sheet is unmistakably favourable to the division and private appropriation of local-level natural resources. But this is only one side of the balance sheet. So far, indeed, we have implicitly assumed that the establishment and protection of property rights are costless operations. Such an assumption is manifestly unrealistic and we need to remove it now. What appears then is that, compared to communal property, private property is costlier from the viewpoint of direct transaction costs, which primarily include set-up and protection expenses. There thus exists a trade-off between two kinds of costs that have a different impact according to the ownership regime considered: the governance costs

that affect communal but not private property, on the one hand, and the direct transaction costs that are smaller under the former than under the latter regime, on the other hand.

The latter conclusion follows from the fact that it is less costly to fence, demarcate and protect a territory of a given size than to do those things for divided portions of that territory. As a corollary, when the surface area of a resource domain is larger, the per capita direct transaction cost of privatizing it increases. To put it in another way, direct transaction costs increase with the physical base of the resource: the more spread the resource base (or the less concentrated the resource) the higher the costs of delimiting and defending the resource territory. Other things being equal, therefore, the more spread a resource is, the less profitable it is to privatize it, and the more compact the resource, the more attractive is its privatization.

A consequence of the above is that we expect private property rights to be established over high-density resources and communal property rights over resources with the opposite characteristic. It is good news for economic theory that such a prediction is systematically verified in reality. Yet, at the same time, one must reckon that practically it is often difficult to disentangle the impact of the density of a resource from that of its value on the probability of privatization. As a matter of fact, high-value resources – for example, fertile lands that are susceptible to being irrigated owing to their favourable location – tend to be divided more often than low-value resources – for example, semi-arid lands that are hardly suitable for any other purpose than extensive grazing. In other words, there exists a strong correlation between the density or compactness of a resource and its value. In fact, in many instances the second characteristic causally determines the first one. What empirical evidence reveals is that compact resources with a high value (for example, intensively cultivated lands or fertile lands located near an important market town) are generally held under private property while resources that stretch over large areas and carry a low value (for example, the immense low-quality grazing areas in Mongolia or the Maasai Mara in Kenya) are jointly held by a local user group or community. Our analytical argument simply shows that these two correlated characteristics of a natural resource – its high value per unit area and its high density – tend to make its privatization relatively profitable.

Two remarks are in order. First, there exist natural resources, the division of which would entail prohibitively high direct transaction costs under the present state of technology. For example, the open sea – or, more exactly, the fish stock contained in it – presents insuperable difficulties for private appropriation. The enforcement of exclusive property rights to individual patches carved up in the ocean would, indeed, be infinitely costly. This is especially evident when fish species are mobile and move over large water

spaces, since exclusive rights are too costly to establish and enforce whether over the resource or the territory in which it moves. The example of wildlife reserves also comes to mind.

Interestingly, even in the case of maritime fisheries, privatization may sometimes be a viable solution. This tends to happen when species are rather sedentary (for example, lobsters, shellfish, molluscs, seaweeds) and live in relatively compact and well-delimited spaces, such as when the fish are found around islands (the Pacific islands, the Shetlands in Scotland, and so on) or in relatively well-sheltered aquatic zones (for example, in deltas or in backwaters) where fishing locations can be easily demarcated and protected against external encroachments. In these conditions, fishing spots are frequently assigned to individuals or families for their exclusive use and these private rights can generally be inherited by future generations as long as they are used effectively (see Platteau, 2000, p. 85, for references to the anthropological literature).

Second, the direct costs of resource division are not exogenously fixed. In the above, we have considered that these costs are determined by the inherent characteristics of each resource, it being understood that a resource may take on various forms and characteristics depending on the precise location and environment in which it is found. Here, we want to point out that, in fact, direct transaction costs may fall with technological progress. One well-known example is the discovery of the barbed wire which proved to be a decisive step in the reduction of the cost of protecting property rights through cheaper fencing of agricultural fields (North, 1981). Another example is the introduction of modern borehole drilling facilities in arid and semi-arid areas where this has the effect of facilitating the privatization of common grazing areas. Before this invention, in a country like Botswana, water extraction was subject to important scale economies as a result of which grazing lands were always the collective property of herders' communities (Peters, 1994). Private appropriation is thus not only facilitated by factors which contribute to enhance the value of a resource, but also by factors which have the effect of reducing the direct cost of partitioning.

The role of opportunity costs

Two types of opportunity costs appear to play an important role in this respect: scale economies and insurance benefits associated with collective ownership.

Let us first consider the impact of scale economies. Resources offering multiple products tend to be subject to scale economies to the extent that they form part of an overall ecosystem. This multiple product character of the resource is a reason often mentioned to argue against the parcelling out

of forest areas into individual holdings. In the case of hunting, on the other hand, wild animals require large territories to survive and reproduce, so that division of a hunting domain into smaller parcels would imply the destruction of the resource.

When discussing the role of scale economies, it is important to bear in mind that they may be present not in the resource itself but in complementary factors. The obvious advantage of coordinating the herding of animals so as to economize on shepherd labour in extensive grazing activities is probably the best illustration of the way scale economies in a complementary factor may prevent the division of a resource domain. Another illustration can be taken from fisheries (maritime or inland). In many cases, indeed, the guarding of privately apportioned fishing spaces is subject to strong scale economies. As a consequence, it may make sense for several individuals or families to get together to enforce a *de facto* right of collective property over a given fishing ground. This is actually what many traditional fishing communities have done in the past when competition around scarce fish resources began to develop, particularly in inland fisheries.

Opportunity costs of privatization also come into the picture when returns to a resource are highly variable across time and space. The need to insure against such variability is then a consideration that militates against resource division (McCloskey, 1976; Dahlman, 1980). When a resource has a low predictability (that is, when the variance in its value per unit of time per unit area is high), indeed, users are generally reluctant to divide it into smaller portions because they would thereby lose the insurance benefits provided by the resource kept whole. In the words of Nugent and Sanchez: 'the lower the quality of land or the more variable the weather, the more important it is that the land be held in communal, that is, tribal form' (Nugent and Sanchez, 1993, p. 107).

The example of extensive grazing and also that of maritime fishing again provide us with good illustrations of the above. Herders (fishermen) typically need to have access to a wide portfolio of pasture lands (fishing spots) insofar as, at any given time, wide spatial variations in yields result from climatic or other environmental factors. Assuming that the probability distributions are not correlated too much across spatial groupings of land or water and that they are not overly correlated over time, a system offering access to a large area within which rightsholding users can move freely appears as highly desirable from a risk-reducing perspective.

It must nevertheless be pointed out that private property rights over portions of the resource could apparently solve the predictability problem. Consider the case of extensive grazing again. Due to the unpredictability of rain-induced growth of grasses within any small region, what is important for herders is to be able to move over large ranges of land and rapidly

change location when the need arises. By holding exclusive ownership rights over widely dispersed patches of pasture lands, they would therefore achieve their objective of risk reduction. Clearly, to account for the maintenance of communal property, one needs to appeal to transaction-cost considerations and the high exclusion costs of a spread resource base (Platteau, 2000, p. 88). Because it would be prohibitively costly to enforce exclusive rights over widely dispersed and infrequently visited ranch patches, the division of the resource domain turns out to be infeasible. The same situation actually obtains in many fisheries.

The additional (direct) transaction costs implied by the necessity to insure against income fluctuations in the event of division or privatization of a resource increase with the variability of incomes and the surface area of the domain of this resource. As a matter of fact, the higher the variability of incomes the larger the number of resource portions that a particular user needs to insure himself and, hence, the higher the costs of establishing and protecting private property rights. On the other hand, the more stretched is the resource base the higher the (direct) transaction costs caused by the demarcation and the guarding of a given portion of the resource.

Conclusion and final considerations about the evolutionary approach to institutions

Two central conclusions emerge from our analysis. First, the transaction-cost economic theory of institutions leads us to expect that an increasing number of village-level natural resources will be divided and individually held as they acquire more value under the combined impact of population growth and market penetration. The important role of governance costs associated with collective ownership goes a long way toward explaining this gradual shift from corporate to private forms of ownership. Nonetheless, and this is our second conclusion, certain resources possess characteristics that make their division and their private appropriation especially costly. The costs involved are those required to establish and protect private property rights (direct transaction costs), or opportunity costs resulting from the loss of benefits provided by communal property.

The second conclusion raises a thorny issue. Indeed, in order that communal property be viable in conditions where private property is infeasible owing to prohibitively high direct transaction costs, or undesirable owing to high opportunity costs, it is essential that governance costs remain within tolerable limits. If this condition is violated, the natural resource concerned will not be regulated in a satisfactory manner and efficiency losses will be significant, perhaps considerable. It will either become a non-regulated common property (meaning that it is characterized by access

rules while rules of use are absent), or it will fall under the open access regime. If the number of users is large, these last two regimes will produce more or less equivalent results in the form of grave inefficiencies both in the static (rent dissipation) and in the dynamic senses (destruction of the stock following a lack of conservation investments or a lack of control of the extraction efforts applied by users).

Note

1. 'A primary function of property rights is that of guiding incentives to achieve a greater internalization of externalities.' (Demsetz, 1967, p. 348).

References

- Alchian, A. and H. Demsetz (1973), 'The Property Right Paradigm', *Journal of Economic History*, 33 (1): 16–27.
- Baland, J.M. and J.P. Platteau (1996), *Halting Degradation of Natural Resources: Is There a Role for Rural Communities?* Oxford: Clarendon Press.
- Baland, J.M. and J.P. Platteau (1997), 'Wealth Inequality and Efficiency in the Commons – Part I: The Unregulated Case', *Oxford Economic Papers*, 49 (4): 451–82.
- Baland, J.M. and J.P. Platteau (1998a), 'Wealth Inequality and Efficiency in the Commons – Part I: The Regulated Case', *Oxford Economic Papers*, 50 (1): 1–22.
- Baland, J.M. and J.P. Platteau (1998b), 'Dividing the Commons: A Partial Assessment of the New Institutional Economics of Property Rights', *American Journal of Agricultural Economics*, 80: 644–50.
- Baland, J.M. and J.P. Platteau (2003), 'Economics of Common Property Management Regimes', in G. Mähler and J. Vincent (eds), *Handbook of Environmental Economics*, Amsterdam: North-Holland, pp. 127–90.
- Baland, J.M. and J.P. Platteau (2007), 'Collective Action and the Commons: The Role of Inequality', in J.M. Baland, P. Bardhan and S. Bowles (eds), *Inequality, Cooperation and Environmental Sustainability*, Princeton, NJ: Princeton University Press, pp. 10–35.
- Dahlman, C.J. (1980), *The Open Field System and Beyond: A Property Rights Analysis of an Economic Institution*, Cambridge: Cambridge University Press.
- Demsetz, H. (1967), 'Toward a Theory of Property Rights', *American Economic Review*, 57 (2): 347–59.
- McCarthy, N., A. de Janvry and E. Sadoulet (1998), 'Land Allocation Under Dual Individual–Collective Use in Mexico', *Journal of Development Economics*, 56 (2): 239–64.
- McCloskey, D.N. (1976), 'English Open Fields as Behavior Towards Risk', in P. Uselding (ed.) *Research in Economic History*, Vol. 1, Greenwich, CT: JAI Press, pp. 124–71.
- Netting, R. (1976), 'What Alpine Peasants Have in Common: Observations on Communal Tenure in a Swiss Village', *Human Ecology*, 4: 135–46.
- Netting, R. (1981), *Balancing on an Alp*, Cambridge: Cambridge University Press.
- North, D.C. (1981), *Structure and Change in Economic History*, New York: W.W. Norton.
- Nugent, J. and N. Sanchez (1993), 'Tribes, Chiefs and Transhumance: A Comparative Institutional Analysis', *Economic Development and Cultural Change*, 42 (1): 87–113.
- Olson, M. (1965), *The Logic of Collective Action*, Cambridge, MA: Harvard University Press.
- Peters, P.E. (1994), *Dividing the Commons: Politics, Policy and Culture in Botswana*, Charlottesville and London: University Press of Virginia.
- Platteau, J.P. (2000), *Institutions, Social Norms and Economic Development*, London: Harwood Academic Publishers.

62 Culture and development

Roland Hoksbergen and Charles K. Wilber

Introduction

Are some cultures more prone to development than others? Does economic development require cultural change? Which is more important, economic development or cultural integrity? These are among the core questions that swirl around the debate on the relationship between culture and development. The debate was initiated in its modern form by Max Weber's classic work, *The Protestant Ethic and the Spirit of Capitalism* (1958), and it has continued ever since. Social scientists past and present have staked out positions on all sides of the issues. Traditional modernization theorists take it for granted that economic development is primary, but they differ on whether cultural change leads development or vice versa. Some see culture functioning as either an instrument of or an obstacle to development, while still others see culture as irrelevant. Post-development localization theorists, on the other hand, disparage modern economic growth and argue that culture is paramount. They are highly resistant to Western-style corporate-led globalization, which they believe destroys local culture. Inbetween are many intermediate positions, like those who hope to maintain cultural integrity while also allowing for the development of market economies, the expansion of trade and economic growth.

Historical background

To begin the exploration of the role of culture in development, it is worthwhile to review briefly the rise of the modern economy in the Western world. Two facts stand out from an examination of the history of modern capitalist development. First, capitalism has succeeded in producing quantities of goods and services unprecedented in history; second, it has done so in a temporally and spatially uneven manner. The economies of some nations take off into a self-sustaining growth, with other nations eventually catching up. Still others, the loosely called 'developing countries', seem to be left hopelessly behind. This pattern occurs across nations as well as across regions within nations.

One of the great economists of the twentieth century, Joseph Schumpeter, captures this dynamic process in his concept of 'creative destruction': 'The fundamental impulse that sets and keeps the capitalist engine in motion comes from the new consumers' goods, the new methods

of production or transportation, the new markets, the new forms of industrial organization that capitalist enterprise creates . . . [These developments] incessantly revolutionize the economic structure *from within*, incessantly destroying the old one, incessantly creating a new one' (Schumpeter, 1950, p. 83).

In both industrial and poor countries this creative-destructive process of development has created socio-political tensions both because of its uneven nature and because of its challenge to traditional values and ways of life. Today this is true particularly in countries with strong Islamic roots and in countries where readily identifiable minorities control the wealth of the society.

A simple policy of free markets and free elections may not advance the welfare of the poor in the world, and it must be remembered that the Western world did not develop that way itself. Development was slow, its major transformation spread over more than a century. For much of that period democracy was limited and countervailing policies were gradually enacted to curb the divisive effects of growing inequality.

Most historians, whether their orientation is political, cultural or economic, recognize that the eighteenth century was a turning point in the nature of the Western world that saw momentous movements and events – intellectual, political, military, social, cultural and economic. The Enlightenment, with its emphasis on reason, natural law and progress, and its *avant-garde* – the philosophes and physiocrats – opened new vistas, even though most of the population of Europe and the remainder of the world scarcely glimpsed those vistas.

In economic affairs, the eighteenth century began with Francois Quesnay's campaign against mercantilism and ended with the completion of the campaign by Adam Smith. In the process, the classical school of economics, a new social science, came into being. Finally, in the course of the century, the agricultural and commercial revolutions of the previous two centuries initiated the Industrial Revolution in England that formed the basis of our modern economies.

Now that self-regulated market capitalism had arrived, what were its characteristics? In its textbook purity, a capitalist market economy is controlled, regulated and directed by markets alone. Socially beneficial outcomes in the production and distribution of goods are entrusted to this self-regulating mechanism, based on the expectation that human beings behave so as to achieve money gains.

In contrast, during preceding historical periods (and even today in many of the poor countries), markets were never more than accessories of economic life. Instead, the economic system was embedded in the social-cultural system. In places like Babylonia and Greece the local markets

(trading centers) were compatible with the established social way of life; markets did not expand at the expense of the society. Even under the mercantile system of the previous two centuries, where markets had expanded to involve a large part of the nation, they were not free markets, for they were subjected to centralized administration. Karl Mannheim argued that the move to self-regulating markets entailed a transformation from a regulated and socially controlled mechanism into the very organizing principle of society itself (Mannheim, 1950, p. 191).

Moreover, such a market-first institutional pattern cannot function unless other aspects of a society's life are subordinated to its requirements, which is what happened over time in today's developed economies. A market economy can only exist in a market society, and the requisite process of social, cultural and institutional change evolved in conjunction with the transition to a market economy. Nations wanting to catch up economically are thus naturally led to the question of how such a market society can be created in countries far removed from Western culture and in a time span shorter than the century or two that was required in Europe. On the heels of this question, another follows close behind, which is whether the creation of such a market society should even be a goal of development. Needless to say, there is substantial disagreement over these questions.

Nailing down a working definition of culture is itself problematic, in part because culture is easier to identify in others than it is in oneself and one's own society, just as most people think others speak with accents, not themselves. That in itself makes talking about, and especially evaluating, culture potentially invidious, because in discussing cultures people are inevitably discussing the ways of life of others. Going way beyond people's culinary, musical preferences, dress and traditions, Rao and Walton understand culture to be:

about relationality – the relationships among individuals within groups, among groups, and between ideas and perspectives. Culture is concerned with identity, aspiration, symbolic exchange, coordination, and structures and practices that serve relational ends, such as ethnicity, ritual heritage, norms, meanings, and beliefs. It is not a set of primordial phenomena permanently embedded within national or religious or other groups, but rather a set of contested attributes, constantly in flux, both shaping and being shaped by social and economic aspects of human interaction. (Rao and Walton, 2004, p. 4)

Culture is thus about our deepest beliefs, values, sense of identity, ways of life and longings, which makes it unsurprising that the discussion of culture and its significance for development generates controversy.

For economists the interest in culture has centered on its support of traits that contribute to economic growth, that is, thrift, hard work and

reinvestment by the middle class; hard work, obedience and contentment for the working class. The key is to discover the historic role of culture in generating the capitalist spirit of entrepreneurship, which for Schumpeter is so basic. For example, what were the psychological conditions – the capitalist spirit – that accompanied and aided the development of a thriving capitalist economy?¹ Greed and the pursuit of riches are nothing new. Money-lending, commercial trading, piracy, plunder and other forms of unrestrained avarice are as old as history. But a way of life based on the rational, calculated pursuit of pecuniary profit through Smith's 'truck, barter and exchange' and its organization into an economic system using free wage labor is a modern phenomenon.

It was only after centuries of struggle that capitalism established its claim to legitimacy, for it involved a code of economic behavior and a system of human relations sharply at variance with traditional religious customs and values. Originality, self-confidence and tenacity of purpose were required to initiate and carry on this struggle. This was the role of entrepreneurs. They emerged partly because changing economic conditions helped the Reformation succeed and helped shape the development of new theologies and creeds. In turn, the emerging religious beliefs helped direct and shape the subsequent economic development. Economic reasons alone are insufficient to account for the extraordinary power of entrepreneurship and rational profit-seeking in the modern world.

Alternative views on the role of culture in development

How does this historical experience of the rise of capitalist economies in the now developed world impact upon our theory and practice of development today? Mainstream views of development today continue to follow in the modernization tradition that arose in the aftermath of World War II. On this view the primary measure of development continues to be a self-reinforcing tendency to economic growth and material progress. It is from this basic developmental reality that other positive features of a good society spring, like better health care, improved education and democratic governance. As Benjamin Friedman says at the close of his recent moral defense of economic growth: 'Only with sustained economic growth, and the sense of confident progress that follows from the advance of living standards for most of its citizens', can a nation hope to achieve 'an open, tolerant, and democratic society' (Friedman, 2005, p. 436). Over time it has become clear that an equitably growing economy requires financial capital, the adoption of efficient technologies, investment, human capital, entrepreneurship and market-promoting policies and institutions, which begs the question of how to get nations to build and use these necessary attributes. For many theorists who see a connection between culture and development,

inquiries into this question lead directly to issues of how cultural adaptation similar to that which occurred historically in the West can be facilitated so as to accommodate the needs of modern market economies throughout the world.

On the other hand, it needs to be said that many economists believe that culture is essentially irrelevant to economic development. Instead, the road to development is paved with market-promoting economic policies and institutions. Hernando De Soto, for example, argues that efforts 'to explain why capitalism fails outside the West remain mired in a mass of unexamined and largely untestable assumptions labeled "culture", whose main effect is to allow too many of those who live in the privileged enclaves of this world to enjoy feeling superior' (de Soto, 2000, p. 225). Along with economist Mancur Olson, De Soto argues that people of all cultures respond similarly when property rights and contract-enforcing institutions are established (Olson, 2000). In his high-profile treatment of twenty-first-century development challenges, Jeffrey Sachs calls the cultural thesis a myth, saying that cultures often follow rather than lead economic change and that culture-based arguments 'are usually made on the basis of prejudice rather than measurable evidence' (Sachs, 2005, p. 317). In his critique of Lawrence Harrison's emphasis on child-rearing, for example, Sachs points out that children are taught the value of hard work more consistently in Nigeria, South Africa and Tanzania than they are in the United States (Sachs, 2005, p. 318). For William Easterly too, the issue is not culture, but the proper structuring of incentives (Easterly, 2006). Summarizing these lines of thought, David Throsby points out that: 'mainstream texts in economic development have no time for culture; taking three such texts more or less at random, an inquisitive reader can find no reference to culture in the subject indexes of any of them' (Throsby, 2001, p. 67). For such theorists, it is no surprise that nations of greatly diverse cultures, like Ireland, South Korea, France, the United States, and now China and India can all enjoy the fruits of economic progress without fundamental changes to their cultures.

And yet there are perhaps a greater number of development theorists, like Weber, who find culture to be of central importance. For many of the early thinkers in this tradition, like economist Bert Hoselitz and sociologist Talcott Parsons, transitions from traditional to modern patterns of life required nothing short of major cultural overhaul.² Major efforts to discover potential sources for such cultural change led David McClelland to identify the significance of a people's psychological 'need for achievement', what he calls 'n-achievement'. McClelland argues that modern societies have been built by innovative entrepreneurial types with high n-achievement. N-achievement, however, is so deeply embedded in people's psyches, which develop gradually

during child-rearing and other social and cultural nurturing practices, that he despairs of offering practical advice for how actually to promote development. He notices, for example, that religious and ideological changes, like the rise of Protestantism in some Mexican communities, were associated with a rise in n-achievement. But he is doubtful this awareness of the relationship will itself lead to acceptable development policies. And though he believes education might be of some help, he doubts it can make a large contribution, because people's personalities are too deeply formed by prior and ongoing child-rearing practices. Ultimately, the best he can do is to encourage policies that facilitate the interaction among entrepreneurs from developed countries with scarce but vital entrepreneurial types engaged in business in 'underdeveloped countries' (McClelland, 1961).³

In a similar vein, Everett Hagen, recognizing both the importance of entrepreneurship and the role of psychological formation and traditional cultural patterns in creating people resistant to progressive change, identifies one major source of change to be a socially deviant group that finds a psychological outlet in violating traditional patterns of life and thus creates its own identity through entrepreneurial change and success in business. Subdominant or threatened minorities, or marginalized but progressive immigrant communities, might be able to provide an impetus to change that breaks through traditional cultural patterns and points the way to modernity (Hagen, 1962).

Still today, many contemporary theorists within the modernization tradition continue to focus on the fundamental importance of cultural change. Perhaps most well known among these is Lawrence Harrison, who after long experience in development work in Latin America has developed his version of the cultural thesis in a series of books written over a twenty year period starting in the mid-1980s.⁴ According to Harrison, the basic thesis is that 'values, beliefs, and attitudes are a key but neglected factor in understanding the evolution of societies and that the neglect of cultural factors may go a long way toward explaining the agonizingly slow progress toward democratic governance, social justice, and prosperity in so many countries' (Harrison, 2006, xiii). In Harrison's view, progress in such basic areas as life, health, liberty, prosperity, education and justice depends on the adoption of a democratic capitalist way of life, which in turn depends on cultural orientations.

Following many of his forbears who have developed lists of the contrasting cultural characteristics of traditional and modern societies, Harrison too identifies cultural traits that either inhibit or advance progress. Over the years, his list has expanded to 25 core cultural traits that make societies either progress-prone or progress-resistant (Harrison, 2006, pp. 36–7). These include religious orientations (for example favorable or

non-favorable attitudes toward material pursuits), values (for example how flexible a society's ethical code is), economic behavior (for example whether people have entrepreneurial inclinations) and social behavior (for example the radius of trust). Societies that value competition instead of fearing it as a threat to equality, for example, are more likely to progress, as are those that focus on success in this world over their place in the next. If progress is to come about in poor countries, then it is culture that must ultimately be changed.

Harrison is the name most often associated with the cultural thesis, but he is not alone. David Landes, in a sweeping study of development patterns across the world, both historical and contemporary, is drawn toward cultural explanations, ultimately assigning a major causal role to culture. Landes argues that: 'just because markets give signals does not mean that people will respond timely or well. Some people do this better than others, and culture can make all the difference' (Landes, 1999, p. 522).

Recent studies lend the cultural thesis some degree of empirical support. In a study of various factors that influence growth, David Weil finds that openness to new ideas, an inclination to work hard and to save, and the level of trust play a significant role in explaining economic growth (Weil, 2005, p. 427). Other studies analyze World Values Survey data with a special focus on the role of religion and find that religious faith and traditions definitely matter. Luigi Guiso et al. conclude that: 'on average . . . religion is good for the development of attitudes that are conducive to economic growth', and that 'on average, Christian religions are more positively associated with attitudes that are conducive to economic growth, while Islam is negatively associated', and that Protestants and Catholics have different mixes of positive and negative factors' (Guiso et al., 2003, p. 280). In neither the Weil nor the Guiso study is the direction of causality firmly established. In an attempt to address this shortcoming, Robert Barro and Rachel McCleary find that religious beliefs, especially as regards the existence of heaven and hell, seem to play a causal role in the achievement of higher economic growth. Like Weber's theories about the role that Calvinism played in Europe's drive toward capitalism and industrialization, Barro and McCleary conjecture that such 'religious beliefs stimulate growth because they help to sustain aspects of individual behavior that enhance productivity', like thrift and a greater work ethic.⁵

Social capital

Another line of thought that has arisen since the 1990s is focused on the causal role played by social capital and civil society. In the economics literature, social capital has come to mean that 'social relations' are important factors in the economy. Francis Fukuyama and Robert Putnam have each

studied the extent to which social relations promote group cooperation, civil society, good governance, trust and productive economic activity. Fukuyama emphasizes interpersonal trust as a key cultural aspect, arguing that: 'one of the most important lessons we can learn from an examination of economic life is that a nation's well-being, as well as its ability to compete, is conditioned by a single, pervasive cultural characteristic: the level of trust inherent in society' (Fukuyama, 1995, p. 7). In a similar vein, Putnam shows that the cultural predilection to work together cooperatively in civic groups is a major well-spring of democratic governance and economic well-being (Putnam, 1993). Fukuyama and Putnam both argue that cultures that foster trusting working relationships outside of narrow family interests, referred to by Fukuyama as 'weak ties' and by Putnam as 'bridging capital', are much more prone to establish successful democratic capitalist societies.

The recent attention to social capital, combined with the renewed focus on religious beliefs, has given rise to the concept of spiritual capital, which refers to the spiritual or religious resources that contribute to a well-functioning community. Putnam argues that religion is by far the largest generator of social capital in the United States, contributing to more than half of the social capital in the country. For developing countries, where religious commitments are generally stronger than they are in the economically developed countries, religious sources of social capital may be even more important. Theodore Malloch claims that: 'In the ultimate sense spiritual capital is the missing leg in the stool of economic development, which includes its better known relatives, social and human capital' (Malloch, 2003, p. 2).

Social capital studies have become common in development economics due, at least in part, to the World Bank which has been working on the concept since the 1990s. On one of its websites it says that: 'Social capital refers to the norms and networks that enable collective action. Increasing evidence shows that social cohesion – social capital – is critical for poverty alleviation and sustainable human and economic development.'⁶ Still, the extent to which social capital and spiritual capital are intellectually valid conceptual categories for economic development is carefully explored in a World Bank-supported volume edited by Anthony Bebbington et al. The presumption throughout the book is that the character of a society's social capital is a culturally defined reality that is central to the empowerment, participation and inclusion of people in the development process. The purpose of the book is to study efforts at the World Bank, populated as it is with technically trained economists, to figure out ways to include social capital in their analysis and in their programming (Bebbington et al., 2006).

In the hands of economists, social capital is typically integrated into neoclassical microeconomic theory, making it a behavioral property of individual actors. The impact of social capital is thus seen to flow from the rational choices of those individuals. The most common ways to model social capital are: (1) as a preference in a utility function; (2) as an individual resource owned by persons or firms; and (3) as an instrument to reduce risk.

Against the instrumentalist view of culture

The neoclassical tendency to consider social capital as a resource for development has generated some strong criticism. Using the term 'capital' is seen as misleading and even ideological, and the empirical support is also seen as weak and circular. As Van Staveren and Knorringa argue: 'one of the most central flaws . . . is a circular explanation of social capital: a group's success is attributed to its social capital, but social capital is measured by group success' (Van Staveren and Knorringa, 2007, p. 110). Moreover, if social capital is a resource for development, then social relations are stripped of their own value and become nothing more than instruments in the service of economic growth. One of the troublesome implications of such a view is that cultures can thus be judged worthy or unworthy, good or bad, based on their ability to generate economic growth. Not surprisingly, such invidious views are not readily accepted by people who have learned to value their own cultures for more than their pecuniary potential.

Amartya Sen, for example, argues that a focus on whether cultures are 'good or bad' fosters prejudicial attitudes, sometimes leading to a 'blame the victim' mentality that can cause great harm. English responses to Irish famines in the nineteenth century, for example, were thought by the British to be the result of Irish cultural deficiencies, which, if true, meant that direct assistance would only exacerbate the problem. What the Irish really needed were civilizing influences, a cultural makeover. Thousands died. By contrast, economic downturns in England were seen by the British as resulting from events beyond the control of the citizenry. Fast forward to the present day and one finds Catholic Ireland's economy growing much faster than Protestant England's. The same is true for India, which with its caste system and other-worldly religious views, was always assumed to have a culture resistant to development. Like China and Ireland, India is now among the fastest-growing countries in the world. Did their cultures change? Or was something else at work? (Sen, 2004, pp. 37–58).

Even some neoclassical economists have trouble with this tendency to instrumentalize culture. Oliver Williamson, responding to the definition of trust as rational expectations of the behavior of others, says 'calculativeness will devalue the [social] relations' because it 'may well be destructive of

atmosphere and lead to a net loss of satisfaction between the parties' (Williamson, 1993, p. 481).

Such disrespect for local culture is even more vehemently rejected by the post-development school, which, with leaders like anthropologist Arturo Escobar, is much to the political and philosophical left of Sen. Escobar, in a postmodern vein, argues that the discourse dominating development thought and practice today is of Western origin and thus embeds the superiority of its cultural orientations in everything it says and does.⁷ When the international development discourse emanates from rich-country governments, universities, non-governmental organizations (NGOs) and multilateral organizations like the World Bank, all dominated by the institutions of Western and/or Northern culture, is it any surprise that its democratic capitalist ways of life and thought are privileged over others? When international economic, political and social institutions are structured on the basis of powerful democratic capitalist countries, is it any wonder that other cultures fail to compete successfully? In almost diametrical opposition to cultural modernizers, who believe Western-style cultural reforms are the way to developmental salvation for the poor world, Escobar believes they will bring destruction, for in disrespecting and ultimately undercutting their own cultures, the requisite cultural changes will cause local cultures to lose their identities, meaning systems and control over their own lives, resources and communities. There is in post-development a strong environmental orientation as well, for often people are dispossessed of their property and their livelihoods to make way for dams, agro-export products and mass production techniques that destroy traditional ways of life. Modernizers like Harrison might think these to be progressive developments, but Escobar believes they lead to domination, dispossession, violence, cultural chaos and poverty.

Unlike modernization theorists like Harrison, whose prescription is to criticize local culture and to enact policies to bring people into the modern world, Escobar exalts local culture, insisting on local solutions, discovered and implemented through locally developed institutions, and understood in terms of local languages and ways of life. Escobar and the post-development movement are thus among the main theoretical supports to the contemporary localization movement, of which David Korten is one of the most prominent supporters, but which also includes popular authors like Wendell Berry and novelist Barbara Kingsolver (Korten, 2001; Kingsolver, 2003).

In addition to the modernizers who discount culture entirely or treat it like one of any number of instruments, and post-development scholars who have a tendency to sacralize local cultures, there are a good number of scholars who take culture seriously while not being so judgmental about it. In these perspectives, culture is seen as a vital and, sometimes, revered

aspect of a people's identity, and as representing the integral and holistic nature of a society that changes in response to internal and external pressures. Culture is not apart from or outside of economic life, but is instead integral to and interactive with it. Lourdes Arizpe, in reference to the UN Commission on Culture and Development, says: 'it is not culture that is embedded in development; it is development that is embedded in culture' (Arizpe, 2004). Gunnar Myrdal was among the early economists to consider culture in this way, emphasizing that the development of a society needed to be fundamentally based on the people's own choices, which would in turn be based on their own values (Myrdal, 1968).

Another leader of this third way was Denis Goulet, who advocated an understanding of development that respected local cultures while at the same time recognizing the need for cultures to change. Goulet thus tried to find a way out of what he termed 'The Cruel Choice', which forced cultures outside the Western mainstream to choose between keeping their local cultural traditions and staying poor, or opting to join modernizing trends and losing their identity and sense of meaning (Goulet, 1971, 1980). His solution was to work within cultures and to find the 'latent dynamisms' that allowed cultural groups to respond constructively to the challenges of modernization. His approach coincides with the views of anthropologist Mary Douglas, who thinks the question of which cultures are 'better' is misguided and dangerous (Douglas, 2004). The common practice of evaluating which religious traditions are more prone to progress, for example, is the wrong way to approach the cultural question. Instead, she promotes a theory that sees every culture as a mixture of four groups and tendencies: the hierarchical, the entrepreneurial, the dissenting and the apathetic. Each of these first three plays crucial roles in the maintenance, protection and growth of cultures in their ever-changing environments. The hierarchical types, often government and religious leaders, want to keep traditions as they are and thus ensure social stability. Entrepreneurs, often from the world of business, are change agents who test the limits of their cultures by trying out and promoting new ways of doing things. Dissenters are typically idealists and visionaries who provide checks and balances on both groups. The apathetic tendency arises out of practices and patterns of life that marginalize whole groups of people or isolate them from the circles of power and decision-making. As the size of the apathetic group grows, so too does cultural distress. In many societies where poverty is prevalent, a study of the interaction of these four groups will reveal much about who holds power, how it is used and how whole groups are consigned to poverty. The analysis is not so much about whether a culture is good or bad as a whole, but about the way power is held and used by the different groups within the culture.

Interaction of development and culture

The notion that economics and culture make up integral parts of a large whole has led David Throsby to hint at the possibility of a new development paradigm that manages to bring these 'two disparate fields closer together', for 'culture is in fact central to and inextricable from the development process, providing both the context within which economic progress occurs and the very object of development from the perspective of individual needs' (Throsby, 2001, pp. 164, 165). This is the project of the two World Bank-spawned volumes cited earlier, and it seems also to be the project of Amartya Sen, who has been a highly influential voice in the development of the Human Development Index that is reported in the United Nations' annual *Human Development Reports*.

As an economist, Amartya Sen is oriented toward matters of efficiency, growth and distribution, but, like Goulet, believes that any decisions to change culture must arise from the people themselves, in processes that ensure and facilitate their active and informed participation. Sen has been in the forefront of a new perspective on development known as the 'capabilities approach'. Based on the idea of individual freedom within cultural contexts, this perspective respects culture by insisting on the legitimacy and inclusion of every voice by allowing the people themselves to evaluate their own cultural institutions and ways of life.⁸ Understanding culture as varied and complex, Sen affirms that culture and economic development are tightly interwoven, arguing also that both cultures and economies evolve in mutually reinforcing patterns. Whether and how economies and cultures should change, however, can only legitimately be determined by the people themselves, which is why Sen is so insistent that the capabilities to make such decisions be widespread throughout society. The capabilities approach thus assesses more whether people have both the individual capabilities (for example, education) and the necessary participatory and inclusive social structures for making informed choices about their own development path, than it does their achievement of predetermined ends, like high gross domestic profit (GDP) growth rates. While respectful of culture, Sen also points out that culturally based arguments are often improperly employed by powerful leaders who invoke the culture argument to defend oppressive systems. He is not persuaded, for example, by the argument that Asian values are more oriented to authoritarian styles of governance than European or other sets of values. Instead, he finds it is the leaders and holders of authoritarian power rather than the masses who support so-called Asian values. He argues that the real purpose of the Asian values argument is not to support local culture, but to legitimize the leaders' hold on power.

For Sen, generalized poverty is often explained by the fact that the poor are disempowered and marginalized and lack key individual and social

capabilities. He is famous for noting the fact that famines do not occur in democratic countries, which is a consequence of the breadth and depth of participation in social, political and economic institutions. Where wealth is widespread, on the other hand, there will also tend to be a widespread diffusion of education, opportunity and voice, which combine to define capability.

Sen's influence at the UN is especially transparent in the 2004 *Human Development Report* entitled *Cultural Liberty in Today's Diverse World* (UNDP, 2004). It respects culture, recognizes that cultures can and should change in the development process and identifies widespread participation as the legitimizing principle for choosing. The report begins with the assertion that: 'cultural liberty is a vital part of human development because being able to choose one's identity – who one is – without losing the respect of others or being excluded from other choices is important to leading a full life'. In contrast to some of the studies cited above, the report argues that 'there is no evidence from statistical analysis or historical studies of a causal relationship between culture and economic progress or democracy', thus rejecting the Weber hypothesis. The report goes on to argue in favor of multicultural democracies that ensure the broadest possible participation in the social choices that affect people, identifying such choices as fundamental human rights. Furthermore, the report does not accept the reality of Goulet's 'cruel choice', for participation guarantees that there need be no trade-offs between cultural integrity and human rights, democracy and economic improvements.⁹

Yet many would argue that Sen and the Bank are too optimistic. Modernizing development strategies have always meant conflict with traditional cultural institutions. Goulet pointed out in one of his last works: 'Under the banner of development, powerful standardizing forces dilute cultures and relegate them to purely ornamental, vestigial or marginal positions in society' (Goulet, 2005, p. 23). He highlights three of these standardizing forces: (1) technology, especially media technology, which spreads the values of individualism, instant gratification and consumerism; (2) the modern state which centralizes everything, including ideas and values; and (3) the managerial ethos which spreads cost-benefit analysis to every sphere of activity as the best way to make decisions.

Moreover, the pervasiveness and the creative destructiveness of these standardizing forces have fomented cultural resistance in many parts of the world, probably best seen in Muslim countries. The present style of development and globalization threatens to generate a whirlwind of cultural and political backlash.

Culture can both hinder and/or aid development, and development can in turn both harm and/or help culture. What drives the change, and whether

development institutions like the World Bank, national aid agencies and NGOs ought to promote such change, are fundamental and controversial questions. Regardless, cultures must and will change. How they will change becomes the issue. Will the changes be destructive, resulting in people losing what gives them meaning in life? Or will people become part of a change process that allows them to adapt gradually and intentionally with minimal social and cultural cost, and without violent and regressive backlashes? Over 30 years ago Jim Lamb pointed out the road that needed to be traveled if that cultural and political backlash is to be avoided: ‘Development should be a *struggle* to create criteria, goals, and means for self-liberation from misery, inequity, and dependency in all forms. Crucially, it should be the process a people choose, which heals them from historical trauma, and enables them to achieve a newness on their own terms’ (Lamb, 1973, p. 20).

Notes

1. The basic sources for this section are: Tawney (1926) and Weber (1958).
2. See Peet and Hartwick (1999), 71–9.
3. See especially the final chapter.
4. See especially *Underdevelopment Is a State of Mind: The Latin American Case* (1985) and *The Central Liberal Truth: How Politics Can Change a Culture and Save it From Itself* (2006). See also *Who Prospers: How Cultural Values Shape Economic and Political Success* (1992), *The Pan-American Dream: Do Latin America’s Cultural Values Discourage True Partnership With the United States and Canada* (1998) and, edited with Samuel P. Huntington, *Culture Matters: How Values Shape Human Progress* (2001).
5. See Barro and McCleary (2003, p. 37). The paper can also be found at http://www.economics.harvard.edu/faculty/barro/papers/Religion_and_Economic_Growth.pdf. Interestingly, Barro and McCleary also find that though certain religious beliefs have a causal impact on economic development, economic development itself also generates a decline in overall religiosity, thus supporting to some degree the secularization thesis.
6. See <http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/EXTSOCIALDEVELOPMENT/EXTSOCIALCAPITAL/0,,contentMDK:20642703~menuPK:401023~pagePK:148956~piPK:216618~theSitePK:401015,00.html>.
7. Escobar (1994). See also Escobar et al. (2002) and Escobar (2004). An overview of the postdevelopment school is found in Peet and Hartwick (1999, 123–62).
8. See Sen (1999) for his most complete explanation of the theory. Another main voice in developing the capabilities approach is philosopher Martha Nussbaum. See Nussbaum (2000) for a discussion of how she integrates the legitimacy of local cultures with an attempt to discover universal values.
9. *Human Development Report 2004*, see the ‘Overview’, pp. 1–12 in UNDP (2004).

References

- Arizpe, Lourdes (2004), ‘The Intellectual History of Culture and Development Institutions’, in Vijayendra Rao and Michael Walton (eds), *Culture and Public Action*, Stanford, CA: Stanford University Press, pp. 163–84.
- Barro, Robert J. and Rachel M. McCleary (2003), ‘Religion and Economic Growth’, NBER Working Paper No. 9682, May.
- Bebbington, Anthony J., Michael Woolcock, Scott Guggenheim and Elizabeth A. Olson (eds) (2006), *The Search for Empowerment: Social Capital as Idea and Practice at the World Bank*, Bloomfield, CT: Kumarian Press.

- de Soto, Hernando (2000), *The Mystery of Capital: Why Capitalism Triumphs in the West and Fails Everywhere Else*, New York: Basic Books.
- Douglas, Mary (2004), 'Traditional Culture: Let's Hear No More about It', in Vijayendra Rao and Michael Walton (eds), *Culture and Public Action*, Stanford, CA: Stanford University Press, pp. 85–109.
- Easterly, William (2006), *The White Man's Burden: Why the West's Efforts to Aid the Rest have Done So Much Ill and So Little Good*, New York: Penguin Press.
- Escobar, Arturo (1994), *Encountering Development*, Princeton, NJ: Princeton University Press.
- Escobar, Arturo (2004), 'Development, Violence and the New Imperial Order', *Development*, 47 (March): 15–21.
- Escobar, Arturo, Dianne Rocheleau and Smitu Kothari (2002), 'Environmental Social Movements and the Politics of Place', *Development*, 45 (March 1): 28–36.
- Friedman, Benjamin M. (2005), *The Moral Consequences of Economic Growth*, New York: Vintage Books.
- Fukuyama, Francis (1995), *Trust: The Social Virtues and the Creation of Prosperity*, New York: Free Press.
- Goulet, Denis (1971), *The Cruel Choice: A New Concept in the Theory of Development*, New York: Atheneum.
- Goulet, Denis (1980), 'Development Experts: The One-Eyed Giants', *World Development*, 8 (7/8): 481–9.
- Goulet, Denis (2005), 'On Culture, Religion, and Development', in Marguerite Mendell (ed.), *Reclaiming Democracy*, Montreal, Quebec and Kingston, Ontario: McGill-Queen's University Press, pp. 21–32.
- Guiso, Luigi, Paola Sapienza and Luigi Zingales (2003), 'People's Opium? Religion and Economic Attitudes', *Journal of Monetary Economics*, 50: 280.
- Hagen, Everett E. (1962), *On the Theory of Social Change: How Economic Growth Begins*, Homewood, IL: Dorsey Press.
- Harrison, Lawrence (1985), *Underdevelopment Is a State of Mind: The Latin American Case*, Cambridge, MA: Harvard Center for International Affairs.
- Harrison, Lawrence (1992), *Who Prospers: How Cultural Values Shape Economic and Political Success*, New York: Basic Books.
- Harrison, Lawrence (1998), *The Pan-American Dream: Do Latin America's Cultural Values Discourage True Partnership With the United States and Canada*, Boulder, CO: Westview Press.
- Harrison, Lawrence (2006), *The Central Liberal Truth: How Politics Can Change a Culture and Save it From Itself*, New York: Oxford University Press.
- Harrison, Lawrence and Samuel P. Huntington (eds) (2001), *Culture Matters: How Values Shape Human Progress*, New York: Basic Books.
- Kingsolver, Barbara (2003), *Small Wonder: Essays*, New York: Harper Perennial.
- Korten, David C. (2001), *When Corporations Rule the World*, 2nd edn, San Francisco, CA: Berrett-Koehler Publishers.
- Lamb, James J. (1973), 'The Third World and the Development Debate', *IDOC North America*, January–February: 18–26.
- Landes, David S. (1999), *The Wealth and Poverty of Nations: Why Some Are So Rich and Some So Poor*, New York: W.W. Norton & Co.
- Malloch, Theodore Roosevelt (2003), 'Social, Human and Spiritual Capital in Economic Development', Spiritual Capital Research Program, http://www.metanexus.net/spiritual_capital/pdf/malloch.pdf
- Mannheim, Karl (1950), *Freedom, Power, and Democratic Planning*, New York: Oxford University Press.
- McClelland, David (1961), *The Achieving Society*, New York: D. Van Nostrand Company.
- Myrdal, Gunnar (1968), *Asian Drama: An Inquiry Into the Poverty of Nations*, 3 vols, New York: Random House.
- Nussbaum, Martha (2000), *Women and Human Development: The Capabilities Approach*, Cambridge: Cambridge University Press.

- Olson, Mancur (2000), *Power and Prosperity: Outgrowing Communist and Capitalist Dictatorships*, New York: Basic Books.
- Peet, Richard with Elaine Hartwick (1999), *Theories of Development*, New York: Guilford Press.
- Putnam, Robert (1993), *Making Democracy Work: Civic Traditions in Modern Italy*, Princeton, NJ: Princeton University Press.
- Rao, Vijayendra and Michael Walton (2004), in Vijayendra Rao and Michael Walton, (eds), *Culture and Public Action*, Stanford, CA: Stanford University Press, pp 3–36.
- Sachs, Jeffrey (2005), *The End of Poverty: Economic Possibilities for Our Time*, New York: Penguin Press.
- Schumpeter, Joseph A. (1950), *Capitalism, Socialism, and Democracy*, 3rd edn, New York: Harper & Brothers.
- Sen, Amartya (1999), *Development as Freedom*, New York: Knopf.
- Sen, Amartya (2004), ‘How Does Culture Work?’ in Vijayendra Rao and Michael Walton (eds), *Culture and Public Action*, Stanford, CA: Stanford University Press, pp. 37–58.
- Tawney, R.H. (1926), *Religion and the Rise of Capitalism*, New York: Harcourt, Brace & World.
- Throsby, David (2001), *Economics and Culture*, Cambridge: Cambridge University Press.
- UNDP (2004), *Human Development Report 2004: Cultural Liberty in Today's Diverse World*, New York: UNDP.
- Van Staveren, Irene and Peter Knorringa (2007), ‘Unpacking Social Capital in Economic Development: How Social Relations Matter’, *Review of Social Economy*, 65 (March): 107–35.
- Weber, Max (1958), *The Protestant Ethic and the Spirit of Capitalism*, New York: Charles Scribner.
- Weil, David N. (2005), *Economic Growth*, Boston, MA: Addison-Wesley.
- Williamson, Oliver (1993), ‘Calculativeness, Trust, and Economic Organization’, *Journal of Law and Economics*, 36: 453–86.

63 The economics of war: causes and consequences

Frances Stewart and Graham Brown

Introduction

The incidence of violent conflict among poor countries is high: seven out of ten of the poorest countries have recently experienced some sort of civil war. Since conflict has a serious adverse impact on development, as well as causing massive human suffering, efforts to promote development and reduce poverty must include policies to prevent conflict and to protect populations during conflict. Moreover, as poorer economies are more likely to experience conflict and conflict inhibits development, a vicious cycle can ensue – of underdevelopment–war–underdevelopment – which it is essential to break if either peace or development is to be sustained, yet to do so is very difficult. This review covers both sides of this cycle: the following section analyses economic causes of contemporary conflicts; the subsequent section explores economic and social consequences.

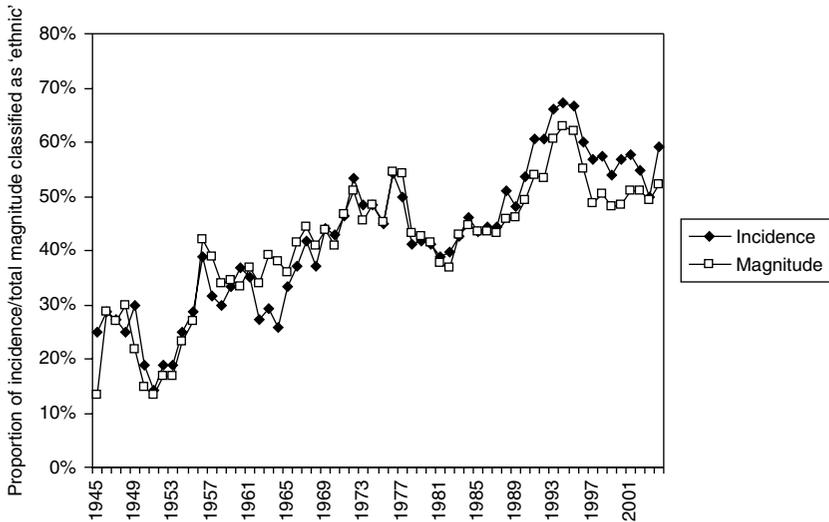
Economic explanations of violent conflict

While some attribute contemporary conflicts to fundamental differences arising from ethnicity or religion (for example Huntington, 1993), such differences are evidently insufficient as an explanation since many multi-ethnic or multi-religious societies live peacefully while others are at peace for decades before experiencing conflict. In fact, the vast majority of multi-ethnic societies are at peace (Fearon and Laitin, 1996). Therefore, we need to look beyond ethnicity to issues of power and economics to understand conflict (Cohen, 1974). Below we consider four explanations that have dominated recent economic analysis of conflict: group motivation and inequalities; private motivations; a failed ‘social contract’; and environmental pressures (‘greenwar’).

Group motivation

Political conflicts consist in fighting between groups – groups that wish to gain independence or take over the state, and others that resist this (Horowitz, 1985). Such groups bring individuals together with a common purpose. While individual motivation is also important, this perspective argues that group motivation and mobilization underlie most political conflicts.

Groups engaged in internal conflict are often united by a common ethnic



Note: The data from which these figures were calculated – the list of ‘major episodes of political violence’ compiled by Monty G. Marshall of the Centre for Systemic Peace – provide a ‘magnitude’ score for each episode, ranging from 1 (mildest) to 10 (severest). Total magnitude here is calculated simply as incidence weighted by magnitude.

Source: Calculated from Marshall (2005).

Figure 63.1 Ethnic violence as a proportion of ‘major political violence’, 1946–2004

or religious identity. Since 1945, the proportion of conflicts attributable to ethnic violence has been steadily increasing (Figure 63.1). While such conflicts are generally presented in religious or ethnic terms, and such identities provide a powerful source of mobilization and unity, underlying differences in access to economic or political resources are generally also present, providing both leaders and followers with a strong motive to fight. Gurr (1970, 1993) terms such group differences ‘relative deprivation’ and Stewart (2000, 2008) defines differences in groups’ access to economic, social and political resources as ‘horizontal inequalities’. Horizontal inequalities consist of inequalities in access to resources between groups differentiated by racial, ethnic, linguistic or religious characteristics, in contrast to vertical inequality which measures inequality among individuals or households. The horizontal inequalities explanation of conflict is based on the view that when such cultural differences coincide with economic and political differences between groups, this can cause deep group resentments that may lead to violent struggles.

Empirical evidence is accumulating that horizontal inequalities constitute a significant cause of violent conflict. Cross-sectional quantitative analyses have shown a significant relationship between various dimensions of socio-economic inequality and conflict. Mancini (2008) shows that differences in infant mortality rates – a broad proxy for levels of socio-economic deprivation – between ethnic or religious groups among districts in Indonesia help explain the location of the communal conflicts that occurred after 1998. A similar relationship – between ‘spatial’ horizontal inequalities and the intensity of insurgency – has been found in Nepal (Gates and Murshed, 2005). Although multi-country studies have been hampered by poor data, supporting evidence has been found by both Østby (2004) and Barrows (1976). There is also substantial case study evidence: Stewart’s (2002) review of the experiences of nine countries shows not only that severe socio-economic horizontal inequalities preceded the emergence of violent conflict, but that reductions in socio-economic horizontal inequalities – such as occurred in Northern Ireland during the 1980s – may contribute to the conditions for a peaceful resolution of such conflicts. However, some societies show severe horizontal inequalities without experiencing conflict. Political inclusiveness is one reason that some societies avoid conflict despite severe economic horizontal inequalities; other reasons are lack of unity among the deprived groups, and state repression.

Political horizontal inequalities – the exclusion or under-representation of groups within the political structure of a state – can provoke violent conflict, especially when they change abruptly. In Côte d’Ivoire, three decades of post-independence rule by Félix Houphouët-Boigny avoided significant conflict, largely due to the policy of balancing representatives of the major groups in positions of importance in the government and bureaucracy. Following Houphouët-Boigny’s death and the introduction of multi-party elections in the early 1990s, political leaders sought to mobilize ethnic sentiments to enforce their grip on power and thus undermined Houphouët-Boigny’s careful balancing act, leading to a spiral of ethnicization, xenophobia and, ultimately, civil war (Langer, 2005).

It is important to note that relatively rich groups may instigate conflict, as well as the relatively poor. The relatively rich do so to preserve their riches (and/or power), while the relatively poor do so out of a sense of injustice with the intention of achieving some redistribution.

Private motivation

People who fight are, of course, individuals with their own private motivation as well as being members of a group. War confers benefits as well as costs on some individuals. Political sociologists (Keen, 1998; Duffield, 1994), and economists (for example Collier and Hoeffler, 2001), have

emphasized private or individual motivation as the fundamental cause of conflict, arguing that the net economic advantages to individuals motivate them to fight. In this approach, which has its basis in 'rational choice' assumptions, group identities are not an independent factor but are instruments, created to help fulfil the private motives of those who fight, especially leaders (Hirshleifer, 1994).

Keen lists many ways in which war confers individual benefit on particular categories of people: it permits people, especially uneducated young men, to gain employment as soldiers; it offers opportunities to loot, to profiteer from shortages and from aid, to trade arms and to carry out illicit production and trade. Where alternative opportunities are few, and the possibilities of enrichment by war are considerable, wars are likely to be more numerous and longer. Conflicts may persist because some powerful actors benefit through the manipulation of scarcity, smuggling, and so forth and have no interest in resolving the conflict. An oft-cited case used to support this view is the role of 'conflict diamonds' in the prolongation of the civil war in Sierra Leone (Collier, 2000, p. 5).

However, case studies suggest that even where natural resources are abundant, private maximizing motives are rarely the full explanation. A study of seven countries in conflict concluded:

very few contemporary conflicts can be adequately captured as pure instances of 'resource wars' . . . Economic incentives have not been the only or even the primary causes of these conflicts. (Ballentine and Sherman, 2003, pp. 259–60)

In most cases of conflict, the risk of death or debilitating injury are high, so that 'rational' actors might be likely to choose another option before engaging in rebellion. This argument may not apply to leaders, who are less often killed or injured, while followers may be coerced into fighting, or persuaded to fight by leaders playing up religious or ethnic differences and grievances:

Grievance is to a rebel organization what image is to a business . . . [A] sense of grievance is deliberately generated by rebel organizations . . . [rebel supporters] are gulled into believing the discourse which self-interested rebel leaders promote. (Collier, 2000, p. 5)

At this point the group explanation and individual explanation of conflict come together. Grievances are hard to sell to the extent of people risking their lives if they are not genuine (that is, unless there is some sort of exclusion or economic horizontal inequalities), while leaders are motivated by political exclusion (that is, political horizontal inequalities) which denies them access to resources and power.

While generally not a sufficient explanation of conflict, it is clear that expected rewards often play a role in the decision to rebel. Econometric evidence confirms that conflict incidence is higher in resource-rich areas (Humphreys and Varshney, 2004). The gains (and motivation) in resource-rich areas may be individual or group, or both. As Collier notes, citing the cases of Aceh (Indonesia), Biafra (Nigeria) and Katanga (Zaire), separatist rebellion often emerges in resource-rich areas of a country (Collier, 2000, p. 10). Yet all these conflicts were framed in ethnic terms. Moreover, in many cases the leaders of the rebellions left lucrative and safe positions to instigate rebellion. For example, Hassan di Tiro left a secure position at the United Nations to instigate the Acehnese uprising. In the case of Colombia, often depicted as a 'greed'-motivated conflict, interviews with both leaders and those mobilized to fight show that generally their economic position worsened as a result of participating in the conflict – most put forward ideological reasons for their actions, especially the issue of land reform (Gutierrez Sanin, 2004).

There are also examples of separatist movements in regions with poor resource endowment such as Eritrea, Bangladesh (then East Pakistan) and the Tamil rebellion in Sri Lanka. Moreover, it is not possible to create an identity out of nothing (Smith, 1991). A common history, language, culture or religion is generally required to generate felt identities powerful enough to mobilize people for conflict.

Failure of the social contract

A third theory of violent conflict derives from the view that social stability is implicitly premised on a social contract between the people and the government. According to this hypothetical contract, people accept state authority so long as the state delivers services and provides reasonable economic conditions in terms of employment and incomes. With economic stagnation, or decline, and worsening state services, the social contract breaks down and violence results. Hence high (and rising) levels of poverty and a decline in state services would be expected to cause conflict (Nafziger and Auvinen, 2000). High vertical inequality might also be associated with such a failure, unless accompanied by populist measures to compensate the deprived. Conversely, political institutions that are able to channel and respond to socio-economic discontents strengthen the social contract, thus reducing the risk of conflict.

Considerable evidence from econometric studies shows that conflict incidence is higher among countries with lower per capita incomes, life expectancy and economic growth (Elbadawi and Sambanis, 2000; Nafziger and Auvinen, 2000; Collier and Hoeffler, 2001). Many analyses have found an inverted U-curve relationship between the extent of democratization in

a country and the risk of conflict (for example Ellingsen, 2000), with the usual interpretation being that 'stable' democracies are able to avert violent conflict through a strong social contract, while strongly authoritarian regimes are able to suppress conflict. However, Reynal-Querol (2002) has argued that it is the particular type of democracy – whether majoritarian, presidential or proportional representation – that affects propensity to conflict, rather than the level of 'democracy' per se.

'Greenwar' and environmental scarcity

The fourth explanation of violent conflict, associated with the work of Homer-Dixon and the 'Toronto Group' (for example Homer-Dixon, 1994; Percival and Homer-Dixon, 1998), is the 'greenwar' or 'environmental scarcity' argument. The essence of this perspective is that contest for control over declining natural resources, often intensified by population pressures, is a major cause of violent conflict around the world. Poorer societies are more at risk because they will be 'less able to buffer themselves' from environmental pressures (Homer-Dixon, 1994, p. 6). Three dimensions of environmental scarcity are identified which may lead to conflict: 'supply-induced scarcity', linked to the 'depletion and degradation of an environmental resource'; 'demand-induced scarcity', linked to population growth and the consequent extra pressures on existing resources; and 'structural scarcity', which 'arises from an unequal distribution of a resource that concentrates it in the hands of a relatively few people' (Percival and Homer-Dixon, 1998, p. 280). Homer-Dixon thus predicts 'an upsurge of violence in the coming decades that will be induced or aggravated by scarcity' (Homer-Dixon, 1994, p. 6).

However, while it is clear that pressures arising from environmental scarcity may play an important role in many conflicts, the environmental scarcity hypothesis is – and really does not claim to be more than – a partial theory that contributes towards our understanding of a set of conflicts, but not the general conditions under which conflict is more likely to arise.

The environmental scarcity hypothesis overlaps substantially with the other hypotheses discussed here. It overlaps with the social contract hypothesis in viewing poverty as the root cause of conflict, although it points to specific environmental causes of such poverty. It also often overlaps with the group motivation approach, as environmental pressures usually lead to conflict where there are 'groups with strong collective identities that can coherently challenge state authority' (Percival and Homer-Dixon, 1998, p. 280). Indeed, the 'structural scarcity' dimension of the greenwar approach is very similar to the group motivation hypothesis, albeit restricted to a particular dimension of inequality.

The environment scarcity view has been criticized by Fairhead (2000) who argues that it is environmental riches, not scarcity, that is associated with conflict because people fight to control lucrative natural resources, as in the Congo. This view fits well into the private motivation or greed hypothesis. In fact both environmental poverty and environmental riches may cause conflict, for different reasons and in different circumstances.

The theories outlined above appear, in their extreme formulations, to be diametrically opposed – as manifest in the ‘greed versus grievance’ debate (Ballantyne and Sherman, 2003). But, as we have seen, proponents of one perspective usually accept in part the insights of other perspectives. Some conflicts fit neatly into one of the explanations, some into others, and some clearly have multiple causes. One rather simple conclusion, therefore – that qualitative analysts of conflict are mostly aware of, but that quantitative analysts tend to overlook – is that each of the broad causal theories discussed above involves a degree of oversimplification and generalization. The causes and dynamics of any single conflict are typically complex, sometimes contradictory, and involve aspects of many, if not all, of the perspectives discussed above.

Yet it is important to understand which explanation dominates in a particular case, since this has important implications for appropriate policy prescriptions for the prevention and resolution of the conflict. There is not space here to discuss policies in detail. In brief, where group motivation is a fundamental cause, policies need to be inclusive and correct large horizontal inequalities (Stewart, 2008). To the extent that private motivation is key, policies need to reduce the profitability of illicit war-sanctioned activities (like drug production and smuggling), and to offer combatants income-earning opportunities as an alternative to fighting. If a failed social contract is the fundamental cause, then the aim should be to improve the functioning of government in relation to security, the economy and the provision of social services. Greenwar conflicts require an attack on the fundamental causes of environmental pressure, both from demand and supply perspectives. Such policies should be applied to all societies at risk of war, which includes all low-income countries, any country which has experienced conflict in the past few decades, and any economy suffering severe horizontal inequalities.

Two points are worth making about this policy set. Firstly, none of these policies are a central part of the current development agenda of the international financial institutions; secondly, they are all desirable in themselves, quite apart from their impact on conflict prevention.

The economic consequences of war

War in general, and civil war in particular, is one of the main causes of human suffering and economic underdevelopment. Most of the human and

economic costs of war do not result directly from battle deaths and injuries, but indirectly from the loss of livelihoods caused by the dislocation of economy and society. An important implication of the substantial indirect costs is that policy might be able to reduce costs substantially if appropriately designed, even while conflict is ongoing.

The economic effects of war are the result of a complex interaction between the particular processes of war and the economy in which it takes place. Figure 63.2 provides an overview of the main relationships likely to be affected, with arrows indicating the direction of causality and expected direction of impact. It is helpful to distinguish between the direct economic consequences of the conflict, and the compensating behaviour of economic agents in their attempt to moderate or offset the negative impacts of war.

Direct effects include: output loss as people move from their place of work because they join the fighting, are killed or flee; the destruction of capital and consequent loss of output; disruption of transport links due to physical destruction; a loss of trust among economic agents, reducing market transactions; disruption of international markets due to frontier closure or embargoes; reduced foreign investment and the diversion of foreign exchange from economic and social needs to military uses.

These effects will tend to reduce aggregate levels of output. Labour markets will be disrupted as many of the unskilled of prime working age become military recruits and some suffer violent deaths, while much skilled labour is likely to leave the country. Reduced agricultural output, and disrupted internal and international markets, are likely to affect exports particularly heavily. Reduced foreign exchange availability for productive inputs results, leading to a shortage of imported inputs and to a further fall in output and exports.

Compensating behaviours which can moderate the negative impact of the effects of war include: increased capacity utilization and import substitution; the provision of international credit or aid, offsetting loss of foreign exchange; the emergence of new forms of social capital compensating for loss of trust in formal institutions – such as enhanced cooperation and trust among members of a group on the same side of the war; and rapid government or community action to reconstruct facilities destroyed by belligerents.

To understand the total impact of these mechanisms on individuals one needs to go beyond money incomes to explore how they affect different types of individual entitlements, including market entitlements (acquired largely through employment and self-employment), direct entitlements (subsistence production), social entitlements (provided by the state), civic entitlements (provided by the community and NGOs) and extra-legal entitlements (see Stewart et al., 2001a, Chapter 1).

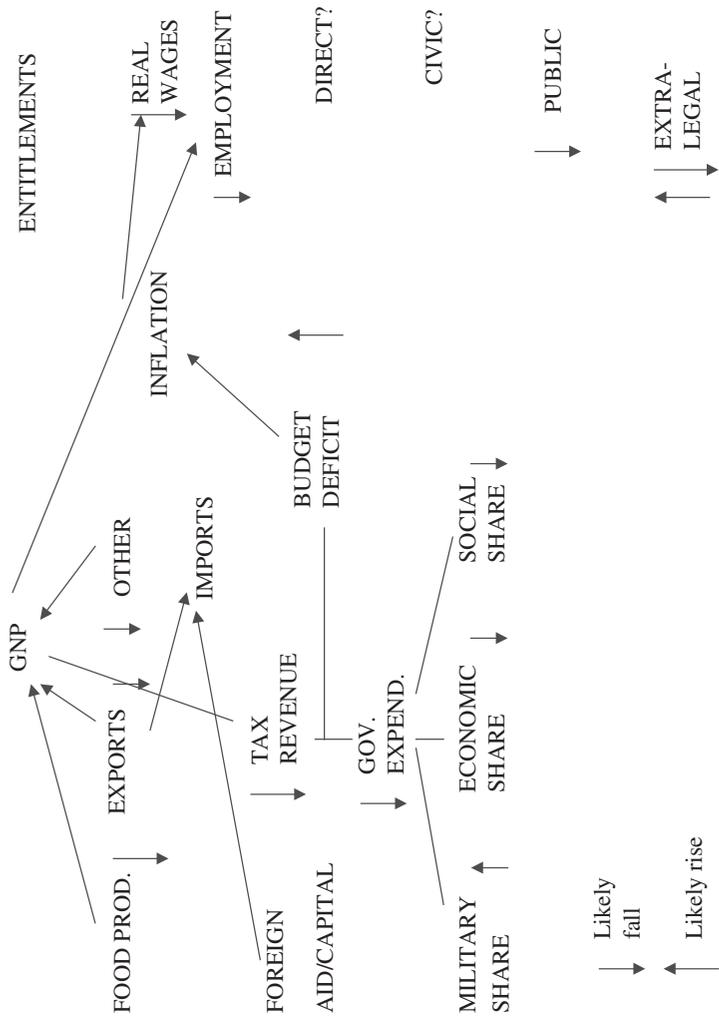


Figure 63.2 Predicting changes in the economy and entitlements during conflict

There are serious methodological problems in estimating the costs of war (Stewart, FitzGerald and Associates, 2001a, Vol. 1, Chapter 1). Below we summarize results adopting a variety of methods.

Macroeconomic consequences

All studies find negative effects on gross domestic product (GDP) growth. One study found a negative impact on GDP per capita growth in 13 out of 14 countries suffering the worst conflicts between 1975 and 1995, with considerable variability in magnitude. The worst losses occurred in long and pervasive wars (Stewart et al., 2001a). Regression analysis of 92 countries, 1960–89, showed an annual loss of 2.2 per cent during the war and in the immediately following years, compared with a no-war situation (Collier, 1999). Other cross-country regression analysis for 1960–99 came to similar conclusions, with an average loss of growth of 2.4 per cent per annum (Hoeffler and Reynal-Querol, 2003), although Imai and Weinstein (2000) suggest somewhat lower costs. The wide range of estimates indicates how dependent they are on the methodology adopted. Nonetheless, all give negative results. Evidence shows that wars with more widespread geographic coverage have a more negative impact (Stewart, FitzGerald and Associates, 2001; Imai and Weinstein, 2000).

Reduced economic growth is the result of capital destruction, lower investment and disrupted markets. All types of capital stock are eroded or destroyed in war. Physical facilities suffer direct attack – roads, ports and energy plants are often targeted. In Mozambique, Brück estimates that there was a two-thirds reduction in operational dams and plant nurseries, with 40 per cent of rural facilities destroyed or eroded. Social infrastructure is also commonly destroyed – again in Mozambique, almost 60 per cent of primary schools were closed or destroyed (Brück, 2001, pp. 64–7). Human capital is killed, or flees. About half the doctors and 80 per cent of the pharmacists left Uganda in the late 1970s (Dodge and Wiebe, 1985). The spread of AIDs that results from the sexual activities of combatants further reduces human capital. Institutions are destroyed – in Uganda, the agricultural extension system virtually disappeared (Matovu and Stewart, 2001). Social capital is weakened, with a severe loss in trust, particularly across groups. Yet new forms of institution and social capital emerge, for example informal banking systems.

Rates of investment and savings fall due to increased uncertainty. Evidence shows falling domestic saving rates (Stewart et al., 2001a; Brück, 2001), while capital flight accelerates. The proportion of private wealth held abroad rose from 9 per cent to 20 per cent in the course of civil wars according to Collier et al. (2004). Voluntary private lending from abroad tends to fall with increased uncertainty, but changes in official foreign lending

depend on political factors. 'Forced' foreign savings may occur, as countries renege on debt servicing obligations. In fact, aggregate foreign savings seem to hold up more than might be expected, evidenced by the huge accumulation of foreign debt during conflict (Stewart et al., 2001c).

Private (domestic and foreign) investment is adversely affected by uncertainty, rising costs of transport and difficulties in securing finance. Foreign investors are likely to be concerned about the safety of their personnel and equipment, and the increased foreign exchange risk. Country studies show a fall in foreign direct investment as expected – indeed this was the most important macroeconomic cost of Sri Lanka's conflict (O'Sullivan, 2001). Imai and Weinstein (2000) show a strong negative impact of civil war on private investment. Government investment is likely to be negatively affected by reduced revenue, and diversion of expenditure to military uses – Ra and Singh (2005) estimate that development expenditure in Nepal fell by one-third, 2001–04. Country studies show that aggregate investment on balance does not fall as much as domestic savings, possibly due to buoyant small-scale investment as the informal sector expands.

Exports are also negatively affected, as a result of the general fall in production, a shift towards domestic markets and disruptions in international markets. In most cases, imports hold up much better than exports, financed by foreign debt. The share of imports going to military items and food imports rises, however, leaving a much smaller share for inputs into the productive sector.

Government revenue is likely to fall absolutely and as a proportion of gross national product (GNP) as the government finds it more difficult to collect taxes and major sources of revenue (for example, from export taxes) fall away. There are sharp divergences across countries. In Uganda (1979–80), revenue as a share of GDP fell dramatically, but in both Mozambique and Nicaragua in the 1980s, the revenue ratio rose. Budget deficits increase as government expenditure rises faster than revenue (Stewart et al., 2001c).

Inflation is expected to accelerate, as governments resort to deficit financing to finance the conflict and other essential services and public confidence in the currency declines. In recent wars, there seems to have been only minor acceleration in price inflation (Stewart et al., 2001c).

Meso-economic consequences

In general, there is a shift from tradeable to non-tradeable sectors, as a consequence of market disruptions, including undermining of formal organizations such as banks, reduced trust and failures of the transport system. One aspect of this is a switch to subsistence and informal activities, including simple production (even arms) and trading (particularly smuggling).

For example, Mozambique experienced rapid growth in the urban informal sector; while the ability to shift from producing marketed crops to subsistence agriculture also helps protect food production and nutritional standards – it did so, for example, during the Amin era in Uganda.¹

The share of government expenditure going to military items invariably increases, making it difficult to sustain social and economic expenditure. On average, it is estimated that military expenditure rises from 2.8 per cent of GDP (average for developing countries in peacetime, 1995) to 5.0 per cent of GDP (Collier et al., 2003). One consequence is likely to be a fall in the share of social expenditure. In Uganda, social expenditure fell by roughly 9 per cent annually during conflict. In exceptional cases such as Nicaragua in the 1980s, countries give increased priority to social expenditure and succeed in increasing public entitlements. Where government revenue collapses there may be a dramatic decline in public entitlements: this was true of Uganda in the late 1970s and mid-1980s and Afghanistan in the early 1990s.

Civic entitlements can substitute for reduced government provision. There was extensive foreign non-governmental organization (NGO) provision of social and economic services, for example, in Afghanistan in the early 1990s (Marsden and Samman, 2001). In Sri Lanka, local NGOs and communities and the Tamil rebel forces provided significant support (O'Sullivan, 2001). But in the worst-affected areas, communities disintegrate as people flee, and NGOs are able to do little – examples are Uganda, Southern Sudan and Sierra Leone.

Human costs

In addition to deaths and injuries, flight and ensuing psychological trauma, human costs result from the changing economy, with worsening entitlements of most types:

- Market entitlements decline as household incomes fall with worsening employment conditions, while the main earners may leave the household to fight or flee. However, extra-legal entitlements rise, with big gains for some households profiting from types of illegal production, but losses for others, subject to theft and looting.
- Worsening nutrition arises from reduced incomes and agricultural output. Calorie consumption fell in over 70 per cent of the countries worst affected by conflict, in 1970–95, falling to below 1700 calories per person per day in Ethiopia, Liberia, Mozambique, Sierra Leone, Somalia and Afghanistan (Stewart et al., 2001c, p. 90). The agricultural sector is typically badly hit in civil wars, as people are forced to move. Rising food prices can have a devastating impact on access to food. Millions of deaths in the Bengal famine of the 1940s have been

attributed to war expenditures leading to food price increases (Sen, 1981). Speculative traders can also contribute to rising food prices (Ravallion, 1987). The impact on nutrition may be offset by food subsidies, food distribution and rations (including food aid). In Nicaragua, nutrition actually improved in war-affected zones as a result of such measures (Utting, 1987), while in Mozambique, food aid seems to have prevented mass starvation (Stewart and Samman, 2001).

- Social entitlements, including health and education services, worsen as a result of destruction of facilities, reduced government resources and flight of personnel. Government expenditure on health and education is estimated to have fallen in ten of 14 war-affected countries (1970–95), by over 40 per cent in Angola, Liberia, Uganda, El Salvador and Iran (Stewart et al., 2001c, p. 87). Primary school enrolment fell substantially in some countries – notably Angola and Mozambique. Doctors per person also fell significantly in about half the war-affected countries. Civil society in some situations is able to substitute for government services but it too disintegrates in the worst cases.
- Health is affected by increased infection rates associated with the mass migration that often accompanies war – as many as one-third of the people in Mozambique were forced to move, while one-third of the Afghanistan population left the country in the 1990s. The spread of AIDs has also become a particular feature of war, as soldiers are notably highly infected and infect others, including sometimes through mass rape. HIV prevalence in the military was estimated to be 40–60 per cent in Angola, the Democratic Republic of Congo and Sierra Leone in the late 1990s (Collier et al., 2003). The extent of health costs is indicated by rising infant mortality rates: in Uganda additional infant deaths, compared with non-war regional trends, amounted to over 2 per cent of the 1995 population. Econometric estimates across countries show an increase in infant mortality rate (IMR) of 13 per cent during conflict (Hoeffler and Reynal-Querol, 2003). Moreover, some case study evidence suggests increases in adult mortality rates sometimes exceed increases in IMR (Guha-Sapir and Van Panhuis, 2002). World Health Organization (WHO) estimates of disability-adjusted life years show a major loss due to increased infectious diseases (Ghobarah et al., 2003), which persists in the post-conflict era.

In summary, while the direction of impact on most variables is demonstrably negative, and most types of entitlement worsen, there are large variations in the magnitude of costs across countries and on the burden of

entitlement loss across groups within a country. Variations in the social and economic impacts of war arise from differences in the nature of the war, the structure of the economy, the character of the government, the response of the international community and people's own actions. Moreover, while the evidence on worsening GDP and average consumption levels suggests that the net situation is one of loss, there are also gains from conflict for particular groups which can form one of the private motives for war or its prolongation as discussed above.

The economic consequences are obviously highly dependent on the nature of the war itself. Firstly, and above all, its duration is important. In a long war, reserves become exhausted, so vulnerability is increased. The geographic spread of the war is also relevant. When confined to one part of the country the war may have only small direct effects on the economy as a whole – conflict in Northern Uganda, which persisted over decades, had much smaller economic effects than the conflict in the mid-1980s which was centrally located. The extent of foreign involvement in the war is another factor affecting vulnerability, since external support may compensate for lost export earnings; it may also, however, contribute to prolonging the conflict.

Secondly, the structure of the economy helps determine the costs. An economy heavily dependent on the agricultural sector will be especially badly affected by widespread disruption of the sector, but may be less affected by reduced imports, since it is possible to retreat into subsistence if markets break down; an inflexible economy with a sizeable industrial sector may be particularly vulnerable to foreign exchange loss, although this can be offset by external grants or credit. An economy with a flexible industrial sector operating at less than full capacity may suffer little, as it is able to make up for loss of imports by domestic production – the UK in World War Two is a classic example.

A third critical factor is the nature of the government: a government that is or becomes very weak loses the ability to provide essential services and relief. Strong governments can sustain services, but they will only do so if they are relatively benevolent, wishing to provide for all the people, despite the war. Some strong governments may deliberately reduce food and basic services to 'enemy' territory, as was the case in Sudan.

Fourthly, international actions contribute to variations in costs. The large supplies of food aid in Mozambique undoubtedly reduced death rates, but in Sudan in the 1980s food aid was too little, too late and its delivery and use were distorted by government policy (Keen, 1994). In Afghanistan in the 1990s, massive support for the refugees in Iran and Pakistan greatly lessened the human costs. In contrast, the international community did little to offset costs in Sierra Leone in the 1990s, while the

USA increased the burden of civil war for vulnerable groups in Nicaragua by trade and aid embargoes.

Finally, people's own actions are important in moderating both human and economic costs. In almost all cases, people found new economic possibilities – many created by the war – which enabled them to survive. The burgeoning of the informal sector in Mozambique is one example (Chingono, 2001). People also protect themselves by fleeing, relocating within the country, or emigrating. People are rarely completely passive victims, but in the worst situations, there is little they can do to protect themselves.

Policies towards economies in conflict

The analysis above suggests that economic and social policies of both governments and donors can be designed to reduce the economic and human costs of conflict, even during the conflict.

An overriding aim of economic and social policies towards countries in conflict should be to maintain entitlements of the vulnerable, especially to food and health services, if possible in a self-sustaining way. In addition, the policies should also aim to tackle the causes of war, following the analysis above.

There can be no generalization about policy irrespective of the actual situation, since it is essential to understand the major cause of entitlement collapse, whether it is loss of employment opportunities, escalating inflation or destruction of key assets; and to understand the nature of the authorities, whether they are so weak that for short-run action they must be bypassed, or are strong but ill-intentioned so resources channelled through them will not reach those in need, and whether there are alternative structures (for example local authorities or rebel authorities) that can handle projects.

Monitoring is essential to identify appropriate actions before the situation becomes critical. Official monitoring is often weak (and politically biased). Greater use could be made of NGOs, local and foreign, who can be well situated to monitor developments and provide early warning of impending adverse changes in human conditions, for example, distress selling of assets, small movements of people or adverse changes in nutrition. In the Sudan, a major reason for the huge rise in death rates in the early 1980s was the failure to take early action, and this partly stemmed from the development community being slow to note early warning signs (Keen, 1994).

Successful macro-policies are generally much more effective in maintaining essential entitlements than direct relief. The aim should be to sustain the economy – which may require external aid and support for export markets – and to prevent escalating inflation. It is vital to sustain revenue to support public entitlements, compensating for declines in the

normal revenue base by devising taxes on war-related activities. The sale of food aid, for example, can be an important source of revenue.

Meso-policies need to be directed towards sustaining the share of aggregate output going to supporting entitlements of the vulnerable. On the social side, the objective is to ensure that everyone has access to adequate food and to sustain public entitlements to basic health and education. Preventive health measures, notably immunization, are particularly vital in war because unusual movement of people causes infections to spread rapidly. Expenditure on basic health and education accounts for only a fraction of social expenditure, so that strong prioritization of these services can ensure their maintenance even if the total is being cut. Yet the problem, of course, is not only one of money. Teachers and doctors may flee, and facilities can be destroyed. A flexible approach is needed. For example, Mozambique introduced mobile clinics and classrooms when Renamo was targeting health and education buildings.

Ensuring food security requires that food prices are monitored and escalating prices prevented, through some combination of increasing supplies (via food aid), controlling prices and rationing – policies which advanced countries have adopted when themselves at war. For the rural population, a combination of ensuring adequate agricultural support (seeds, fertilizer and so on), employment schemes, and the provision of food in schools and clinics, can achieve wide food access.

As well as domestic policies, there are important spheres of international action which can help (or worsen) the situation – including long-term policies to support development (such as improved terms of trade and aid flows), and short-term policies to reduce human costs (such as welcoming refugees, and providing food aid), as well as policies to reduce the financing of conflict (including policies towards trade in conflict commodities: for example, diamonds). International policies need to take into account the impact on the livelihoods and survival of poor populations. Economic sanctions, for example, are frequently harmful to the poor and often ineffective in achieving political objectives (Clark, 1996; O'Sullivan, 2003).

Conclusion

This overview of the economics of war and development suggests three major conclusions. Firstly, economic analysis of causes and consequences of conflict is essential to design appropriate policies. Secondly, both causes and consequences are diverse, varying across countries, so in-depth knowledge of the particular case is vital. Thirdly, there is a rich menu of appropriate policies which can help reduce the incidence and costs of conflict – yet in most cases these policies currently form a part of neither the normal development agenda nor the normal relief agenda.

Note

1. The burgeoning of subsistence and informal activities means that official statistics can greatly understate production, so that the aggregate costs of conflict may not be as great as they appear from official data.

References

- Ballentine, K. and J. Sherman (2003), *The Political Economy of Armed Conflict: Beyond Greed and Grievance*. Boulder, CO: Lynne Rienner.
- Barrows, W.L. (1976), 'Ethnic Diversity and Political Instability in Black Africa', *Comparative Political Studies*, 9 (2): 139–70.
- Brück, T. (2001), *Mozambique: The Economic Effects of War*, Oxford: Oxford University Press.
- Chingono, M. (2001), 'War, Economic Change and Development in Manica Province, 1982–1992', in F. Stewart and E.V.K. Fitzgerald & Associates (eds), *War and Underdevelopment*, Vol. 2, Oxford: Oxford University Press, pp. 89–118.
- Clark, R. (1996), *The Children are Dying: The Impact of Sanctions on Iraq*, New York: World View Forum.
- Cohen, A. (1974), *Two-Dimensional Man: An Essay on the Anthropology of Power and Symbolism in Complex Society*, Berkeley, CA: University of California Press.
- Collier, P. (1999), 'On the Economic Consequences of Civil War', *Oxford Economic Papers*, 51 (1): 168–83.
- Collier, P. (2000), 'Economic Causes of Conflict and their Implications for Policy', Washington, DC: World Bank
- Collier, P. and A. Hoeffler (2001), 'Greed and Grievance in Civil War', Washington, DC: World Bank Policy Series Working Paper 2355.
- Collier, P., H. Hegre, A. Hoeffler, M. Reynal-Querol and N. Sambanis (2003), *Breaking the Conflict Trap: Civil War and Development Policy*, Washington, DC and New York: World Bank and Oxford University Press.
- Collier, P., A. Hoeffler and C. Patillo (2004), 'Africa's Exodus: Capital Flight and the Brain Drain as Portfolio Decisions', *Journal of African Economics*, 13, Supplement 2: 1115–1154.
- Dodge, C.P. and P.D. Wiebe (eds) (1985), *Uganda: the Breakdown of Health Services*, Oxford: Pergamon Press.
- Duffield, M.R. (1994), 'The Political Economy of Internal War: Asset Transfer, Complex Emergencies and International Aid', in J. Macrae and A. Zwi (eds), *War and Hunger: Rethinking International Responses to International Emergencies*, London: Zed Books, pp. 50–70.
- Elbadawi, I. and N. Sambanis (2000), 'How Much War Will We See? Estimating the Incidence of Civil War in 161 countries', Washington, DC: World Bank.
- Ellingsen, T. (2000), 'Colorful Community or Ethnic Witches' Brew? Multiethnicity and Domestic Conflict During and After the Cold War', *Journal of Conflict Resolution*, 44 (2): 228–49.
- Fairhead, J. (2000), 'The Conflict over Natural and Environmental Resources', in E.W. Nafziger, F. Stewart and R. Vayrynen (eds), *War, Hunger and Displacement: The Origin of Humanitarian Emergencies*, Oxford: Oxford University Press, pp. 147–78.
- Fearon, J. and D. Laitin (1996), 'Explaining Interethnic Cooperation', *American Political Science Review*, 90 (4): 715–35.
- Gates, S. and M.S. Murshed (2005), 'Spatial–Horizontal Inequality and the Maoist Insurgency in Nepal', *Review of Development Economics*, 9 (1): 121–34.
- Ghobarah, H.A., P. Huth and B. Russett (2003), 'Civil Wars Kill and Maim People: Long after Shooting Stops', *American Political Science Review*, 97 (2): 189–202.
- Guha-Sapir, D. and W. Van Panhuis (2002), 'Mortality Risks in Recent Civil Conflicts: A Comparative Analysis', Brussels: Centre for Research on the Epidemiology of Disasters.
- Gurr, T.R. (1970), *Why Men Rebel*, Princeton, NJ: Princeton University Press.
- Gurr, T.R. (1993), *Minorities at Risk: A Global View of Ethnopolitical Conflicts*, Washington, DC: United States Institute of Peace Press.

- Gutierrez Sanin, F. (2004), 'Criminal Rebels? A Discussion of Civil War and Criminality from the Colombian Experience', *Politics and Society*, 32 (2): 257–85
- Hirshleifer, J. (1994), 'The Dark Side of the Force', *Economic Inquiry*, 32: 1–10.
- Hoeffler, A. and M. Reynal-Querol (2003), *Measuring the Costs of Conflict*, Oxford: Centre for the Study of African Economies.
- Homer-Dixon, T. (1994), 'Environmental Scarcities and Violent Conflict: Evidence from Cases', *International Security*, 19 (1): 5–40.
- Horowitz, D.L. (1985), *Ethnic Groups in Conflict*, Berkeley, CA: University of California Press.
- Humphreys, M. and A. Varshney (2004), 'Violent conflict and the Millennium Development Goals: Diagnosis and Recommendations', prepared for the Millennium Development Goals Task Force Workshop, Bangkok, June.
- Huntington, Samuel (1993), 'The Clash of Civilizations', *Foreign Affairs*, 72 (3): 22–49.
- Imai, K. and J. Weinstein (2000), 'Measuring the Impact of Civil War', Center for International Development at Harvard University Working Papers: 27, Cambridge, MA.
- Keen, D. (1998), *The Economic Functions of Violence in Civil Wars*, Oxford: Oxford University Press for the International Institute for Strategic Studies.
- Keen, D. (1994), *The Benefits of Famine: A Political Economy of Famine Relief in Southwestern Sudan 1883–1989*, Princeton, NJ: Princeton University Press.
- Langer, A. (2005), 'Horizontal Inequalities and Violent Group Mobilization in Côte d'Ivoire', *Oxford Development Studies*, 33 (1): 25–45.
- Mancini, L. (2008), 'Horizontal Inequality and Communal Violence: Evidence from Indonesian Districts', in F. Stewart (ed.), *Horizontal Inequalities and Conflict*, London: Palgrave, pp. 106–35.
- Marsden, P. and E. Samman (2001), 'Afghanistan: The Economic and Social Impact of Conflict', in F. Stewart et al. (eds), *War and Underdevelopment*, Vol. 2, *Country Experiences*, Oxford: Oxford University Press, pp. 21–55.
- Marshall, M.G. (2005), 'Major Episodes of Political Violence, 1946–2004', Centre for Systemic Peace, <http://members.aol.com/cspm/mg/warlist/htm>, accessed 15 July 2005.
- Matovu, J.-M. and F. Stewart (2001), 'The Social and Economic Consequences of Conflict: A Case Study of Uganda', in F. Stewart et al. (eds), *War and Underdevelopment: Country Experience*, Oxford: Oxford University Press pp. 240–88.
- Nafziger, E.W. and J. Auvinen (2000), 'The Economic Causes of Humanitarian Emergencies', in E.W. Nafziger, F. Stewart and R. Väyrynen (eds) *War, Hunger and Displacement: The Origin of Humanitarian Emergencies*, Vol. 1, Oxford: Oxford University Press, pp. 91–146.
- Østby, G. (2004), 'Do Horizontal Inequalities Matter for Civil Conflict?' Oslo: Centre for Study of Civil War (CSCW), International Peace Research Institute (PRIO).
- O'Sullivan, M. (2001), Sri Lanka: Civil Strife, Civil Society and the State 1983–1995, in F. Stewart et al. (eds), *War and Underdevelopment*, Vol. 2, Oxford: Oxford University Press.
- O'Sullivan, M. (2003), 'Shrewd Sanctions: Statecraft and State Sponsors of Terrorism', Washington, DC: Brookings Institution.
- Percival, V. and T. Homer-Dixon (1998), 'Environmental Scarcity and Violent Conflict: The Case of South Africa', *Journal of Peace Research*, 35 (3): 279–98.
- Ra, S. and B. Singh (2005), 'Measuring the Economic Costs of Conflict: The Effect of Declining Development Expenditures', Manila: Asian Development Bank.
- Ravallion, M. (1987), *Markets and Famines*, Oxford: Clarendon Press.
- Reynal-Querol, M. (2002), 'Political Systems, Stability and Civil Wars', *Defence and Peace Economics*, 13 (6): 465–83.
- Sen, A.K. (1981), *Poverty and Famines: An Essay on Entitlement and Deprivation*, Oxford: Clarendon Press.
- Smith, A.D. (1991), 'The Nation: Invented, Imagined, Reconstructed?' *Millennium: Journal of International Studies*, 20: 353–68.
- Stewart, F. (2000), 'Crisis Prevention: Tackling Horizontal Inequalities', *Oxford Development Studies*, 28 (3): 245–62.
- Stewart, F. (2002), 'Horizontal Inequality: a Neglected Dimension of Development', Helsinki: WIDER Annual Development Lecture.

- Stewart, F., (ed.) (2008), *Horizontal Inequalities and Conflict: Understanding Group Violence in Multiethnic Societies*, London: Palgrave.
- Stewart, F. and E. Samman (2001), 'Food Aid During Civil War: Conflicting Conclusions from Alternative Approaches', in F. Stewart et al. (eds), *War and Underdevelopment*, Vol. 1, *The Economic and Social Consequences of Conflict*, Oxford: Oxford University Press, pp. 169–204.
- Stewart, F., V. Fitzgerald, D. Keen, C. Huang, M. Wang, E. Greitens and E. Samman (2001a), *War and Underdevelopment, Volume 1: The Economic and Social Consequences of Conflict*, Oxford: Oxford University Press.
- Stewart, F., V. Fitzgerald, P. Marsden, E. Samman, T. Brück, M. Chingono, D. Keen and J. Matovu (2001b), *War and Underdevelopment, Volume 2: Country Experiences*, Oxford: Oxford University Press.
- Stewart, F., C. Huang and M. Wang (2001c), 'Internal Wars: An Empirical Overview of the Economic and Social Consequences. In F. Stewart et al. (eds), *War and Underdevelopment*, Vol. 1, Oxford: Oxford University Press, pp. 67–103.
- Utting, P. (1987), *Economic Adjustment under the Sandanistas: Policy Reform, Food Security and Livelihood in Nicaragua*, Geneva: UNRISD.

PART IX

ECONOMIC DEVELOPMENT EXPERIENCE

64 Economic development in the Third World: an international perspective

Richard Kozul-Wright and Paul Rayment

Introduction

According to conventional economic logic the surest way for poorer countries and communities to achieve lasting gains in economic welfare is by getting rid of state-created 'distortions' that impede market forces in allocating existing, and mobilizing future, resources. In the recommended reform package, opening up to international markets and firms offers the surest way of bringing about the right set of incentives to help realize those gains. The liberalization of international trade will enable countries to exploit their comparative advantages better, with big gains expected in the South given their legacy of protectionist policies. Where finance is a constraint on growth, the liberalization of international capital markets will ensure that investment funds flow from the capital-abundant North to the capital-scarce developing countries. Attracting foreign direct investment (FDI), including through the sale of state-owned assets, will help gain quick access to new technologies and management practices as well as providing ready-made export opportunities. Entrusting development to these international market forces is seen as having the additional advantage of their being less vulnerable to capture by local interest groups and rent-seeking coalitions. A plethora of econometric studies have backed up this case for hitching development to open markets, and a string of popular economic pundits have been ready at hand to explain just how their 'win-win' logic has already begun to flatten out the global economy, raising expectations of a swift eradication of extreme poverty, a narrowing of income gaps, and the emergence of a truly global middle class.

There is little doubting that, in the wake of the debt and development crisis of the 1980s, such thinking, oftentimes with the strong backing of the international financial institutions, affected a radical shift in policy-making in many developing countries. During the 1990s, new technologies and international business practices (closely identified, if not synonymous, with globalization) were added to the reform mix, ending, on many accounts, any further debate on development strategy. Efforts to direct globalization were deemed futile and resistance would result only in marginalization, or perhaps worse.¹ Developing countries were, accordingly, advised to adapt

to this new reality by fully relinquishing economic sovereignty to mobile capital and the forces of international competition.

This chapter questions such advice. It argues that economic trends during the past quarter-century do not support the utopian pronouncements of many globalization enthusiasts; that the destructive impulses released by the radical turn in policy after the debt crisis have, in many countries, outweighed the creative impulses; and that expunging the historical detail and structural diversity from the catch-up process is unlikely to provide the direction on policy advice and institutional reform needed to narrow income gaps worldwide.

Openness, accumulation and structural change in an interdependent world

Most development economists agree that a strong productivity performance is essential to any successful catch-up growth path, not only because it translates (though not always directly) into rising living standards, but also because it enables poorer countries to manage better the various adjustments, trade-offs and distributional conflicts that taking such a path will generate. Most would also agree that this performance is more about galvanizing dynamic economic forces than it is about maximizing the static gains from an improved allocation of existing resources.

Certainly, recent efforts to revive the idea of market-driven convergence for a globalizing world have tried, by introducing a broader conception of capital (including human capital and other less tangible wealth-creating assets, and an expanded role for foreign capital), more variegated technological trajectories and additional behavioural parameters, to accommodate more dynamic forces in their blend of (old) trade and (new) growth models. The empirics of this new convergence literature have already provoked much debate and controversy and methodological questions surround the idea of 'conditional' convergence.² However, what is more revealing about this literature is its enduring commitment to a set of standardized policies derived from the identification of distortionary market impediments within mathematically tractable equilibrium models. This approach continues to trump historical detail and structural differences in the design of development strategy (Kenny and Williams, 2001).³

An alternative place to begin thinking about development strategy is with the empirical regularities identified by economic historians and classical development economists linking industrialization, and more particularly its manufacturing component, to strong productivity and income growth.⁴ This leading role is due less to any uniquely intrinsic qualities of industrial activity and more to the confluence of growth impulses that accompany its evolution, including the presence of increasing returns whether at the plant, firm or industry levels, high elasticities of supply and demand for

manufactured goods, strong externalities linked to a high rate of technological innovation, and complementarities between production and consumption. Successful industrialization paths have, moreover, been closely identified with a 'well-filled input-output matrix' with an increasingly dense set of links between sectors (a high level of sectoral articulation between, for example, rural and urban, and consumer goods and intermediate goods), and a structure of demand such that a high proportion of domestic production is sold to domestic wage earners (Wade, 2003, p. xlvi).

In the interplay of elements making up this strong growth regime, capital accumulation plays a pivotal role. A given pace of accumulation can of course generate different growth rates, depending on its nature and composition, as well as the efficiency with which production capacity is utilized. This is one of the main reasons why econometric studies have failed to establish a one-to-one relation between the rate of investment and economic growth.⁵ Unfortunately, systematic discussion of the forces that govern the process of capital accumulation has long been a stumbling block in the development literature (Hirschman, 1958, p. 35), and the recent growth literature is no exception.⁶ In particular, the neglected role of profits for financing investment in developed and developing countries alike, cuts off conventional economic analysis from a careful examination of how dynamic industrial activity can provide abundant opportunities to create rents whose reinvestment is key to perpetuating a dynamic growth regime. This profit-investment nexus has been identified in the recent fast-growth experiences in East Asia (Akyüz and Gore, 1996; Singh, 1999; Ros, 2000) and provides an initial guide to why and how policy interventions might make a difference in stimulating catch-up growth (UNCTAD, 1997).

While industrialization can generate the kinds of cumulative impulses that describe a successful growth process, these are not automatically self-sustaining. Various constraints, traps and coordination failures can upset the process, making it generally impossible to rely on market forces to establish the linkages associated with internal integration and to move economies through the various stages of industrialization. This provides further rationale for policy intervention and institutional learning. The variety of linkages and their local specificity are now much better appreciated. Among these, insufficient domestic demand to absorb the growing industrial output can be just as important as failures on the supply side. The close links between industrialization and 'external' integration have, of course, been a familiar feature of the development process since Adam Smith saw the size of the market as a constraint on the division of labour. By broadening the size of the market, exports allow scale economies to be exploited; they also provide the foreign exchange needed for capital accumulation, in view of the dependence of most developing countries on

imported capital and intermediate goods. At the same time, investment improves export potential by adding to production capacity and improving competitiveness through productivity growth. As such, a successful industrialization path is usually characterized by rising investment, exports and manufacturing value added, both in absolute terms and as shares of gross domestic product (GDP).

The historical evidence provides little support for an independent role for rapid liberalization of foreign trade and finance and the deregulation of domestic markets in stimulating and sustaining catch-up growth. In most cases, the state has provided a necessary complement to, and sometimes a corrective influence on, the market, particularly by promoting a rapid pace of capital accumulation and technological progress linked to expanding industrial output, employment and exports. Accordingly, establishing a robust nexus between investment, profits and exports remains key to the design of development strategy, helping identify the cumulative links that describe a successful industrialization path, as well as providing a framework in which to explore the institutional and policy challenges thrown up by the catch-up process.

The international economic environment: open for business

Following the debt crisis of the early 1980s, deregulation of domestic economic activity and its opening up to international firms and market forces became the leitmotif of economic policy design in many developing countries, more often than not with the overt support of the international financial institutions. Success has tended to be measured in terms of monetary and fiscal discipline, an increasing volume of international trade and capital flows, and rising ratios of trade and FDI to GDP. On these measures, many poorer countries, and the world economy more generally, have since the late 1980s registered a good deal of success. Trade has consistently outpaced global output, with the pace of expansion much faster in the 1990s, and with developing countries in the vanguard (UNCTAD, 2003, pp. 41–4). As a result there has been a rapid and ubiquitous rise in the share of exports and imports in GDP in developing countries, as well as a rapid increase in the share of these countries in global trade – from about 24 per cent of total exports in 1980 to 28 per cent in 1995 and 34 per cent in 2004.

The rise in capital flows has been even more dramatic. The global stock of financial assets rose more than elevenfold between 1980 and the end of 2004, from \$12 trillion to \$136 trillion; in 1980 they were roughly equal to global GDP but by the end of 2004 were more than three times higher. Much of the increase in flows has been among developed countries, but the 1990s saw a strong surge of financial flows to developing countries, following a sharp dip

in the 1980s (UNCTAD, 2003, pp. 23–31). The greater part of these flows consisted of short-term investments: bank loans, equities and short-dated government securities, inter-bank and other deposits. However, beginning in the early 1990s, FDI in developing countries also rose sharply, more than doubling between 1995 and 2000, when it reached over \$250 billion, rising from a quarter to close to one-third of global flows.

These numbers do not fully capture the changes in the workings of the international economy. In a world where a good deal of trade takes place between affiliates, where technology transfer is tightly controlled from corporate headquarters and where credit is extended by firms to their customers, corporate governance has become a much more prominent factor coordinating international economic relations. Innovation has also characterized financial institutions as a growing share of their activity is distributed across multiple locations and these institutions tend to offer a variety of services and undertake multiple activities to an extent that breaks with past practice.

Multilateral institutions, rules and arrangements have, over this same period, also gone through a series of mutations in line with a more open and liberal system of trade and capital movements. The International Monetary Fund (IMF), since the mid-1970s, has abandoned its objective of seeking to ensure stable exchange rates in an orderly international financial system and, instead, openly encouraged the opening up of capital accounts as a way of attracting international financial flows. The World Bank has also given up its emphasis on longer-term infrastructure project lending, concentrating instead on adjustment lending and poverty reduction. With the Uruguay Round negotiations, the governance of international trade has moved towards a single-tier system of rights and obligations, in which trade liberalization has been given priority over economic growth and full employment, and linked to a series of ‘related issues’ which take the multilateral rule-making process much deeper into the workings of national economies.

Trade liberalization and the investment climate

Conventional trade theory does not distinguish between different activities in terms of their differential impact on economic growth. In practice, however, most economic policy-makers worry about the composition of trade for this reason. There have been some positive trends in recent years. In particular, the sharp rise in the share of manufactures in exports from developing countries, from one-fifth in the early 1980s to around 70 per cent by 2004, points to an important change in global trade dynamics and one that most proponents of market-led globalization have been quick to highlight as a further measure of its success. However, this is hardly an

exhaustive picture of the changes in the trading system. In the first place, the rise of manufacturing exports is down to just a handful of countries, largely from East Asia. Indeed, simply taking out the first-tier newly industrialized economies (NIEs) from the developing-country group, their share of world trade drops from around 22 per cent in the early 1980s to below 15 per cent by the end of that decade, but has now (in 2005) climbed back to the earlier figure; however, taking out China leaves the developing countries with less than one-sixth of total world trade. By implication, other developing regions have exhibited a stagnant or declining share since 1980, with exports very heavily skewed towards natural resources; these account for more than 30 per cent of export earnings in over 80 countries, often from just one or two products. Together these trends imply that many countries remain stuck with undynamic trading regimes.

Another, and less reported, trend in the trading system is that while many countries have been trading more, they have been earning less from doing so. In a recent study of 127 developed and developing countries, Dowrick and Golley (2004) found that between 1960 and 1980, increased trade helped productivity to grow in poorer countries at double the rate in richer countries, but that this gain was reversed in the period of more open trade between 1980 and 2000, when the marginal impact of trade on productivity growth favoured the richer countries, and indeed turned negative for poorer countries.⁷ Wacziarg and Welch (2003), using the Sachs–Warner methodology for determining openness, found that its links to faster growth were period-sensitive, with much weaker links in the 1990s than in earlier decades, and with more open economies if anything benefiting less than relatively closed economies. Finally, increases in the developing countries' share of world manufactured exports since 1980 have not been matched by a corresponding rise in their share of global value added, and in a number of cases shares of global manufacturing income have actually fallen over the past decade or so, even as their share of world exports of manufactures was rising, while for others it has risen by much less than that share.

One possible explanation for these trends is biases in the liberalization process which have prejudiced growth prospects in developing countries by discriminating against sectors where they had or could build strong export sectors, even as domestic producers were forced to confront strong competitive rivals on home markets (UNCTAD, 1999; ILO, 2004). However, the fact that many countries were trading more but earning less during the 1990s suggests more deep-seated biases in the operations of the trading system (UNCTAD, 2002). These seem likely to be linked to the lopsided reliance on external demand as the basis of sustained growth (Heintz, 2003).

The risk of falling export prices, resulting from too many producers chasing too few markets, has of course been familiar to commodity exporters, where returns have long been asymmetrically skewed through organized markets in rich countries and more disorganized and fragmented markets in poorer countries.⁸ However, the structure of some key markets for developing-country manufactured exports seems to point to similar pressures emerging even where trade expansion has been rapid (UNCTAD, 2002, pp. 121–4).⁹ In this respect, the suggestion that in the new era of globalization developing-country exporters of some high-technology products have been able to bypass technological constraints is often misleading. Even when the final product has been classified as high-tech, many developing countries have in reality only been involved in low-skill assembly activities, using imported capital and intermediate goods, and whose contribution to value added is determined by the cost of the least-scarce and weakest factor, namely unskilled labour.

The confusion over what constitutes a dynamic export is linked, in part, to the increase in FDI flows through international production networks (IPNs), in which corporations slice up their value chain, relocating or outsourcing the different parts, from product design to final delivery, in a way that enhances overall profits.¹⁰ In some cases, production is organized by large transnational corporations (TNCs) producing a standardized set of goods in several locations (as in electronics and transport industries). In others, production involves groups of small and medium-sized enterprises located in different countries and linked through international subcontracting (as in clothing). Under these conditions, FDI could bring macroeconomic benefits to the host through a positive impact on its investment climate and on its balance-of-payments position. Certainly the expectation after the debt crisis, and particularly following the Brady Plan, was that liberalization, particularly when export-oriented FDI was attracted, would bring such benefits (Camdessus, 1997).

Although the bulk of FDI continues to flow between advanced countries, developing countries have, indeed, experienced a fifteenfold rise in average annual inflows of FDI since the mid-1980s, which has translated into it taking a significantly higher share of capital accumulation in most developing countries and regions. Moreover, between 1990 and 2003, the share of manufacturing in the FDI stock of the group of developing countries rose from 25 per cent to 37 per cent while the share of developing countries in the global stock of manufacturing FDI increased from one-fifth to close to one-third during the same period. However, FDI surges have often failed to stimulate a domestic investment recovery and there is evidence that it has even crowded out local investment (Gosh, 2004). The extent to which these flows have eased the balance of payments constraint has varied

unpredictably with the share of TNC profits in value added, the degree of import dependence, and the proportion of the final good sold in domestic markets (Akyüz, 2004). Certainly where the propensity to import of foreign firms has been much higher than that of domestic firms, and their export propensities similar, increased FDI has added to balance-of-payments pressures (Chudnovsky and López, 2002). In fact, the broad body of evidence points to this constraint remaining a tight bind on economic growth in many developing countries (UNCTAD, 1999).

In the absence of favourable macroeconomic effects from attracting FDI, much depends on technological and other spillovers that might strengthen internal integration. The broad body of evidence suggests that such effects are plant-, sector- and country-specific, though generally significant effects seem to depend on local absorptive capacity already being in place.¹¹ As such, FDI tends to lag rather than lead the growth process (UNDESA, 2006). From this perspective, production networks may well increase the risk of TNCs being attracted into enclaves with limited linkages to the domestic economy; taking account of the different methodologies employed, there is little evidence to date of positive spillovers from the recent surge in FDI to developing countries, and virtually none on a significant scale.

Financial liberalization and capital flows

Financial markets have undergone a dramatic transformation since the early 1980s, thanks to a combination of deregulation, internationalization and innovation. While the impulse for this transformation came from the advanced countries, economic logic promised much for the world's poorest countries. Deregulated and open financial markets would not only increase the availability of investment finance, both domestic and foreign, but they would also help create a more stable and disciplined investment climate, and free deficit countries from the unpredictable politics of ODA flows (Camdessus, 1997; Mishkin, 2006).

An upsurge in flows began in the 1990s, albeit in part a return to trend after the blighted years of the 1980s. Still, a plethora of new financial instruments promised to mitigate risk, particularly in those emerging markets that greatly excited investors after the Berlin Wall collapsed, providing arbitrage opportunities but also encouraging herding behaviour to become a more significant influence on the direction of flows.¹² Consequently, these flows were very unevenly distributed, increasingly concentrated in a small group of 20 or so emerging markets which received over 90 per cent of total inflows of capital in the 1990s, compared to some 50 per cent before the outbreak of the debt crisis. Thus, the share of low-income countries in net private capital flows to developing countries has been steadily declining

since the second half of the 1980s, from 20 per cent (around \$6 billion) to just 6 per cent in the second half of the 1990s (around \$14 billion), with an attendant rise in the share of middle-income emerging markets.¹³ In recent years more than three-quarters of bond issues have been accounted for by less than ten borrowers in Latin America and Asia and much of the syndicated bank lending has gone to half a dozen countries in Asia. These same countries were also the main recipients of international equity investment.

Despite the ever louder beating of the financial liberalization drum during the 1990s, including efforts to put capital account openness in the IMF's Articles of Agreement, these flows proved increasingly difficult to manage in a way consistent with faster and more inclusive economic growth. For countries seeking re-entry into international financial markets after the debt crisis, higher real interest rates and a stable exchange rate were prerequisites for attracting new inflows. While the resulting financial stringency could prove attractive to foreign investors, a tight monetary and fiscal stance, oftentimes with an appreciating currency, did little to stimulate domestic investment or to improve export prospects. Indeed, increased debt-servicing obligations resulting from higher interest rates, along with import surges, ran the risk of reproducing an unsustainable debt burden. In many cases, a combination of capital outflows, profit remittances and the accumulation of exchange reserves greatly reduced net inflows, and of these a growing proportion was absorbed by activities which added little to productive capacity (UNDESA, 2005). Particularly in the form of short-term loans and portfolio equity, these inflows, often highly leveraged through derivative contracts and hedge funds, could be very unstable and an unreliable source of development finance.

As the workings of financial markets became increasingly disconnected from the longer-term demands of productive investment and industrialization, unregulated financial flows triggered boom–bust cycles, which became a recurrent feature of the developing world during the 1990s. The precise circumstances in which the vulnerability to the reversal of capital inflows arose, and the subsequent impact on growth, varied from region to region. An early warning was given by the Mexican peso crisis of 1994. However, the full force of unregulated financial flows was revealed by the financial crises in East Asia, a region with a long-standing record of strong growth and fiscal discipline. As in other episodes of financial crisis and currency turmoil, the crisis in East Asia was preceded by financial liberalization and deregulation which, in some cases, constituted a major break with past practice.¹⁴ Moreover, the extremes of collapse were amplified by unnecessarily tight monetary policies which deepened the debt deflation process, served to depress output and employment, and caused serious dislocations in the corporate and financial sectors (Stiglitz, 2002).

Global slowdown and instability

The debt crisis and its aftermath generated a lost decade for many developing countries in the 1980s, with incomes contracting in many cases. On some assessments this has given way to a much more benign macroeconomic climate with lower real interest rates, less-volatile growth in the leading markets and diminished inflationary threats. Despite this, the slowing global trend persisted for much of the 1990s. Indeed, even as recovery in the USA turned into more sustained growth from the mid-1990s and growth in China and India accelerated, the average global growth rate for the decade was still not above that of the 1980s, and remained below that of the 1960s and 1970s.

This slowing trend has been associated with a good deal less stability in the growth performance of many developing countries. Some of this (as in the case of transition economies) can be traced to political shocks. But, if history is any guide, it should not come as a surprise that a relaxation of regulations put in place in response to earlier excesses generates 'manias, panics and crashes' (Kindelberger, 1984), or that these would hit weaker economies the hardest. On one estimate, during the 1990s the financial system was in crisis for 40 out of 120 months (Plender, 2003, p. 57). According to Barry Eichengreen (2002), the period since the collapse of Bretton Woods has seen a sharp increase in the incidence of financial crises, principally in the form of currency crises but also in conjunction with banking crises. Most of these have been in the developing world.

The global slowdown and increased incidence of financial crises has coincided with a declining frequency of strong growth episodes and a rising frequency of negative growth episodes. According to Ocampo and Parra (2005), in the 1960s and 1970s about 40 per cent of developing countries had successful growth episodes (with annual average rates of per capita GDP growth greater than 3 per cent over at least a five-year period) but this proportion fell to less than 20 per cent through most of the past quarter-century, while in 40 per cent of countries there were negative growth episodes compared with just 15 per cent in the earlier period. Similarly, Hausmann et al. (2004) searched for episodes of rapid acceleration in economic growth that were sustained for at least eight years and found that while there were 23 and 30 such episodes in the 1960s and 1970s respectively, there were only 14 in each decade of the 1980s and 1990s.

Fragmented development

While the influence of radical market-based reforms on policy-making in advanced countries was heavily qualified by checks and balances in their political systems (Krugman, 2007), these, thanks in no small part to the support of the international financial institutions, quickly became a fully

fledged development blueprint showing how poor countries could extricate themselves from the problems of the debt crisis and establish a new growth path. It promised to remove structural and institutional impediments to growth, improve productive capacity and trade performance, and put an end to stop-go development associated with excessive indebtedness and periodic payments crises. Above all it promised strong convergence in the global economy as growth in poorer countries outpaced that in richer countries and income gaps across the world economy narrowed sharply. The previous section has raised some initial doubts about blindly trusting in international market forces and firms to achieve this outcome. This section looks in more detail at how the interplay of external and internal integration forces has played out across the developing world.

Growth, gaps and international inequality

Contrary to much contemporary rhetoric, the growth in cross-border trade and financial flows since the mid-1980s has failed to stop the growing gap in real per capita incomes between rich and poor countries. Economic divergence is, here, the simple product of the higher starting income of richer countries and their faster average annual per capita growth rates (Milanovic, 2005). This has happened despite the sharp slowdown in the advanced countries since the mid-1970s, a slowdown which, as noted earlier, has resulted in a general slowing of the world economy over the same period.

Such a high level of aggregation can, however, be misleading and there are important differences across regions within the developing world. Asia has persistently maintained a much stronger growth performance than other developing regions, enjoying catch-up growth with the advanced countries since the early 1990s, and for a good deal longer in East Asia. By contrast Africa, particularly south of the Sahara, and Latin America have continued to fall further behind. Given that, among developing countries, Latin America was the richest developing region by some margin at the end of the 1960s, subsequent growth trends have generated convergence across the developing world. There are also some noteworthy differences in growth performance over time, with developing countries outpacing advanced countries in the 1970s and again over the decade since the mid-1990s, with a lost decade sandwiched between in the 1980s when developing-country growth performance was below its own historical average and that of the more advanced economies.

In both popular and scholarly accounts, the actual and projected tilting of the world's economic axis towards Asia is often taken as synonymous with globalization. In fact, that shift has been ongoing since the 1950s, when Japan entered a period of rapid and sustained growth, joined soon

after by the small Asian tigers, notably Korea and Taiwan, China.¹⁵ A second tier (of more variegated economies) from South-East Asia joined in during the 1980s. However, it has been the emergence of China and India, particularly the former, that has added a new dimension to the Asian catch-up story.¹⁶ While both are still a long way from being middle-income countries, they have a large number of citizens who would qualify as middle class. Their emergence has already had a significant impact on global growth dynamics as well as on the wider regional performance, with growth in South Asia, since 1980, improving sharply over the preceding two decades, and high growth rates maintained in East Asia in the 1990s despite the onset of a severe financial crisis in 1997, though the pace did not match that of the 1970s. There is also little doubt that the impact is being felt beyond the region, though just how the balance between their growing demand for natural resources and their addition to the world's unskilled labour force will play out elsewhere is a subject of much dispute.¹⁷

After the lost decade of the 1980s, Latin American economies enjoyed a brief renaissance in the early 1990s when the intensification of structural reforms enabled them to return to the international capital markets; but after 1997 developments again turned sour and produced another 'lost half-decade' (Ocampo, 2002). Growth performance in the 1990s was similar to that in the 1960s but well below that in the 1970s. Sub-Saharan Africa, like Latin America, also suffered a 'lost decade' of development in the 1980s, but with a weaker (if less erratic) recovery in the 1990s; there was generally little or no inflow of private capital into the region, and no abrupt reversal at the end of the decade.¹⁸

Against the backdrop of a global slowdown, tight macroeconomic policies and the increasing frequency of financial crises, many developing countries have seen a weakening of growth relative to their own past performance. Of a total of 124 developing countries, growth in 95 of them (that is, over three-quarters) was faster in the period 1960–78 than between 1978 and 1998 (Milanovic, 2002), and only a handful of countries have been able to hit the kind of growth targets needed to address their economic and social deficits:¹⁹ between 1980 and 2000, of 140 developing countries only 20 grew at annual rates above 5 per cent, a number that rises to 30 for the period 1990–2000, but if a rate of 7 per cent is taken as the benchmark, the number of success stories falls to just five and six respectively.²⁰

Countries at the very bottom of the income scale appear to have lost most ground. This has led some to talk of a 'twin peaks' global income distribution (Quah, 1996), with a 'hollowing-out' of the middle-income range of countries (UNDESA, 2006b, p. 8). Milanovic and Yitzhaki (2001), for example, estimate that just 8 per cent of the world's population

fall into that category. Polarization can, in large part, be explained by the fact that the general slowing of global growth since the 1980s has hit poorer countries particularly hard. According to Milanovic (2005, p. 5), the average annual per capita growth rate of the group of LDCs was just 0.1 per cent between 1980 and 2002, compared to 1.9 per cent in the 'old' Organisation for Economic Co-operation and Development (OECD) economies, and while the former figure hides a very wide dispersion of performance, there were no stellar growth performance of the kind found in East Asia. Growth collapses were, not surprisingly, much more frequent among this group of countries than other developing-country groupings.

Given the close (though not direct) connection between income growth and poverty reduction, there have been some clear differences in poverty trends across regions over the past two decades. Overall the number of people living on less than \$1 a day has dropped since the early 1980s by around 400 million. However, this headline figure hides very large falls in some countries, notably China, which pulls down the regional figure for Asia, along with little change or increases in almost all other regions. Indeed, the improvement in the world poverty figure can be attributed almost exclusively to China.

Just how income and poverty trends translate into a picture of global inequality is an ongoing source of controversy among economists. The Gini coefficient is often used as a more integral measure that aims to take all members of a chosen set into account, although measurement problems abound.²¹ On this measure, a number of studies have reported a reduction in international inequality since 1980, linking this in turn to a more open world economy. However, these findings are open to serious questioning on both methodological and empirical grounds. Indeed, as Sutcliffe (2006) has noted in his balanced assessment of the debate, there is little agreement on how significant the change has actually been or just when the reversal might have begun. More importantly still, the result hinges on the performance of a single outlier; removing China from the country set reverses the trend to one of rising international inequality, and even more sharply than before 1980. The Theil decomposition of international inequality shows a similar result (UNDESA, 2006, p. 14). Given its size, China's performance is obviously central to the analysis of global trends. However, from a comparative perspective what happens in a single country (however large) should not be used to obscure what is in fact a highly variegated picture across the global economy (Berry and Serieux, 2004).

The variety of development experiences in the 1990s

A recent assessment by the World Bank (2005, p. 30) of its own policy advice during the 1990s has acknowledged that it persistently overestimated

growth prospects in regions implementing adjustment programmes, and underestimated growth performance in those that were not. This is a belated acknowledgment of a good deal of academic research which finds little evidence that adjustment programmes have had a positive impact on growth, employment or poverty.²² However, there is a reluctance to accept that the adjustment path that these policies helped fashion has actually set back development prospects, in many cases by distorting the process of internal integration, focusing instead on what might have been left out of the package in terms of social policies, good governance, institutional detail, and so on.

In fact, regions that pursued extensive adjustment programmes in the 1980s found it particularly difficult to reverse the sharp drop in the share of investment triggered by the debt crisis. In some countries, this dipped to below the levels needed to replace depreciated capital, and where adjustment programmes persisted, the investment cycle remained volatile, even after the immediate disturbances of the crisis subsided. The downsizing of public investment has been a prominent part of this story, and in many cases this has crowded out domestic private investment.²³ But, as important, has been the approach to macroeconomic fundamentals guiding policy design, which has ignored the mutually reinforcing links between aggregate demand, capacity utilization and investment. Tight monetary conditions and accompanying currency gyrations have further discouraged growth based on capital accumulation, adding to the difficulties encountered by domestic industry in introducing the kind of restructuring that would help raise productivity, even as it faced stiff competition from lower trade barriers. In part as a consequence of these difficulties in the productive sectors of the economy, and in part as a result of ongoing liberalization of the financial sector, 'rentier' investments (including in government bonds) and 'capital flight' have become much more attractive options. In many cases this financialization of the domestic investment regime has coincided with and reinforced commodity dependence, often linked to increased FDI inflows (Magalhães Prates and Paulani, 2007).

Under these conditions, the cumulative nature of industrial development can very easily go into reverse gear. Certainly, where industrial stagnation was the norm in the 1980s, as in Latin America and Africa, most countries found it difficult to reverse the trend in the 1990s, with premature deindustrialization a visible trend in some cases (UNCTAD, 2003, pp. 92–9). Even where this has not been a dominant trend, there is little sign of the diversification and upgrading which describes a healthy internal integration dynamic. This is often associated with enclaves of specialized development. In the case of many Latin American economies these enclaves have been in the primary sector, though in some cases manufacturing activities

have been able to carve out niche markets. Where productive capacities are even more rudimentary, as in much of sub-Saharan Africa (SSA), the likelihood of enclave development in the primary sector generating a natural resource curse seems to be even greater.

Moreover, in economies with declining shares of investment and manufacturing value added, a stagnant or falling share of manufactures in total exports was often the norm, even as the overall composition of developing-country exports was shifting rapidly towards manufactures, including more skill- and technology-intensive goods. Labour market performance, key to tackling poverty, has also been caught up in these divergent developments. The impact of liberalization on labour market performance is a complex matter, with the effects differing between countries depending on the domestic and international conditions prevailing when liberalization occurred, as well as on the broader development strategy being pursued by policy-makers. However, a good deal of evidence from Latin America and sub-Saharan Africa, where neoliberal policies have been pursued most intensively, suggests that increases in unemployment and/or increasing wage inequality have often accompanied rapid trade liberalization.²⁴ In many cases, adjustment policies and the downsizing of the public sector have led to a hollowing-out of the middle class. The counterpart of these trends has been an expanding informal economy, which by the end of the 1990s accounted for anywhere between one-third and three-fifths of the labour force in Africa and Latin America (Schneider, 2002).

By contrast, the countries in East and South Asia that bucked the declining investment trend in the 1980s were able to maintain a reasonably stable or even rising pace of capital accumulation for most of the 1990s, with the share of public investment holding up or rising further (UNCTAD, 2003, pp. 65–73). But just as importantly, strong investment has supported structural change. In countries where investment and industrial output expanded in the 1980s, this continued in the 1990s. In these cases rising and fully utilized capacity levels have had a knock-on effect on productivity growth through both the demand and supply sides. Indeed, Asian development since the 1970s confirms the key role of industrialization in establishing a rapid growth path, with positive cumulative effects through strong export drives linked to rising levels of productivity (UNDESA, 2006). During the 1990s, just eight East Asian countries accounted for 70 per cent of developing countries' trade in manufactures. Outside this group, export strategies relied on low (and in some cases falling) wages or currency depreciation rather than strong productivity growth, and while this stimulated recoveries in some countries, few were able to reach a threshold level of exports consistent with a vibrant industrialization path (UNCTAD, 2003, pp. 99–102).

Taking stock of trends since the debt crisis of the early 1980s, it is clear that by itself the degree of exposure to global market forces is not what distinguishes ‘winners’ and ‘losers’. Rather, the main difference, particularly between the East Asian NIEs and most other developing countries, was that liberalization followed the successful implementation of industrial and trade policies; protection and support were removed in large part because they were no longer needed. In the latter, on the contrary, liberalization has largely been triggered by the failure to establish efficient, competitive industries in labour- and/or skill-intensive sectors. From this perspective, the different ways in which internal and external integration have taken shape in developing countries have failed to trigger convergence since the early 1980s, but have instead created a very uneven economic landscape of diverse experiences.

Mature industrializers This group includes the first-tier NIEs, notably the Republic of Korea and Taiwan Province of China, which achieved industrial maturity through rapid and sustained accumulation of capital, and growth in industrial employment, productivity and output, as well as manufactured exports. In the 1990s these economies enjoyed a share of industrial output in GDP above the levels of advanced countries, exports had shifted to more capital- and technology-intensive goods, and industrial growth was starting to slow down as resources shifted towards the service sector.

Rapid industrializers A number of countries saw a rising share of manufactures in total output, employment and exports, based on strong investment in resource-based and labour-intensive activities, and were beginning to upgrade to middle-range technology products. This group included the second-tier Asian NIEs, but also isolated success stories from other regions, as well as the waking giants of China and (albeit more cautiously) India.

Enclave industrializers Some countries moved away from dependence on commodity exports by linking to international production chains, often by attracting large amounts of FDI and with heavy reliance on imported inputs and machinery. Export growth was often very fast, as in the Philippines, Mexico and, to some extent, Morocco. However, overall performance in terms of investment, value added and productivity growth was often quite weak.

Premature deindustrializers This group included most countries in Latin America, which had achieved a certain degree of industrialization but were unable to sustain a dynamic process of structural change through rapid

accumulation and growth. In a context of rapid liberalization, declining shares of manufacturing employment and output and a downgrading to less technology-intensive activities were common trends.

Commodity-dependent exporters Many poorer economies, particularly in sub-Saharan Africa, remained heavily dependent on one or two commodity exports. In the face of relatively stagnant markets, volatile prices and declining terms of trade, investment dropped further, diversification stalled and productivity remained stagnant. In some cases enclaves of faster export growth emerged in the extractive sectors, usually tied to FDI, but with weak linkages to the rest of the economy. However, some wealthier developing countries, notably Chile, did achieve a faster pace of investment and growth based on their natural resource endowments.

Conclusion

The historical experience of advanced countries, including recent graduates in East Asia, establishes that a broad and robust domestic industrial base remains key to successful development, because of its potential for strong productivity and income growth. Success hinges on building such a base from local conditions and overcoming local constraints on its expansion and evolution. This is still the big policy challenge for most developing countries. The experience of the 1980s and 1990s suggests that the policy direction launched in many parts of the developing world after the debt crisis failed to establish flourishing alternatives. Despite the widely shared belief that a more open economic environment would demonstrate the benefits of unrestricted capital mobility and the superiority of markets over government intervention, the period since the collapse of the Bretton Woods system has instead been marked by very heavily concentrated flows to comparatively wealthy countries and by an increasing incidence of financial crises, and their growing virulence in terms of lost output and jobs. Recent efforts to salvage the strategy have emphasized omissions: not enough good governance, not enough market access, not enough FDI or aid. Doing so continues to overlook the damage to macroeconomic growth fundamentals and to the direction of structural change that have accompanied the turn to neoliberal policy reform.

By contrast, success stories in the 1990s built steadily on improving performance established in the 1980s or earlier. In most cases they followed what Birdsall et al. (2005) have called 'heterodox gradualism', using an array of policy options to manage integration into the global economy and ensure that more of the value added linked to trade stayed at home. Doing so involved experimenting with a range of more strategic measures to encourage strong capital formation, expand domestic markets and support

technological upgrading. These measures have been rooted in specific institutional settings reflecting national political and social cultures and consistent with the bureaucratic and entrepreneurial capacities of local elites. In their different ways, all have eschewed a softening of the state and instead premised their structural transformation on a harder development state, that exhibits both 'adaptive efficiency' and the effective creation and utilization of 'policy space' (Kozul-Wright and Rayment, 2007). These same features will need to be recovered in countries that have fallen back since the debt crisis, if catch-up growth is to be reignited. This will require active policies, particularly on such matters as industrial support, technological progress and public infrastructure, all of which will have to be tailored to the particular circumstances of the countries concerned.

Notes

1. In an article deploring the slow progress in the Doha round of trade negotiations, the *Financial Times* journalist Martin Wolf described French President Chirac's scepticism as to the virtues of unlimited free trade as 'foolish, even depraved' since if the negotiations 'should fail, disorder alone should triumph', Martin Wolf, 'The World has Everything to Lose if Trade Liberalisation Fails', *Financial Times*, 2 November 2005.
2. Barro and Sala-i-Martin (1992) was a seminal study on the new growth empirics and convergence, and Sachs and Warner (1995) revived the links between openness and growth; for an assessment of their accounting exercises see Rodriguez (2007).
3. The Nobel Prize-winning economist Douglas North (1994) has candidly admitted that the aversion of neoclassical economics to historical processes and structural discontinuities precludes a proper understanding of the development process.
4. These regularities are associated with the empirical work of Verdoorn, Lewis and Kaldor. For a review see Toner (1999). See also Rodrik (2006) for a recent assessment of why manufacturing still matters to the development process.
5. See Levine and Renelt (1992); Ros (2000) and Bosworth and Collins (2004).
6. This conclusion applies, we believe, to much of the recent discussion of the importance of the 'investment climate' in developing countries.
7. Dowrick and Golley call for further research to understand these findings. Their own tentative suggestions are that the nature of technology transfer through MNCs has changed in the latter period and that the range of complementary policies that supported successful liberalization in the earlier period have gone missing in the 'one policy fits all' approach of the latter period. Both suggestions are in line with the arguments presented in this and subsequent chapters.
8. Certainly, the period of rapid export expansion in the advanced countries in the three decades after World War II coincided with strong wage growth and the development of deeper domestic consumer markets (Armstrong et al., 1984). However, intra-industry trade, strong productivity growth and high rates of capital formation were not the automatic outcome of market forces in these countries but were closely interrelated components of a politically fashioned socio-economic compromise.
9. For empirical evidence on this trend, see Maizels et al. (1998), UNCTAD (1999, 2002), and Maizels (2000). On the related idea of immiserizing growth see Kaplinsky and Morris (2002).
10. Such networks are not a new development, dating back to the 1960s in parts of East Asia, and becoming a more prominent feature of the international division of labour in the 1970s, see Helleiner (1973) and Henderson (1991).
11. See Gorg and Greenaway (2001), for a comprehensive review of the spillover literature. Also Glass et al. (1999) and Blomström and Kokko (2003).

12. The annual capital inflow in the 1990s was around 5 per cent of GNP, which was roughly the level prevailing in 1975–82. If China is excluded, the ratio is actually lower than in the earlier period by one percentage point.
13. Although the call for financial liberalization was heeded across the developing world in the 1990s, the majority of countries, and particularly those in Africa, attracted little private flows, and certainly not enough to offset declining aid flows during the 1990s. FDI was also very unevenly distributed: three-quarters of the total in the 1990s went to just ten emerging market economies, and China, Brazil and Mexico together took nearly one-half. Indeed, China accounted for about one-third of all FDI in the developing countries in the 1990s and about two-thirds of the total went to East Asia. Other developing regions – sub-Saharan Africa, South Asia, North Africa and the Middle East – have only marginal shares of the total and in the case of sub-Saharan Africa a falling one, mainly in extractive sectors or in public utilities.
14. For accounts of the origins of the Asian crisis see UNCTAD (1998, pp. 53–77), UNCTAD (2000), and also Chang et al. (2001) and Wade (2003).
15. The city states of Hong Kong and Singapore are fascinating stories in their own right. However, it should be noted that their per capita incomes in the early 1950s were comparable to Japan, and higher than much of the European periphery. Recognizing this goes a long way in dissociating their transformation from the current phase of globalization.
16. China's break with its earlier, highly erratic growth path occurred in the mid-1970s and has been building strength since the early 1980s. There are some significant differences between these two awakening giants, notably the much slower pace of urbanization in India, and with this a much slower pace of industrialization and capital accumulation. There is a good deal of dispute among economists as to which of the two development paths is likely to be the most sustainable.
17. See Freeman (2005) and Rowthorn (2006).
18. The recent pick-up in African growth rates is largely a result of an increase in commodity prices linked in no small part to growing demand in China. According to a recent IMF Working Paper, a dozen middle-income and oil-producing African countries have been the principal beneficiaries of a stronger growth performance since 1997, see Tahari et al. (2004).
19. Obviously in light of what has already been said it would be wrong to identify any hard and fast benchmark for all developing countries. Still, many regard a 6–8 per cent target as what is needed to tackle those deficits and visibly to close income gaps with the more developed countries.
20. Identifying potential growth rates, and the related challenge of suggesting growth targets, must be seen as a somewhat hit-and-miss business. But assuming that a country's labour force is growing on average at 2–3 per cent per year and that productivity growth needs to match that rate to maintain internal and external balance, then 5–6 per cent would be a minimum growth rate for GDP. Many, including ourselves, would regard a 7–8 per cent target as more appropriate if the policy objective is to make real progress in tackling the massive social deficits that have built up in most developing countries, and to start closing the income gaps with the more developed countries.
21. The Gini coefficient derives from the Lorenz curve which shows the cumulative share of the income received by the cumulative shares of the population, starting from the poorest income-receiving units. The coefficient measures the area between this curve and the diagonal of perfect equality and varies from 0 (maximum equality) to 1 (maximum inequality or from 0 to 100 when expressed in percentages. Milanovic's review of the literature notes that there is little disagreement on the degree of international inequality, with most studies finding a Gini coefficient in the range of 63 to 68, a figure that is higher than the coefficient in highly unequal countries such as Brazil or South Africa (Milanovic, 2006, pp. 140–42).
22. For critical assessments of the impact of these policies see Killick (1995), Mosley (1999), Przeworski and Vreeland (2000), Barro and Lee (2002), Vreeland (2003), Rodrik (2004), and Evrensel (2005).

23. See further UNCTAD (2003, pp. 74–6), also Ramirez and Namzi (2003).
 24. See Arbache et al. (2004) for review of evidence, also ILO (2004, pp. 40–45 and Akyüz et al. (2005).

Bibliography

- Adelman, I. (2000), 'The Role of Government in Economic Development', in F. Tarp (ed.), *Foreign Aid and Development*, London: Routledge.
- Agosin, M. and R. Mayer (2000), 'Foreign Investment in Developing Countries: Does it Crowd in Domestic Investment?' UNCTAD discussion paper no. 146, United Nations Conference on Trade and Development, Geneva.
- Ahluwalia, M. (1999), 'The IMF and the World Bank: Are Overlapping Roles a Problem?' *International Monetary and Financial Issues for the 1990s*, Vol. 11. United Nations publication, New York and Geneva: UNCTAD.
- Akyüz, Y. (2004), 'Trade, Growth and Industrialisation in Developing Countries: Issues, Experience and Policy Challenges', Penang, Malaysia, UNDP Asia-Pacific Conference on Trade.
- Akyüz, Y. (2006), 'Reforming the IMF: Back to the Drawing Board', G-24 Discussion Paper Series, no. 38, Geneva, UNCTAD.
- Akyüz, Y. and C. Gore (1996), 'The Investment-Profits Nexus in East Asian Industrialization', *World Development*, 24 (3).
- Akyüz, Y., H. Flassbeck and R. Kozul-Wright (2005), 'Globalization, Inequality and the Labour Market', Mimeo, UNCTAD, Geneva.
- Amsden, Alice (2001), *The Rise of the 'Rest': Challenges to the West from Late Industrializing Economies*, New York: Oxford University Press.
- Arbache, Jorge S., Andy Dickerson and Francis Green (2004), 'Trade Liberalisation and Wages in Developing Countries', *The Economic Journal*, 114.
- Armstrong, P., A. Glyn and J. Harrison (1984), *Capitalism Since World War II: The Making and Break-up of the Great Boom*, London: Fontana.
- Barro, R. and J.W. Lee (2002), 'IMF Programs: Who is Chosen and What are the Effects?' NBER Working Paper Series, no. 8951, May.
- Barro, R., and X. Sala-i-Martin (1992), 'Convergence', *Journal of Political Economy*, 100 (2): 223–51.
- Baumol, W. et al. (1991), *Productivity and American Leadership*, Cambridge, MA: MIT Press.
- Berger, S. and R. Dore (1996), *National Diversity and Global Capitalism*, Ithaca, NY: Cornell University Press.
- Bernard, A. and C. Jones (1996), 'Comparing Apples to Oranges: Productivity Convergence and Measurement across Industries and Countries', *American Economic Review*, 86 (5): 1216–38.
- Berry, A. and J. Serieux (2004), 'All about the Giants: Probing the Influences of World Growth on Income Distribution at the End of the 20th Century', *CESifo Economic Studies*, 50 (1): 133–70.
- Bhaskar, V. and Glyn, A. (1992), 'Investment and Profitability: The Evidence from the Advanced Countries', Economic Series Working Papers 99144, University of Oxford, Department of Economics.
- Birdsall, Nancy, Dani Rodrik and Arvind Subramanian (2005), 'How to Help Poor Countries', *Foreign Affairs*, July/August.
- Blomström, M. and A. Kokko (2003), 'Human Capital and Inward FDI', CEPR Working Paper, No. 167, CEPR, London.
- Bosworth, B. and S. Collins (2004), 'The Empirics of Growth: An Update', Brookings Institute, Washington, September.
- Braunstein, E. and G. Epstein (2004), 'Bargaining Power and Foreign Direct Investment in China: Can 1.3 billion Consumers Tame the Multinationals?' in Will Milberg (ed.), *Labor and the Globalization of Production: Causes and Consequences of Industrial Upgrading*, New York: Palgrave, pp. 209–48.
- Camdessus, M. (1997), 'Global Capital Flows: Raising the Returns and Reducing the Risks', speech to Los Angeles World Affairs Council, 17 June.

- Carcovic, M. and R. Levine (2002), 'Does Foreign Direct Investment Accelerate Economic Growth?' University of Minnesota, Minneapolis, MN, May.
- Chang, H.-J., G. Palma and H. Whittaker (eds) (2001), *Financial Liberalization and the Asian Crisis*, London: Palgrave.
- Chudnovsky, D. and A. López (2002), 'Estrategias de las empresas transnacionales en la Argentina en los años 1990', *Revista de la Cepal*, no. 76.
- Crotty, J. (1990), 'Owner-Manager Conflict and Financial Theories of Investment Instability: A Critical Assessment of Keynes, Tobin and Minsky', *Journal of Post Keynesian Economics*, Summer.
- Darity, W. and L. Davis (2005), 'Growth, Trade and Uneven Development', *Cambridge Journal of Economics*, 29 (1): 141–70.
- Dowrick, S. and J. Golley (2004), 'Trade Openness and Growth: Who Benefits?' *Oxford Review of Economic Policy*, 20 (Spring): 38–56.
- ECLAC/UNCTAD (2002), 'Summary and Conclusions', Joint Regional Seminar on FDI policies in Latin America, Santiago, Chile, 7–9 January.
- Eichengreen, B. (2002), *Financial Crises: And What to Do About Them*, Oxford: Oxford University Press.
- Evrensel, A. (2005), 'IMF Programmes in Emerging Countries', *Comparative Economic Studies*, 47.
- Freeman, R. (2005), 'What Really Ails Europe (and America): The Doubling of Global Workforce', *The Globalist*, 3 June.
- Glass, A. et al. (1999), 'Linkages, Multinationals, and Industrial Development', Working Paper No. 99–16, Ohio State University, Department of Economics, August.
- Gorg, H. and D. Greenaway (2001), 'Foreign Direct Investment and Intra-Industry Spillovers', paper prepared for the UNECE/EBRD Expert Meeting Financing for Development, Geneva, 3 December.
- Gosh, A. (2004), 'Capital Inflows and Investment in Developing Countries', Employment Strategy Papers, November, International Labour Organisation, Geneva.
- Harrison, A. and M. Macmillan (2003), 'Does Direct Foreign Investment Affect Domestic Credit Growth?' *Journal of International Economics*, 61 (1): 73–100.
- Hausmann, R., L. Pritchett and D. Rodrik (2004), 'Growth Accelerations', NBER Working Paper No. 10566, Cambridge, National Bureau of Economic Research.
- Heintz, J. (2003), 'The New Face of Unequal Exchange: Low-Wage Manufacturing, Commodity Chains and Global Inequality', PERI Working Paper Series No. 59, Political Economy Research Institute, University of Massachusetts, Amherst, MA.
- Helleiner, G. (1973), 'Manufacturing Exports from Less Developed Countries and Multinational Firms', *Economic Journal*, 83.
- Helleiner, G. (2000), 'Markets, Politics and Globalization: Can the Global Economy be Civilised?' Tenth Raúl Prebisch Lecture, United Nations Conference on Trade and Development, Geneva.
- Henderson, J. (1991), *The Globalisation of High Technology Production*, London: Routledge.
- Hirschman, A. (1958), *The Strategy of Economic Development*, New Haven, CT: Yale University Press.
- ILO (2004), *A Fair Globalization: Creating Opportunities for All*, Geneva: ILO.
- Imbs, J. and R. Wacziarg (2000), 'Stages of Diversification', CEPR discussion paper no. 2642, Centre for Economic Policy Research, London.
- Kahn, M. and Jomo, K.S. (eds) (2002), *Rents, Rent-seeking and Economic Development: Theory and Evidence in Asia*, New York: Cambridge University Press.
- Kaplinsky, R. and M. Morris (2002), 'The Globalization of Product Markets and Immiserising Growth: Lessons from the South African Furniture Industry', *World Development*, 30 (7): 1159–78.
- Kenny, C. and D. Williams (2001), 'What Do We Know about Economic Growth? Or, Why Don't We Know very Much?', *World Development*, 29 (1): 1–22.
- Killick, T. (1995), *IMF Programmes in Developing Countries: Design and Impact*, London: Routledge.
- Kindleberger, C. (1984), *Manias, Panics and Crashes*, New York: Basic Books.

- Kozul-Wright, R. and P. Rayment (2007), *The Resistible Rise of Market Fundamentalism: Rethinking Development Policy in an Unbalanced World*, London: Zed Books.
- Krugman, P. (2007), 'Who was Milton Friedman?', *New York Review of Books*, 54 (2), 15 February.
- Kuznets, S. (1971), 'Modern Economic Growth: Findings and Reflections', Nobel Lecture, Oslo, December.
- Levine, R. and D. Renelt (1992), 'A Sensitivity Analysis of Cross-Country Growth Regressions', *American Economic Review*, September.
- Lucas, R. (2000), 'Some Macroeconomics for the 21st Century', *Journal of Economic Perspectives*, 14 (1): 159–68.
- Magalhães Prates, Daniela and Leda Maria Paulani (2007), 'The Financial Globalization of Brazil under Lula', *Monthly Review*, 58 (8).
- Maizels, A. (2000), 'The Manufactures Terms of Trade of Developing Countries with the United States (1981–1997)', Working paper 36, Queen Elizabeth House, Oxford University.
- Maizels, A., T. Palaskas and T. Crowe (1998), 'The Prebisch–Singer Hypothesis Revisited', in D. Sapsford and J. Chen (eds), *Development Economics and Policy*, London: Macmillan Press.
- Mankiw G. (1995), 'The Growth of Nations', *Brookings Papers on Economic Activity*, 1.
- Milanovic, B. (2002), 'True World Income Distribution 1988 and 1993: First Calculations based on Household Surveys Alone', *Economic Journal*, 112 (476).
- Milanovic, B. (2005), 'Why Did the Poorest Countries Fail to Catch Up?' Carnegie Papers no. 62, November, Carnegie Endowment for International Peace, Washington, DC.
- Milanovic, B. (2006), 'Global Income Inequality: What It Is and Why It Matters', Policy Research Working Paper Series 3865, World Bank.
- Milanovic, B. and S. Yitzhaki (2001), 'Decomposing World Income Distribution: Does the World Have a Middle Class?' World Bank Policy Research Working Paper no. 2562, February, World Bank, Washington, DC.
- Mishkin, F. (2006), *The Next Great Globalization: How Disadvantages Nations Can Harness Their Financial Systems to Get Rich*, Princeton, NJ: Princeton University Press.
- Mody, A. (2004), 'Is FDI Integrating the World Economy?' *World Economy*, 27 (8): 1195–1222.
- Mosley, P. (1999), 'Globalization, Economic Policy and Growth Performance: International Monetary and Financial Issues for the 1990s', UNCTAD. New York and Geneva: United Nations.
- Myrdal, G. (1957), *Economic Theory and Underdeveloped Regions*, London: Duckworth.
- North, D. (1994), 'Economic Performance Through Time', *American Economic Review*, 83 (3).
- Ocampo, J.A. (2002), 'The Lost Half-Decade', WIDER Angle, Helsinki: WIDER, No. 2.
- Ocampo, J.A. and M.A. Parra (2005), 'Explaining the Dual Divergence: Successes and Collapses in the Developing World since 1980', background paper for World Economic and Social Survey, New York, UNDESA.
- Ocampo, J.A. and L. Taylor (1998), 'Trade Liberalisation in Developing Economies: Modest Benefits but Problems with Productivity Growth, Macro Prices, and Income Distribution', *Economic Journal*, 108, (450).
- Plender, J. (2003), *Going off the Rails. Global Capital and the Crisis of Legitimacy*, Chichester: Wiley.
- Przeworski, A. and J. Vreeland (2000), 'The Effects of IMF Programs on Economic Growth', *Journal of Development Economics*, 62.
- Quah, D. (1996), 'Twin Peaks: Growth and Convergence in Models of Distribution Dynamics', *Economic Journal*, 106.
- Ramirez, M. and N. Namzi (2003), 'Public Investment and Economic Growth in Latin America: An Empirical Investigation Test', *Review of Economic Development*, 17 (1).
- Reati, A. (2001), 'Total Factor Productivity: A Misleading Concept', *BNL Quarterly Review*, 218 (September).
- Rodriguez, F. (2007), 'Openness and Growth: What Have We Learned?' DESA Working Paper No. 51, United Nations, New York.
- Rodriguez, F. and D. Rodrik (1999), 'Trade Policy and Economic Growth: A Skeptic's Guide to the Cross-National Evidence', NBER Working Paper no. 7081, National Bureau of Economic Research, Cambridge, MA.

- Rodrik, D. (1999), *The New Global Economy and Developing Countries: Making Openness Work*, Washington, DC: Overseas Development Council.
- Rodrik, D. (2002), 'Feasible globalizations', Mimeo, Kennedy School of Government, Harvard University, Cambridge, MA, July.
- Rodrik, D. (2003), 'Growth Strategies', NBER Working Paper Series no. 10050, October.
- Rodrik, D. (2004), 'Rethinking Growth Policies in the Developing World', Luca D'Agliano Lecture in Development Economics, Turin, October.
- Rodrik, D. (2006), 'Industrial Development: Stylized Facts and Policies', Mimeo, Kennedy School of Government, Harvard University, Cambridge, MA.
- Rogoff, K. (1999), 'International Institutions for Reducing Global Financial Instability', *Journal of Economic Perspectives*, 13 (Fall): 21–42.
- Rogoff, K., A. Kose, E. Prasad and S.-J. Wei (2006), 'Financial Globalization: A Reappraisal', NBER Working Paper no. 12484, National Bureau of Economic Research, Cambridge, MA.
- Ros, J. (2000), *Development Theory and the Economics of Growth*, Ann Arbor, MI: University of Michigan Press.
- Rowthorn, R. (2006), 'The Renaissance of China and India: Implications for the Advanced Economies', Discussion Paper No. 182, UNCTAD Geneva, October.
- Rowthorn, R. and J. Wells (1987), *Deindustrialisation and Foreign Trade*, Cambridge: Cambridge University Press.
- Sachs, J. and A. Warner (1995), 'Economic Reform and the Process of Global Integration', *Brookings Papers on Economic Activity*, 1.
- Sala-i-Martin, X. (1996), 'The Classical Approach to Convergence Analysis', *Economic Journal*, 106 (437): 1019–36.
- Schneider, F. (2002), 'Size and Measurement of the Informal Economy in 110 Countries around the World', paper presented at an Australian National Tax Centre workshop, Canberra, 17 July.
- Singh A. (1999), 'Savings, Investment and the Corporation in the East Asian Miracle', in Y. Akyüz (ed.), *East Asian Development – New Perspectives*, London: Frank Cass.
- Stiglitz J. (2002), *Globalisation and its Discontents*, London: Allen Lane.
- Stiglitz, J. (2005), 'The Ethical Economist', *Foreign Affairs*, November/December.
- Sutcliffe, R. (2006), 'A Converging or Diverging World?' in Nomo K.S. (ed.), *Flat World, Big Gaps*, London: Zed Books.
- Svennilsson, I. (1954), *Growth and Stagnation in the European Economy*, Geneva: United Nations.
- Tahari, A. et al. (2004), 'Sources of Growth in sub-Saharan Africa', IMF Working Paper 176, International Monetary Fund, Washington, DC.
- Toner, P. (1999), *Main Currents in Cumulative Causation: The Dynamics of Growth and Development*, London: Macmillan.
- UNCTAD (1994), *Trade and Development Report 1994*, New York and Geneva: United Nations.
- UNCTAD (1997), *Trade and Development Report 1997*, New York and Geneva: United Nations.
- UNCTAD (1998), *Trade and Development Report 1998*, New York and Geneva: United Nations.
- UNCTAD (1999), *Trade and Development Report 1999*, New York and Geneva: United Nations.
- UNCTAD (2000), *Trade and Development Report 2000*, New York and Geneva: United Nations.
- UNCTAD (2002), *Trade and Development Report 2002*, New York and Geneva: United Nations.
- UNCTAD (2003), *Trade and Development Report 2003*, New York and Geneva: United Nations.
- UNCTAD (2006a), *Economic Development in Africa: Performance, Prospects and Policy Issues*, New York and Geneva: United Nations.
- UNCTAD (2006b), *Trade and Development Report 2006*, New York and Geneva: United Nations.

- UNDESA (2005), *World Economic and Social Survey 2005*, New York: United Nations.
- UNDESA (2006), *World Economic and Social Survey 2006*, New York: United Nations.
- Vreeland, J. (2003), *The IMF and Economic Development*, Cambridge: Cambridge University Press.
- Wacziarg, R. and K. Welch (2003), 'Trade Liberalization and Growth: New Evidence', NBER Working Paper no. 10152, National Bureau of Economic Research, Cambridge, MA.
- Wade, R. (2003), 'Preface', *Governing the Market*, new edn, Princeton, NJ: Princeton University Press.
- Wieczorek-Zeul, H. (2004), 'Statement to Development Committee', IMF–World Bank meetings, Washington, DC, 2 October.
- Winters, L.A. (2004), 'Trade Liberalization and Economic Performance: An Overview', *Economic Journal*, 114 (February): F4–F21.
- World Bank (1995), *Global Economic Prospects and Development Countries*, Washington, DC: World Bank.
- World Bank (2002), *Globalization, Growth and Poverty: Building an Inclusive World Economy*, Oxford: Oxford University Press.
- World Bank (2003), *Global Economic Prospects*, Washington, DC: World Bank.
- World Bank (2005), *Economic Growth in the 1990s, Learning From a Decade of Reform*, Washington, DC: World Bank.
- Young, A. (1928), 'Increasing Returns and Economic Progress', *Economic Journal*, 37 (December).

65 Latin America and the Caribbean

Albert Fishlow

Introduction¹

Economic development in Latin America, finally, seems on the verge of beginning again. After close to 25 years of poor performance, despite dramatic policy changes in fiscal policy, and hence inflation, and expansion in international trade, the region seems to be expanding. Results both in 2004 and 2005 have been better than initially foreseen. The key question is whether this process will continue over the next decades, or whether the current advance will soon collapse – as it has done in the past – with declining commodity prices, limited foreign demand for exports, less foreign investment and more inflationary domestic policy.

Much has altered within the region's economies since the 1990s. Domestic politics have also changed profoundly, as democracy has continuously spread and deepened. Yet a high degree of political dissatisfaction has evolved, and successive elections have accurately reflected the population's unhappiness. Although inflation rates are now virtually at a record low, benefiting substantially the bottom third of the income distribution, people do not vote for those that promise fiscal discipline. The Left has run well, and won, virtually everywhere since the beginning of the 21st century. But, with the exception of Hugo Chavez, President of Venezuela, and possibly Nestor Kirchner, President of Argentina, policies have remained responsible and even responsive to the international markets. Evo Morales's election in Bolivia introduces a new element, control over the trade in drugs, while recapitulating once more the question of state control over energy resources that has grown in significance in recent years with the rising price of petroleum in world markets.

These contemporary events, both economic as well as political, are better understood within a historical context. Latin America was a latecomer to continuous economic expansion. Only since the last decades of the nineteenth century did economic growth become a regular process. That was the era when the region began its increasing participation in external markets, taking advantage of its natural resources, and simultaneously sought to implement, late, its own industrial revolution. That is where we start the story in the next section. It extends through the Great Depression, with its dramatic decline in external demand, up to the recovery of the world economy at the end of World War II.

Next, we analyze the rise and decline of import substitution in the three following decades until the debt crisis that began in 1982. That history simultaneously comprehends a surge of military dictatorships within virtually the entire region. At the same time, that experience encompasses a period of above-average growth in Latin America. Finally, this interval also represents growing dependence on external finance to finance the increased domestic investment that drove expansion. This combination proved too fragile to survive the successive oil shocks of 1973 and 1979.

In the next section we take up the economic performance of the region during a subsequent period of profound retrenchment, altering the previous large role of the state almost everywhere. Fiscal deficits were sharply curtailed, much privatization of state-owned activities occurred, and protection against international imports was sharply curtailed. Latin America in the 1990s gave signs of joining Asia and Eastern Europe in rapid expansion. But economic development soon proceeded much less rapidly than had been hoped, giving rise to increasing criticism of the Washington Consensus and sparking a clear move to the political Left that continues today.

A final part takes up the central problems and policy issues that confront virtually all the countries of the region today, in the midst of a record number of presidential elections. Politics and economics are intrinsically linked. That combination sometimes leads to an inability to focus effectively upon the longer run. But Latin America needs to continue to change if it is to keep up with global advance. It requires more and better educational and health outlays, both to improve the quality of the labor force as well as to improve the lamentable degree of inequality that plagues almost all countries. Macroeconomic reform has to continue, both in governmental tax collection as well as more efficient expenditure. And that subject necessarily includes the indispensable increase in domestic savings and investment. A third matter is the continuing increase, as well as diversification, of export earnings and thus reduced sensitivity to variation in external demand. Finally, the region will have to recognize the continuity of present economic policy as a virtue rather than a vice, even as the electoral process becomes more inclusive and civil society flourishes. Institutional change, making permanent these structural advances, requires continuing attention and additional major effort.

A brief conclusion extends the lessons of the Latin American and Caribbean experience to the broader agenda of contemporary development economics.

Boom, bust and war recovery: 1900 to 1945

Latin America, despite the achievement of political independence early in the nineteenth century, missed out on the initial Industrial Revolution. As

a consequence, overall, its standard of living probably improved little between 1820 and 1870. Those were years of political instability and modest engagement in international trade. Only during the period 1870 to 1913, when an initial era of globalization fueled economic activity, did the region expand its relative export position, helped by an increasing number of immigrants and much greater foreign investment. That investment underwrote great expansion of the rail infrastructure and construction of port facilities to facilitate export of wheat and meat, coffee, sugar, rubber, nitrates and other primary products demanded by the United States and Europe.

Such were the mounting receipts from exports after 1900 that not only Argentina returned to the gold standard, but Brazil and Chile also joined, then still a signal of successful developing-country adaptation within the expanding world economy. The region impressively focused on exporting primary products – something of the order of 97 percent of exports were classified as such in 1913 – even while high levels of tariff protection assured a growing market for domestic industrial production. Import substitution generally began in the larger countries before the Great Depression of the 1930s, focusing on the elemental, but substantial, areas of foodstuff and textile and clothing production.

What is important to note is the great difference in incomes among countries of the region just before World War I. Data are presented in Table 65.1. Per capita income in 1913 in Argentina was comparable with that of Western Europe as a whole, and some three-quarters of that in the United States. Chile followed, about a third below Argentina's pace-setting level. The rest of the region trailed much farther behind. At the very bottom was Brazil, burdened by the poor north-east that offset the rising, and more diversified, economy to the south.

These regional income differences reflected a variable commitment to exports, as can easily be seen in the same table. Argentina and Chile were the leaders in their share of income generated by exports. They were, as well, producers of non-tropical exports more in competition with developed countries than the products from other parts of the region, and hence, as W.A. Lewis has shown, able to gain higher incomes.² Both of these countries also benefited from high levels of capital inflow from abroad, contributing to higher levels of domestic investment. Finally, the labor force in the case of both countries was more literate than elsewhere, and the movement toward local non-agricultural activities was equally more substantial. Bulmer-Thomas runs a regression for 13 countries, relating economic growth to exports over this interval, and finds a high degree of explanation; all seven of the countries separated correspond well.³

Table 65.1 *Latin America before World War I*

	Per capita income 1913 1990 Geary–Khamis \$	Growth per capita 1900–1913 in %	Exports as % of income	Literacy %	% Labor force agriculture
Argentina	3797	38	0.36	63	34.2
Brazil	839	19	0.3	35	66.7
Chile	2653	36	0.54	56	37.7
Colombia	1236	27	0.13	41	70.5
Mexico	1467	27	0.17		63.7
Peru	1037	27	0.24	27	
Venezuela	1104	34		30	72
Seven countries	1134	39			
Relative to USA	0.28				

Sources: Per capita income and growth: Angus Maddison, *Monitoring the World Economy*, OECD Development Centre, 1995
Exports as % of income, literacy and % labor force: *An Economic History of Twentieth Century Latin America*, ed. Rosemary Thorp, Enrique Cardenas and Jose Antonio Ocampo 2000, Vol 1, p 26

The World War in 1914 had variable effects within the region.⁴ No longer is it viewed as an unmixed blessing. The sudden reduction of imports that the conflict imposed did offer opportunities for domestic producers in several countries in the industrial sector. Yet capacity for production of capital goods was still rudimentary, and that constraint limited possibilities for significant output gains. Foreign investment from Europe ceased, and that of the United States was still directed primarily to the north of the region. Argentina, in particular, disappointed. With lower imports almost everywhere, governmental receipts flagged, and inflation began to expand in many countries. What helped was the relative brevity of the conflict, allowing many of the countries to go back to what they had been doing before: exporting their primary products.

Thus the 1920s featured a return to previous history. Imports of manufactured products went up, as a consequence of rising primary exports. Even Chile again managed to expand its sales of nitrates by 1929 to three-quarters its 1913 real level. National performance, as before, was related closely to success in international trade. There was a new entrant with much larger income upon the stage, Venezuela, whose petroleum exports found an expanding market within the United States. Colombia and Peru also improved their positions.

But there were also some important differences. Two stand out. First was the reality of major external change.⁵ Globalization, and British leadership, was a phenomenon of the past. There was increasing financial dominance of the United States in the world, and also at the regional level. Much of the renewed surge in investment flows went to local and state sources instead of, as earlier, to national governments. As it would soon turn out, the new American investment banks competing for business were more effective in finding a supply of finance than identifying productive applications of the money. The weakened position of Europe had consequences not only for investment, but also for international trade. The gold standard had vanished, leaving individual countries confronting hyperinflation and internal political disruption. The League of Nations, an attempt at supranational economic guidance, provided little assistance in the midst of these new circumstances.

Second, the renewed rise in export markets in the 1920s was weaker than it had earlier been. Countries varied in their ability to cope. Brazil, for example, despite losing out in exports of rubber and facing a slowing market for coffee, managed to continue a growth predicated upon continuing diversification of its industrial structure. Others, like Argentina and Chile, saw lesser rises in expansion in the period from 1913 to 1929 than they had previously. This modest deviation from the historic export-oriented model of growth was, of course, soon to become the beginning of a new commitment to import substitution after the start of the Great Depression.

Much of Latin America definitively changed its development style in the 1930s. There was little option. Prices of the region's raw material exports plunged with declining demand. Quantities also fell. This time, unlike the earlier decline after World War I, recovery did not occur within the industrial center. As a result, it was necessary to cease full payment of interest on the public debt, which had much risen as a proportion of export earnings. Exchange rate controls were applied to limit imports of consumption goods capable of being supplied internally. Real exchange rate devaluations occurred almost everywhere. They most frequently took the form of multiple exchange rates enhanced by direct controls.

The deterioration of the balance of payments meant a simultaneous decline in government revenues: some countries were still dependent upon import duties and export taxes for as much as half of their total receipts. As a natural consequence government deficits showed a tendency to rise, financed through increases in the money supply, and thereby stimulating domestic demand. Inflation did not increase much as domestic supply instead responded well to the lack of import competition. Special circumstances, such as the potential Brazilian civil war with a seceding Sao Paulo

in 1932, or the Leticia War between Colombia and Peru in 1933–34, occasionally helped in establishing such proto-Keynesian policies and contributed to earlier and stronger recovery.

With a rise in export earnings after 1932 for most countries, rising imports of intermediate and capital goods became available to satisfy increasing demand for such inputs as could not be supplied locally. These sometimes consisted of used machinery rather than that newly produced. Labor migrated to the cities to meet the demands. For this period, unlike the later decade of import substitution in the 1950s, production increases were dependent upon increased labor, rather than capital, input. In some instances, extra shifts were employed to satisfy an expanding market. Noteworthy is the greater rise in industrial output than in gross domestic product that occurred virtually everywhere in the region.⁶

Such expansion was not without a negative side. Military leaders emerged in many countries in response to the new economic challenges being confronted. Constitutions were rewritten, or reinterpreted. Authoritarianism had a preview before its re-enactment in the 1960s and 1970s. There was an increasing degree of state intervention and regulation, not always efficiently. Markets were not again to operate in an unfettered fashion as they had somewhat done during the years before 1930.

With the approach of a new war in the late 1930s, economic conditions began to worsen slowly. Current dollar trade growth reversed after 1937, giving further stimulus to domestic production where it was possible. When the European war emerged in 1939, Latin America, like the United States, stayed out. With Pearl Harbor in December 1941, and United States engagement, several of the countries declared war. But only Brazil was actively engaged in the conflict, with a detail of some 25 000 troops. Indeed, the southern cone countries, with a long and strong German influence upon their military, were late in their formal espousal of the Allied side. That indecision did not affect their trade flows much. Between 1940 and 1945 there was a major reversal: all countries suddenly found their exports dominantly flowing northward to the United States, with portions going to other neighboring countries within the region.

The war saw slower expansion in Latin America as a result of much lesser imports. Reserves accumulated, substantially in some countries. Inflation too accelerated domestically, leading to real appreciation of unchanged nominal exchange rates. When peace returned, repressed demand and favorable prices led to substantial import flows. There was now a domestic sector in many countries resistant to such foreign competition and a return to the experience of the 1920s. Not surprisingly, within the region a new group of young economists emerged calling for a conscious strategy of import substitution. They soon took form around the Economic

Table 65.2 *Latin American per capita income (1990 Geary–Khamis dollars)*

	1913	1929	1950
Argentina	3797	4367	4987
Brazil	839	1106	1676
Chile	2653	3396	3827
Colombia	1236	1505	2089
Mexico	1467	1489	2085
Peru	1037	1619	2263
Venezuela	1104	3426	7424
Regional average	1575	1929	2614
Relative to USA	0.29	0.28	0.27

Source: Angus Maddison, *Monitoring the World Economy*, OECD Development Centre, 1995

Commission for Latin America (CEPAL) in Santiago under the leadership of Raul Prebisch.⁷ Their views were given powerful reinforcement by the failure of the Latin American countries to obtain their own Marshall Plan, as had the European countries, thereby assuring integration into the wider global economy.

Table 65.2 sets out the growth experience that the principal regional countries had achieved since 1913. What these data reveal are three characteristics. First is the extraordinary rise in income per capita of Venezuela: it moves from a third of the Argentine level in 1913 to one that exceeds it by half as much again. Note as well that the second-largest rate of advance is attained by Peru. That provides the basis for those countries' choice not to move to import substitution. Second is the significant slowing in the performance of Argentina and Chile, whose leadership had been so predominant earlier on. Third is the relative rise of Brazil, Colombia and Mexico as a result of the move toward domestic industrial production.

But, ultimately, and alas, there is the unfortunate regional relative aggregate. It moves from 29 percent of the per capita level of the United States in 1913 to 27 percent in 1950. Even with the better performance of Latin America during the 1930s, the rapid recovery of the United States during the war years and immediately thereafter gave proof to the continuing circumstance of regional backwardness.

Post-war growth

Latin America, or at least a goodly number of countries in the region, launched itself toward a continuation of industrialization, impelled by a

more active state policy in the 1950s. Helped by the rise in primary product prices in the wake of the Korean War, and thus sufficient export receipts, countries – both large and small – imposed high tariffs, quotas and even multiple exchange rates, to permit giving substantial incentives, and subsidies, to the intermediate and capital goods sectors where imports retained a large percentage of the domestic market. As the world gradually moved away from the limitations to world commerce imposed during the Depression, Latin American countries largely ignored the opportunities afforded for greater diversification as well as market expansion of their traditional export products.

In this attempt, most countries badly failed, including the two regional leaders, Argentina and Chile. Three countries, Mexico, Brazil and Colombia, managed the process better during the 1950s. They had started farther behind, with large internal markets, and managed to focus on less-inefficient undertakings. They also benefited from lesser balance-of-payments shocks during the expansion of the 1950s, assisted by more foreign investment. Two countries, Venezuela and Peru, continued their past emphasis on petroleum and mineral exports, respectively, and emerged well.

What characterized the CEPAL model were three marked deviations from standard economic theory. First, the state role in the process of economic performance was much enhanced. Priorities were a matter of public policy: planning groups were established, production was directly encouraged, special tariff protection was granted, subsidies appeared, and so on. Matters during the 1950s in Latin America were less formal than the Indian Mahalanobis scheme of targeted expansion, or the planned efforts of the Soviet Union and its followers in Eastern Europe. Nonetheless, they represented new and important interventions, then and subsequently, designed to achieve desired results within the industrial sector.

Second, at the macroeconomic level, there was clear understanding of the recurrent need for external support of the balance of payments. International markets would not by themselves, in the short run, absorb a sufficient quantity of primary exports to support the large quantity of imports of intermediate and capital goods required to allow import substitution to proceed. Neither, in the midst of the Marshall Plan and World Bank efforts directed toward Western Europe, were there public funds available. Foreign direct investment, although frequently criticized as part of the greater degree of nationalism characteristic of the period, helped in two ways: it provided the needed finance, but also a direct access to the newer technology needed for more advanced stages of industrialization to proceed. This tension was to continue through subsequent years.

Third, at the microeconomic level, all efforts to enhance industrial expansion were provided. Public investment in urban centers occurred to

meet the needs of a rapidly expanding population, many now coming from the rural areas. New highways were built, as was occurring in the United States and Europe, to extend the transportation network. Older railways, established to facilitate primary exports, languished. The national political balance was frequently transformed in a variety of countries as this new effort was launched: labor unions became more relevant, and wage policy frequently surged to the national level for decision. At this time, minimum wages were a new instrument of government policy.

These characteristics, and their strong element of directing the market, rather than reacting to it, stood in sharp contrast to the capitalist style of earlier export orientation. But the import-substitution model also gave rise to important internal contradictions that very soon – almost within a decade – represented the beginning of the end.

One of the problems with the strategy was exactly its focus upon encouraging domestic production to substitute for continuing imports. Alas, that necessarily required still other imports of capital goods and other inputs in order to accomplish it. So net savings of foreign exchange were not as great as initially were conceived. Frequently, imports were initially permitted as an incentive to establish the industry domestically. Moreover, to ensure that those imported inputs were inexpensive, an overvalued exchange rate was the rule. In turn, this ensured reliance only on export of the basic primary commodities that had been present from the start. There was no motivation, nor was there initial efficiency, enabling exports to emerge from the new manufacturing firms that had been established. Instead, there was great emphasis on import protection, equivalent to a tax on the few agricultural exports able to compete. What saved the day, as noted above, was foreign direct investment and thus greater dependence on the outside world instead of the independence so much vaunted.

Another difficulty encountered was a rising government deficit, soon resulting in higher rates of inflation. Government expenditures and employment rose, while taxes lagged behind. An increased money supply was the solution. A debate ensued in most countries, between the structuralists – seen to be progressive – and the monetarists – conservative and crotchety. It is true that one can transfer some resources to the government sector through an inflation tax of limited amount. At that time, advocates of a bolder governmental strategy were speaking only of something like price increases of 10 to 15 percent, a rate that was soon exceeded in most cases. The problem was that the process of inflation acceleration was auto-generated. Continuously larger price increases – to make possible the needed transfer – were necessary as the public reduced their holdings of money in order to limit the loss of income.⁸ Another, and important, consequence was deterioration of the income distribution. Those at the bottom

of the income distribution, half and even more of the population, were affected negatively; they had no alternative but to use cash whose value began to deteriorate more rapidly.

Additionally, the very success in establishing industry had its negative counterpart in the lack of expansion of traditional sectors. Agriculture and mining lagged behind, unaccompanied by application of new technology or capital applied to enhance productivity. Because industry was capital-intensive, the growing labor force could not find increased employment there, but rather in urban services and governmental employment. Additionally there were pressures placed on agricultural producers to keep prices of foodstuffs low for the expanding urban areas. A sectoral imbalance emerged, one that prejudiced policy in many countries. Argentina is perhaps the best example of this, with political appeals to the military to prevent relative deterioration of agriculture, whose large exports of foodstuffs were regularly impeded and reallocated to domestic consumption. Everywhere, internal relative prices favored industry and penalized agriculture.

By the end of this first decade, even CEPAL could see that its initial hopes were unrealized, and that some change in policy was necessary.⁹ Raul Prebisch opted for a slightly altered strategy, focusing this time upon import substitution at the regional level. This, by allowing trade with neighbors, extended the basis for a viable industrialization through economies of scale. Despite the creation of a Latin American Free Trade Area in 1961, that undertaking made scant progress at the time. There were ambitious, but fortunately unrealized, plans for new geographic specialization in a variety of industries. Later, in 1967, the United States even accepted waiver of the 'most favored nation' principle: greater protection could be legitimately applied against its exports than those of Latin American neighbors.

Another option, emerging from the Left, was for much more fundamental internal reform, and less reliance on trade with the outside world. This view, in a few years to emerge more fully in the guise of dependency theory, saw the problem as a continuing commitment to the market system. There was too little focus on wide-ranging national reforms. Deficiencies of the import-substitution period began to be recognized. But the solution was now to come through greater centralization and state engagement, not exactly of the Soviet type, but more sympathetic to planning as well as more committed to elimination of rank income inequality.

A third alternative, and the one actually pursued, but only for a short time, was the Alliance for Progress, joined by the new Inter-American Development Bank. Both were to be new sources of public foreign capital, capable, together with private investment, of carrying Latin America to a stage of higher rates of economic growth and internal reform: land

redistribution, a new more equitable tax code, commitment to wider education at the basic level, and so on. Planning agencies were made universal, and approval of national plans was necessary to receive resources. Large groups of technical personnel made their way to Latin America to offer assistance. The region, atypically and suddenly, was momentarily at the center of attention.

This bold notion of democratic reform throughout the hemisphere was a United States response to the Cuban Revolution. It had barely begun its operation before another option soon gained force and spread through much of the region. That was explicit military intervention, first apparent in the case of Brazil in 1964, but rapidly spreading to other countries thereafter. The Alliance continued for additional years, but largely in name alone.

The military role was now somewhat a repeat of the Depression decade, although the economic circumstances were far different. The major objective was to preserve capitalism in the midst of the Cold War, and to eliminate the more radical options that threatened. Not all countries in the region were affected: Colombia and Venezuela in South America, Costa Rica in Central America, and Mexico retained civilian leadership. But the decade of the 1970s was later to be recalled as a decade of infamy for the violations of human rights that occurred in the Southern Cone.

Complicating matters still more, the petroleum crisis in 1973 exposed a major weakness in the model of Latin American development. Countries were substantially dependent upon petroleum imports, even those with national oil companies. Venezuela alone was able to benefit greatly from the fourfold rise in price from \$3 to \$12 a barrel in that year. Others coped by relying on a great surge of borrowing. At first, because interest rates remained low while commodity prices rose, such debt appeared a sensible way to finance the increased trade deficits. Unfortunately, debt-led development soon gave way to debt-led debt, a cumulating problem whose magnitude further multiplied with the outbreak of the war between Iraq and Iran in 1979, and a doubling once more of the price of petroleum.

Some in the region initially benefited. On this positive side, Venezuela was joined by Mexico, which had, during the 1970s, expanded its production considerably. Others, like Brazil, Chile and Argentina, had immediate adverse effects, the more so as international interest rates rose to new highs, for Paul Volcker's Federal Reserve now sought much more actively to restrain inflation in the United States. As events would unfold, it was one of the oil beneficiaries, Mexico, that first signaled an inability to pay. That happened in July 1982, when an appreciated exchange rate and lack of international reserves forced it to seek help from the International Monetary Fund (IMF) and the United States. Thereafter, soon came many

Table 65.3 *Latin American per capita income (1990 Geary–Khamis dollars)*

	1950	1981
Argentina	4987	7655
Brazil	1673	4984
Chile	3827	5933
Colombia	2089	4272
Mexico	2085	5582
Peru	2263	4292
Venezuela	7424	9637
Regional average	2614	5528
Relative to USA	0.27	0.30

Source: Angus Maddison, *Monitoring the World Economy*, OECD Development Centre, 1995

others. By the mid-1980s, more than a dozen countries in Latin America were receiving financial assistance, and advice, from the IMF. Even with that help, imports had to be drastically curtailed. The international banks that had been so eager to lend in the 1970s had completely withdrawn. Debt-led debt had come to a disastrous end.

At the virtual peak in 1981, as Table 65.3 reveals, many of the countries in the region had shown a rapid expansion from 1950: per capita income had doubled, and for Brazil, almost tripled. Venezuela retained its position as the leader in per capita income, but failed to alter its dependence on oil exports alone as the source of wealth. Argentina and Chile only grew modestly, indicative of the failures of import substitution, compared to the relative successes of Brazil, Colombia and Mexico. More impressively, average Latin American income per capita had finally gained on the United States. It is no wonder that many look back to those three decades with continuing longing for the levels of high tariff protection, for substantial government assistance and for a clear emphasis upon industrial expansion.¹⁰

Two decades of reconstruction

In the 1980s, in the midst of coping with the debt crisis, came a new beginning for the region, both politically and economically. The ubiquity of military government that had begun in the 1960s and reached its height with the Pinochet intervention in Chile in 1973, faded extraordinarily rapidly. Interestingly, the problem of managing the debt – whose accumulation had occurred under military regimes – was a decisive element in speeding their demise. New constitutions and new civilian governments appeared

throughout the region. Leaders sought support in fairer elections in which larger proportions of the population figured than had previously. Democracy, and inevitably a negative response to continuing inflation, whose inevitable consequence was redistribution of income away from the poor, emerged and strengthened over this period.

The ongoing debt problems of the region in the 1980s cried out for solution. All that initially emerged were low growth, negative capital flows, real devaluations and postponement of external payments. Finally, by the decade's end, the principle of debtors paying less was accepted. Private banks in the USA and elsewhere had finally emerged from the danger of their own failure, and now, in conjunction with the IMF, were amenable to the substitution of new Brady bonds – so named after the US Secretary of the Treasury – that saw reductions of some 40 percent of initial value. Mexico was the first country to proceed along these lines; Brazil was the last. Commercial banks got out of the business of extensive lending for capital investment, and back into providing shorter-term finance for international trade and other needs.

As this solution of the external problem evolved, there emerged a parallel recognition that fiscal deficits were the prime causal force in explaining Latin American inflation. Only if that governmental excess were credibly, and visibly, curtailed would prices stop their continuing acceleration. Some stabilization plan was necessary, and one that was more immediate in impact than past exercises. National outcomes varied in their particular style, their timing and the number of attempts required. Sometimes, as in Argentina and Brazil, multiple efforts were necessary. Sometimes, as in Nicaragua and Peru, populist measures were vainly instituted as alternative strategies. And sometimes, as in Bolivia in 1985, after achieving rates as high as an annualized 50 000 percent, results were virtually immediate after curtailing government deficits.

This was the essential key. In almost every instance, a firm price anchor was initially provided by a fixed exchange rate, and abetted by freer access to imports that helped to restrain future price increases. Ultimately, however, the key variable to brake inflation expectations remained the fiscal surplus, and when that was not realized, very high real interest rates and reliance on external capital flows were forced temporarily to substitute. These could work, but only for a short period. That is what we learned from the 'Tequila' crisis in Mexico at the very end of 1994, the Brazil crisis in 1999, and the Argentine collapse at the end of 2001.

What impresses is the extent to which past inflationary experience had to be overcome. Cutting back dramatically on inflation in the short term was not too difficult. The problem was ensuring continuity. Modest governmental deficits – by comparison with those of Europe, Japan and the

United States – gave rise to very high rates of price increase in Latin America. Internal credibility was slow to be achieved. Today, despite the variety of political leadership found in the region, the battle against inflation seems to have been achieved. That success had a notable and positive effect on the distribution of income, even when other factors intervened to worsen the aggregate.

Concern about fiscal deficits had a direct spillover upon two related subjects. The first of these was privatization. The impulse to sell off substantial state assets in energy, telephones, railways, steel and other intermediate sectors, petroleum exploration and production, airlines, banks, and so on came less from a philosophical commitment than from financial necessity. Initial levels of tax revenue, post-reform efforts to stem inflation, were generally insufficient to cover continuing current expenses. Virtually the only feasible solution was massive disposition of state property; the revenues thereby gained immediately eased the problem. When the need for revenue was very large, the terms of privatization were necessarily more favorable to the private buyers. Employment was almost always directly affected. There was excess labor, at higher wages than paid in the private sector, and almost every instance of privatization saw reductions in jobs, reflected in increased productivity, but also increases in formal sector unemployment.

Additionally, there was renewed attention to collecting larger revenues on a regular basis. Reform of the public sector to reduce current expenses was always announced, but rarely implemented. As a result, revenue collection generally increased much more than outlays were reduced. New taxes were frequently imposed: Brazil is at the regional extreme, with a total inflow of something like 37 percent of product, compared to an initial imposition of 25 percent. In other countries, beginning with Chile in the 1980s, attention turned to privatization of the social security system as a means of reducing future expenses. But these conversions required a surplus of current revenues as an enabling mechanism, thereby providing further incentives to ensuring adequacy of tax receipts.

Still a third area of great change during this period was a movement away from domestic protection and toward greater engagement with enhanced trade flows. Import flows, actual and potential, were an important competitive force capable of checking domestic price increases. Liberalization was therefore frequently utilized as part of the anti-inflationary effort. As countries privatized, external inflows to purchase, partially or entirely, the former nationalized assets became substantial. Currencies appreciated. The familiar problem of balance-of-payments limitations to growth soon reasserted itself. Numerous countries experienced serious crises as a consequence. There are the cases of Mexico in 1994, Brazil in 1999 and Argentina in 2002. But the commitment to openness, although occasionally frayed,

has remained. Tariff levels declined from an average in excess of 60 percent at the end of the 1980s to something between 15 and 20 percent in the mid-1990s.

These dramatic changes in the region inspired much criticism from the Left. As they progressively took place over the decade from 1985 to 1995, there was vocal opposition that took form in the electoral process. But newly elected presidents, frequently seemingly critical of such neoliberal policies, soon adopted the same measures. There was an initial positive consequence upon economic growth almost everywhere. But the new circumstances of the Tequila crisis, declines in Asia, the strong US dollar, failure in Russia and finally, Brazilian devaluation in 1999 took their toll: the promise that such reforms – termed neoliberal by their critics – might permit a resumption of sustained expansion was frustrated.

This is no way better seen than by viewing the data in Table 65.4. For virtually every country, with the prominent exception of Chile, and to some degree Mexico, the lost decade of the 1980s has transformed into a virtual repetition in the 1990s and even beyond. Note especially the dramatic decline in the ratio of income relative to that of the United States. Over the course of the entire twentieth century, there has been a dramatic decline; instead of convergence, there has been a disappointing divergence.

The present

Political opposition to these ‘Washington Consensus’ reforms has continued to increase in recent years. Much of that opposition – within, as well as outside, Latin America – emanates from intellectuals. This opposition

Table 65.4 Latin American per capita income (1990 Geary–Khamis dollars)

	1980	1985	1990	1995	2000
Argentina	8206	6834	6436	8005	8544
Brazil	5198	4917	4923	5296	5556
Chile	5738	5168	6402	8612	9841
Colombia	4265	4282	4840	5418	5096
Mexico	6289	6218	6119	6027	7218
Peru	4205	3631	2955	3505	3686
Venezuela	10139	8521	8313	8950	8415
LA average	5412	5052	5053	5460	5838
Relative to USA	0.29	0.24	0.22	0.22	0.21

Source: Angus Maddison, *The World Economy: Historical Statistics*, OECD Development Centre, 2003

starts from a legitimate concern about highly unequal income distributions and excessively high rates of unemployment that have been the counterpart of low growth. They then place blame upon a macroeconomic policy that has seemingly been too much in search of price stability, and too little in pursuit of economic expansion. The IMF has become an easy target for its insistence upon stabilization: reduced government outlays, higher rates of taxes, but above all, high real rates of interest. Privatization has been lamented, and its reversal sought, largely ineffectively. Now the argument has shifted to demands for greater regulatory control over private operation and decision. Greater protection against imports has been sought by industrial producers, eager to re-establish their former degree of dominance in internal supply. Finally, the initial aim of the 34 countries of the hemisphere – excluding Cuba – for a comprehensive free trade area, opening trade in goods and services, has not met its goal. Instead, there have been only an increasing number of bilateral agreements with the United States, with overt opposition from Venezuela, Mercosur and Bolivia.

But outright reversal of the great changes since the 1990s hardly guarantees the future development of Latin America and the Caribbean. Much that has occurred has been necessary, and much overdue. Market forces and wider trade are now a prominent feature almost everywhere in the world. Ignoring them has a real cost. Better to build upon what has been accomplished, and to recognize that more reforms, not fewer, are needed. Recent high rates of growth in the region since 2003 provide a singular opportunity to seize the moment.¹¹

As essential first step is a substantial and ongoing commitment to educational reform and expansion.¹² Latin America, for a variety of reasons, has been very late to recognize the importance of universal education to economic growth as well as to a more equal income distribution. In recent years, catch-up has begun to occur: the number of years of education available to the young has expanded, but still lags behind the impressive efforts in Asia. Repetition of beginning years of schooling continues, poor quality of teaching remains, and free public universities absorb significant budgetary resources. Past mistakes inevitably linger as individuals with limited literacy persist in the labor force. Reform is further complicated by its necessary continuity: one requires consistent policies over a generation, not over a single presidential term. So despite expenditures as a percentage of national product that approach 5 percent, few countries stand out positively. Comparable international test results confirm this regional backwardness.

Advance in coping with inequality requires continuing advance in the educational system. That is clear even from the extensive comparative report put together by the World Bank that correctly stresses other

factors.¹³ Education is hardly the only element influencing the distribution of income, but it frequently comes out as a principal one. Most relevantly, the quality of education is likely dramatically different for individuals whose parents are poor versus those who are wealthier. The appropriate conclusion is that 'educational disparities account for an important share of Latin America's high income-related inequality, but are not the only explanatory factor'.¹⁴

A second area requiring attention is continued macroeconomic reform. In particular, the extraordinary regional difference in public finance seems to suggest a fundamental area of needed advance. Brazil, at one extreme, has revenues that exceed 38 percent of product, while Mexico, excluding oil receipts, barely enters into double digits. Neither level is likely to be efficient or effective. Even when large expenditures are made, governmental investment is small, and an inadequate infrastructure (physical, as well as individual, where health, nutrition, and so on weigh heavily) results. Countries, for example, have invested little in research and technology. Consequently, they have less capability to follow on closely the advances achieved elsewhere. Solving the inflation problem, as most countries have, does not mean the end of concern; instead, it is just the beginning.

Regional countries equally must expand their savings rates if they are again to achieve in the twenty-first century the per capita expansion accomplished in the 1960s and 1970s. Something of the order of 25 percent of national product is required, of which the foreign component should not exceed a small proportion. An inversion is needed. Historically, the private sector saved, allowing the public sector to invest. Now primary surpluses should become the rule, financing not only public capital formation, but permitting private firms to borrow at lower interest rates and for longer terms.

Trade advance should continue, on a more geographically as well as product-diversified basis. In the midst of the present rise in the terms of trade, and strong trade surpluses, there is greater enthusiasm about international trade within the region. Latin America continues to be a region where trade has been less significant than objective indicators suggest it should.¹⁵ Chile and Mexico are now the exceptions rather than the rule. The real test will come if, and when, the commodity boom begins to slacken. Latin America has had that experience before, and the inevitable result was to emphasize the internal market and strengthen protection against import competition, rather than to upgrade the quality of its exports and compete more effectively.

Whether an increasing commitment to the international market will persist, is the question. Globalization is seen in many countries as a hindrance rather than an opportunity. Last time, in the expansion of income

through 1980, that option was largely ignored. Now one hears increasing doubts about whether an open market should persist. Latin America needs the chance to enter, however belatedly, into the international marketplace. For it to choose an alternative path will again deter rather than enhance the region's emergence.

A final, but important, subject is the need for institutional reform. This refers not merely to reform of the judiciary and the need for persistent and independent standards, but also to the area of regulation of the newly privatized activities. As natural monopolies, compounded by the reality of extensive foreign participation, these activities require supervision and reasonable certainty. Otherwise, investment required in infrastructure will continue to lag behind. Not only economic institutions are involved. Political reform is likewise needed in so many countries, where electoral rules are inadequate, and fundamental changes in structure are required.¹⁶

Conclusion

The countries of Latin America and the Caribbean find themselves challenged once again, just as they have been over the entire post-World War II period. This time, after having lost out to South Korea and Taiwan in the 1980s and 1990s, the task is much greater. China, India, South Africa and still other Asian countries are now the new competitors in world markets for trade and foreign investment. These are much larger challengers to cope with. And, over the last several years, they have been impressive performers.

In the face of this new challenge, some have advocated greater South-South exchange, as was the mantra in the 1960s and 1970s. Some see the present World Trade Organization (WTO) Doha Round as the place for large reductions in agricultural protection that will assure fairer trade and better opportunities for Latin American advance. Some have called for a resurgent state to replicate the advances achieved in the earlier era, a return to subsidies and import substitution, but this time fully committed to reduction of inequality in the distribution of income. Second-stage reform, as advocated here, is rejected in favor of a different approach.

That route of denial, attractive as it sometimes seems, runs a substantial risk of casting away the advances of almost a generation of reforms. International competition is inevitable, and increasing, in the present age of globalization, as even the developed nations have been discovering. Rather than try to deny that reality, the countries of the region must seek to confront it. Only through a process of internal productivity advance, fueled by advances in technology and greater domestic savings, will Latin America be able to compete, and thereby achieve greater continuing growth.

There is not much time left to choose.

Statistical appendix

There are now four comprehensive series of estimates of national income dating back to the nineteenth century, and encompassing the countries of Argentina, Brazil, Chile, Colombia, Mexico and Venezuela. Peru is the additional country incorporated in the Maddison estimates. Uruguay is also sometimes found. Although these series do differ with the particular base year selected, as well as sources utilized, which does explain some difference in absolute value relative to the United States or OECD countries, they are quite similar in their reported longer-term rates of growth.¹⁷ Table 65A.1 provides these results. Shorter-term movements in individual countries do differ, however, and sometimes importantly.

I will not enter here into the index number problems presented by such long-term comparisons, and the familiar Laspayres bias found as growth incorporates new sectors into production over time. These, and other methodological issues, are discussed extensively by Prados de la Escosura.¹⁸ But less time seems to have been spent in searching out and considering critically the national historical estimates that underlie the aggregates. For Brazil, for example, three different series seem to have been chosen by the authors. Ironically, the most recent estimates by Prados de la Escosura choose a series of Raymond Goldsmith, put together in the 1960s, well before recent research on historical Brazilian development flourished.

Table 65A.1 Comparative per capita rates of growth

	(1) LA 6	(2) LA 6	(3) LA 6	(4) LA 7
1900–13	2.2	2.5	2.2 (2.6)	2.4
1913–29	1.0	1.5	2.4 (1.5)	1.6
1929–50	1.4	1.6 ^a	1.8 (1.8)	1.5
1950–80	3.1	2.7 ^b	2.5 (3.0)	2.8

Notes:

LA 6 is Argentina, Brazil, Chile, Colombia, Mexico and Venezuela

LA 7 includes Peru

a 1929–45

b 1945–81

Sources:

Col (1) Prados de la Escosura (2004)

Col (2) Thorp (1998, Statistical Appendix), calculated from Total GDP minus population growth

Col (3) Hofman, (2000, p. 169)

Col (4) Maddison (2001)

I have chosen to use the Maddison estimates for Latin America – despite his use of the Goldsmith estimates for Brazil – principally because of their greater comparability with those for other regions. That database extends back to 1820 and incorporates information not only on the present developed countries but also on many developing nations in Africa, Asia and Southern and Eastern Europe. In all, there are 56 countries utilized, far larger than in any other source.

Notes

1. There has been a proliferation of new research covering Latin American economic history over the course of the last two decades. Much of this work is quantitative and a large part has been the product of Rosemary Thorp, who merits special credit. She authored *Progress, Poverty and Exclusion: an Economic History of Latin America in the Twentieth Century* for the Inter-American Development Bank in 1998, and co-edited three volumes of background papers in the series *An Economic History of Twentieth Century Latin America*, published by Palgrave in 2000. I have made much use of these volumes, without explicit citation. But additionally, there are the recent contributions of Leandro Prados de la Escosura, as well as that of Andre Hofman. And, of course, there has been much independent research done on a national level. In the brief ‘Statistical appendix’, I explain my choice of Angus Maddison’s national income estimates as the basis for long-term comparisons.
2. Lewis (1978).
3. Bulmer-Thomas (2003, p. 144), finds an R-squared of 0.82.
4. Cf. Albert (1988).
5. See Eichengreen and Fishlow (1996).
6. Thorp (1998, p. 114).
7. The classic exposition is found in CEPAL (1951).
8. This can be easily seen from the formula for revenue from the inflation tax: $p(M/P)$, the rate of inflation times the public holdings of money. Of course, as inflation occurs, desired holdings of money decline, which is why the process tends to become cumulative.
9. The Economic Bulletins issued by CEPAL at the beginning of the 1960s give evidence of this shift, as do the *Annual Economic Surveys*.
10. Ironically, many sometimes forget the degree to which the gains were associated with military governments, as well as substantial debt accumulation.
11. Nancy Birdsall has put together another 11 measures beyond the Washington Consensus, ten that are seen as required for the future. They stress the need for great social equity in the region. There is some overlap with my modest four; Birdsall et al. (2001).
12. PREAL, *Quantity without Quality, 2006 Report Card*, provides extensive details on the reforms required, and the poor marks countries continue to receive for only partial commitment.
13. de Ferranti et al. (2003).
14. Ibid., p. 200.
15. This is true of various gravity models using physical characteristics such as distance from markets as well as income levels. See Carillo-Tudela and Li (2004).
16. See the recent 2006 Economic and Social Progress Report of the Inter-American Development Bank, *The Politics of Policies* for a full discussion of the roles of Politics.
17. For some reason, Andre Hofman (2000) has chosen to present his aggregate results for per capita income, as well as for other measures, as a simple average of the individual country results, rather than weighting by the relevant magnitudes. This is appropriate were one interested exclusively in differences among countries, rather than also concerned with group, that is, Latin American, performance. In Table 65A.1, I have provided his original averages in parentheses, and recalculated the LA 6 totals.
18. See Prados de la Escosura (2000).

References

- Albert, B. (1988), *South America and the First World War: the Impact of the War on Brazil, Argentina, Peru and Chile*, Cambridge: Cambridge University Press.
- Birdsall, N. and Augusto de la Torre with Rachel Menezes (2001), 'Washington Contentious: Economic Policies for Social Equity in Latin America', Washington, DC: Carnegie Endowment and Inter-American Dialogue.
- Bulmer-Thomas, V. (2003), *The Economic History of Latin America since Independence*, 2nd edn, Cambridge: Cambridge University Press.
- Carillo-Tudela, Carlos and Carmen A. Li (2004), 'Trade Blocks and the Gravity Model: Evidence from Latin American Countries', *Journal of Economic Integration*, 19 (4), December: 667–89.
- CEPAL (1951), *Economic Survey of Latin America, 1949*, New York: UN.
- de Ferranti, D., Guillermo Perry, Francisco H.G. Ferreira and Michael Walton (2003), 'Inequality in Latin America and the Caribbean: Breaking with History?' World Bank Latin American and Caribbean Studies.
- Eichengreen, B. and A. Fishlow (1996), 'Contending with Capital Flows: What is Different About the 1990s?', Council on Foreign Relations.
- Hofman, A. (2000), *The Economic Development of Latin America in the Twentieth Century*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.
- Inter-American Development Bank (various years), *Economic and Social Progress Report*, Washington, DC: IDB.
- Lewis, W.A. (1978), *Growth and Fluctuations, 1870–1913*, London: Allen & Unwin.
- Maddison, Angus (2001), *The World Economy*, Paris: OECD.
- Prados de la Escosura, L. (2000), 'International Comparisons of Real Product, 1820–1990: An Alternative Data Set', *Explorations in Economic History*, 37: 1–41.
- Prados de la Escosura, Leandro (2004), 'When Did Latin America Fall Behind? Evidence from Long Run International Inequality', *Working Papers in Economic History*, Universidad Carlos III.
- PREAL (2006), *Quantity without Quality, 2006 Report Card: A Report of the PREAL Advisory Board*, PREAL.
- Thorp, R. (1998), *Progress, Poverty and Exclusion*, Baltimore, MD: Johns Hopkins University Press.
- Thorp, R., E. Cardenas and J.A. Ocampo (2000), *An Economic History of Twentieth Century Latin America*, 3 vols, Palgrave.

66 Sub-Saharan Africa

Benno J. Ndulu and Stephen A. O'Connell

Introduction

In 1960, economic development was the mandate and keenly sought-after province of a founding generation of African political leaders. Visions of economic progress were central to liberation rhetoric (Ake, 1996) and were widely embraced within the broader development community (for example, Karmarck, 1971). Yet development failed, and it failed uniquely. Over the next 40 years, as extreme poverty fell dramatically on a worldwide basis,¹ poverty in sub-Saharan Africa (henceforth SSA or 'Africa') increased. At the turn of the millennium, nearly one in two Africans or 300 million in total consumed less than \$1 a day, a proportion twice as high as the global rate and double the number prevalent in Africa in 1970. Home to 10 percent of the world's population, the region now accommodates 30 percent of the world's poor – who spend a quarter less than the Asian poor on their livelihood (World Bank, 2005). The African development challenge has become the global development challenge. How and why did it become so, and what can we expect, looking ahead?

We approach these questions through the lens of economic growth. The second section describes the African record from 1820 to the present, focusing on the period since 1960 and emphasizing demographic and other features that differentiate African growth patterns from those of other developing regions. In the third section we outline two main structural explanations for the African experience, based in turn on governance and resource endowments. We develop the former theme in a detailed exploration of the political economy of African growth. The fourth section documents the revival of growth that got under way in the mid-1990s; we interpret this revival as a lagged response to the economic and political reforms of the late 1980s and early 1990s. We argue that Africa now faces a window of opportunity, with politically stable countries facing the prospect of mutually reinforcing declines in fertility rates and increases in capital formation and growth. We conclude with a summary of lessons from the African experience, both for the region itself and for development economics.

The growth record

Any adequate account of Africa's development experience must acknowledge the overall weakness of Africa's growth record, its juxtaposition with

population explosion, and its variability over time and across countries. We begin with Africa's long-run growth experience, drawing first on the painstaking historical research of Maddison (2001) and then on World Bank data for 100 developing countries since 1960.² Our central concern here is the relative stagnation of economic growth in SSA in the period since 1950. This performance is not an outlier in historical terms: in Table 66.1, Africa's per capita growth rate of 0.7 percent after 1950 matches that of the rest of the developing world over the previous century. But African populations missed out on the economic transformation that took place in the developing world – particularly in Asia – in the second half of the twentieth century. The result was that by the 1950s, African incomes, which had gained considerable ground in relative terms since 1913, had begun to diverge powerfully from incomes elsewhere in the developing world. By comparison with East Asia and Pacific, a shortfall of less than 50 percent in purchasing power parity (PPP)-adjusted terms around 1960 rose to well over 300 percent by the end of the century. The consequences of this growth failure are apparent in Table 66.2, which provides a snapshot comparison of human development, first at the outset of the 1960–2000 period and then at the end. With the exception of the primary enrollment rate, which was already high outside of Africa in the early 1960s, Africa fell further behind the rest of the developing world. Regress was not as severe for non-income measures as it was on income or poverty: by 2000 Africa exceeded the levels of primary enrollment, adult literacy and life expectancy that had prevailed elsewhere in 1960. But the failure to raise per capita incomes significantly had critically undermined these achievements.

At the sectoral level, slow growth has gone hand-in-hand with limited structural diversification. Traditional agriculture continues to absorb the majority of the labor force in many African countries, a feature no longer observed in any other region of the world (O'Connell and Ndulu, 2000). Irrigation is expensive and extremely sparse, with the result that African agriculture remains largely rain-fed and subject to periodic drought. Exports have tended to remain concentrated in a narrow band of primary commodities (Berthélemy and Soderling, 2001, 2002), including exploitation of mineral resources. Collier and O'Connell (2007) use global data to identify 'resource-rich' economies as those that exceed threshold values for the ratios of primary commodity rents (from energy, mineral and forest resources) to gross domestic product (GDP) and primary commodity exports to total exports on a sustained basis. Comparing SSA with other developing regions (excluding the Middle East and North Africa), they find that a stark difference already existed in 1960, with 12.5 percent of the SSA sample classified as resource-rich and only 7 percent of the non-SSA sample. This difference expanded over time, with another 16.7 percent of

Table 66.1 *Long-run growth rates by developing region.*

	Other developing regions				
	SSA	Total	of which:		
			LAC	ASIA	MENAT
<i>Population</i>					
1820–70	0.3	0.2	1.3	0.1	0.5
1870–1913	0.7	0.6	1.6	0.6	0.8
1913–50	1.7	1.0	2.0	0.9	1.3
1950–2001	2.6	2.0	2.3	1.9	2.6
1950–60	2.2	2.1	2.8	2.0	2.6
1960–2001	2.7	2.0	2.2	1.9	2.6
Peak year*	1983	1971	1960	1971	1981
Peak rate*	3.00	2.52	2.87	2.52	3.34
<i>Real GDP (PPP-adjusted 1990 dollars)</i>					
1820–70	0.6	0.1	1.2	0.0	1.0
1870–1913	1.1	1.3	3.5	1.1	1.6
1913–50	2.7	1.5	3.4	0.9	2.5
1950–2001	3.3	5.0	4.0	5.4	4.9
1950–60	4.1	5.6	5.1	5.8	5.6
1960–2001	3.1	4.9	3.7	5.3	4.7
<i>Real GDP per capita (PPP-adjusted 1990 dollars)</i>					
1820–70	0.2	-0.1	0.0	-0.1	0.5
1870–1913	0.4	0.7	1.8	0.5	0.8
1913–50	1.1	0.5	1.4	-0.1	1.1
1950–2001	0.7	3.0	1.7	3.4	2.2
1950–60	1.9	3.4	2.3	3.7	2.9
1960–2001	0.4	2.9	1.5	3.4	2.0
<i>No. of countries</i>	53	90	27	42	21

Notes:

SSA is Maddison's 'Africa', excluding Algeria, Egypt, Morocco, and Tunisia, but including Libya (see below). LAC comprises 'Latin America' and includes the Caribbean. ASIA is Maddison's 'East Asia'. MENAT (Middle East, North Africa and Turkey) corresponds to 'West Asia' plus North Africa. Libya could not be separated out and we therefore include it here in SSA rather than in MENAT. Note that these growth rates are based on regional totals and are therefore dominated by the large countries in each region. More specifically, the population, real GDP and real GDP per capita growth rates correspond to population-, GDP- and (approximately) GDP-weighted individual-country growth rates. Note also that data before 1950 unavoidably contain major 'guesstimates'. For example, the African growth rate for 1870–1950 is based on case study work on Algeria, Egypt, Ghana, Morocco, South Africa and Tunisia. Maddison applies the average growth rate of GDP per capita for this group of six (= 0.90) to all countries in SSA other than Ghana and South Africa. Our SSA aggregate for 1870–1913 combines this indirect estimate for 51 countries with Maddison's direct estimates of 2.07 and 1.25 for Ghana and South Africa.

* Peak year and peak rate correspond to the earliest year after which all subsequent population growth rates are lower.

Source: Ndulu and O'Connell (2007), Table 1.4. Calculations are based on country and regional data from Maddison (2001).

Table 66.2 Regional growth comparisons

Region	N	Initial values (1960 or earliest year before 1965, or as indicated)				End-to-end annual growth rates (earliest year before 1965 to latest year between 1995 and 2000)				Ending values (latest year between 1995 and 2000)				
		Real GDP per capita (1996 PPP\$)	Gross primary enroll- ment rate, 1970	Adult illiteracy rate, 1970	Life expec- tancy at birth	Real GDP per capita	Popu- lation	Total	Real GDP per capita		Real GDP per capita (1996 PPP\$)	Gross primary enroll- ment rate	Adult illiteracy rate	Life expec- tancy at birth
									Workers per capita	per worker				
SSA	35	1278.1 (15.0)	53.8 (52.8)	55.8 (58.5)	41.1 (58.5)	3.20	2.63	0.56	-0.15	0.72	2047.5 (8.4)	90.1 (86.4)	41.2 (61.1)	47.8 (61.1)
OTHER DEV'ING	43	2591.5 (30.5)	90.5 (88.9)	26.5 (75.8)	53.2 (75.8)	4.28	2.16	2.12	0.23	1.90	6409.1 (26.2)	107.4 (103.0)	17.9 (89.1)	69.7 (89.1)
LAC	22	3338.4 (39.2)	99.1 (97.3)	17.4 (80.3)	56.4 (80.3)	3.52	2.08	1.44	0.42	1.03	6268 (25.6)	113 (108.3)	11.1 (90.5)	70.8 (90.5)
SASIA	5	934.4 (11.0)	58.6 (57.6)	55.5 (64.5)	45.3 (64.5)	4.34	2.23	2.10	-0.32	2.42	2186.3 (8.9)	100.1 (96.0)	45.2 (81.6)	63.8 (81.6)
EAP	9	1833.1 (21.5)	94 (92.3)	20.4 (72.1)	50.6 (72.1)	5.48	2.07	3.41	0.16	3.29	8691.0 (35.5)	101.3 (97.1)	11.4 (89.0)	69.6 (89.0)
MENAT	7	2402.5 (28.2)	81.9 (80.5)	42.3 (73.6)	51.7 (73.6)	5.09	2.48	2.61	0.13	2.48	6934.7 (28.3)	103 (98.8)	27.8 (90.3)	70.6 (90.3)

Table 66.2 (continued)

Region	Initial values (1960 or earliest year before 1965, or as indicated)		End-to-end annual growth rates (earliest year before 1965 to latest year between 1995 and 2000)				Ending values (latest year between 1995 and 2000)				
	Real GDP per capita (1996 PPP\$)	Gross primary enroll- ment rate, 1970	Adult illiteracy rate, 1970	Life expec- tancy at birth	Real GDP per capita worker	Workers per capita	Total Popu- lation	Real GDP per capita (1996 PPP\$)	Gross primary enroll- ment rate	Adult illiteracy rate	Life expec- tancy at birth
INDUST	22 8507.6	101.8		70.2	3.45	0.71	2.74	24489.2	104.3		78.2
Total	100 3433.3	80.4	38.1	52.7	3.72	2.00	1.71	8860.2	100.6	27.1	63.9
SSA v SASIA	(136.8)	(91.8)	(100.5)	(90.7)				(93.7)	(90.0)	(91.2)	(74.9)
SSA v Other Dev	(49.3)	(59.4)	(210.6)	(77.3)				(31.9)	(83.9)	(230.2)	(68.6)

Notes:

Except in the final 2 rows, the numbers in parentheses give the relevant developing-country mean as a percentage of the industrial-country mean. The final 2 rows show the SSA mean relative to the SASIA mean and the mean for all non-SSA developing regions.

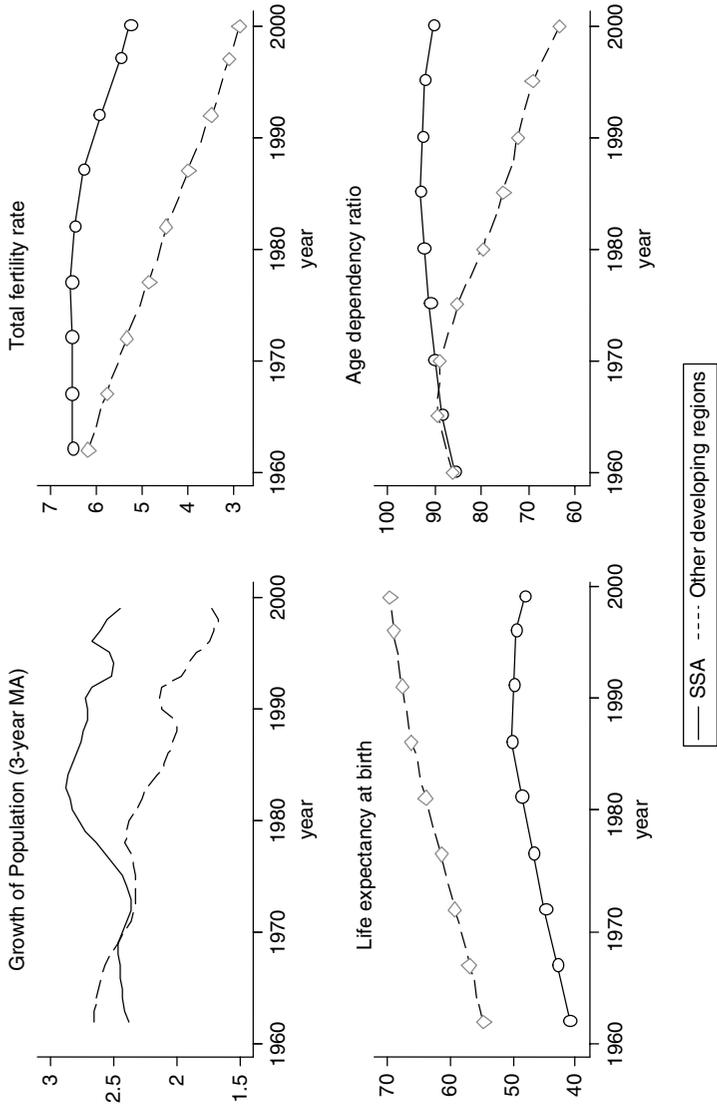
Regions: SSA = Sub-Saharan Africa; OTHER DEVING = Other Developing (LAC = Latin America and Caribbean, SASIA = South Asia, EAP = East Asia and Pacific, MENAT = Middle East, North Africa and Turkey); INDUST = Industrial countries.

Source: Ndulu and O'Connell (2007), using PWT6.1 and *World Development Indicators*.

the African sample acquiring resource-rich status by 1990, as compared with only 10.5 percent of the non-African. A consequence of Africa's delayed structural transformation has been the continued vulnerability of its population to shocks to rainfall and world commodity markets.

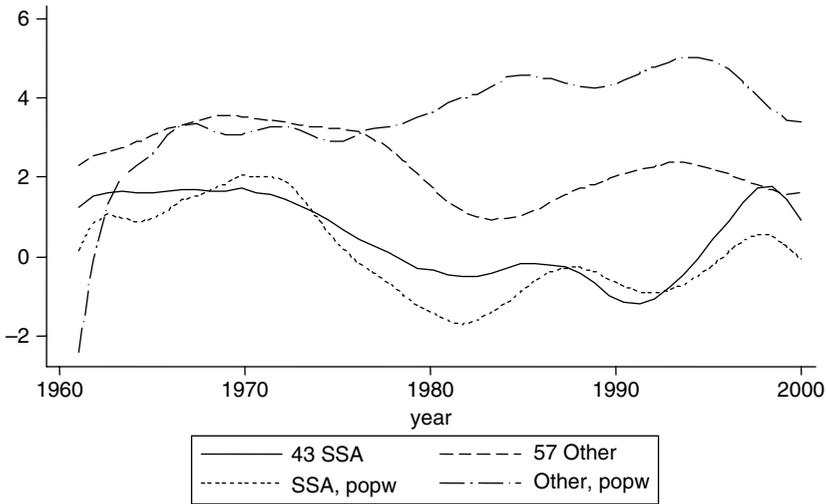
Tables 66.1 and 66.2 document the demographic explosion that is a striking correlate of Africa's economic stagnation over the post-1960 period (O'Connell and Ndulu, 2000; Lucas, 2003). With the exception of life expectancy rates, standard indicators of demographic pressure differed only modestly between Africa and the rest of the developing world in the early 1960s (Figure 66.1). But the demographic transition was already well underway in Latin America, where population growth rates had peaked in 1960 (Table 66.1). For the next 25 years, total fertility rates fell sharply outside of Africa while remaining virtually unchanged within Africa. Population growth rates therefore diverged sharply, and from the early 1970s through the remainder of the century, the population of SSA grew more rapidly than the non-African developing-country population had grown at its peak. The ratio of (overwhelmingly young) dependents to working-age population grew steadily, exceeding historical developing-country norms by 1970 and remaining above these through 2000. The fertility rate began to fall in Africa in the mid-1980s, suggesting entry into the final phase of the demographic transition. We will return to this observation, which is complicated by the huge impact of HIV/AIDS on life expectancies starting in the late 1980s.

The aggregate growth record conceals considerable variability, both over time and across countries. Table 66.1 indicates that a sharp deceleration of growth took place in Africa between the end of the colonial period (1950–60) and the remainder of the century. Within the latter period, a further distinction can be made (Figure 66.2), between the moderate growth rates of the 1960s and late 1990s and the deep contraction of 1974–94. African economies continued to grow in the 1960s. This growth was already weaker than that of other developing regions, however, and the global deceleration of the 1970s therefore took substantial portions of the continent into outright contraction. Between 1960 and 1994, nearly half of African countries with comparable data suffered per capita income losses exceeding 20 percent in constant domestic currency (Rodrik, 1998). The bulk of these losses occurred between 1974 and 1994, a period that began with a set of shocks to energy and tropical commodity markets (1974–79) and ended with a concentrated wave of African democratic reforms (1989–94). As indicated in Figure 66.2, the growth shortfall between 1974 and 1994 is much larger on a population-weighted basis than in the simple averages. This is partly a large-country phenomenon within Africa,³ but after 1980 it is driven primarily by the dramatic growth performance of



Source: Ndulu and O'Connell (2007), using World Development Indicators 2005. The figures show simple averages of country observations, for all countries with continuously available data.

Figure 66.1 Demographic pressure: SSA versus Other developing regions, 1960–2000



Source: Ndulu and O'Connell (2007), using World Bank data. The figure shows averages of country observations, for all countries with continuously available data. The suffix 'popw' refers to Population-Weighted averages. The data are smoothed using a Epanechnikov kernel-weighted polynomial smoother of degree 1; alternative approaches yield similar results.

Figure 66.2 Smoothed average growth in real GDP per capita (countries with full set of growth observations)

China, India and Indonesia. The revival of African growth in the mid-1990s lends a distinct U-shape to the region's overall growth record for 1960–2000.

The averages discussed here also mask wide variation in the growth performance of individual countries. Many African countries have experienced limited episodes of robust growth. Nearly half of the African economies studied by Pritchett (1998), for example, exceeded a per capita income growth threshold of 1.5 percent through the mid-1970s. Ghana and more strongly Uganda have consistently exceeded a 2.5 percent threshold (roughly the long-run median for developing countries) since the mid-1980s. Botswana and Mauritius grew spectacularly; their long-run records compare favorably with those of the East Asian miracle economies. A diversity of outcomes also characterizes the period after 1994, during which 15 African countries have seen growth rates of total GDP in excess of 5 percent. Focusing on the most recent five-year period (1999–2004) and excluding the oil countries, median per capita growth rates in the fastest-growing, middle, and slowest-growing thirds of the African sample – each comprising 12 or 13 countries – were 2.8, 1.0 and –1.8 percent respectively.⁴

Outside of the mineral-exporting group, rapid growth after 1990 has also been associated with substantial diversification of production and exports.

Finally, the weak contribution of increases in measured physical and human capital per worker to African growth is yet another significant feature. Weak investment effort is part of this story (see below): standard growth accounting exercises assign about half of the post-1960 growth shortfall relative to other developing regions to a shortfall in measured capital deepening per worker. But the data also suggest profound problems in translating investment effort into effectively utilized capital (Pritchett, 2000). In order to reconcile observed growth outcomes with measured capital inputs, one must conclude that the average productivity of African inputs per worker fell considerably over time, not just relative to productivity elsewhere but in absolute terms. Ndulu and O'Connell (2007), for example, find that the country-level cumulative changes in total factor productivity over 1960–2000 were as likely to be negative within Africa as positive (see also Hall and Jones, 1999). The correlation of physical capital accumulation with growth, moreover, is considerably lower within Africa than in the rest of the developing world, even over periods of a decade or longer. The latest example of this is Africa's recovery starting in the mid-1990s, which was not accompanied by a commensurate boom in aggregate investment (Berthélemy and Soderling, 2001).

Explaining poor growth performance in Africa

Why did the African environment prove hostile to economic growth after 1950? Two broad lines of argument dominate the literature, based in turn on economic mismanagement and structural impediments to growth. With some imprecision we refer to these below as the 'governance' and 'geography' views. Demographic trends pose distinct structural challenges, and we treat these separately. A synthesis of these strands has yet to be achieved, though we make a tentative approach after reviewing the main arguments.

The governance critique

A critique of economic management has dominated the literature on African economic performance since the early 1980s. The central themes of what we will call the 'governance critique' were laid out in the World Bank's 1981 *Accelerated Development in Sub-Saharan Africa* (the 'Berg Report', after its chief author Elliot Berg), Robert Bates's classic 1981 *Markets and States in Tropical Africa* and, with a ten-year lag, Paul Collier's 1991 *African Affairs* article on agencies of restraint.⁵

In its initial and most influential form, the governance critique sought to explain what it characterized as excessive intervention by African governments in economic markets. Bates (1981) argued that African policy-makers

had sacrificed both the agricultural sector and industrial efficiency in order to divert resources to favored interests. The 1980 Berg Report had located African economic stagnation in overtaxation of export agriculture, overprotection of import-competing industries, and oppressive state control of finance, industry and agricultural marketing. Bates (1981) argued that while these policies were economically inefficient, the 'urban bias' they displayed was rational from the point of view of political elites. Farmers would lose out, and their losses would exceed the gains to the political elites and to the urban interests more generally (the civil service, the military, and labor and capital in the formal private sector or state enterprise sector). But farmers faced deep collective action problems; while numerous, they were too poorly organized to constitute the 'selectorate' on which the government depended to retain power (Bates and Devarajan, 2001).

Governments would therefore penalize agriculture and support urban-based industrialization far beyond what could be justified by the correction of market failures. Moreover, they would do so using inefficient quantity-based instruments rather than price-based interventions (that is, quotas, exchange controls and marketing monopolies rather than tariffs and explicit export taxes), because the distribution of policy-generated rents was central to their political security. Growth would fail, but the protected urban electorate would be well served. Exceptions to the urban bias pattern could be explained, in the Bates analysis, by appealing to the rural political roots and business interests of founding political leaders in countries like Kenya, Malawi and Côte d'Ivoire.

The Berg–Bates contribution provided a description of African policy biases, a link from these to growth outcomes, and a grounding of these choices in the interests of African political elites. Each element of the argument provoked important voices of dissent or moderation.⁶ But the empirical content of the critique gave it substantial appeal to Africa's donors and to economists seeking to understand the continent's lagging growth performance. For donors, the governance critique provided a rationale for using conditional lending to push market-based reforms – a tendency undoubtedly strengthened by the ascendancy of conservative governments in the United States and Europe in the early 1980s. Donors could act as the agents of disenfranchised African populations, imposing conditionality on easily monitored policy reforms like trade liberalization and exchange rate unification. Since the existing policies penalized the poorer rural sector, their removal would simultaneously improve growth and distribution.⁷ For economists, the view that policies represented the largely autonomous choices by a self-interested political elite provided a causal interpretation of regression evidence linking measures of policy distortion with economic growth (for example, Sachs and Warner, 1995).

A generalization of the critique to misaligned interests

Subsequent contributions developed the governance critique beyond its initial application by restating it in terms of a conflict of interest between African policy-makers and African populations. The basic argument is that under authoritarian rule, development policy in a large number of African countries has tended to be captured by a narrow political elite operating under relatively weak institutional constraints. Characterized by weak legitimacy and tenuous bureaucratic control, autocratic regimes in Africa until the 1990s did not function as agents of the public interest, tending instead to sacrifice growth in favor of patronage-based redistributive politics or outright predation (see also Adam and O'Connell, 1999; Ndulu and O'Connell, 1999; Humphreys and Bates, 2001).

An anti-growth bias may become severe if policy-makers undervalue future interests relative to present. Policy-makers may discount the future excessively if they have little expectation of remaining in power or being held accountable. Using African data from the 1980s, McMillan (2001) finds that high rates of presidential turnover predict inefficiently high rates of export taxation in Africa, particularly for crops with large fixed inputs.⁸ Fosu (2002) documents the frequency of attempted coups in Africa and finds a strongly negative causal link to overall growth. In an extreme case of misaligned interests, the expected tenure of an authoritarian leader or oligarchy may therefore be inversely related to growth performance. Diamond (1977) appeals to global evidence, for example, to argue that economic success tends to bring democratization. If autocratic leaders internalize such a link, then development may represent a threat rather than an investment in future rents (Robinson, 1997). Political elites may then actively oppose development – as in some cases the colonial powers themselves did, fearing the contestability development might create in the economic and political spheres.⁹

A combination of autocracy and tight economic controls during 1960–85 presented a particularly potent environment for patronage and predation. Using data on the type of African leadership Ndulu (2007) shows that 80 percent of autocratic regimes since independence imposed soft or hard controls (see Collier and O'Connell 2007) for a substantial portion of their tenure.

Economic controls were not new to Africa at the time of independence. Colonial administrations had embraced an activist, developmental role by the late 1940s, reflecting the suddenly temporary nature of their remaining trusteeship and the allure of state intervention following depression, wartime mobilization and the emergence of the Soviet Union as a great power. A mentality of market regulation was therefore in place well before the formal transfer of sovereignty in Africa, as were some of its key institutional mechanisms including monopoly export marketing boards,

exchange restrictions and economic plans (Fieldhouse, 1986). But the period from 1960 to 1975 saw a dramatic expansion of the regulatory presence of the African state. Although this sharp expansion was part of a global phenomenon and grounded in the global development paradigm of the day, it was propelled further by rent-seeking behavior.¹⁰

Late in the colonial period, the colonial powers had begun to introduce the rudiments of democratic self-government, and independence constitutions reflected the institutional structures of Western democracies, with contested multi-party elections, checks and balances, and substantial civil liberties. But the degree of *de facto* democracy deteriorated steadily following independence. As in the case of economic controls, authoritarian government was not an African innovation. Its growth impact in SSA, however, appears to have been far from beneficent, in contrast to the broad Asian experience with authoritarian rule during the same period (Alesina and Perotti, 1994). The governance critique asserts that in the African context, causality runs from government institutions to growth.

Limitations and extensions of the governance critique

While the governance critique had instinctive appeal both to Africa's donors and to economists seeking to understand the continent's lagging growth performance, it also raised serious conundrums.

Institutions Van de Walle (2001) argues that the unifying feature of African political economy is not the power of urban interests but rather the absence of coherent domestic political interest groups of any kind. In this view, domestic interests are everywhere too weak to restrain the behavior of the tiny political elite that holds or shares power in its own interest. To explain the persistent 'choice' of stagnation, he appeals to the low capability of African public bureaucracies and the dysfunctional influence of strategically motivated external donors. These conspired to undermine the medium-run coherence of virtually any policy initiative, he argues, while simultaneously protecting high-level African elites from personal responsibility for economic decline.

In an influential extension of the governance critique, Collier (1991) similarly appealed to an institutional vacuum that left too much discretion and too little accountability in the hands of political elites. Collier focused on agencies of restraint – public institutions designed to protect national assets (including privately held ones) from predation. Before 1960, this role was provided by colonial administrations operating under tight mandates of internal security, fiscal solvency, and openness to metropolitan trade and finance. Immediately following independence, the new counterparts to colonial institutions – export marketing boards, national central banks,

multiparty parliamentary systems, independent judiciaries – struck a balance between flexibility and restraint that reflected the conservative mandates of their colonial predecessors. But political leaders sought to consolidate power and could appeal to the need for ambitious development programs. By the mid-1970s, political leaders had systematically relieved national ‘agencies of restraint’ of their powers to restrain executive action (see also Collier, 1982; Bratton and van de Walle, 1997). Some palpable gains in policy flexibility were observed – Botswana, for example, left the Rand Monetary Area (RMA) and subsequently achieved lower inflation than the RMA countries. But in most cases such gains appear to have been overwhelmed by increased macroeconomic instability and deterioration in the protections afforded to private investment (points developed in detail in Collier and Pattillo, 1999).¹¹

In the mid-1990s, the cross-country growth literature began to develop a broader institutional version of the governance critique, based on the tendency of African countries to cluster in the lower ranks of institutional performance measures that are correlated with growth on a global basis. The attentions of policy reformers meanwhile began to shift from conventional economic reforms to problems of improving public service delivery, reducing bureaucratic corruption and strengthening the rule of law. While economists still know relatively little about how durable improvements in public sector performance are achieved, three observations seem relevant to the African situation. First, when institutions are initially weak, the initiative of high-quality political leaders is critically important in determining how well or poorly existing institutions actually perform (Reinikka and Collier, 1999; Glaeser et al., 2004). Second, as suggested above (and as measured by shares in formal sector employment or total investment), African governments have tended to be large and overbearing rather than small and efficient. Third, institutions are known to display persistence: learning takes place and interests form around existing patterns of behavior (North, 1990). These observations suggest that much of the logic of the governance critique applies directly to the performance of public sector institutions over the 1960–2000 period. Wherever political leaders were unable to reconcile the benefits of a market-friendly institutional environment with their own priorities, institutional performance deteriorated and growth suffered.

Did Africa’s abrupt democratization between 1988 and 1994 improve the institutional ground of policy-making? Contested elections and a free press are among potentially important agencies of restraint cited by Collier (1991). Partly for this reason, our overall answer is a positive one.

Fractionalization, polarization and nation-building We next look at the impact on policy of the patterns of sub-national identity that existed at the

Table 66.3 *Dates of political independence, developing countries*

Region	<i>n</i>	Proportion of countries politically independent				
		10%	25%	50%	75%	All
SSA	46	1957	1960	1961	1966	1993
Other Developing	66	1830	1830	1946	1961	1981
of which:						
LAC	27	1818	1825	1840	1962	1981
ASIA	20	1816	1933	1948	1956	1975
MENAT	19	1816	1932	1948	1962	1971
Total	112	1822	1907	1960	1964	1993

Source: Gleditsch and Ward (1999) database, as compiled by Ndulu and O'Connell (2007), Table 1.2.

time of independence and that in many cases became the dominant mode of political mobilization and conflict. The countries of SSA came to political independence both later and more rapidly than those of other developing regions (Table 66.3). While only Ethiopia, Liberia and South Africa existed as independent states at the end of 1955, fully three-quarters of colonial Africa, representing the vast bulk of its population and GDP, had achieved political independence by 1966. In 1966 the average independent state in SSA had held sovereignty for fewer than ten years; its counterpart in the rest of the developing world had been independent for the better part of a century.

Colonial structures of political control were both arbitrary – with boundaries cutting across historical patterns of politics and trade – and effective. Their abrupt departure meant that the challenge of economic development was in many cases confounded from the outset with an acute problem of nation-building. Nigeria provides a telling example of the impact of *ex ante* regional polarization on political and economic development. But similar patterns of internal polarization, often created or reinforced in the encounter with conquering European powers, existed throughout the continent in 1960. Azam (2007) emphasizes the salience of coastal–interior cleavages in West Africa, operating as in Nigeria on a North–South axis and tending to separate a nomadic, pastoralist Muslim interior from a more sedentary, educated, Christianized coast. In the Horn of Africa, both Sudan (Arab North, Christian and Animist South) and Ethiopia (federated with richer Eritrea after World War II, to guarantee Ethiopia's access to the coast) have engaged in ethno-regionally based civil wars since the early 1960s. In Central Africa, Belgian favoritism towards

the Tutsi minority produced enduring cleavages that erupted in civil war in the 1990s. In South Africa, and in those portions of colonial Africa with large settler populations – including Kenya and Zimbabwe – race-based geographical discrimination was a matter of state policy; in these cases the pattern of *ex ante* polarization would eventually require determining the status of settler minorities.

While the salience of ethno-regional polarization was clear to political scientists in the early 1960s (for example, Carter, 1966), economists have only recently begun to come to grips with the implications of nation-building for African economic growth. Two approaches have been important. The first is due to Easterly and Levine (1997), who focused on fractionalization rather than on polarization *per se*. Easterly and Levine noted that the degree of ethno-linguistic fractionalization – measured by the probability that two randomly chosen individuals in a given country spoke a different first language – was extremely high in many African countries, by comparison with global norms. Moreover, on a global basis ethno-linguistically heterogeneous countries tended to grow more slowly, as a result of weaker public sector performance. Miguel (2004) reports a similar finding for Kenya and Tanzania, using data on local provision of public services. Collier (2000) finds, however, that the adverse impact of heterogeneity is strongly contingent on political institutions. In democracies, ethno-linguistic heterogeneity has no impact either on overall growth or on microeconomic efficiency (as measured by the economic return on World Bank projects), while in dictatorial regimes the adverse impact is strong.

Azam (1995, 2007) focuses directly on polarization, defining a polarized society as one like Nigeria's, in which there are two or three large sub-national ethnic groups that dominate population and politics in separate regions. Azam argues that in a situation of *ex ante* ethno-regional polarization, regionally-based redistribution may be required to buy off the threat of armed conflict. The existence of such a risk is consistent with the global evidence of Collier and Hoeffler (2004), who find the risk of civil war maximized under conditions of polarization: homogeneous societies have low exposure to civil war, but so do heterogeneous societies. In cases of *ex ante* polarization, then, the Azam analysis may force a reinterpretation of what is conventionally viewed, within the governance critique, as distortionary redistribution. If the absence of redistribution invites armed conflict and economic collapse, then a program that distorts efficiency relative to an irrelevant peaceful counterfactual may in fact be growth-promoting relative to the true counterfactual of civil war. This shifts the ground of the governance critique from redistribution *per se* to the instruments that are employed to achieve it. Political elites attempting to 'buy the peace' should be observed doing so transparently and credibly

(perhaps via constitutional means), and with a minimum of distortion; and they should simultaneously employ instruments directly targeted at reducing polarization.

Geography

Starting in the mid-1990s, Jeffrey Sachs and Adrian Wood began to build an empirical case for the adverse influence of resource endowments and geography on African economic growth. Wood argued that in a world of capital mobility, comparative advantage was determined by endowments of immobile factors: primarily unskilled labor, human capital and natural resources. Africa's rich endowment of natural resources relative to human capital implied a deep comparative advantage in the production and export of primary commodities (Wood and Berge, 1997). The failure of African countries to achieve competitiveness in manufacturing was therefore largely independent of trade policy or the quality of governance, although these factors may have affected the long-term evolution of factor endowments. Sachs argued that high transport costs and a hostile disease environment conspired to make capital accumulation and productivity growth much more expensive in Africa than elsewhere in the developing world (Sachs and Warner, 1995, 1997, 2001; Bloom and Sachs, 1998; Gallup and Sachs, 1999).

Distance and landlockedness African populations are internally fragmented and isolated from world trade by unusually large land distances, unhealthy lowland coastlines, a sparse network of ocean-navigable rivers, and multiple political borders (Gallup and Sachs, 1998; Faye et al., 2004). With its 48 economies, the region has by far the highest density of countries per land area of any developing region; on average, each country shares a border with four neighbors (Ndulu, 2004). Nearly 40 percent of the African population lives in countries that are landlocked or virtually so.¹² The unusual distance of African population concentrations from coastlines and ocean-navigable rivers seems to follow in part from the inland locations of water resources critical to agriculture, including the Great Lakes, major non-ocean-navigable rivers, and fertile rain-fed uplands.

The remoteness of African population concentrations may have severely limited the scope for Asian-style growth patterns based on proximity to global markets, scale economies and agglomeration. Africa's relatively more sparse distribution of population, significantly low population density and relatively lower rate of urbanization raises significantly the transport intensity of its economic activities (C. Kessides, 2005). Furthermore, the lower population density and urbanization tend to increase the amount of infrastructure investment required to produce similar levels of income (Esfahani and Ramirez, 2003).

Unusually high internal transport costs accentuate this remoteness. Limão and Venables (2001) estimate that it costs nearly twice as much for the median African country to move a 40-foot container from a coastal port to its in-country destination, as it does for countries in other developing regions.¹³ These costs are particularly damaging for manufacturing, where the share of traded intermediate inputs is relatively large. They also penalize physical capital accumulation by raising the relative price of investment. Investment in Africa is unusually expensive in terms of local income, so that a given national saving rate delivers a lower increment to real capital accumulation in Africa than in other regions. The average relative price of investment goods for sub-Saharan Africa was 70 percent higher than for Organisation for Economic Co-operation and Development (OECD) countries or East Asia. Artadi and Sala-i-Martin (2003) find that the high relative price of investment goods reduces Africa's predicted growth rate by 0.44 percent on an annual basis, holding saving effort constant.

Notwithstanding variations across countries in the region, for most African countries distance from their primary markets and the high transport intensities of their products (low value, high weight and sparsely produced) are major impediments for production and trade (Esfahani and Ramirez, 2003). Using a gravity model Limão and Venables (2001) estimated the elasticity of trade with respect to transport costs, and found it typically to be quite high at -3 . Distance to key markets is an important impediment to trade as expected, but in their model poor infrastructure (measured by an index combining road, rail and telecom density) accounts for 40 percent of the predicted transport cost for coastal countries and up to 60 percent for landlocked countries. The median landlocked country has only 30 percent of the trade volume of a median coastal country. What is also striking from this study is that, holding activity levels and direct distances between trading partners constant, improving internal infrastructure within the landlocked country itself is as important as improving the infrastructure in the transit country.

Landlockedness adds a political dimension to remoteness. Transport costs now depend crucially on the infrastructure investments and pricing policies of coastal neighbors. These neighbors may also be important but unreliable hosts for the export of labor services. Sachs and Warner (1997) and others find that landlocked status reduces predicted growth by up to 1 percent per year on a global basis.

Disease burden Sachs and Warner (2001) and Masters and McMillan (2001) emphasize the high burden of human and animal disease in tropical climates and its impact on life expectancy, human capital formation, labor

force participation and economic growth. Ninety-two percent of SSA lies within the tropics, as compared to 60 percent for East Asia. Following Sachs and Warner (2001), Artadi and Sala-i-Martin estimate the foregone growth in Africa as a result of malaria prevalence at 1.25 percent per annum, a figure that surely reflects the influence of other highly correlated aspects of the health environment. Acemoglu et al. (2001) take a very different, institutions-based approach to linking disease burden with growth. They argue that the quality of contemporary institutions reflects the nature of the institutions introduced by European powers during the colonial period. Where the local disease environment was inhospitable, Europeans introduced extractive institutions, leaving a legacy of predation and violence that continues to undermine the rule of law and the security of property. Where the disease environment was favorable to European settlement, colonial regimes set up institutions conducive to long-term growth (see also Easterly and Levine, 2003).

Demography

In contrast to the experience of other regions, a fertility transition has not happened in Africa despite sharp reductions in infant mortality since the late colonial period and (until the HIV/AIDS epidemic starting in the 1990s) gradual improvements in life expectancy across the age distribution. The distinctive demographic features of African countries weigh unusually heavily on national saving and undermine the building up of the human capital needed for growth (Bloom and Sachs, 1998). Indeed, although we saw earlier that human development indicators have not diverged as strongly as income levels when comparing SSA averages with averages for other developing regions, these indicators have nonetheless deteriorated in relative terms. Enhancing human capacity by increasing the longevity of working life and improving skills and organizational effectiveness are important components of a strategy to close the growth differential with other regions.

Two distinct consequences of continued high fertility stand out. Firstly, the average population growth rate is at least one full percentage point above that for other developing regions. This increases the amount of national saving required to achieve any given increase in human and physical capital stocks per capita. It also increases the age dependency ratio, which reduces the per capita purchasing power associated with any given level of output per worker. High dependency ratios may also undermine the quality of human capital accumulation by spreading educational resources more thinly.

Secondly, until the early 1990s, rapid population growth produced not just a high but also a rising age dependency ratio. As we saw earlier, as Africa's age dependency ratio gradually increased, the rest of the developing world

experienced a fertility transition that lowered population growth rates sharply and gradually reduced the ratio of dependents to working population over time.

In a regression-based counterfactual exercise, O'Connell and Ndulu (2000) estimate that Africa's average growth is reduced by 0.85 percentage points relative to the sample mean, and by nearly 1.5 percentage points relative to East Asia as a consequence of its distinctive demographic patterns. This situation is made worse by the fact that HIV/AIDS has become epidemic and added to the burden particularly for survivors. HIV/AIDS patients in Africa account for 60 percent of the world's people living with HIV/AIDS. This has a profound social and economic impact due to the large number of premature deaths of people in their prime age of employment and parenting.

A window of opportunity and sustained growth since the mid-1990s

Over the decade since 1995, 16 countries have had annual GDP growth of 5 percent or higher, by comparison with only five during the previous decade (Table 66.4).¹⁴ These countries account for 35 percent of the sub-Saharan Africa population. Some of the fastest-growing countries have also done relatively well in terms of poverty reduction, as demonstrated by a group of eight low-income African countries that grew at an average rate of 2.9 per capita per year and reduced poverty headcounts at an annual rate of 1.5 percentage points (World Bank, 2005).¹⁵ Excluding the oil-producing countries, income per capita in the fastest-growing one-third of African countries grew at a median rate of 2.8 percent over the five years 1999–2004. During the same period, the slowest-growing countries – predominantly those affected by conflict – saw their economies contract at a median rate of 1.8 percent. These host 21 percent of the region's population. There has nonetheless been a striking decline in the number of countries posting negative growth rates of total GDP: four during the most recent period, down from 13 during the first half of the 1990s. In the middle of the growth distribution (accounting for some 13 percent of Africa's population) the median growth rate was 1.0 percent per capita over 1999–2004.

These trends reflect important changes that are taking place across the continent. Policies and institutions are improving, peace and security is returning to the region, and African governments are increasingly taking control of their own economic destiny. Increased political participation and competition are giving Africans a greater stake in their own future. Demographic trends appear to have taken the first turn towards a transition that will reduce pressures on fiscal resources, encourage savings and support productivity growth. We elaborate below on each of these areas of progress.

Table 66.4 Average GDP growth rates: sub-Saharan Africa

	85/94	95/04
Angola	-0.9	7.7
Benin	2.9	5.0
Botswana	8.2	5.5
Burkina Faso	3.6	4.7
Burundi	2.7	0.0
Cameroon	-1.6	4.6
Cape Verde	4.4	5.8
Central African Republic	0.2	1.8
Chad	4.2	7.4
Comoros	1.1	1.7
Congo, Dem. Rep.	-3.6	-0.6
Congo, Rep.	-0.4	3.1
Côte d'Ivoire	1.1	2.1
Equatorial Guinea	2.9	22.3
Eritrea	-	2.3
Ethiopia	1.6	5.2
Gabon	1.5	2.3
Gambia, The	2.9	4.1
Ghana	4.6	4.5
Guinea	4.0	3.9
Guinea-Bissau	3.5	0.5
Kenya	3.6	2.0
Lesotho	5.0	3.3
Liberia	-18.7	16.6
Madagascar	1.2	2.9
Malawi	1.7	4.1
Mali	1.7	5.9
Mauritania	3.3	4.3
Mauritius	6.4	5.0
Mozambique	4.2	8.0
Namibia	3.5	3.4
Niger	2.2	3.7
Nigeria	4.7	3.6
Rwanda	-4.3	10.9
Sao Tome and Principe	1.7	3.2
Senegal	2.2	5.1
Seychelles	5.4	2.2
Sierra Leone	-1.7	0.0
Somalia	-	-
South Africa	0.6	2.7
Sudan	3.6	6.2
Swaziland	6.6	2.9

Table 66.4 (continued)

	85/94	95/04
Tanzania	3.1	5.1
Togo	1.3	4.0
Uganda	4.6	6.7
Zambia	0.6	2.8
Zimbabwe	3.7	-1.5

Source: World Bank data.

A significant and durable improvement in the policy and institutional environment

During the last decade a large number of reforming African countries have re-established sustained macroeconomic stability, committed credibly to more open trade regimes, and consolidated market-based economic reforms that have improved the conditions for private sector involvement in the economy.

Sustained macroeconomic stability has returned to a large number of countries in the region, as evidenced by significantly lower inflation, narrower fiscal and external trade deficits, and a widespread move to Article VIII status in the International Monetary Fund (IMF), implying a commitment to currency convertibility for current account transactions – a status hardly conceivable only a decade earlier.

Consumer price inflation has persistently and sharply fallen within a decade, from 27 percent in 1995 to about 6 percent by 2004. This has resulted from a combination of significantly stronger fiscal controls across a wide range of countries and a shift by central banks towards a focus on price stability as the primary goal (Ndulu, 2004). In the median African country, government spending as a proportion of GDP fell sharply in the past decade, as it has in other developing countries in the world. During the 1990s, fiscal deficits among the 31 low-income African countries for which comparable data exist dropped from double digits to 5.2 percent before grants and only 2.5 percent after grants (World Bank, 2000). The residual deficit has been financed largely through non-inflationary sources. Where financing through external grants is assured, fiscal deficits (before grants) have tended to be higher, but without destabilizing the macroeconomic environment. At the turn of the century black market premia were extremely low across the continent, averaging just 4 percent outside of a few countries like Zimbabwe in acute political turmoil. Through unilateral trade reforms, African countries have also compressed both the tariff rates and categories with average tariff rates of 15 percent. In these respects the

continent now more resembles other developing regions, where reforms have been pursued in earnest for prolonged periods.

We also use here the Country Policy and Institutional Assessment (CPIA) data from the World Bank to show the extent of improvement in the policy and institutional environment since the mid-1990s. Annually the World Bank assesses the quality of borrowers' policy and institutional performance in areas relevant to economic growth and poverty reduction. These assessments began in the late 1970s but the criteria of good performance have evolved over time. While in earlier years assessments focused mainly on macroeconomic policies, they now include other aspects such as institutional quality and governance. CPIA scores for individual countries have been found to be highly correlated with private-sector country ratings, such as the Institutional Investor, International Country Risk Guide (ICRG) and Euromoney ratings (Easterly, 1993). The CPIA scores range between 0 and 6 and are benchmarked globally. A score of 3.5 is broadly considered as a threshold of good performance. The average CPIA score for African countries rose sharply starting in the mid-1990s, and between 1997 and 2004 the number of countries scoring above the 3.5 threshold tripled, from 5 to 15 – the latter number accounting for nearly a third of all countries in the region.

Peace and security is spreading in the region

After protracted periods of conflict, peace and security has returned to many parts of the region. Southern Africa entered the millennium as a region of stability following the resolution of long-standing conflicts in Mozambique, Namibia, Zimbabwe and Angola (2002) and the transition to majority rule in South Africa. West Africa likewise has seen the end of conflicts in Chad, Sierra Leone and Liberia. In Eastern Africa and the Horn of Africa, the resolutions of conflicts in Burundi and more tentatively in Sudan and Somalia hold out the prospect of durable change. Africa's progress is significant relative to other regions of the world, as reflected in global data on the incidence of civil war. The incidence of civil war shows a sharp increase in both absolute and relative terms in SSA in the early 1990s, a period that coincided with the wave of first democratization documented above. Since the mid-1990s, however, there has been a sharp decline in the proportion of countries under civil conflict in the region; and the proportion of African population in such countries has fallen even more sharply, dipping below the average of other developing regions at the turn of the century. This positive change is corroborated by a new global database on the incidence of violent conflicts (Gleditsch, 2004). Collier and Hoeffler (2004), using this data, show that immediately after the end of the Cold War the incidence of wars declined. The number

of wars peaked in Africa in 1992 and since then the number of civil wars seems to be declining.

Increased political participation gives Africans a greater stake in their own future, laying a stronger foundation for domestic accountability and policy restraint

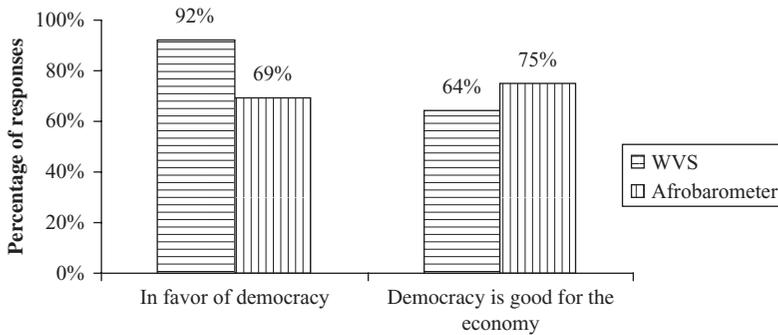
During 1990–94, SSA experienced a sharp wave of democratization (Bates, 2007). By the mid-1990s, this episode had fully made up for nearly three decades of absolute and relative deterioration, placing the region well above the norms prevalent in the rest of the developing world during 1960–85.

Political competition and participatory processes improved by more in Africa during the 1990s than in other regions. In 1982, only one-tenth of African countries and two-tenths of other developing countries had competitively elected executives. As late as 1991, Africa showed virtually no improvement, while other developing countries had doubled their figure to 40 percent. By 1995, however, the gap was nearly closed, despite continuing increases in other regions. In 2002, Africa was ahead of the other regions by about eight percentage points.

The political democratization drive in Africa has created space for peaceful regime changes, deeper debates about societal development visions, and greater respect for human rights. Although we observed earlier that the initial wave of democratization was associated with a sharp rise in civil conflict in the region, this situation appears to have reversed itself as democratic practices have taken hold. The democratization process has also raised the expectations from citizens throughout the region. Analysis of recent Afrobarometer surveys and the World Values Survey show that the majority of Africans believe democracy is good for the economy. They also prefer democratic political systems to authoritarian alternatives, as indicated in Figure 66.3. The African public expects democracy to deliver access to the basic necessities of life, including food, water, shelter, and also education. These ‘values’ surveys also show that Africans care about equity and public action to reduce poverty. They report discomfort with wide wealth differentials and a strong commitment to political equality (‘voting not a privilege of the better-educated’).

African governments are pursuing collective action to improve the region’s attractiveness and taking increasing control of their economic destiny

There is now a strong revival in regional integration initiatives in Africa, with a change in focus from preoccupation with preferential trade arrangements to an approach that emphasizes market integration and promoting the region as an attractive investment destination for foreign and African



Source: World Value Survey (1999–2002 wave), covering six countries (<http://www.worldvaluessurvey.org>), and Afrobarometer survey 2001–2003, covering 12 countries (<http://www.afrobarometer.org>).

Figure 66.3 African Political Values

capital. The African Union and the New Partnership for Africa's Development (NEPAD) have embraced the latter two objectives. These objectives can be met if Africa as a region can: (1) achieve a critical mass of countries with a policy environment friendly to capital accumulation and private business; (2) improve cross-country infrastructure links; (3) moderate the risks faced by domestic and foreign capital, and employ risk-mitigating instruments more effectively; and (4) strengthen as well as retain its pool of human skills.

As part of the effort to improve the region's collective reputation and attractiveness, African governments are taking action to improve governance and connectivity under the African Union (AU) and the NEPAD initiatives. These initiatives are designed to: (1) push African countries to be assertive about ownership and to assume leadership and accountability for their development programs; (2) improve the reputation of the region through certification of good practices in governance for a critical mass of African countries under the African Peer Review mechanism; (3) increase regional connectivity to improve capacity to trade within the region and with the outside world through regional initiatives to scale up collaborative effort in improving infrastructure; and (4) enhance the capacity of a rationalized system of regional bodies to provide regional public goods – such as cross-country transportation and power-sharing networks, coordination in managing pandemics such as HIV/AIDS and malaria, and protection of regional commons such as the Nile river basin and the Great Lakes.

Implications for future development strategies

We have anchored the analysis of Africa's development experience in the overall weakness of the region's growth record, the juxtaposition of this record with a population explosion, and the variability of experience across countries and over time. If recent improvements are to be sustained and deepened, a combination of addressing the fundamental bottlenecks to scaling-up growth, and accelerating the demographic transition now slowly under way, is fundamental. The wide diversity of opportunities and constraints in the region rules out generalizations about the country-level growth strategies appropriate to these goals. Nevertheless there are a few broad lessons worth highlighting by way of concluding remarks.

The development constraints African countries face are neither static nor decisive. Geographical disadvantages and natural-resource dependence are not destiny, as their effects can be offset or ameliorated. Botswana, the fastest-growing economy in Africa (and among the fastest globally) since its independence in 1966, presents a striking example. It is landlocked and natural-resource dependent. Arguably, the strength of its state capacity, together with its being part of Southern Africa's relatively effective infrastructure system, customs union and monetary area (for a long period), helped offset the negative effects of remoteness and served as a commitment instrument against rent-seeking.

The historical analysis showed clearly that Africa virtually missed the rapid development that has taken place in other developing regions in the closing four decades of the twentieth century and indeed the region can be considered the last frontier of the global development challenge. Being a late starter has its advantages and disadvantages. Access to knowledge from development experience and technological progress presents an opportunity to fast-track the development process and leapfrog. At the same time, Asia's cumulative success presents a challenge to the competitiveness of late-starting Africa, as trade preferences are eroded and opportunities to learn before facing intense competition shrink. Among the resource-poor coastal economies, the incumbency of the successful Asian coastal economies has probably created a more challenging playing field for African export diversification starting in the 1990s than existed in earlier decades.¹⁶

We can categorize the constraints to growth discussed above into three groups: risks, transactions costs and capacity. Risk is largely associated with macroeconomic instability and absence of credibility to commit due to weaknesses in governance and institutions. The higher transactions costs in the region are largely associated with its unfriendly geography, climate and bureaucratic processes. Capacity constraints relate primarily to low human capital, partly associated with the late demographic transition and weak institutions.

There is no doubt that raising the level and efficiency of investment is critical if Africa is to close the gap in growth with other regions of the world. Getting the right policy environment in place and sustaining such an environment is a key signaling instrument for credibility to investors (both local and foreign). Indeed, given the central role of the modern state in defining the incentive environment for private economic activity, the failure to engender sustained growth must ultimately be traced to unsuccessful policy choices. Collier and O'Connell (2007) estimate the contribution of anti-growth policies to the growth differential between Africa and other regions (p. 24); taking 40 years of African growth experience as a whole and controlling for differences in the composition of opportunities, they conclude that policies inimicable to growth account for more than half of the overall growth differential with the rest of the developing world, or as much as 1.8 percentage points out of an overall (population-weighted) differential of 3.5 percent. This result is confirmed by a regression analysis in which they estimate the effect of anti-growth policies, controlling for shocks and differences in growth opportunities across countries. This magnitude of relative importance of policy mistakes in explaining the growth performance differential with other regions is corroborated by evidence from cross-country growth studies reviewed extensively in O'Connell and Ndulu (2000).¹⁷

Notwithstanding the importance of policy choices in explaining Africa's growth differential with other developing regions, however, a substantial portion of the growth differential is accounted for by geographically based proxies for differential growth opportunities (Collier and O'Connell, 2007, pp. 81–8). Separating landlocked and resource-poor, coastal and resource-poor, and resource-rich countries, Collier and O'Connell find that while African countries tended to underperform in each category, nearly a third of Africa's overall growth shortfall is associated with the unusually high share of African countries that are either landlocked and resource-scarce or resource-rich – both relatively low opportunity categories as compared to the coastal and resource-scarce group, based on the global evidence. The underperformance was most severe for Africa's coastal resource-scarce economies (suggesting that missed opportunities dominate the story) and least severe for Africa's land-locked resource-scarce economies (suggesting poor potential). Two sources of this differential are important – proneness to policy errors and the higher cost of doing development business, discussed above.

Proneness to policy mistakes is likely to be more important in resource-rich and coastal countries where the rent-seeking stakes are higher (from resource and trade rents), while higher transactions costs of economic activities from geographical disadvantages are likely to be more important

in landlocked countries. The convergence of resource wealth and ethno-regional fragmentation in many of the countries of SSA presents a particularly high risk of adoption of anti-growth policy regimes. The management of resource rents under conditions of ethnic diversity is one of the most important and distinct policy challenges for SSA.

A third challenge is dealing with the consequences of a dramatically delayed demographic transition in Africa compared to other regions. The fertility rate began to fall in Africa in the mid-1980s, suggesting entry into the final phase of the demographic transition. But unlike the experience in other regions, the pace of this transition is very slow; and the HIV/AIDS pandemic complicates the situation further. Although population changes are longer-term phenomena, as Srinivasan (1988) urges, it may be worthwhile to look at potential policy responses in the shorter term to help accelerate the demographic transition. These include education policy, population policy and gender equality to induce greater attention to the quality of children.

Much of the above discussion has employed *ceteris paribus* counterfactuals in order to isolate the individual effects on growth of policy mistakes, underprovision of public goods or weaknesses in human capital. If there are critical synergies across these factors, or if there are sharply increasing returns to individual state variables over some interval (as suggested by Azariadis and Drazen, 1990 and Berthélemy, 2005 in the case of human capital and Sachs et al., 2004 for both human capital and public infrastructures), then the achievement of rapid growth may require complementing policy improvement with a 'big push' to deal with poor governance, conflict and insecurity, and inadequate infrastructure. In terms of their ambition, contemporary versions of the big push are not novel; the record of Africa's early decades is replete with the ambitions of visionary leaders who sought to engineer a sharp break from patterns of specialization and distribution inherited from the colonial era (Ndulu, 2007). What is different this time around, and provides a critically important window of opportunity for stakeholders, is that the core functions of market-friendly governance are widely understood to be among the binding constraints.

Acknowledgment

We thank Lopa Chakraborti for research assistance.

Notes

1. Sala-i-Martin (2002), for example, documents a reduction by nearly two-thirds between 1970 and 1998 alone.
2. Here and throughout the chapter, we confine group totals to countries with continuously available data. This eliminates potentially misleading compositional effects within

- groups, but it also means that group composition can differ by table or variable (for example, World Bank data are available for only 35 countries in SSA, while Maddison provides estimates for 52). For each variable (or table, as indicated), we report group totals for the full set of countries with continuously available data.
3. Nigeria, Ethiopia, Democratic Republic of the Congo and South Africa together account for 45 percent of African population among countries with available data; these economies contracted at an average rate of 1.8 percent over 1974–94.
 4. We exclude oil countries, which grew at a median per capita rate of 3.4 percent. The upper, middle and lower thirds of the growth distribution comprise 35, 13 and 21 percent of Africa's population, respectively.
 5. A recent contribution in this line is van de Walle's *The Politics of Permanent Crisis (2001)*, discussed further below.
 6. On within-Africa variation in agricultural policy and export taxation, see Mkandawire and Soludo (2000). On the weakness of within-country links from policy to long-term growth, see Easterly et al. (1993). On the ideological (as opposed to self-interested) motivations of African policy-makers, see Ndulu (2007), who notes the profound influence of Fabian socialism and dependency theory on founding African leaders. On the relative weakness of initial urban interests see Ndulu and O'Connell (1999), who observe that urban interests were nonetheless created by policy in some cases, and thereby came to undermine subsequent prospects for reform (see also the discussion of van de Walle, 2001 in the text).
 7. Sahn (1996) argues cautiously, on the basis of calibrated general equilibrium models, that if market-based reforms had been fully implemented in the 1980s, their impact would indeed have been mildly progressive.
 8. In McMillan's analysis, policy-makers seek to maximize the revenue from taxing agricultural exports. They face a time-consistency problem, however: once farmers have sunk planting costs in the hope of receiving high producer prices, there is an incentive for policy-makers to pay very low producer prices that cover only the costs of harvesting. Evidence on crop- and country-specific export tax rates confirms that this incentive is strongest when policy-makers have urgent revenue needs and short planning horizons, and when the ratio of harvest to planting costs is small.
 9. In a celebrated article, Acemoglu et al. (2001) argued that colonial regimes brought development-oriented institutions only where local health conditions supported the establishment of a large settler presence (see p. 507).
 10. Migdal (1988) describes the development paradigm spanning this period as one in which the state is the '*primum mobile*' of socio-economic progress. The idea of 'developmentalism' and the idea of state intervention were seen as inseparable, and policies and planning were seen as offering boundless possibilities for social engineering. It was taken for granted by multilateral and bilateral development agencies that the state had a pivotal role to play in transforming societies from backwardness to modernity (Ljunggren, 1993, pp. 7–8).
 11. Collier argued that in the resulting situation of executive dominance, effective restraints in the areas of trade and monetary policy would have to mimic the reciprocal and supra-national structure of international trade agreements, where countries reciprocally commit to growth-promoting policies and to penalty structures capable of enforcing them. Donor conditionality, he argued, was ill-suited to fill the institutional vacuum; donors had their own constituencies and could not credibly threaten to terminate aid based on poor policy performance. As examples of partially successful supra-national arrangements he cited the maintenance of low inflation in the CFA countries (the 13 members of 2 monetary zones issuing respectively the West African CFA – Communauté financière d'Afrique – franc and the Central African CFA – Coopération financière en Afrique centrale – franc) and Rand Monetary Area, and the avoidance of highly distorted trade regimes by members of the Southern African Customs Union.
 12. The Democratic Republic of Congo belongs in the latter category; perhaps also the Sudan, with its vast internal territory and limited access to its Red Sea coastline, and Ethiopia before the independence of Eritrea in 1994.

13. Limão and Venables (2001) estimate the median transport cost for a 40-foot container, from coastal port to destination (including transshipment), at \$7600 for African countries. The comparable figures for Latin America and the Caribbean, East and South Asia, and the Middle East and North Africa are \$4600, \$3900, and \$2100.
14. The 16 countries are Angola, Benin, Botswana, Cape Verde, Ethiopia, Mali, Mauritius, Mozambique, Rwanda, Senegal, Tanzania, Uganda, Chad, Equatorial Guinea, Liberia and Sudan (see Table 66.4)
15. These countries are Senegal, Mozambique, Burkina Faso, Cameroon, Uganda, Ghana and Cape Verde.
16. Collier and O'Connell (2007) find, however, that African coastal economies that maintained market-friendly policy environments for longer periods after 1980 achieved substantially greater export diversification into manufactures and services.
17. There is a wide range of other studies that corroborate this same conclusion. Ndulu (1998) reported the results of counterfactual simulations of Africa's growth performance if conditions obtaining in East Asia were present in the region. Using results earlier obtained by Elbadawi and Ndulu (1995), Easterly and Levine (1997) and Elbadawi et al. (1997), a better policy environment as obtaining in East Asia would have raised growth by an additional 1.5–2.6 percent.

References

- Acemoglu, Daron, Simon Johnson and James Robinson (2001), 'The Colonial Origins of Comparative Development: An Empirical Investigation', *American Economic Review*, 91 (December): 1369–1401.
- Adam, Christopher S. and Stephen A. O'Connell (1999), 'Aid, Taxation and Development in Sub-Saharan Africa', *Journal of Economics and Politics*, 2 (3): 225–53.
- Ake, Claude (1996), *Democracy and Development in Africa*, Washington, DC: Brookings Institution.
- Alesina, Alberto and Roberto Perotti (1994), 'The Political Economy of Growth: A Critical Survey of the Recent Literature', *World Bank Economic Review*, 8 (3), September: 351–71.
- Artadi, Elsa V. and Xavier Sala-i-Martin (2003), 'The Economic Tragedy of the XXth Century: Growth in Africa', National Bureau of Economic Research Working Paper 9865, July.
- Azam, Jean-Paul (1995), 'How to Pay for the Peace? A Theoretical Framework with Reference to African Countries', *Public Choice*, 83 (1/2): 173–84.
- Azam, Jean-Paul (2007), 'The Political Economy of Redistribution', in B. Ndulu, S.A. O'Connell, R.H. Bates, P. Collier and C.C. Soludo (eds), *The Political Economy of Economic Growth in Africa, 1960–2000, Volume 1*, Cambridge: Cambridge University Press, pp. 225–48.
- Azariadis, Costas and Allan Drazen (1990), 'Threshold Externalities in Economic Development', *Quarterly Journal of Economics*, 105: 501–26.
- Bates, Robert (1981), *Markets and States in Tropical Africa*, Berkeley, CA: University of California Press.
- Bates, Robert (2007), 'Domestic Interests and Control Regimes', in B.J. Ndulu, S.A. O'Connell, R.H. Bates, P. Collier and C.C. Soludo (eds), *The Political Economy of Economic Growth in Africa, 1960–2000, Volume 1*, Cambridge: Cambridge University Press, pp. 176–207.
- Bates, Robert and Shantayanan Devarajan (2001), 'Framework Paper on the Political Economy of African Growth', AERC Working Paper Series.
- Berthélemy, Jean-Claude (2005), 'Conférence du G8: Pour chaque dollar de dette annulée, autant sera supprimé dans l'aide', *Marché tropicaux et méditerranéens*, 60 (3112), July: 6–8 [interview].
- Berthélemy, Jean-Claude and Ludvig Soderling (2001), 'The Role of Capital Accumulation, Adjustment and Structural Change for Economic Take-Off: Empirical Evidence from African Growth Episodes', *World Development*, 29 (2): 323–43.
- Berthélemy, Jean-Claude and Ludvig Soderling (2002), 'Will There Be New Emerging-Market Economies in Africa by the Year 2020?' IMF Working Paper WP/02/131, August.

- Bloom, David and Jeffrey D. Sachs (1998), 'Geography, Demography, and Economic Growth in Africa', *Brookings Papers on Economic Activity*, 2: 207–73.
- Bratton, Michael and Nicolas van de Walle (1997), *Democratic Experiments in Africa*, Cambridge: Cambridge University Press.
- Carter, Gwendolyn (1966), *National Unity and Regionalism in Eight African States*, Ithaca, NY: Cornell University Press.
- Collier, Paul (1991), 'Africa's External Relations, 1960–1990', *African Affairs*, 90 (3): 339–56.
- Collier, Paul (2000), 'Ethnicity, Politics and Economic Performance', *Economics and Politics*, 12 (3), November: 225–45.
- Collier, Paul and Anke Hoeffler (2004), 'Greed and Grievance in Civil War', *Oxford Economic Papers*, 56 (4): 563–95.
- Collier, Paul and Stephen A. O'Connell (2007), 'Opportunities and Choices', in B.J. Ndulu, S.A. O'Connell, R.H. Bates, P. Collier and C.C. Soludo (eds), *The Political Economy of Economic Growth in Africa, 1960–2000, Volume 1*, Cambridge: Cambridge University Press, pp. 76–136.
- Collier, Paul and Catherine Pattillo (eds) (1999), *Investment and Risk in Africa*, London: Macmillan.
- Collier, Ruth Berins (1982), *Regimes in Tropical Africa: Changing Forms of Supremacy 1945–75*, Berkeley, CA: University of California Press.
- Diamond, M. (1977), 'Ethics and Politics: The American Way', in R.H. Horowitz (ed.), *The Moral Foundations of the American Republic*, Charlottesville, VA: University Press of Virginia.
- Easterly, William (1993), 'How Much Do Distortions Affect Growth?' *Journal of Monetary Economics*, 32 (2): 187–212.
- Easterly, William and Ross Levine (1997), 'Africa's Growth Tragedy: Policies and Ethnic Divisions', *Quarterly Journal of Economics*, 112: 1203–50.
- Easterly, William and Ross Levine (2003), 'Tropics, Germs, and Crops: How Endowments Influence Economic Development', *Journal of Monetary Economics*, 50 (1): 3–39.
- Easterly, William, Michael Kremer, Lant Pritchett and Lawrence Summers (1993), 'Good Policy or Good Luck? Country Growth Performance and Temporary Shocks', *Journal of Monetary Economics*, 32 (3): 459–84.
- Elbadawi, I.A. and B.J. Ndulu (1995), 'Growth and Development in Sub-Saharan Africa: Evidence on Key Factors', paper presented at the 1995 World Congress of the International Economic Association, Tunis, 17–22 December.
- Elbadawi, I.A., B.J. Ndulu and N.S. Ndung'u (1997), 'Macroeconomic Performance in Sub-Saharan Africa in a Comparative Setting', presented at AERC Workshop on Comparative African and East Asian Development Experiences, in South Africa, 3–6 November.
- Esfahani, Hodi Salehi and Maria Teresa Ramirez (2003), 'Institutions, Infrastructure, and Economic Growth', *Journal of Development Economics*, 70 (2): 443–77.
- Faye, Michael, John McArthur, Jeffrey Sachs and Thomas Snow (2004), 'The Challenges Facing Landlocked Developing Countries', *Journal of Human Development*, 5 (1): 31–68.
- Fieldhouse, David K. (1986), *Black Africa, 1945–1980: Economic Decolonization and Arrested Development*, London: Allen & Unwin.
- Fosu, Augustin K. (2002), 'Political Instability and Economic Growth: Implications of Coup Events in Sub-Saharan Africa', *American Journal of Economics and Sociology*, 61 (1): 329–48.
- Gallup, John and Jeffrey Sachs (1998), 'Development and Poverty in a Global Age: Location, Location: Geography and Economic Development', *Harvard International Review*, 21 (1): 127–78.
- Gallup, John Luke and Jeffrey D. Sachs with Andrew Mellinger (1999), 'Geography and Economic Development', in Boris Pleskovic and Joseph E. Stiglitz (eds), *Annual World Bank Conference on Development Economics 1998*, Washington, DC: The World Bank.
- Glaeser, Edward L., Rafael La Porta, Florencio López-de-Silanes and Andrei Shleifer (2004), 'Explaining Growth: Institutions, Human Capital, and Leaders', paper presented at the Brookings Panel, March.
- Gleditsch, Kristian Skrede (2004), 'A Revisited List of Wars Between and Within Independent States, 1816–2002', *International Interactions*, 30: 231–62.

- Gleditsch, Kristian S. and Michael D. Ward (1999), 'Interstate System Membership: A Revised List of the Independent States Since 1816', *International Interactions*, 25: 393–413.
- Hall, Robert and Charles Jones (1999), 'Why Do Some Countries Produce So Much More Output Per Worker Than Others?' *Quarterly Journal of Economics*, 114 (1): 83–116.
- Humphreys, M. and R. Bates (2001), 'Economics Policies in Africa: The Good, the Bad and the Ugly', paper presented at the 97th Annual Meeting of the American Political Science Association.
- Karmarck, Andrew M. (1971), *The Economics of African Development*, New York: Praeger Publishers.
- Kessides, Christine (2005), 'The Urban Transition in Sub-Saharan Africa: Implications for Economic Growth and Poverty Reduction', *Africa Region Working Paper Series No. 97*, December, Washington, DC: World Bank.
- Kessides, Ioannis N. (2005), 'Infrastructure Privatization and Regulation: Promises and Perils', *World Bank Research Observer*, 20 (1): 81–108.
- Limão, N. and A.J. Venables (2001), 'Infrastructure, Geographical Disadvantage, Transport Costs, and Trade', *World Bank Economic Review*, 15 (3): 451–79.
- Ljunggren, Borje (1993), 'Market Economies under Communist Regimes: Reform in Vietnam, Laos and Cambodia', in B. Ljunggren (ed.), *The Challenge of Reform in Indochina*, Cambridge, MA: Harvard University Press.
- Lucas, Robert E. (2003), 'The Industrial Revolution: Past and Future', *The Region*, Minneapolis, MN: Annual Report of the Federal Reserve Bank of Minneapolis.
- Maddison, Angus (2001), 'The World Economy: A Millennial Perspective', Development Centre Seminars. Paris and Washington, DC: Organisation for Economic Co-operation and Development.
- Masters, William and Margaret S. McMillan (2001), 'Climate, Scale and Economic Growth', *Journal of Economic Growth*, 6 (3): 167–86.
- McMillan, Margaret (2001), 'Why Kill The Golden Goose? A Political-Economy Model Of Export Taxation', *Review of Economics and Statistics*, 83 (1): 170–84.
- Migdal, Joel S. (1988), *Strong Societies and Weak States: State–Society Relations and State Capabilities in the Third World*, Princeton, NJ: Princeton University Press.
- Miguel, E. (2004), 'Tribe or Nation? Nation Building and Public Goods in Kenya versus Tanzania', *World Politics*, 56 (3): 327–62.
- Mkandawire, Thandika and Charles C. Soludo (2000), *Our Continent, our Future: African Perspectives on Structural Adjustment*, Trenton, NJ: Africa World Press Inc.
- Ndulu, B.J. (1998), 'From Stabilization to Sustained Growth in Africa', *International Politics and Society*, Online, Vol. 1.
- Ndulu, B.J. (2004), 'Infrastructure, Regional Integration and Growth in Sub-Saharan Africa', AERC Plenary Session, November.
- Ndulu, B.J. (2007), 'The Evolution of Global Development Paradigms and their Influence on African Economic Growth', in B.J. Ndulu, S.A. O'Connell, R.H. Bates, P. Collier and C.C. Soludo (eds), *The Political Economy of Economic Growth in Africa, 1960–2000, Volume 1*, Cambridge: Cambridge University Press, pp. 315–47.
- Ndulu, B.J. and S.A. O'Connell (1999), 'Governance and Growth in Sub-Saharan Africa', *Journal of Economic Perspectives*, 13 (3): 41–66.
- Ndulu, B.J. and S.A. O'Connell (2007), 'Policy Plus: African Growth Performance, 1960–2000', in B.J. Ndulu, S.A. O'Connell, R.H. Bates, P. Collier and C.C. Soludo (eds), *The Political Economy of Economic Growth in Africa, 1960–2000, Volume 1*, Cambridge: Cambridge University Press, pp. 3–75.
- North, D.C. (1990), *Institutions, Institutional Change and Economic Performance*, Cambridge: Cambridge University Press.
- O'Connell, Stephen A. and Benno J. Ndulu (2000), 'Explaining African Economic Growth: A Focus on Sources of Growth', AERC Working Paper Series.
- Pritchett, Lant (1998), 'Patterns of Economic Growth: Hills, Plateaus, Mountains, and Plains', Policy Research Working Paper Series 1947, World Bank.
- Pritchett, Lant (2000), 'Understanding Patterns of Economic Growth: Searching for Hills among Plateaus, Mountains, and Plains', *World Bank Economic Review*, 14 (2): 221–50.

- Reinikka, R. and P. Collier (eds) (1999), *Uganda's Recovery: The Role of Farms, Firms, and Government*, Washington, DC: World Bank.
- Robinson, James A. (1997), 'When is a State Predatory?' mimeo, Department of Economics, University of Southern California, Los Angeles, October.
- Rodrik, Dani (1998), 'Where Did All the Growth Go? External Shocks, Social Conflict, and Growth Collapses', mimeo, John F. Kennedy School of Government, Harvard University, August.
- Sachs, Jeffrey D., John W. McArthur, Guido Schmidt-Traub, Margaret Kruk, Chandrika Bahadur, Michael Faye and Gordon McCord (2004), 'Ending Africa's Poverty Trap', *Brookings Papers on Economic Activity*, 1: 117–241.
- Sachs, Jeffrey D. and Andrew Warner (1995), 'Economic Reform and the Process of Global Integration', *Brookings Papers on Economic Activity*, 1: 1–95.
- Sachs, Jeffrey D. and Andrew Warner (1997), 'Sources of Slow Growth in African Economies', *Journal of African Economies*, 6 (3): 335–76.
- Sachs, Jeffrey D. and Andrew Warner (2001), 'The Curse of Natural Resources', *European Economic Review*, 45 (4–6): 827–38.
- Sahn, David (ed.) (1996), *Economic Reform and the Poor in Africa*, Oxford: Clarendon Press.
- Sala-i-Martin, Xavier (2002), 'The World Distribution of Income (estimated from Individual Country Distributions)', NBER Working Papers, 8933.
- Srinivasan, T.N. (1988), 'Population Growth and Economic Development', *Journal of Policy Modeling*, 10 (1): pp 7–28.
- van de Walle, Nicolas (2001), *African Economics and the Politics of Permanent Crisis*, New York: Cambridge University Press.
- Wood, Adrian and Kersti Berge (1997), 'Exporting Manufactures: Human Resources, and Natural Resources and Trade Policy', *Journal of Development Studies*, 34 (1): 35–9.
- World Bank (1981), *Accelerated Development in Sub-Saharan Africa: An Agenda for Action*, Washington, DC: World Bank.
- World Bank (2000), *Can Africa Claim the 21st Century?*, Washington, DC: World Bank.
- World Bank (2005), *World Development Report 2005: A Better Investment Climate for Everyone*, Washington, DC: World Bank and Oxford University Press.

67 The Middle East and North Africa

Omar S. Dahi and Firat Demir

Introduction

The growth and development performance of the Middle East and North Africa (MENA) region presents one of the major anomalies that current economics literature seeks to resolve, which is how to reconcile the existence of massive natural resources with the high unemployment, low growth and general underdevelopment of the region. In this debate, much attention is focused on the problems arising from: (1) state-oriented inward-looking economic policies; (2) lack of 'integration' with the world economy; (3) underdeveloped financial sectors and chilling investment climate; and (4) low levels of human capital development. In this chapter, we attempt to present a summarized yet more balanced and hopefully more insightful analysis of the growth and development experience of the countries in the region, with special attention given to the existing bottlenecks hindering future development prospects.

While discussing the MENA region as a whole we will divide the countries into five subgroups: (1) oil-rich labor-importing states (Bahrain, Kuwait, Oman, Libya, the United Arab Emirates, Qatar and Saudi Arabia); (2) oil-rich labor-abundant states (Algeria, the Islamic Republic of Iran, Iraq, Syria); (3) oil-poor labor-abundant NICs (Egypt, Morocco, Turkey); (4) oil-poor limited natural resource states (Israel, Tunisia, the West Bank and Gaza, Jordan, Lebanon);¹ and (5) natural resource-poor states (Sudan, Yemen) (Richards and Waterbury, 1996).² Although the inclusion of Turkey, Israel and Iran is controversial as the trajectory of the Arab and other Middle Eastern countries constitute a more appropriate whole, they share many commonalities as well. However, unless stated otherwise, the general statements will exclude Turkey and Israel.

The economic history of the MENA region is characterized by several cycles of growth and accumulation. In retrospect, the region formerly enjoyed higher levels of economic development and prosperity compared to its counterparts in Europe. While Istanbul with its 700 000 inhabitants in the sixteenth century was the largest city in the world, North Africa overall was much more urbanized than Europe (Paris with 125 000 inhabitants versus Cairo with 450 000 in around 1500) (Bairoch, 1997, pp. 517–37). However, in the last of these cycles, the region experienced a decline in its growth and development indicators starting from the early eighteenth century, with the

factors that precipitated this decline remaining a source of continuing debate.³ The current chapter will focus on the region's most recent cycle, namely that after the 1950s following the gaining of political independence and control over natural resources by the countries in the region.

Growth and development in the MENA

From the late 1950s to the late 1970s (and in some cases, till the late 1980s and early 1990s) the economic structure of the region was characterized by an import-substituting industrialization (ISI) regime, the main features of which (as elsewhere) included strict quantitative controls on international trade, overvalued exchange rates and severe rationing in foreign exchange and credit markets.

Following the hikes in petroleum prices in the early 1970s, growth and development indicators in the MENA region improved rapidly. The sudden increase in investment and growth rates in the oil-exporting countries spread to the rest of the region through increases in worker remittances, and capital flows. In addition, gross capital formation jumped to exceptionally high rates, generating a locomotive effect on growth rates and overall standards of living. On the financial front, considerable amounts of financial savings were accumulated abroad, resulting in the famous expansion of the Eurodollar market through the recycling of petrodollars.

In contrast, the downside of the above picture has been the high level of volatility of gross domestic product (GDP) growth since the 1970s: the average volatility of GDP growth in the region as a whole has been twice that of the developing-country average, and twice more volatile in the oil-rich economies than the rest of the region (Abed and Davoodi, 2003; Hirata et al., 2004, pp. 62–3).

The single most important determinant of growth in the MENA (where fuel products account for about half of the region's GDP and around 90 percent of total exports in the oil-rich countries) has been the fluctuations in international oil prices. In addition to growth volatility, as a result of high dependence on oil revenues, fiscal policy in the oil-rich countries is also volatile and procyclical. Likewise, the oil-poor labor-abundant countries are also oil price-sensitive because a large part of their economies is dependent on worker remittances as well as on development aid and tourism revenues from the oil-rich labor-poor countries. The non-oil-producing sectors, on the other hand, suffer from the 'Dutch disease' where the continuous flow of large oil revenues result in an appreciation of the real exchange rate, making it less competitive.

In large part due to the collapse in oil prices during the 1980s and 1990s the growth rates in the region experienced marked declines. Low growth rates failed to provide the rapidly expanding labor force with sufficient

employment opportunities and led to a deterioration of living standards and a rise in poverty rates. As a result, growth performance and GDP per worker and total factor productivity growth rates in the MENA region since the early 1980s has been near zero and negative, closer to Latin America and lagging far behind East Asia. Furthermore, despite substantial improvements since the gaining of political independence, the region lags behind both East Asia and Latin America in the UN Human Development Index (that is, in adult illiteracy rate, life expectancy at birth, and so on) (Bosworth and Collins, 2003; UNDP, 2002).

Nevertheless, there is considerable heterogeneity in performance across countries within the region. While the per capita incomes of the oil-producing countries declined at an average annual rate of -0.79 percent between 1980 and 2000, those in the non-oil-producing countries increased by around 2 percent over the same period. Looking at the oil boom years we get a similar picture where non-oil-producing countries grew almost twice as much as oil-producing ones. What are the reasons behind this diverse yet overall poor performance? We will turn to this question in the coming sections.

State, institutions and development

Despite the diversity in state structures, resource endowments and economic performance a characteristic shared virtually by all countries in the region (including Turkey and Israel) is the dominant role played by the public sector in the development process (Richards and Waterbury, 1996).

The majority of the states also experimented with the usual sequencing in terms of development models, going through an ISI period (accompanied by land reform) which would be disrupted (usually following an economic and/or political crisis, such as Egypt in 1967, 1974; Turkey in 1980) and be replaced by an outward-oriented development model where the role of the state is attempted to be downsized through domestic and external liberalization programs and public sector restructuring.

The first ISI attempts in the region were launched by Turkey in its first five-year plan of 1934, emphasizing the establishment of state enterprises in textiles, primary commodities and minerals, ceramic and glass, paper, chemicals and cement, and iron and steel as well as state banks for financing these enterprises. The Turkish model would provide a guide for the other MENA countries and be replicated throughout the region (Richards and Waterbury, 1996). In the oil-rich labor-abundant countries, the availability of massive oil rents, or what are termed 'soft-budget constraints', allowed the large countries to launch ambitious ISI projects. This led to a proliferation of capital-intensive (often turnkey) industries producing protected intermediate and final products for the domestic market. However, tariff

protection and credit access was often granted wholesale and no technological upgrading or other performance measures were required as was the case in East Asian countries.

During this period, despite the presence of a generally hostile attitude by the state bureaucracy, the private sector in many countries benefited largely from intermediate products supplied by the state enterprises at discounted prices or from other subsidies in the form of cheap credits or foreign exchange.⁴ One of the characteristics of the ISI era was that during this period the accumulation process was highly dependent on politics rather than markets. The political and economic environment thus created opportunities for wide-ranging rent-seeking behavior within the business community, as businesses competed for the special set of incentives (subsidized credit and foreign exchange, import licenses, and so on) provided by the state. The pre-liberalization ISI era thus gave rise to a narrow distributional coalition between the state bureaucracy and the business community.

Regarding financial and banking sector development, the region continues to suffer from the lack of an efficient banking system with long-term credit availability for private investment projects (excluding Israel). Furthermore, the use of state banks for political rent distribution in the form of distributing cheap credits on non-economic grounds manifests itself in the accounts of large 'duty losses' of these banks (OECD, 2001; Mitchell, 1999, pp. 29–30). In the case of capital market deepening, the money markets are mostly dominated by short-term government securities, while capital markets in private securities remain underdeveloped throughout the region.

Moreover, the tax system of the countries in the region has been characterized by the inability or unwillingness of policy-makers to implement an efficient and fair tax scheme which, in addition to resulting in a narrow tax base and high tax evasion, further contributes to the unequal distribution of the tax burden on low-income groups. As a result in virtually all countries in the region the business environment for private firms with no political ties with the state bureaucracy is not encouraging. Informal tax traps are common and lack of clear-cut and consistent tax laws result in incorrect assessments of tax obligations (for example EIU, 2004).

On the other hand, following independence the survival of these mostly autocratic regimes required distribution of economic rents to a wider group of supporters. Indeed, what is common in the region is that the continuous flow of revenues (mostly from oil rents) has helped postpone economic and political reforms, since the region as a whole (excluding Turkey and to some extent Israel) could manage to avoid the balance-of-payments crises that other developing countries faced at the last stage of their ISI experience.

Natural resources and development

After gaining full independence and national sovereignty in the post-colonial era, oil-producing MENA countries (which account for about three-quarters of the world's proven crude oil reserves and 35 percent of global oil production) reaped the benefits of increasing oil revenues. Once national governments secured control over their oil production and pricing, oil revenues started to flow in cascades. For example, in the case of Saudi Arabia, crude oil revenues increased from around \$10.4 million in 1946 to \$104.2 billion in 1980 (Owen and Pamuk, 1998, p. 210).

However, this development has led to a Dutch disease with destructive effects on non-oil industrial sectors while retarding economic diversification and growth (Sachs and Warner, 2001). Appreciating domestic currency resulting from large oil-related foreign exchange inflows created an unsuitable environment for the development of domestic industries by making non-oil exports less competitive. Furthermore spending on massive construction projects further turned the terms of trade against manufacturing. Another major reason for the misalignment is the pegged or fixed exchange rate regimes adopted in the region as a whole (excluding Turkey) (World Bank, 2003, p. 110). Nabli and Veganzones-Varoudakis (2002) argued that MENA countries experienced overvaluation of more than 20 percent a year in their real exchange rates from the mid-1970s to 1999. They also suggest that the exchange rate policy explains losses in competitiveness and in manufactured exports in the region as a whole where real exchange rate overvaluation has decreased the ratio of manufactured goods to GDP by 18 percent a year.

On the other hand, in contrast to the abundance of oil resources, with 5 percent of the world's population, the MENA countries have only 1 percent of the world's renewable fresh water. According to the World Bank the region's per capita supply 'stands at only one-third of its 1960 level, and water availability is expected to halve over the next 25 years if the present pattern of use continues' (World Bank, 2004, p. 4). Water shortage means that in addition to the strain of providing clean water to a rapidly increasing population, the countries are also increasingly dependent on food imports. Moreover, conflicts over water distribution and sharing have been exacerbated due to a lack of adequate regional conflict resolution mechanisms.

Trade and development

Historically, the MENA region was a thriving center of trade both originating within the region and as a crossroads for trade routes between Europe, East Asia and southern Africa. However, the shift in the balance of power between the MENA and Europe over the seventeenth and eighteenth centuries and Europe's subsequent industrialization instituted a new

pattern of trade, that of manufactures exports from Europe in return for primary products and raw materials, and led to the subsequent decline and decimation of existing manufactures and crafts production that the region had enjoyed. During this period, any attempt by the region to industrialize was forcefully prevented (most notably by Britain and France), such as the industrialization efforts by Muhammed Ali in Egypt (Issawi, 1966, p. 363). This not only significantly shifted the pattern of production and trade, but also served to disrupt intra-regional trade in agriculture and manufactured goods, which had expanded under the consolidation of the region under the Ottoman rule (Owen, 1993).

More recently, the fortunes of the region (excluding Turkey and Israel) in the post-World War II period have been dependent on two types of trade. The first is the inter-regional export of fuels and other primary products (for example natural gas, iron phosphates), which during oil price booms reached almost 50 percent of GDP in the oil-exporting countries, 'with 35 to 40 percent of GDP "spillover" effects for the region as a whole' (Shafik, 1998). The spillover was mainly due to the second, intra-regional trade in labor, which has been a vehicle of transmitting the rents throughout the region, reaching as high as 20 percent of GDP for some countries such as Jordan and around 5–10 percent of GDP or higher for several countries such as Egypt, Syria, Morocco, Jordan and Tunisia (Galal, 2000).

Both the oil-rich labor-poor and labor-abundant countries have fuel exports that made up around 85 percent of their total exports as of 2000. The oil-poor small states, on the other hand, have successfully diversified their exports whereby manufactures made up around 75 percent of total exports in 2000. Furthermore, Israel and Turkey have highly diversified exports compared to the rest of the region, with Israel emerging as a world leader in high-technology exports.

Although regional integration or 'Pan-Arab unity' has been a prominent topic in the region, intra-regional trade in the MENA has never exceeded 8 percent of exports and is the lowest of any region in the world (Galal, 2000). The lack of diversified production structures has undoubtedly been a hindrance for intra-regional trade; gravity model estimations reveal that MENA countries trade about a third less than otherwise identical countries (Rose, 2002).⁵

In order to reverse this trend, the Arab Free Trade Agreement has been established, with 18 countries signing the agreement in 1997. Furthermore, as a sign of expanding regionalism, in addition to intra-Arab treaties, several MENA countries have signed bilateral association agreements with the EU, with others to follow (Fawzy, 2003).

The uncertainty in gains from regional integration is a risk for regimes that are wary of engaging in potentially destabilizing reforms. Moreover,

the availability of windfall rents has allowed the regimes to appease domestic constituencies, and prevented the formation of coalitions pressuring integration or other deep structural transformations (Carkoglu et al., 1998).

Labor markets and human capital in the MENA

Regarding demographical challenges, the MENA region has the second-highest population growth rate in the world after sub-Saharan Africa, exacerbating labor market problems. Although the rate has been decreasing in recent years, average annual growth in the labor force is still expected to be 3.4 percent a year in 2000–2010, which is twice that of other developing countries, with adverse effects on per capita incomes (World Bank, 2003, p. 19).

The presence of a disproportionately high share of the young in the population (under-30-year olds constitute almost two-thirds of the population on average), low growth rates and lack of skill development has resulted in high unemployment throughout the region.

It is estimated that 16 MENA countries that represent 60 percent of the regional population need to provide 47 million new jobs between 2002 and 2012 just to keep up with the increasing labor supply (Keller and Nabli, 2002). As a result, the unemployment (and underemployment) rate is quite high in the region, and despite underestimated official figures stands at around 15 percent in the Arab countries (UNDP, 2002). Furthermore, since 1981 the labor force has grown faster than population growth and can be expected to increase further with increasing female participation rates.

In the case of oil-poor countries, another problem lies in their dependence for job growth on the oil-producing countries. As of 1997, for example, foreigners in Kuwait held 99 percent of private sector and 42 percent of public sector jobs. The ratio for the foreign to domestic workforce is 90 percent in the UAE, 83 percent in Qatar and 69 percent in Saudi Arabia (McMurray, 1999, p. 19).

Regarding human capital, following political independence the MENA countries faced a daunting task to educate their population, with adult illiteracy of 70 percent in Syria and 85 percent in Algeria, Iraq and Libya around independence (El-Ghonemy, 1998). The colonial powers had established parallel systems of education and the systemic discrimination in education left the majority of the population, especially in rural areas, with dilapidated and low-quality public schools while the expatriates, the urban elites and sectors friendly to colonial powers enjoyed high-quality educational establishments (El-Ghonemy, 1998).

Since the post-independence period MENA countries have invested a high proportion of their GDP towards education and health, and have

made remarkable gains on both counts. Average illiteracy rate dropped from 60 percent in 1980 to about 43 percent in the mid-1990s, while enrollment at all levels went up from 31 million to 56 million during the same period (UNDP, 2003). However there is still widespread illiteracy among youth and adults and even higher rates among women and the rural poor. A side-effect of the industrialization attempts by the MENA countries was an allocation of resources towards secondary and higher education, which typically have lower social rates of return than primary education. The result has been the oddity of unemployed highly educated workers, while having large numbers of illiterate adults and youth (Richards and Waterbury, 1996).

International conflicts and socio-political instability

The region has been plagued with ongoing conflicts since the eighteenth century, starting with the decline and the following collapse of the Ottoman Empire and the erection of colonial regimes. Having borders drawn by the colonial powers based on politics rather than historical, cultural or ethnic backgrounds or social consensus led to subsequent ethnic and religious civil conflicts (for a detailed list of these conflicts, see for example Elbadawi, 2005, pp. 306–7).

In addition, since 1948 the Middle East has witnessed: four wars between Israel and several of its Arab neighbors; three wars with Western countries; the full occupation of Iraq and Palestine and the partial occupation of Egypt, Lebanon and Syria; extended periods of economic sanctions on Syria, Iraq, Sudan and Libya; and several *coups d'état* instigated from within and outside the region. The Iran–Iraq war alone left around 1 million dead and 2.5 million refugees, with an estimated cost of \$200 billion. These conflicts have had a direct impact on state structure and overall trajectory of development

The artificial mapping of the region with sovereign borders overlapping with different ethnic and religious groups further fed into the authoritarian state structure, thanks to the excuse that the survival of the unity of the country is dependent on the suppression of popular demands by different groups.

The majority of publications on socio-political risk and the investment and growth relationship find a negative correlation between these variables. Venieris and Gupta (1986), Alesina and Perotti (1996) and others find an inverse relationship between political instability and growth or investment, or savings rate. In addition, Asteriou and Price (2001) found that socio-political instability not only negatively affects the growth rate but also increases its volatility. Similarly, Rodrik finds a significant negative relationship between external shocks and growth in countries where there

are latent domestic social conflicts and poor conflict management institutions, as in the case of the MENA (Rodrik, 1998). The idea behind the above research is that socio-political unrest and instability disrupts market activities and investment decisions by increasing uncertainty and risk while directing limited resources to non-productive security-related expenditure.

As a result (or on the pretext) of non-stop civil or military conflicts the existing regimes have devoted a sizable portion of their budgets to military spending. The average military expenditure to GDP ratio in the region is 6.6 between 1990 and 2004 with a maximum of 21.8 in Kuwait and minimum of 1.8 in Tunisia. Comparatively, the averages were 1.4, 0.5 and 2.5 in Argentina, Mexico, and Malaysia for the same period (SIPRI, 2005). Such military spending creates a substantial potential for peace dividend in the region. However, for the peace dividend to materialize, the peace must be 'real and durable, and perceived as such' (Fischer et al., 1993).

Economic reform in the MENA

Despite the presence of a general consensus among policy-makers and economists on the need for reform, the question regarding which path to follow remains unanswered. Several countries in the region have embarked on structural adjustment programs (SAPs) under the guidance of the International Monetary Fund (IMF) and the World Bank. The reforms included standard policy packages by the twin institutions such as fiscal reform (introducing value-added taxes – VATs; eliminating state subsidies; increasing transparency in public expenditures), liberalizing trade and capital accounts, and shifting to more flexible foreign exchange regimes. Despite differences, the countries that have enjoyed higher rates of growth since the early 1990s have been those that implemented reform programs (that is, Egypt, Jordan, Morocco, Tunisia) (Hirata et al., 2004).

Nevertheless, it is difficult to make a generalized statement on the success of the SAPs in the MENA due to credits and debt reliefs extended to certain reforming countries based on political considerations during the adjustment period (for example Egypt, for its support to the first Gulf War) (Gray, 1998).

Despite the implementation of comprehensive trade and financial liberalization programs including tariff reductions, privatization, tax breaks and eased restrictions on foreign ownership, as well as establishment of free trade zones and other incentives to encourage foreign direct investment (FDI), capital flows to the region remain minimal. The region's share of FDI fell to 0.7 percent in 2000 from 2.5 percent in 1980 (Hirata et al., 2004). In addition, most of the capital flowing into the region appears to be in short-term funds. In the case of Turkey, capital account liberalization has

exposed it to the uncertainties and instabilities associated with short-term capital flows, which have demonstrated themselves in three major crises in 1994, 2000 and 2001. Furthermore, as shown in the case of banking sector crises and subsequent cost of bank defaults resulting from endemic corruption as well as rent-seeking promoting IMF engineered policies (such as 100 percent state insurance on private bank deposits in Turkey), the countries in the region need a major restructuring in their financial systems.

On the other hand, income inequality and poverty rates have increased since the implementation of reform policies (Ali and Elbadawi, 2002; Fergany, 1998). The region had previously enjoyed the lowest incidence of poverty and income inequality of any region in the developing world (Adams and Page, 2003). As a result, only 5.6 percent of the population in the region lived on a less than the \$1 a day benchmark compared with 14.7 percent in East Asia and 28.8 percent in Latin America (Shafik, 1995). Adams and Page (2003) pointed out two statistically and economically significant reasons for this: international remittances, and public sector employment and the welfare state. The policy of public sector employment and subsidized public services and pricing to generate popular support for the survival of the political regimes appears to be the common element in the region. However, the slow-down in growth as well as neoliberal reforms, which have scaled back the role of the state, have reversed the trend of lowered inequality (Ali and Elbadawi, 2002; Fergany, 1998).

Conclusion and policy suggestions

Contrary to the view that finds 'little reason for gradualism', as in World Bank (2003, p. 7), there may arise significant socio-political costs to a 'big bang' approach in the region. In a majority of MENA countries, certain sectors and groups of people (that is, peasantry, civil servants, organized labor) will stand as absolute losers from the reform programs, at least in the short run. Furthermore, the economic and political failures of the past have created an unstable environment pregnant with socio-political fault-lines, which are further exposed by slow growth rates, increasing unemployment, and increasing income inequality and poverty among different income groups and different regions.

The worsening economic performance has radicalized the divide between urban and rural, secular and Islamist, and ethnic identity groups, and these politicized fault-lines have, in turn, been accompanied by increasing authoritarian governance in the region (Lubeck, 1998, p. 299).

As a result, increasing hegemony of neoliberal policies along with economic liberalization and deregulation of markets may have the opposite effect on political liberalization and consolidation of democracy in the

region, by further deepening such divisions through increasing economic insecurity and social dislocation among the public.

The experiences of MENA countries suggest that historically determined institutional characteristics and the political environment of a country are of crucial importance in determining both the nature of the adjustment process and subsequent economic performance. Given that the market-led, outward-oriented reform programs have not produced the anticipated results so far, there is a continuing debate among economists about the underlying reasons. This chapter follows the line that developing countries share common structural problems in their institutional settings, and that policies that are designed to liberalize their economy (and political and civil life) may also generate serious instabilities without necessarily eliminating the previously existing ones. The existence of strong state hegemony in the form of military, legislative and economic institutions with a lack of clear-cut lines between private and public spheres resulted in a lack of democratic accountability and transparency during the design and implementation of reform packages. In addition, previously formed rent-seeking coalitions have prevented the implementation of a comprehensive reform program designed according to the needs of the countries in the region.

As a result, instead of removing the state, the neoliberal reform programs helped the state become instrumental in distributing rents to a new group of rentiers that make their living from financial rents (Mitchell, 1999, p. 30; Yeldan, 2001; Demir, 2004, 2005). Hence the state(s) 'now subsidizes financiers instead of factories, speculators instead of schools' (Mitchell, 1999, p. 31).

The recent experience of MENA countries suggests the state and the institutional infrastructure need to be reformed before embarking on reform programs that may undermine the legitimacy of the state structure and lead to socio-political instabilities. In other words, sequencing of reforms is a must both for the sustainability of the reforms and for their further deepening. The future of the region in terms of economic and political outcomes depends on the following:

1. Institutional reform in the form of judicial, legal, administrative and prudential regulation including the rule of law should be established. In addition, rent-seeking groups need to be controlled if any economic reform is to be successful.
2. Providing social safety nets for the disadvantaged and the losers during transition.
3. Political liberalization including reforming the state and making it democratically accountable rather than populist.

4. Privatization of the ownership or the management of State Economic Enterprises (SEEs) and public banks are needed to avoid corruption, rent-seeking and subsidized credit distribution to a few wealth groups based on political considerations.
5. Boom–bust cycles need to be stabilized especially in the case of resource-rich countries, which are dependent on the changes in the oil markets. As also argued by the World Bank (2003, p. 10), the countries need to: establish rules that shield fiscal spending from fluctuations in oil revenues; create deposit accounts for oil revenues to be set aside for future generations; and avoid misalignment in exchange rates. This may have solved the ‘resource curse problem’ in the oil-rich countries by offering an alternative to investing revenues in non-profitable and non-competitive domestic investment projects. This may also pave the way to avoid currency appreciation and support competitive domestic sectors. Also, this may provide an outlet for intergenerational resource distribution for future generations, given the limited supply of oil reserves.⁶

Notes

1. Israel is considered an industrialized country and in that sense is in a separate category to other MENA countries.
2. Our classification here is slightly different from Richards and Waterbury and is meant to highlight export structure, intra-regional labor migration, patterns of industrialization and dependence on oil revenues. For example, the proven oil reserves of Syria are negligible; however it was highly dependent on oil exports for revenues during the ISI period and until today as fuel exports made up 76 percent of total exports in year 2000.
3. Some recent scholars attempted to explain the lagged performance in the region with the legacy of Islam. Kuran (2004) for example, blames the Islamic *waqf* or trusts that locked capital into a dysfunctional institution, Islamic inheritance law which dispersed inheritance among multiple heirs, and the individualism of Islamic law as preventing capital accumulation à la Europe. However, the lagging performance of the region *vis-à-vis* Europe came far too recently on a historical scale to be pinned on the influence of religious (or cultural) institutions. Moreover, as Inalcik (1969) emphasized, Islamic society and law ‘shaped themselves from the very first in accordance with the ideas and aims of a rising merchant class’ (Inalcik, 1969, p. 101). Finally, any attempt to explain the decline in economic performance of the region after the eighteenth century with the religious and cultural factors or institutions should also be able to explain how the same institutions could create the opposite results prior to that date.
4. In Turkey, unlike others, the state assumed a direct role in creating and supporting the development of a national business class; Bugra (1994).
5. The colonial legacy on regionalism warrants greater attention. As Ventura-Dias (1989) argues, colonial powers promoted intra-regional trade in Asian countries which allowed ‘permanent marketing channels to be established’ in contrast with both MENA and Latin America, where colonial intervention disrupted intra-regional trade.
6. The only country in the region with a definite plan to limit the harmful effects of the oil curse is Kuwait, which preferred to utilize its oil revenues on investments abroad (Owen and Pamuk, 1998, p. 216).

References

- Abed, G.T. and H. Davoodi (2003), *Challenges of Growth and Globalization in the Middle East and North Africa*, Washington, DC: IMF Publications.
- Adams, R.H., Jr. and J. Page (2003), 'Poverty, Inequality, and Growth in Selected Middle East and North Africa Countries, 1980–2000', *World Development*, 31 (12): 2027–48.
- Alesina, A. and R. Perotti (1996), 'Income Distribution, Political Instability and Investment', *European Economic Review*, 40 (6): 1203–28.
- Ali, A.A. and I.A. Elbadawi (2002), 'Poverty in the Arab World: The Role of Inequality and Growth', in I. Sirageldin (ed.), *Human Capital: Population Economics in the Middle East*, Cairo: American University in Cairo Press, pp. 62–95.
- Asteriou, D. and S. Price (2001), 'Political Instability and Economic Growth: UK Time Series Evidence', *Scottish Journal of Political Economy*, 48 (4): 383–99.
- Bairoch, P. (1997), *Victoires et déboires II: Histoire économique et sociale du monde du XVI^e siècle à nos jours*. Paris: Gallimard. Quoted in A.G. Frank (1997), *ReOrient*, University of California Press, p. 12.
- Bosworth, B.P. and S.M. Collins (2003), 'The Empirics of Growth: An Update', *Brookings Papers on Economic Activity*, 2: 113–79.
- Bugra, A. (1994), *State and Business in Modern Turkey*, Albany, NY: State University of New York Press.
- Carkoglu, A., M. Eder and K. Kirisci (1998), *The Political Economy of Regional Cooperation in the Middle East*, London: Routledge Press.
- Demir, F. (2004), 'A Failure Story: Politics and Financial Liberalization in Turkey, Revisiting the Revolving Door Hypothesis', *World Development*, 32 (5): 851–69.
- Demir, F. (2005), 'Militarization of the Market and Rent Seeking Coalitions in Turkey', *Development and Change*, 36 (4): 667–90.
- Economic Intelligence Unit (EIU) (2004), *Country Finance Egypt*, London: Economic Intelligence Unit.
- Elbadawi, I.A. (2005), 'Reviving Growth in the Arab World', *Economic Development and Cultural Change*, 53 (2): 293–326.
- El-Ghonemy, M.R. (1998), *Affluence and Poverty in the Middle East*, New York: Routledge.
- Fawzy, S. (2003), 'The Economics and Politics of Arab Economic Integration', in A. Galal and B. Hoekman (eds), *Arab Economic Integration: Between Hope and Reality*, Washington, DC: Brookings Institution Press, pp. 13–37.
- Fergany, N. (1998), 'Human Impact of Capitalist Restructuring in Arab Countries', *Journal of Development and Economic Policies*, 1 (1): 9–46.
- Fischer, S., D. Rodrik and E. Tuma (eds) (1993), *The Economics of Middle East Peace: Views from the Region*, Cambridge, MA: MIT Press.
- Galal, A. (2000), 'Incentives for Economic Integration in the Middle East' in B. Hoekman and H. Kheir El-Din (eds), *Trade Policy Developments in the Middle East and North Africa*, Washington, DC: World Bank Institute, pp. 51–67.
- Gray, M. (1998), 'Economic Reform, Privatization and Tourism in Egypt', *Middle Eastern Studies*, 34 (2): 91–113.
- Hirata, H., H.K. Sunghyun and M.A. Kose (2004), 'Integration and Fluctuations: The Case of MENA', *Emerging Markets Finance and Trade*, 40 (6): 48–67.
- Inalcik, H. (1969), 'Capital Formation in the Ottoman Empire', *Journal of Economic History*, 29 (1): 97–140.
- Issawi, C. (1966), 'The Economic Development of Egypt, 1800–1960, A Book of Readings', in C. Issawi, *The Economic History of the Middle East 1800–1914*, Chicago, IL: University of Chicago Press, pp. 359–374.
- Keller, J. and M.K. Nabli (2002), 'The Macroeconomics of Labor Market Outcomes in the MENA Region over the 1990s', World Bank, <http://www.worldbank.org/wbi/mdf/mdf4/papers/keller-nabli.pdf>
- Kuran, T. (2004), 'Why the Middle East is Economically Underdeveloped: Historical Mechanisms of Institutional Stagnation', *Journal of Economic Perspectives*, 18 (3), pp. 71–90.

- Lubeck, P.M. (1998), 'Islamist Responses to Globalization: Cultural Conflict in Egypt, Algeria, and Malaysia', in B. Crawford and R.D. Lipschutz (eds), *The Myth of 'Ethnic Conflict'*, University of California Press/University of California International and Area Studies Digital Collection, Edited Volume 98, pp. 293–319.
- McMurray, D. (1999), 'Recent Trends in Middle Eastern Migration', *Middle East Report*, 211 (Summer): 16–19.
- Mitchell, T. (1999), 'Dreamland: the Neoliberalism of Your Desires', *Middle East Report*, 210: 28–33.
- Nabli, M.K. and M.A. Veganzones-Varoudakis (2002), 'Exchange Rate Regime and Competitiveness of Manufactured Exports. The Case of MENA Countries', MENA Working Paper 27, World Bank, Washington, DC.
- Organisation for Economic Co-operation and Development (OECD) (2001), *OECD Economic Surveys: Turkey*, Paris: OECD.
- Owen, R. (1993), *The Middle East in the World Economy 1800–1914*, New York: St Martin's Press.
- Owen, R. and S. Pamuk (1998), *A History of Middle East Economies in the 20th Century*, Cambridge, MA: Harvard University Press.
- Richards, A. and J. Waterbury (1996), *A Political Economy of the Middle East*, Boulder, CO: Westview Press.
- Rodrik, D. (1998), 'Where Did All the Growth Go? External Shocks, Social Conflicts, and Growth Collapses', unpublished paper.
- Rose, A. (2002), 'Estimating Protectionism Through Residuals from the Gravity Model', unpublished paper.
- Sachs, J.D. and A.M. Warner (2001), 'Natural Resources and Economic Development: the Curse of Natural Resources', *European Economic Review*, 45: 827–38.
- Shafik, N. (1995), *Claiming the Future: Choosing Prosperity in the Middle East and North Africa*, Washington, DC: World Bank.
- Shafik, N. (ed.) (1998), *Prospects for Middle Eastern and North African Economies*, New York: St Martin's Press.
- Stockholm International Peace Research Institute (SIPRI) (2005), Database on Military Expenditure, available at http://www.sipri.org/contents/milap/mex_database1.html.
- United Nations Development Programme (UNDP) (2002), *Arab Human Development Report 2002: Creating Opportunities for Future Generations*, New York: United Nations.
- United Nations Development Programme (UNDP) (2003), *Arab Human Development Report 2003: Creating A Knowledge Society*, New York: United Nations.
- Venieris, Y. and D. Gupta (1986), 'Income Distribution and Sociopolitical Instability as Determinants of Savings: A Cross-sectional Model', *Journal of Political Economy*, 94 (4): 873–83.
- Ventura-Dias, V. (1989), 'The Structure of South–South trade', in V. Ventura-Dias (ed.), *South–South Trade: Trends, Issues, and Obstacles to its Growth*, New York: Praeger, pp. 25–81.
- World Bank (2003), *MENA Development Report: Trade Investment and Development in the Middle East and North Africa, Engaging with the World*, Washington, DC: World Bank.
- World Bank (2004), *MENA Regional Brief*, Washington, DC: World Bank.
- Yeldan, E. (2001), *Kuresellesme Surecinde Türkiye Ekonomisi Bolusum, Birikim ve Buyume (Turkish Economy in the Process of Globalization, Accumulation and Growth)*, Istanbul: Iletisim Yayinlari.

68 China

*Yasheng Huang*¹

Many economists believe that China today is largely a market economy. Yingyi Qian (2003), a leading economist on China, observes: 'In the last 22 years of the 20th century, China transformed itself from a poor, centrally planned economy to a lower-middle-income, emerging market economy.' Barry Naughton, another leading economist on China, echoes this view in his recent textbook (Naughton, 2007). This chapter counters this claim with evidence that shows that China today is far from a market economy – defined as one predominantly based on private ownership.

This chapter first reviews China's growth experience since 1978. I will note briefly the enormous gains China has made, but the emphasis here is to highlight aspects of Chinese performance that are less well known, such as the increasingly investment-driven growth, the slowdown of productivity growth, and some notable microeconomic inefficiencies (as compared with, say, India). Social performance also deteriorated in the 1990s.

I then turn to ask the question, 'Just how capitalist is China?' The answer, surprisingly, after nearly 30 years of reforms, is not at all clear. There is evidence that China today is a commanding-heights economy similar to some of the most statist economies of the 1970s. I use data on fixed-asset investments by what I call the registered domestic private sector to show that in the 1980s the domestic private sector developed vigorously but in the 1990s the pace of liberalization and denationalization slowed down. The final section concludes with some broad implications of this analysis.

China's growth experience

China's economic success is indisputable. Its gross domestic product (GDP) growth has led the world in the growth table. Between 1978 and 2004, according to the World Bank, real GDP growth per annum averaged 9.73 percent, the fastest in the world. Poverty has fallen dramatically since the onset of the reforms in the late 1970s. By one estimate, the overall poverty level in China – measured as the percentage share of the population living under the poverty line – declined from 53 percent in 1981 to only 7.97 percent in 2001 (Ravallion and Chen, 2004). In the 1970s, the Chinese economy was plagued by shortages; today, it is often blamed as the source of the worldwide deflation as the economy has become a powerful export engine.

I do not repeat these well-known success indicators. Rather, this is a focused treatment of those topics that have received less attention in the academic literature.² These include the non-economic foundation of China's rapid growth, the heterogeneous growth record during the reform era, and the divergence between the GDP indicators and other development indicators. The purpose of the discussion is to highlight the complexities involved in trying to understand China's growth experience.

The social and political foundations of growth

Economists assign a heavy weight to the reforms when explaining China's economic success. This is not wrong, but it is important to point out that cross-country evidence on the linkage between policy reforms and economic growth is not nearly as clear-cut. Yet, it is uncontroversial that China's impressive growth clearly followed the policy changes introduced in the late 1970s. Even if we acknowledge the importance of the reforms, the appropriate way to frame the discussion is to ask whether the policy changes introduced in the late 1970s interacted with some of the favorable pre-existing conditions in the country and whether it was an interaction effect, rather than just the policy changes alone, that spurred China's growth.

In this context, it is useful to compare China with India. When examining the details of India's growth record, Hausmann et al. (2004) correctly note that India's growth rate began to pick up relative to the historic benchmark in the 1980s and that this acceleration of growth coincided with a number of modest policy changes. In addition, they note that China's growth also followed seemingly modest policy changes in the 1980s. If we accept this characterization of the respective policy changes in China and India, we are still left to explain why China's growth was so much stronger than India's growth, as indicated in Table 68.1.³ For the period on which Hausmann et al. focus, that is, from 1978 to 1990, the average annual GDP growth in China was 9.28 percent, almost twice India's growth rate of 5.01 percent.

This differential suggests the importance of identifying some initial differences between China and India that might have contributed to their divergent growth performances in the 1980s. The most important initial difference in favor of China has to do with the social conditions. For complex reasons, during the orthodox socialist period (1949–78) China invested heavily in health and education, especially in the rural areas. For example, as early as 1965 the life expectancy of Chinese women at birth was 55 years, compared with only 44 years in the case of Indian women. In the 1980s, infant mortality in China was substantially lower than infant mortality in India. (In the mid-1980s, China's infant mortality rate was 54 per

Table 68.1 GDP growth and GDP per capita growth, China and India, 1978–2004 (%)

	Panel A: Average annual % growth of real GDP			
	1978–2004	1978–90	1991–97	1998–2004
China	9.73	9.28	11.53	8.76
India	5.37	5.01	5.43	5.97
Ratio of China to India	1.81	1.85	2.12	1.47
	Panel B: Average annual % growth of real GDP per capita			
	1978–2004	1978–90	1991–97	1998–2004
China	8.44	7.73	10.26	7.95
India	3.37	2.78	3.53	4.29
Ratio of China to India	2.51	2.78	2.91	1.85

Notes: GDP data are calculated on an exchange rate basis.

Source: Data are downloaded from *World Development Indicators* at <http://devdata.worldbank.org/dataonline/>.

1000, compared with 122 per 1000 in India.) Furthermore, China's primary education enrollment ratio was far higher than that of India as early as 1975. China also had a more equal initial distribution of income as compared with that in India.⁴

Panel (A) of Table 68.1 gives the average annual growth percentages of real GDP for China and India. Panel (B) gives the average annual growth percentages of real GDP per capita. Comparing the data in Panel (A) with the data in Panel (B) reveals an interesting pattern: Although China's aggregate GDP performance was better than that of India, its per capita GDP performance was even better. For the period from 1978 to 2004 as a whole, China's GDP growth rate was 1.81 times that of India (9.37 percent compared with 5.37 percent). However, the ratio of Chinese GDP per capita growth to that of India was 2.51 for this period (8.44 percent compared with 3.37 percent), substantially exceeding the GDP growth ratio of 1.81 of the two countries. It is important to understand the source of this differential.

One hypothesis centers on population control. China has a draconian population control program, but for political reasons India cannot replicate this aspect of China's 'development strategy'. Because both countries have a chronic surplus of labor, it is plausible to argue that the higher GDP growth per capita in China – far in excess of its aggregate GDP growth – is

partially due to the coercive capacity of the state. Many economists neglect this fact when they tout China's supposedly superior record of economic reforms; in reality it is the political, not the economic, management that matters.

Heterogeneous growth experience

It is interesting to explore the substantial heterogeneity in China's growth experience during the long time-span since reforms (1978). I deal with three respects: the importance of investments for China's recent growth performance, productivity development and social performance.

Let us revisit Table 68.1, which breaks down the reform era into three sub-periods. GDP growth was fastest pace during the 1991–97 period, averaging 11.5 percent per year. The period from 1978 to 1990 came second, at an annual 9.28 percent on average. The most recent period from 1998 to 2004 turned out to be the least impressive, averaging 8.76 percent per year.

These fluctuations in the GDP growth coincided with huge changes in the investment levels. In the 1980s, the gross fixed capital formation as a percentage ratio to GDP averaged around 30 percent. During the recessionary years of 1989 and 1990, this ratio declined to 25 percent and then it surged to 35 percent in 1993. Starting in 1997, the ratio increased sharply, reaching 40 percent in 2004 and then 48 percent in 2005, another new high.⁵

So there was a change in the drivers of growth over time. Of the three sub-periods presented in Table 68.1, China's GDP performance is the least impressive during the 1998–2004 period, but this is also a period when China was investing at its highest level. Thus, at 8.76 percent per year, although China was still leading the world in GDP growth, it was achieving this performance at a substantially higher level of investments than when it was growing faster in the 1980s and the early 1990s.

Is China's latest growth spurt as sustainable as the one it experienced in the 1980s and the early 1990s? The answer depends on the productivity of the latest investment surge. If the high levels of investments are accompanied by or directly lead to technological progress, then an investment-driven growth pattern can be sustainable. Research on Western economies has shown a sustained, long-term positive correlation between productivity growth and capital deepening (Wolf, 1991).

The evidence, however, suggests that the latest investment growth did not lead to improved productivity. An exhaustive survey of the various studies on China's total factor productivity (TFP) reveals one consistent pattern: TFP performance declined beginning in the late 1990s relative to earlier periods. The key findings are summarized in Table 68.2. These studies differ on the TFP estimates but they converge on trend developments: TFP growth during the last period, that is, in the late 1990s or early 2000s, was

Table 68.2 Estimates of the annual TFP growth in the Chinese economy (%)

Sources of estimates	Level of data	Reference periods						
		1980s			1990s			2000s
		First half	Second half	First half	Second half	First half	Second half	
Heytens and Zebregs (2003)	National-level GDP data	2.78 (1979–84)	2.11 (1985–89)	2.81 (1990–94)	2.30 (1995–98)			
Zheng and Hu (2004)	National-level GDP data	3.26 (1978–95)			0.32 (1995–2001)			
Kuijs and Wang (2005)	National-level GDP data	3.74 ^a (1978–93)			2.7 (1993–2004)			
Miyamoto and Liu (2005)	Provincial-level GDP data	5.45 (1981–85)	1.73 (1986–90)	6.28 (1991–95)	2.91 (1996–2000)			
Wu (2004)	Provincial-level GDP data	n/a	n/a	1.88 (1993–97)	1.19 (1993–2002)			
Wu (2003)	Provincial-level GDP data	2.35 (1982–85)	0.43 (1986–91)	1.75 (1992–97)				
Ren and Sun (2006)	Industry-level data	6.45 (1981–84)	3.14 (1984–88)	3.83 (1988–94)	0.52 (1994–2000)			

Note: a Kuijs and Wang (2005) estimate TFP growth to be 3 percent during the 1978–2004 period and 2.7 percent during the 1993–2004 period. I have calculated TFP growth to be 3.74 percent during the 1978–93 period on the basis of their estimates.

Sources: See the table for citation information.

considerably more moderate than TFP growth in the 1980s and the early 1990s. For example, Zheng and Hu (2004) report that TFP grew annually by 3.26 percent between 1978 and 1995, but during the 1995–2001 period TFP growth virtually disappeared (0.32 percent). Focusing only on Chinese industry, Ren and Sun (2006) report a reduction in TFP growth of a similar magnitude.⁶

Although during the reform era as a whole China made substantial progress in eradicating poverty, progress was uneven across both space and time. The largest gains in poverty reduction occurred in the first five years of the 1980s. According to Ravallion and Chen (2004), between 1980 and 1985 poverty declined drastically. In the rural areas, the incidence of poverty – measured by the headcount of those living below the poverty line – declined from 75.7 percent to only 22.7 percent. Income distribution improved in the early 1980s, as indicated by a reduction of the Gini coefficient.

Since then, the pace of poverty reduction has been considerably more measured and there have been episodic setbacks. In 1998 rural poverty was at 11.6 percent, and in 2000 it rose to 13 percent and in 2001 to 12.5 percent. Income distribution also deteriorated. In 2001 China had a Gini coefficient of 39.45, compared with 27.98 in 1980. This remarkable speed of reversal, according to two experts, is ‘almost unheard of in the developing world’ (Khan and Riskin, 2001).

The regressions in the late 1990s warrant special scrutiny. First, productivity indicators for this period deteriorated as well, so there was no economic and social trade-off. Second, according to the data provided by Ravallion and Chen (2004), the poverty level increased in the late 1980s as well. But the circumstances of the late 1980s differed substantially from those of the late 1990s. The surge in poverty in the late 1980s can be easily explained by the macroeconomic shocks – the Chinese economy went into a severe contraction in the late 1980s. It is more difficult to explain the increase in poverty during the economic boom period of the late 1990s.

The adverse social developments in the late 1990s are a sign that Chinese growth may have acquired an inherent anti-poor bias. The magnitude of the effect is substantial. Consider the increase in poverty incidence from 11.4 percent in 1999 to 12.96 percent in 2000. In percentage terms this increase may not be striking, but because China has a huge rural population, such a seemingly small rise in poverty incidence in fact corresponds to 11.24 million rural residents newly thrust into poverty, equivalent to the entire population of Greece. Other – far less known – indicators are also telling. For example, the World Bank has documented that during the reform era China has underperformed – both against other countries such as India and against its own economic potentials – in terms of reducing

infant mortality. The World Bank has also reported that China is one of only seven countries in the world to have a higher infant mortality rate among girls than among boys.⁷

The most recent evidence is even more alarming. The World Bank has just reported that the income of China's poorest 10 percent of the population declined by 2.4 percent between 2001 and 2003 (McGregor, 2006). This is the first documented evidence that a large number of Chinese people – about 130 million people – has actually experienced an absolute reduction in their living standards. The issue is no longer one of 'relative deprivation', about which economists tend not to be as concerned, but one of 'absolute deprivation'. If this trend continues, it will have serious implications for the prospects for China's growth and political stability.

Creating output vis-à-vis creating value

Bai et al. (1997) draw the distinction between the technical capabilities on the part of the SOEs to produce outputs and their economic capabilities to create value. State-owned enterprises (SOEs) can produce a massive quantity of a product, but the product itself may lack demand. In this case, there is a divergence between technical and economic measures of efficiency.

We can apply the same reasoning to the Chinese economy as a whole. It is well known that India's GDP growth has lagged that of China for much of the last two decades of the 1980s and 1990s. But a little-known fact is that the Indian economy is able to create more value and wealth for a given unit of GDP than is the Chinese economy. It is intriguing to note that India has a higher manufacturing value added per worker than China. The value added per worker in manufacturing was 2885 dollars per year during the 1995–99 period for China, but 3118 dollars per year during the same period for India. In fact, the value added in manufacturing declined between the mid-1980s and the mid-1990s in the case of China but increased in the case of India.⁸

An International Monetary Fund (IMF) paper shows that India 'overly' specialized in highly skilled industries at the expense of low-skilled industries.⁹ But even if the two countries differed in their initial starting points, it still begs the question why the manufacturing value added declined over time in China. The aforementioned IMF paper shows that between 1981 and 1996 China's share of output in skill-intensive industries was not only lower than that of India, but it was also declining over time. The declining value added in Chinese production suggests that Chinese firms were not climbing up the value chain during a period of massive boom.

Just how capitalist is China?

The conceptual framework explaining China's transition to a market economy is gradualism – the idea that the reforms are an endogenous

process whereby ‘unhooking a single key connection can cause the entire fabric to unravel’ (Naughton, 1996, p. 311). A critical empirical benchmark is the size of the private sector. China started out with a very small private sector, but due to the increasingly supportive policy environment, the private sector grew and overtook the previously dominant state sector.

I do not question the logic of a gradualist approach nor the empirical basis for applying a gradualist approach to China in the 1980s.¹⁰ The findings of a deterioration of TFP performance, worsening social performance and the increasingly investment-driven nature of the high growth since the mid-1990s raise a question whether gradualism is still an accurate empirical characterization of China today. Apart from the fact that it sheds light on the sustainability of China’s growth, the TFP performance can be a proxy measure of the reforms.¹¹ In this regard, the across-the-board finding that TFP growth began to deteriorate in the late 1990s is significant. Did the reforms stall in the 1990s?

I examine a key benchmark in the gradualist interpretation of China’s reforms – private sector development. I ask, ‘After nearly 30 years of transition, just how capitalist is the Chinese economy?’ Surprisingly, the answer is not straightforward. I distinguish between an output-based measure and an input-based measure of the size of the private sector and I show that there is a huge discrepancy between these two measures. First, while the output-based measure shows the size of China’s private sector to be quite large, the input-based measure gives rise to a far smaller estimate. Second, while the output-based measure shows a continuously rising private sector, the input-based measure indicates severe setbacks – and even retrogressions – in private sector development in the 1990s. The latter finding directly contradicts the predictions of the gradualist framework.

In the following paragraphs, I will first describe the conceptual and methodological problems underlying many of the output-based measures of the Chinese private sector. I will then present detailed statistics based on a critical input – fixed-asset investment – on the evolution of the registered private sector in China. This measure shows a robust development of the private sector in the 1980s and a sharp contraction in the 1990s.

The registered private sector firms, either individual businesses (*getihu*) or privately operated enterprises (*siying qiye*), refer to newly established private businesses registered as such with the government. For sure, this is a narrow and conservative measure of the private sector but, as I will explain, an examination of the development of this sector provides an important insight into the course of the economic reforms in the 1980s and 1990s.

Output-based measures

The most frequent measure of private sector development used by economists is the share of the private sector in production. By this measure, China's private sector has made huge strides. For example, Naughton (2007, p. 300) shows that the domestic private sector accounted for 19 percent of industrial output value in 1996, up from zero in 1978. This type of data is often used to support the argument that the policy environment improved steadily for the domestic private sector.

An output-based measure incorporates two very different effects. One is the 'policy effect': the increase in the private sector share that resulted from a more favorable policy environment. But this measure also incorporates what might be called an 'efficiency effect'. The private firms are more efficient than the SOEs and, therefore, even given a very narrow business space, they can outcompete the SOEs. This suggests that the ratio of the private to the state sector can rise without any improvement in the policy environment for private sector firms and with rising inefficiencies of the SOEs. Thus, this measure tells us as much about China's policy environment as about the huge inefficiencies of the state sector, and we cannot distinguish which of the two dynamics is driving this ratio. As an illustration, in 1985 the industrial output of the private sector was about 2.9 percent that of the state sector; by 1997, this ratio had risen to 70.2 percent. Even if the argument is correct that the policy environment improved between 1985 and 1997, it would be highly misleading to conclude that the policy environment facing the private and state sectors converged at about 70.2 percent in 1997.

China economists use a broader measure than the registered private sector to showcase China's transition success. The Chinese style of reforms has spawned a variety of hybrid and highly ambiguous ownership forms, such as SOEs with some private revenue rights, collective firms controlled at the local levels, and private–state joint-ownership firms. It is difficult to sort out who actually controls these myriad firms. The most careful analysis to decompose the ownership of Chinese firms has been carried out by two economists at the Organisation for Economic Co-operation and Development (OECD) (Dougherty and Herd, 2005). For their estimation, they use a detailed industrial firm data set maintained by the National Statistical Bureau (NSB).

One feature of the NSB dataset makes this estimation possible: the data set identifies the controlling shareholder of the firm. The OECD economists then use the shareholding structure information to generate estimates of the size of the Chinese private economy. In the NSB industrial data sets, the shareholders are classified among the following categories: (1) state (direct or indirect); (2) collective (that is, local governments); (3) individuals;

(4) domestic legal persons; and (5) foreign companies. The OECD economists make the crucial assumption that individuals, domestic legal-persons and foreign companies comprise the private ownership. They conclude that the private economy accounted for 52.3 percent of industrial value added in 2003, compared with 27.9 percent in 1998.

The most serious problem with the OECD study is the assumption that domestic legal-person shareholders are private. Legal-person shareholding refers to cross-shareholding by firms.¹² The term implies a type of *keiretsu* arrangement whereby firms own one another's stocks. The difference with Japan, however, is that in China much of the legal-person share capital originates in the state sector, that is, SOEs establishing or holding significant equity stakes in other firms. The subsidiaries of the SOEs, on account of their final ownership, should be classified as SOEs. However, the OECD calculation classifies the entire output of these firms as 'private'. As an example, the OECD methodology would classify Shanghai Automotive Industry Corporation (SAIC) as a private firm. SAIC is a quintessential SOE but its largest shareholder is a legal-person shareholder, an investment company of the Shanghai government.¹³

But let us take at face value the claim that the Chinese private sector – inclusive of the foreign firms – is producing 52.3 percent of industrial production and let us place that claim in perspective. Here a comparison with India in the 1970s is revealing. India at that time was at the apex of its commanding heights after Indira Gandhi had nationalized all major banks, significantly expanded the scope of the 'License Raj', and created numerous barriers for the private sector. But even at the height of the 'License Raj', the importance of the Indian private sector far exceeded the level of the Chinese private sector in 2003. One estimate puts the share of private sector firms in total manufacturing GDP at 93 percent in the early 1960s and at 69 percent in 1983–84. The share of fixed-asset investments of the private sector was around 58 percent,¹⁴ a ratio that is several multiples of the Chinese level today. Thus, even a generous accounting of the current size of the Chinese private sector puts China roughly in the same league as some of the world's most statist economies of the 1970s.¹⁵

An input-based measure of the domestic private sector

I focus only on the registered domestic private sector firms. This definition covers newly established private businesses (exclusive of foreign firms). This is a narrow definition of the private sector and, by construction, it understates its true size. The usefulness of this measure is that it is a test of a dynamic claim in the gradualist framework – that China chose the politically and economically prudent path to encourage the entry and the organic growth of new private businesses over time. Comparing this measure across

different time periods provides a way to assess the claim that the size of newly established private businesses has grown over time.

The input we focus on is capital allocated for fixed-asset investments. Fixed-asset investments, compared with similar activities in a market economy and with other economic activities in the Chinese economy, are heavily controlled by the government.¹⁶ All investment projects above a fairly low threshold require government scrutiny and approval. For this reason, fixed-asset investments are a superior indicator of the ownership policies of the state as compared with the output measure, because they are not subject to the confounding influences of efficiency differentials between the state and the private sectors.¹⁷

Table 68.3 provides the fixed-asset investment data on four ownership types: SOEs, collective firms, the 'individual economy',¹⁸ and firms of other ownership. (The Chinese statistical system records fixed-asset investments by the registered private sector under the 'individual economy'.) These four categories are exhaustive and mutually exclusive and thus their totals add up to 100 under Column (6).

One striking pattern emerges from this table. The investment share of the individual economy in the 1990s was actually smaller than it was in the 1980s. In the first six years of the 1980s, between 1980 and 1985, the 'individual economy' accounted for 20.7 percent of the total fixed-asset investments. This share climbed slightly in the second half of the 1980s, to an average of 21.9 percent between 1986 and 1990. In contrast, during the 1991–95 period, the 'individual economy' share declined to 13.2 percent and during the 1996–2000 period it was 13.9 percent. Panel (B) of Table 68.3 provides annual data for selected years. In 1993, the 'individual economy' only accounted for 11.9 percent of total fixed-asset investments, a full 10 percent drop from that prevailing in the second half of the 1980s (at 21.9 percent). After 1993, this ratio climbed slowly to 15 percent in 2002 and then fell back to 14.2 percent in 2004, just one percentage point higher than that at the very onset of the reforms in 1980.

Because our measure only covers fixed-asset investment activities in the registered private sector, and because of the possibility that the individual economy category may not fully record activities by the established private firms, we must return to the question of whether this measure is too narrow. In particular, the 'other' ownership category exploded from effectively zero in the second half of the 1980s to 11 percent in the 1991–95 period and then to 18.7 percent during the 1996–2000 period. If the 'other' ownership category encompasses mainly private sector firms, then the 'true' investment share of the private sector should be the sum of the individual economy and the 'other' ownership. That would put the investment share of the private sector during the 2001–2003 period at 43.2

Table 68.3 Ownership composition of fixed-asset investment (%)

Year	(1) SOEs	(2) Collective firms	(3) Individual economy	(4) Of individual economy: urban only	(5) 'Other' ownership	(6) Total
Panel (A): Period data						
1980–85	66.7	12.7	20.7	1.6	0.0*	100.0
1986–90	64.8	13.4	21.9	2.9	0.0*	100.0
1991–95	59.0	16.3	13.2	2.7	11.0	100.0
1996–2000**	52.5	15.0	13.9	4.1	18.7	100.0
2001–03	42.7	14.1	14.4	7.6	28.8	100.0
Panel (B): Annual data						
1993	61.5	17.9	11.9	2.7	8.8	100.0
1997	52.5	15.4	13.8	3.0	18.3	100.0
2000	50.1	14.6	14.3	5.5	21.0	100.0
2001	47.3	14.2	14.6	6.6	23.9	100.0
2002	43.4	13.8	15.0	7.8	27.9	100.0
2003	39.0	14.4	13.9	8.1	32.7	100.0
2004		14.1	14.2			100.0

Notes:

*: Constructed as zero since this category did not exist prior to the 1991–95 period.

** : In 1997 the government changed the investment reporting and approval procedure. The investment reporting threshold was revised from 50 000 yuan to 500 000 yuan, but this change only applied to SOEs and urban collective firms. The effect of this change is that the published amount of fixed asset investments in the state and urban collective sectors is smaller than the actual amount. For 1996, the government published both the revised and unrevised data. In the unrevised data, the SOEs invested 1205.6 billion yuan in fixed assets and the collective firms invested 366 billion yuan. In the revised data, the SOEs invested 1200.6 billion yuan and the collective sector invested 365.2 billion yuan. This is about a 0.4 percent and 0.2 percent difference, respectively.

Sources: Based on various sources on fixed asset investments compiled by the NSB. See the text for a detailed explanation.

percent (14.4 + 28.8), roughly double the level of the early 1980s at 20.7 percent.

But assigning all the firms in the 'other' ownership category to the private sector is a massive overstatement. The 'other' ownership category encompasses four types of firms: (1) joint-ownership firms; (2) shareholding firms; (3) foreign-invested enterprises (FIEs); and (4) unclassified firms. Shareholding firms and FIEs dominate this category of firms. These two types of firms accounted for 95.6 percent of the fixed-asset investments in this category during the 1996–2000 period and 97.2 percent during the 2001–03 period.

A critical empirical issue is whether shareholding firms are private. During the 2001–03 period, they accounted for 68.2 percent of the fixed-asset investments in the ‘other’ ownership category. The largest shareholding firms are SOEs that have issued shares on the stock market; and only 6.97 percent of the shareholding firms were private Initial Public Offerings (IPOs) between 1990 and 2003.¹⁹ These shareholding firms are firmly in the hands of the state although they have some private revenue rights. According to a detailed study of over 600 firms on the Shanghai Stock Exchange and the Shenzhen Stock Exchange, the three main groups of shareholders – the government, legal-persons and private individual investors – each controlled about 30 percent of the outstanding shares (Xu and Wang, 1997). But the control rights are far less dispersed. According to the same study cited above, on average individual shareholders controlled only 0.3 percent of the board seats of those firms, whereas the government retained 50 percent of the board seats and state-owned institutions controlled the remainder.²⁰

Another test of the gradualist claim is to compare firms that have clear, straightforward ownership rights at the extreme ends of the ownership spectrum. One useful indicator is the ratio of fixed-asset investments in the registered private sector relative to the state sector. If, as commonly alleged, ownership biases against the private sector declined over time, it must be the case that the ownership biases against registered private sector firms relative to the treatment of the explicit state sector firms must have declined. A decrease in the ownership biases should be associated with a rising ratio; an increase in the ownership biases should be associated with a declining ratio.

The ratio in fact declined over time. Table 68.4 presents data on fixed-asset investment in the registered private sector as a ratio of fixed-asset investment in the state sector, collective sector and firms of ‘other’ ownership under Columns (1a), (2a) and (3a). In the 1980s, the fixed-asset investments undertaken by the registered private sector in both urban and rural areas already amounted to about one-third of the fixed-asset investments in the state sector. The ratio of the individual economy to SOEs, under Column (1a), was 0.31 during the 1980–85 period and 0.34 during the 1986–90 period. But this ratio declined sharply between 1991 and 1995, to only 0.22. Between 1996 and 2000, the ratio rose moderately, to 0.27. Between 2001 and 2003, despite a period of rapid growth and economic reforms widely perceived as bold, the ratio of fixed-asset investment by purely private to state firms only managed to recover to the level prevailing at the very onset of the reform era. For much of the 1990s there is no evidence that the ownership biases abated by this measure. In fact, the evidence suggests the opposite. This pattern holds when comparing the

Table 68.4 Fixed asset investment ratios of the individual economy to 'other' firms

Year	(1) Individual economy/ SOE ratios		(2) Individual economy/Collective firm ratio		(3) Individual economy/'other' ownership ratio	
	(1a) Urban & rural	(1b) Urban only	(2a) Urban & rural	(2b) Urban only (urban collective only)	(3a) Urban & rural	(3b) Urban only
	Panel (A): Period data					
1980–85	0.31	0.024	1.64	0.13 (0.41)	–	–
1986–90	0.34	0.045	1.64	0.22 (0.51)	–	–
1991–95	0.22	0.045	0.80	0.16 (0.58)	1.15	0.23
1996–2000	0.27	0.078	0.93	0.27 (1.24)	0.74	0.22
2001–03	0.34	0.18	1.02	0.54 (2.9)	0.50	0.27
Panel (B): Annual data						
1993	0.19	0.044	0.66	0.15 (0.56)	1.35	0.31
1997	0.26	0.056	0.89	0.19 (0.93)	0.75	0.16
2000	0.29	0.11	0.98	0.38 (1.8)	0.68	0.26
2001	0.31	0.14	1.03	0.47 (2.4)	0.61	0.28
2002	0.35	0.18	1.09	0.57 (3.1)	0.54	0.28
2003	0.36	0.21	0.96	0.56 (3.1)	0.43	0.25

Sources: Based on various sources on fixed asset investments compiled by the NSB. See the text for a detailed explanation.

individual economy with the collective sector and with firms in the 'other' ownership category.

How do we reconcile the above findings with the widespread view that China's private sector developed rapidly in the 1990s? Part of the answer is provided in Tables 68.3 and 68.4 where it is clear that the investment share of the urban private sector rose dramatically in the 1990s. Thus, the boom in the private sector occurred in the easily observable urban sector. But it is important to stress that private sector development in China was overwhelmingly rural in origin. In the 1990s, the private sector in the rural areas faced increasing difficulties and because of their sheer weight the problems in the rural areas weighed down the total size of the private sector.²¹

Conclusion

According to a famous formulation, the Chinese reform has been ‘Pareto-optimal’ in that it has created winners without creating losers (Lau et al., 2000). It is time to reassess this claim about the Chinese reforms.²² The decline in income of 130 million Chinese in recent years and the massive forcible land seizures are at direct odds with the view that the Chinese reforms have been ‘Pareto-optimal’. A substantial portion of the Chinese population may have lost absolutely.

It is also time to reassess a central claim in the gradualist framework – that the Chinese reforms have continuously deepened over time and have succeeded in establishing a market economy. Many economists marvel at China’s speed of transition but, as I have shown, by an investment measure China may have retreated in establishing a market economy in the 1990s, and by an output measure the Chinese economy today is less private than the Indian economy in the early 1980s. Let us also keep in mind that in 2008 China will celebrate the thirtieth anniversary of its reform program (1978–2008). This is one year longer than the duration of the orthodox phase of central planning in China (1949–78). While some argue that China has chosen a different reform strategy as compared with other transitional economies (Stiglitz, 1999), a more plausible hypothesis is that the Chinese may have a different goal: they have reformed to preserve socialism, not to institute capitalism.

Notes

1. I thank Professors Jaime Ros and Amitava Dutt for comments on an earlier draft and Nancy Hearst for editorial assistance. The usual caveats apply. The empirical component of this chapter is based mainly on my book (see Huang, 2008).
2. For a comprehensive account of the Chinese economy, see Naughton (2007).
3. The GDP data are calculated on a foreign exchange basis and the source of the data is the *World Development Indicators*, compiled by the World Bank (available at <http://devdata.worldbank.org/dataonline>).
4. The data on social development in China and India can be accessed in the *World Development Indicators* (available at <http://devdata.worldbank.org/dataonline>).
5. The data on fixed-asset investments are reported in the *China Statistical Yearbook*, various years. For the latest data, see National Statistical Bureau (2006).
6. One exception to the findings reported here is Wang and Meng (2001), who report that TFP growth averaged 7.3 percent during the 1992–97 period but only 2.5 percent during the 1978–91 period. However, the authors themselves dismiss this finding as ‘a statistical error’ since they cannot locate the sources of this dramatic acceleration of TFP growth.
7. It is not known whether this is because of female infanticide. If it is, it is useful to determine whether female infanticide increased or decreased in the 1990s. Demographers believe that economic hard times tend to be associated with a higher rate of female infanticide. See the findings by the World Bank in research notes on China’s health sector in World Bank (2005a) and World Bank (2005b).
8. See World Bank (2001, pp. 60–61).
9. The output measures here refer to the ratios of output in labor-intensive (skill-intensive) industries to the output in less labor-intensive (less skill-intensive) industries. High (low) labor-intensive industries are those industries above (below) the median value of labor

- intensity. The skill-intensity measure is similarly derived. For details, see Kochhar et al. (2006).
10. On the logic of gradualism, see Roland (2000). Naughton (1996) provides a convincing explanation of the reforms in the 1980s based on a gradualist framework. For a rejoinder of the gradualist approach to China, see Woo (1999).
 11. On TFP in China, see Chen et al. (1988), Borensztein and Ostry (1996) and Hu and Khan (1997). Groves et al. (1994) link specific reform measures to some aspects of the operating improvements of the SOEs. The disagreements over TFP center around the relative importance of productivity improvement *vis-à-vis* factor accumulation to explain China's growth, and evidence on whether the state sector also experienced productivity growth. See the debates between Jefferson et al. (1992) and Woo and Fan (1994).
 12. The other problem is that the study treats domestic private sector firms and foreign-invested enterprises (FIEs) as a single homogenous category. This treatment does not recognize that China has favored foreign firms at the expense of domestic private sector firms. Thus, the estimate implicitly incorporates a substitution effect between FIEs and domestic private sector firms.
 13. An analogy would be those firms owned and controlled by Temasek, the holding and investment arm of the Singaporean government. Whether Temasek behaves as if it is a private firm is a separate question, but, from an accounting point of view, because Temasek itself is state-owned, the firms controlled by Temasek ought to be classified as state-owned as well.
 14. From World Bank (1989, p. 91).
 15. One huge difference separates China today from India of the late 1970s – the role of foreign direct investment. The government of Indira Gandhi severely restricted FDI whereas China today welcomes FDI. It should be stressed that the OECD's estimate of China's private sector at 52.3 percent is inclusive of foreign firms and a comparison of only domestic private firms would be even more unfavorable to China.
 16. The data on fixed-asset investments used in this section come mainly from a series of NSB publications specifically devoted to covering fixed-asset investment activities. We have checked the data in these specialized publications with those published in the annual *China Statistical Yearbooks*. In comparison with the Chinese data on output, the Chinese data on fixed-asset investments are remarkably consistent across a number of publications. The data used in the text come from NSB (1987), NSB (1989b; National Statistical Bureau 1989a), NSB (1991), NSB (1992), NSB (1997) and NSB (1999). The data for some of the later years are from NSB (2003).
 17. On the extent of state controls of fixed-asset investments, see Rawski (2001b).
 18. The individual economy includes households, that is, self-employment proprietorships.
 19. See http://www.baidu.com/s?cl=3&wd=http://news.xinhuanet.com/stock/2004-09/07/content_1952118.htm, accessed 5 June 2006.
 20. Another type of shareholding firm – known as a shareholding cooperative – is genuinely private. These were converted from township and village enterprises (TVEs) or small SOEs and are typically majority-owned by their employees. But they are very small. As of 2002, the shareholding cooperatives accounted for only 2.89 percent of China's industrial output by value, as compared with 11.7 percent for the privately-operated enterprises (*siying qiye*). Therefore, this is not a serious downward bias. It should be noted that the NSB no longer uses the 'individual economy' in its data series on industrial output, although it still uses the 'individual economy' category for its fixed-asset investment reporting. The 11.7 percent quoted in the text refers only to *siying qiye* and presumably does not include industrial *getihu*. See NSB (2003).
 21. I explore this topic in detail (Huang, 2008).
 22. One issue that clouds an assessment of Chinese performance is whether the Chinese growth rate is as truly impressive as the official figures suggest. Young (2000) and Rawski (2001a) raise questions about the veracity of the Chinese data. This is a complicated issue that I will not deal with here, except to note that better research is needed to reconcile the well-documented microeconomic inefficiencies in the Chinese economy with the apparent macroeconomic success.

References

- Bai, Chongen, David D. Li and Yijiang Wang (1997), 'Enterprise Productivity and Efficiency: When is Up really Down?' *Journal of Comparative Economics*, 24: 265–80.
- Borensztein, Eduardo and Jonathan D. Ostry (1996), 'Accounting for China's Growth Performance', *American Economic Review*, 86 (2): 224–8.
- Chen, Kuan, Hongchang Wang, Yuxin Zheng, Gary H. Jefferson and Thomas G. Rawski (1988), 'Productivity Change in Chinese Industry: 1953–1985', *Journal of Comparative Economics*, 12 (4): 570–91.
- Dougherty, Sean and Richard Herd (2005), 'Fast-Falling Barriers and Growing Concentration: The Emergence of a Private Economy in China', Paris: Economics Department, Organisation for Economic Cooperation and Development.
- Groves, Theodore, Yongmiao Hong, John McMillan and Barry Naughton (1994), 'Autonomy and Incentives in Chinese State Enterprises', *Quarterly Journal of Economics*, 109 (1), 183–209.
- Hausmann, Ricardo, Lant Pritchett and Dani Rodrik (2004), 'Growth accelerations', NBER Working Paper No. 10566.
- Heytens, Paul and Harm Zebregs (2003), 'How Fast can China Grow?' in W. Tseng and M. Rodlauer (eds), *China: Competing in the Global Economy*, Washington, DC: International Monetary Fund, pp. 8–29.
- Hu, Zuli and M.S. Khan (1997), 'Why is China Growing so Fast?', *IMF Staff Paper 44*, Washington, DC: International Monetary Fund.
- Huang, Yasheng (2008), *Capitalism with Chinese Characteristics*, New York: Cambridge University Press.
- Jefferson, Gary H., Thomas G. Rawski and Yuxin Zheng (1992), 'Growth, Efficiency and Convergence in China's State and Collective Industry', *Economic Development and Cultural Change*, 40 (2): 239–66.
- Khan, Azizur Rahman and Carl Riskin (2001), *Inequality and Poverty in China in the Age of Globalization*, Oxford: Oxford University Press.
- Kochhar, Kalpana, Utsav Kumar, Raghuram Rajan, Arvind Subramanian and Ioannis Tokatlidis (2006), 'India's Pattern of Development: What Happened, What Follows?' Washington, DC: International Monetary Fund.
- Kuijs, Louis and Tao Wang (2005), 'China's Pattern of Growth: Moving to Sustainability and Reducing Inequality', Washington, DC: World Bank.
- Lau, Lawrence J., Yingyi Qian and Gerard Roland (2000), 'Reform without Losers: An Interpretation of China's Dual-Track Approach to Transition', *Journal of Political Economy*, 108 (1), 120–43.
- McGregor, Richard (2006), 'China's Poorest Worse Off after Boom', *Financial Times*, 21 November.
- Miyamoto, Katsuhiro and Huangjin Liu (2005), 'An Analysis of the Determinants of Provincial-Level Performance in China's Economy', *Comparative Economic Studies*, September: 520–42.
- National Statistical Bureau (NSB) (1987), *Zhongguo guding zichan touzi tongji ziliao (1950–1985) (Statistical Materials on Chinese Fixed Asset Investment (1950–1985))*, Beijing: Zhongguo tongji chubanshe.
- National Statistical Bureau (NSB) (1989a), *Zhongguo guding zichan touzi tongji ziliao, 1986–1987 (Statistical Materials on Chinese Fixed Asset Investment, 1986–1987)*, Beijing: Zhongguo tongji chubanshe.
- National Statistical Bureau (NSB) (1989b), *Zhongguo guding zichan touzi tongji ziliao (1986–1987) (Statistical Materials on Fixed-Asset Investment in China (1986–1987))*, Beijing: Zhongguo tongji chubanshe.
- National Statistical Bureau (NSB) (1991), *Zhongguo guding zichan touzi tongji ziliao (1988–1989) (Statistical Materials on Fixed-Asset Investment in China (1988–1989))*, Beijing: Zhongguo tongji chubanshe.
- National Statistical Bureau (NSB) (1992), *Zhongguo guding zichan touzi tongji ziliao (1990–1991) (Statistical Materials on Fixed-Asset Investment in China (1990–1991))*, Beijing: Zhongguo tongji chubanshe.

- National Statistical Bureau (NSB) (1997), *Zhongguo guding zichan touzi tongji 1950–1995 (China Statistical Yearbook on Fixed Asset Investments 1950–1985)*, Beijing: Zhongguo tongji chubanshe.
- National Statistical Bureau (NSB) (1999), *Zhongguo guding zichan touzi tongji 1999 (China Statistical Yearbook on Fixed Asset Investments 1999)*, Beijing: Zhongguo tongji chubanshe.
- National Statistical Bureau (NSB) (2003), *China Statistical Yearbook 2003*, Beijing: Zhonggong tongji chubanshe.
- National Statistical Bureau (NSB) (2006), *China Statistical Yearbook 2006*, Beijing: China Statistics Press.
- Naughton, Barry (1996), *Growing Out of the Plan: Chinese Economic Reform, 1978–1993*, New York: Cambridge University Press.
- Naughton, Barry (2007), *The Chinese Economy: Transitions and Growth*, Cambridge: MIT Press.
- Qian, Yingyi (2003), 'How reform worked in China' in D. Rodrik (ed.), *In Search of Prosperity: Analytic Narratives on Economic Growth*, Princeton, NJ: Princeton University Press, pp. 297–333.
- Ravallion, Martin and Shaohua Chen (2004), 'China's (Uneven) Progress Against Poverty', World Bank Policy Research Working Paper No. 3408, Washington, DC: World Bank.
- Rawski, Thomas G. (2001a), 'What is Happening to China's GDP Statistics', *China Economic Review*, 12: 347–54.
- Rawski, Thomas G. (2001b), 'Will Investment Behavior Constrain China's Growth?' *China Perspectives*, 12 (4), 28–35.
- Ren, Ruoan and Lin Lin Sun (2006), 'Total Factor Productivity Growth in Chinese Industries, 1981–2000', Beijing: Beihang University.
- Roland, Gerard (2000), *Transition and Economics: Politics, Markets, and Firms*, Cambridge, MA: MIT Press.
- Stiglitz, Joseph E. (1999), 'Whither Reform? Ten Years of Transition', paper read at Annual Conference on Development Economics, 28–30 April, Washington, DC.
- Wang, Xiaolu and Lian Meng (2001), 'A Reevaluation of China's Economic Growth' *China Economic Review*, 12 (4), 338–46.
- Wolf, Edward N. (1991), 'Capital Formation and Productivity Convergence over the Long Term', *American Economic Review*, 81 (3): 565–79.
- Woo, Wing Thye (1999), 'The Real Reasons for China's Growth', *China Journal*, 41: 115–37.
- Woo, Wing Thye and Gang Fan (1994), 'How Successful has Chinese Enterprise Reform been? Pitfalls in Opposite Biases and Focus', *Journal of Comparative Economics*, 18 (3): 410–37.
- World Bank (1989), *India: An Industrializing Economy in Transition*, Washington, DC: World Bank.
- World Bank (2001), *World Development Indicators 2001*, Washington, DC: World Bank.
- World Bank (2005a), 'China's Progress Towards the Health MDGs', Briefing Note No. 2, March.
- World Bank (2005b), 'Rural health insurance: Rising to the challenge', Briefing Note No. 6, May.
- Wu, Yanrui (2003), 'Has Productivity Contributed to China's Growth?', *Pacific Economic Review*, 8 (1): 15–30.
- Wu, Yanrui (2004), 'Productivity and Sustainable Growth', in R. Garnaut and L. Song (ed.), *China: Is Rapid Growth Sustainable*, Canberra: Asia Pacific Press, pp. 181–96.
- Xu, Xiaonian and Yang Wang (1997), 'Ownership Structure, Corporate Governance, and Corporate Performance', Washington, DC: World Bank.
- Young, Alwyn (2000), 'The Razor's Edge: Distortions and Incremental Reform in the People's Republic of China', *Quarterly Journal of Economics*, 115 (4): 1091–135.
- Zheng, Jinghai and Angang Hu (2004), 'An Empirical Analysis of Provincial Productivity in China (1979–2001)', Goteberg, Sweden: Department of Economics, Goteberg University.

69 South Asia

Prabhat Patnaik and Jayati Ghosh

Similarities among South Asian economies

The countries of the South Asian region differ in size, resource endowment, the specificities of class configurations and the nature of the ruling regimes. Nonetheless they share certain common structural characteristics: high degrees of inequality of asset ownership, especially in Pakistan and India; the presence of substantial underemployment; a strong dualism between organized and unorganized sectors, especially in manufacturing, which sometimes (but not always) translates into the dualism between large-scale and small-scale economic activities; the continuing significance of agriculture as a major employer; the recent emergence of service activities as the largest incremental employers; and the involvement of the dominant share of the workforce in what are essentially low-productivity activities, often in the form of self-employment.

There is also an apparent synchronicity of policies and processes across the region, despite very differing social and political pressures. All the economies of the region had import-substituting industrialization strategies and substantial state regulation over economic activity for the first few decades after independence. From the 1980s onwards, all of them moved, in varying degrees, to a strategy of development based on export-orientation, internal deregulation, trade liberalization and privatization. The process started in Sri Lanka, as the Jayawardene government in 1977 moved towards liberalization and dismantling of the earlier universal food security system. Thereafter, especially from the early 1990s, all the governments in the South Asian region introduced policies of internal and external liberalization and privatization (Mahmud, 2000).

There was reduction in state control in terms of administered prices, regulation of economic activity and direct responsibility for a range of goods and services. Along with internal deregulation there was trade liberalization, entailing shifts from quantitative restrictions to tariffs and sharp reductions in the average rate of tariff protection. Financial liberalization involved reductions in directed credit (especially to agriculture and small industries), freeing of interest rate ceilings and other measures which raised the cost of borrowing for governments, peasants and petty producers. There was privatization of state assets, often in controversial circumstances.

All the economies moved towards 'market-determined' exchange rates, liberalization of current account transactions, and some degree of capital account liberalization, such as easing rules for foreign direct investment, allowing non-residents to hold domestic financial assets and making it easier for domestic firms to access foreign commercial borrowing. The implications of such external liberalization were very different in traditionally aid-dependent economies such as Pakistan and Bangladesh compared to India or Sri Lanka: in the latter, net capital inflows kept real exchange rates at levels that generated current account deficits, while in the former capital inflows were a substitute for aid.

In fiscal policy, most countries experienced some degree of 'rationalization' (a euphemism for reduction) of direct and indirect tax rates. This was associated with declining tax-GDP (gross domestic product) ratios in several cases, as tax buoyancy failed to meet the optimistic expectations that had justified tax rate cuts. In particular, the cuts in import tariffs (and the associated cuts in domestic duties required to establish 'level playing fields') involved lower aggregate collections relative to GDP in most of these countries, except Bangladesh. Attempts to reduce fiscal deficits typically involved cutting back public productive investment and social expenditure, reducing subsidies to farmers and increasing user charges for public services and utilities. Ironically, fiscal deficits in most cases did not fall relative to GDP, as the largest increases in expenditure came about in interest payments, partly due to the burden of past debt and partly because of the increased costs of public sector borrowing. In addition to falls in public investment as a proportion of GDP, there was also a resource crunch at regional and lower levels of government. This tended to reduce per capita spending in important areas such as basic infrastructure development, health and education (as in India), and affected the viability and legitimacy of local government institutions (as in Nepal and Pakistan).

Results of the transition from *dirigisme* to neoliberalism

The results of this process had similar outcomes in most of these economies, despite their very different initial conditions. On the positive side, growth rates on average have increased and this has generally been associated with greater macroeconomic stability in terms of lower inflation and avoidance of balance-of-payments difficulties. There has been some increase in private investment in all of these countries due to the immediate effects of liberalization and increased export orientation. However, aggregate investment rates have not increased much except in Bangladesh and Sri Lanka, and very recently in India.

However, income inequalities have increased in all the economies of the region. Growing economic inequalities are evident between rural and

urban residents; between households in various size-classes of expenditure; between sub-regions and provinces within countries. This has been associated with increased social and political tensions in the region, which have often been expressed not so much in direct demands for redressal of income imbalances, but in terms of other ethnic, social, cultural or regional demands.

There has generally been a deceleration of employment growth, compared to the *dirigiste* period. This has occurred despite an improvement, or at least the same trend level, of growth in aggregate economic activity. In general, employment has not kept pace with the increase in population, resulting in higher rates of unemployment and underemployment, and also in declining labour force participation (which is not fully explained by increased involvement in education). The quality of employment also appears to have deteriorated, with declines in regular work and increases in either casual contracts or self-employment in adverse conditions. Wage shares of income have typically declined; and real wage rates have stagnated or declined.

The relative decline of manufacturing, especially in the small-scale sector, and the stagnation or decline of manufacturing employment, is marked across the region, with the exception of the garments industries in Bangladesh and Sri Lanka. Agriculture and/or services appear to have become residual refuge sectors for workers who cannot find productive employment in industry. Across the region, there appears to have been relatively little link between rates of aggregate economic growth and employment generation.

The decline of institutional credit has been a major factor affecting the viability of agriculture and small-scale industrial development in most countries of the region. This has become a particularly severe problem in the recent past, when trade liberalization in the context of stagnant or declining world trade prices for South Asian crops has put additional competitive pressure on farmers, and contributed to an agrarian crisis across all South Asian countries.

Standard indicators of human development have improved on the whole, but the rate of improvement is much lower than desirable, and some indicators have even worsened in some cases. Literacy rates and primary enrolment have improved across the region, but the progress is uneven, with some pockets of backwardness according to region and social group; and average levels remain low everywhere except Sri Lanka. Human development indicators have improved faster only in Bangladesh, where public expenditure in social sectors has been high. Nutritional improvements have been slow or minimal, and per capita calorie consumption has fallen even in supposedly fast-growing countries like India. Infant mortality rates

have actually increased in recent years in Pakistan and in some states of India.

Notwithstanding some of these outcomes, the transition from *dirigisme* to neoliberalism has been seen by many as a vindication of the latter, especially in the context of India where the retreat from *dirigisme* has been followed by a fairly creditable export performance in some spheres and a higher rate of economic growth than most of the developing world since the mid-1990s (Joshi and Little, 1996). However, such sanguineness about neoliberalism is unwarranted. This is evident from the analytical reconstruction, within a political economy context, of India's development story, which we attempt below, and from our brief sketches of trends in other South Asian economies that follow.

India

The economic policy regime erected in the 1950s in India had its roots in the freedom struggle. The economy had been dominated by metropolitan capital and metropolitan commodities before independence. Freedom meant freedom from this domination; and this could not be ensured without giving the state in independent India a major role in building infrastructure, expanding and strengthening the productive base of the economy, setting up new financial institutions and regulating and coordinating economic activity. This was necessary for building capitalism itself, although some saw it as a means of transition to socialism. State capitalism and state intervention were essential instruments for the development of a relatively autonomous Indian capitalism, displacing metropolitan capital from the pre-eminent position it had occupied in the colonial economy.

This intervention however occurred within a certain context. Despite talk of land reform, of providing 'land-to-the-tiller', and curbing the concentration of economic power, little was done to attack or redress asset and income inequality. The worst forms of absentee landlordism were done away with, but concentration of landownership remained largely intact. And while some monopolistic practices were curbed, asset concentration in the industrial sector was never really challenged. India's monopolists were thus able to use state intervention as a device to consolidate and expand their positions.

The persistence of asset and income inequality imposed a constraint on the expansion of the market for mass consumption goods in the country. The absence of any radical land redistribution meant that the domestic market, especially for manufactured goods, remained socially narrowly based. It also meant that the growth of agricultural output, though far greater than in the colonial period (where the last half-century had witnessed virtual stagnation), remained well below potential. For the country

as a whole, the benefits of such agricultural growth as did occur were largely confined to a relatively narrow stratum of landlords-turned-capitalists and sections of rich peasants who had improved their economic status. Meanwhile, industrial growth was not sufficiently employment-generating to create large increases in demand from this source.

In this context, continuous growth in government spending became crucial for imparting a dynamic to the system. Import-substitution policies provided domestic capitalists with a large initial market for manufactures, which the government sought to expand through its current and capital expenditures. At the same time the state supported the domestic capitalist class through its infrastructure investment and through the creation of a number of specialized institutions for financing private investment. While this strategy succeeded in overcoming the long stagnation in agriculture, established a range of basic and capital goods industries, and achieved on average a rate of industrial growth in excess of 7 per cent during the 1950s and the early 1960s, by the mid-1960s it was clearly running out of steam. Not only was the initial stimulus offered by import substitution exhausted, but the capacity of the state to continue to provide the stimulus to growth was undermined by its inability to raise adequate resources.

This reflected the contradictory nature of the roles the State was required to fulfil (Patnaik, 1994). While increasing government expenditure was necessary to keep the domestic market expanding, the state also became the most important instrument for what Marx had called 'primary accumulation' by the domestic bourgeoisie, which was reflected in the fiscal crisis of the state. This implied that the government had either to cut back the tempo of its investment, or to maintain this tempo through increased borrowing with inflationary consequences, since the *dirigiste* economy was not a demand-constrained one.

The state's inability to impose a measure of discipline (essential for viable capitalist functioning) on the capitalists, which made tax evasion rampant and contributed to the fiscal crisis, also made the attempts at regulation through licensing and other instruments quite ineffective. The imbalance between the pattern of domestic production envisaged in the plans, and the pattern of demand emanating from the relatively affluent sections who were the main beneficiaries of growth, gave rise to unutilized capacity through the illicit diversion of resources towards unplanned end-uses. The consequent 'irrationality' reflected *inter alia* in an increase in the capital-output ratio, compounded the problem arising from the increasing fiscal constraints upon the state.

This basic contradiction of the *dirigiste* regime, namely that it could sustain the tempo of development only at the cost of accentuating inflation

which undermined its own social stability, became apparent after the mid-1960s. After 15 years of rapid industrial expansion until the mid-1960s, industrial growth slowed over the next decade. This was mainly because of a cutback in public investment in the face of strong inflationary pressures, accentuated by the oil price hike of 1973. Public investment recovered somewhat after the mid-1970s, by putting a check on inflation, first by turning the terms of trade against agriculture and then, after the mid-1980s, by going in for external borrowings. But such borrowings made the economy vulnerable to capital flight. One such episode, triggered by the payments difficulties arising from the Gulf War in 1991, led to the explicit introduction of a neoliberal economic regime.

Growth had already picked up from the mid-1980s. The annual rate of growth of real GDP rose from an average of over 3.5 per cent per annum in the three decades after independence, to between 5 and 6 per cent in the 1980s and 1990s. Rather than import liberalization per se, the growth of the 1980s was related to a strong revival of agriculture through the spread of the 'Green Revolution' to Eastern India. It also owed much to higher government spending, financed increasingly by external commercial borrowing in the late 1980s. While domestic demand was kept up by higher government spending, import liberalization (especially of capital goods and components for manufacturing) reduced inflationary pressures and imparted an impetus to final good production. Inflation was also kept under control by lower relative prices in agriculture.

The inter-sectoral terms of trade for agriculture in the 1980s and early 1990s remained low compared to the early 1970s; but, for almost a decade prior to the mid-1990s, Indian agriculturists were protected from the far more adverse international movement of terms of trade against primary products. Relative prices for agriculture worsened again only in the late 1990s, when trade liberalization exposed farmers to declining world prices. This marked the onset of the period of agrarian crisis and increase in peasant suicides.

While the internal contradictions of the *dirigiste* regime were bringing it to a dead-end, major international changes also undermined *dirigisme*. The easier access to private capital flows negated the basic assumption of a binding foreign exchange constraint. More fundamentally, the process of globalization of finance sought to break down all barriers to the free flow of finance capital. It has been argued that the economic reforms after 1991 reflected an acceptance of the validity of the neoliberal economic paradigm (Bhagwati, 1993). It is more correct to locate the policy shift in the totality of circumstances produced by the interplay between the changing external context and the accentuating domestic contradictions within the earlier regime.

These contradictions encompassed other elements too (Chandrasekhar and Ghosh, 2004). A comparatively narrow social segment had provided the main source of growth in domestic demand under *dirigisme*. This social segment, eager to emulate the lifestyles and consumption patterns of the metropolitan centre, was not satisfied with having more domestically produced goods; its demand was increasingly for the new goods produced in the metropolitan centres, which could not be locally produced using only indigenous resources and technology. This imbalance increased over time because of further innovations in the metropolitan economies, creating a powerful and growing pressure from the more affluent groups for a dismantling of controls.

The emergence of newer strata, through a process of proliferation and diversification of the Indian capitalist class during the years of *dirigisme*, was another contributory factor. New capitalists operating outside the traditional bases of existing monopolistic groups, such as trade, finance, services of various kinds, had come into being, and operations abroad by non-resident Indian groups had become significant. These new entrants sought to diversify into manufacturing, and therefore welcomed deregulation. Because of access to newer technology, they were less averse to import competition. The established big businesses too which were, to start with, beneficiaries of state controls of various kinds, began to chafe against these controls at a certain stage in their search for new avenues for investment and production.

Support for economic liberalization also came from other quarters: new businessmen involved in what were essentially 'parallel market' transactions; a section of the top bureaucracy; the large and politically powerful urban middle classes; the more prosperous farmers whose real incomes increased in the 1980s. The technological and media revolutions imparted a significant impetus to the international demonstration effect, further fuelling neoliberal and consumerist demands. The 'Non-Resident Indian (NRI) phenomenon', whereby a significant number of people from the Indian elites and middle classes actually became resident abroad, also contributed to demands for opening up the economy.

The neoliberal reform strategy did not have much impact on growth to start with. The growth rate during the decade of the 1990s was scarcely any higher than during the 1980s; in the material commodity-producing sectors, agriculture and industry taken together, it was lower. Agricultural growth in particular decelerated dramatically during the 1990s, which became the first decade since independence to witness a decline in food-grain output per capita. So great however was the compression in domestic demand, especially in rural India where government expenditure was sharply reduced, that per capita foodgrain absorption fell even more

dramatically, resulting in the accumulation of 63 million tonnes of food-grain stocks by mid-2002, of which 41 million tonnes were excess stocks. Per capita absorption of foodgrains for the country as a whole fell from the post-war triennial peak of 178 kg per annum during 1988–91 to 157 kg by 2001–04, which was the level in ‘British India’ on the eve of World War II. The excess stocks were largely dumped on the international market, where they were bought up to be used as animal feed for the rich countries. This was because neoliberal orthodoxy was hostile to the use of foodstocks on employment schemes, lest it enlarge the fiscal deficit (even though such deficit entailed no inflationary consequences in a demand-constrained system, and would not even raise the government’s net indebtedness much, as the foodstocks were with the government-owned Food Corporation of India).

GDP growth accelerated only after 2001, reaching more than 8 per cent between 2003–04 and 2005–06, mainly because of certain ‘newer’ services and in some export-oriented manufacturing of garments and chemicals. This period however also saw an even sharper divergence between agricultural and non-agricultural growth. Foodgrain production showed a zero trend, so that notwithstanding fiscally caused demand compression in the countryside, demand-pull inflation resurfaced by 2006. This was worsened by the government’s policy of stock reduction and winding down of public procurement operations in foodgrains.

Does liberalization per se account for India’s more successful recent export performance? If we leave aside ‘gems and jewellery’ where India was a successful exporter even before ‘liberalization’, and garments where India and her low-wage neighbours have been traditionally powerful, the one new area of strength that has emerged recently is information technology (IT)-related services and business process outsourcing (BPO). Here, the existence of a substantial educated manpower, whether with high skills as in the case of software exports, or with low skills as in the case of BPO, has been an important contributory factor. But the credit for this must go to the earlier *dirigiste* regime which defied conventional wisdom in setting up institutions of higher education, including several of excellent quality. Likewise, even the recent surge in exports of cement, steel and construction material are forays into the world market of industries that were set up and strengthened during the *dirigiste* period. All these avenues of export success would have eluded the Indian economy had it adopted neoliberal policies from the beginning, and not broken decisively out of the inherited colonial pattern of international division of labour.

Indeed, this was precisely the idea of P.C. Mahalanobis, the architect, along with Jawaharlal Nehru, of the *dirigiste* strategy in India. Mahalanobis assumed a closed economy in his 1950s plan model (inspired

by Soviet experience) because of his export pessimism on the basis of the then existing production structure. However, he strongly believed that by breaking out of this production structure through protectionism and state intervention, and imparting divergence and depth to it, the economy would be able to emerge as a more successful exporter at a later date. Recent Indian experience appears to vindicate him more than his neoliberal critics.

The recent acceleration in growth however cannot be explained by export performance. True, the export performance, especially in the service sector, has boosted domestic incomes and consumption to an extent at the upper end of the spectrum. But much of this boost is self-propelling, giving a lift to 'animal spirits' à la Keynes, which explains the recent increase in investment ratio. Underlying this boom are: a consumption splurge by the upper income groups, driven by the international demonstration effect of metropolitan living standards; a construction-cum-real estate boom, led by an enormous appreciation in land values; and an explosion in corporate profits. Two of the most palpable features of the contemporary Indian economy fall into place here: the increase in the unemployment rate even in the midst of this unprecedented boom, since structural-cum-technological change is in the direction of high-productivity sectors; and the rampant drive to displace peasants from land in the name of 'infrastructure projects', 'Special Economic Zones' and such like, all of which camouflage land speculation.

This growth has been accompanied by significant increases in inequality – both across regions of India and within regions across different economic and social categories. There is a widening gap between incomes in agriculture and non-agriculture: the ratio of per-worker domestic product in non-agriculture to that in agriculture which was about 2 in the 1950s has increased to well over 4 in the early 2000s. The wage share of national income has also fallen sharply and the wage share of value added in organized manufacturing declined to only around 10 per cent in 2004. Consumption surveys show absolute declines in the income and consumption of a substantial share of the population. During the last two decades of the twentieth century, the urban top 20 per cent of households experienced historically unprecedented increases in per capita consumption in real terms, while the per capita consumption of the bottom 40 per cent of the rural population actually declined.

Agriculture, which continues to employ the bulk of the workforce, has been afflicted by an intense and prolonged crisis. Its problems are closely related to more open trade combined with domestic deflationary policies. From the mid-1990s, the financial viability of cultivation has been eroded, owing not only to longer-term problems such as declining soil fertility, changing weather conditions and excessive dependence on depleted

groundwater, but also to a combination of sharply rising input prices and volatile and declining output prices. Indian farmers were encouraged to shift to cash crops, and exposed to international competition from the highly subsidized crop production in the North, even as they were facing reduced support from central and state governments in the form of inadequate availability of institutional credit, decline in public extension services, insufficient regulation of some inputs and reduced subsidies for other inputs such as fertilizers, reduction in timely crop price support, and poor storage and marketing facilities.

Poor employment generation remains a critical issue. Agricultural employment has fallen, due to both agricultural stagnation and technological and cropping pattern changes that reduced labour demand in agriculture. Such changes were also hastened by the growth of landlessness (as peasant cultivation became less viable because of rising input costs and falling or stagnant crop prices) accompanied by a shift to commercial cultivation with hired labour. In urban India, manufacturing is increasingly characterized by more capital-intensive techniques, and therefore declining employment elasticity of production. Even the 'newer' and more dynamic services such as IT-enabled activities that have increased their share of output still remain minuscule in employment terms. All IT-related activities currently employ less than 0.2 per cent of the total workforce. Recent employment increases have been mainly in the form of subsistence self-employment in low-value services, despite the economic boom.

Until the late 1980s, there was a secular trend towards declining poverty.¹ Subsequently, while official estimates show a decline in poverty, nutritional indicators suggest the contrary. The coverage and quality of public services has worsened, which has had particular impact upon the condition of women and girl children (Dreze and Sen, 1994). The persistence of illiteracy, especially among females; the inability to ensure even primary education to all children and high drop-out rates over successive years of schooling; the poor indicators of health and the recent stagnation of infant and maternal mortality rates; the absence of proper sanitation for a large proportion of the population – all these provide an indication of the current state of the development project in India.

The external sector, however, has displayed a degree of overall stability in the balance of payments and a relative absence of the boom-and-bust cycles that marked some other emerging markets. To some extent this reflects the relatively limited extent of capital account liberalization over much of the period, and the fact that the Indian economy was not really 'chosen' to be a favourite of international financial markets until 2002. The greatest stability to the balance of payments was imparted by the substantial inflows of workers' remittances from temporary migrant workers in the

Gulf and other regions, which has amounted to more than all forms of capital inflow put together. Since 2004 there has been a sharp increase in portfolio capital flows and external borrowing, but FDI remains relatively small.

The alternative to neoliberalism in India consists of policies to ensure the viability and sustainability of agriculture, and greater emphasis on public expenditure with high direct and indirect effects on employment generation, especially in infrastructure, health and education. This requires higher resource mobilization from the rich. It is also necessary to counter some of the adverse effects of trade liberalization on employment, apart from more directly addressing the basic structural issues of asset and income inequality and the persistence of low-productivity employment. Of course, such a policy shift requires political will and therefore a change in political configurations.

South Asian scenario

The political economy of the transition from *dirigisme* to neoliberalism in other South Asian countries, while differing in detail from that of India, was roughly similar: it represented everywhere the abandonment of the domestic bourgeoisie's quest for a relatively autonomous trajectory of capitalist development, and the pursuit of an alternative trajectory with much closer integration between domestic and metropolitan capital, between finance and industry, and between the local and the global. It therefore meant a growing hiatus between the bourgeoisie and the urban upper income groups on the one side and the bulk of workers (facing insecure employment), peasants (facing agrarian crisis), petty producers and small businessmen (facing closures), and agricultural labourers (facing shrinking employment) on the other.

Pakistan

Despite a respectable per capita growth of around 5 per cent per annum over the second half of the twentieth century, Pakistan systematically underperformed on most social and political indicators, including education, health, sanitation, fertility, gender equality and political instability. In general, output growth has been associated with very low employment growth, at the trend rate of only 2 per cent per annum for the long period 1960–99. Employment growth worsened after the imposition of an International Monetary Fund (IMF) structural adjustment programme in 1987–88. In the 1990s, economic growth reduced on average and also became much more volatile. This was associated with historically low rates of investment, as private investment failed to revive or compensate for the decline in public spending. The investment–GDP ratio declined from 17.3

per cent in 1998–89 to 14.7 per cent in 2000–01, largely due to the collapse in public investment. Industrial growth rates almost halved from 8.2 per cent to 4.8 per cent per annum. Further, the earlier success at reducing poverty was reversed in the 1990s, as the percentage of households living in absolute poverty increased from 21.4 per cent in 1990–91 to 40.1 per cent in 2000–01 (Hussain, 2004).

The initial years after Pakistan's third military coup witnessed a worsening of the macroeconomic situation, with increasing poverty and unemployment, falling real wages and worsening income distribution. However, recent geopolitics has impacted in some positive ways upon Pakistan's economy, mainly because of the willingness of the Musharraf regime to be a key ally of the USA. This caused the waiver or rescheduling of more than one-third of Pakistan's external debt, an increase in foreign aid flowing back to Pakistan, and the reinstating of export quotas in textiles and garments. Pakistani workers abroad have contributed to the massive recent increase in remittances, to as much as 14 per cent of GDP. However, since the domestic investment rate is still below the savings rate, the inflow of aid and remittances has not really contributed to economic activity, but is simply stored as foreign exchange reserves.

Bangladesh

Bangladesh shows a different and somewhat more optimistic economic trajectory than other South Asian countries. Since independence in 1971, there has been a slow but accelerating improvement in living standards. GDP grew at an average rate of 3.7 per annum in the 1980s, 4.8 per cent in the 1990s, and 5 per cent in the 2000s. Per capita income has grown even faster as population growth has slowed down in the recent years. With the fertility rate reduced from 6.3 children in 1975 to 3.3 in the mid-1990s, the rate of population growth has fallen from over 3 per cent to less than 1.5 per cent in a space of three decades. Per capita income growth, consequently, has doubled from about 1.6 per cent in the 1980s to over 3 per cent in the following decade and a half (Muqtada et al., 2002). This has been associated with a reasonable degree of macroeconomic stability, with inflation rates falling from over 10 per cent in the 1980s to just over 5 per cent in the subsequent 15 years, and moderate falls in the fiscal deficit and trade deficit ratios to GDP. Strong export performance in the garments sector and a steep increase in workers' remittances from abroad have played an important role in improving both external and internal balances of the economy, as the inflow of foreign aid has experienced a secular decline.

The rate of investment in Bangladesh increased slightly from 17 per cent of GDP in the 1980s to close to 20 per cent in the subsequent 15 years, with

both public and private sectors contributing to this increase. A major factor was the increase in public expenditure over this period, despite declining foreign aid. This was partly because of the rising tax–GDP ratio, reflecting higher indirect tax revenue after the imposition of a uniform value-added tax. The share of social sectors, such as education and health, increased from 15 per cent of the growing total budgetary expenditure to over 20 per cent. The share of physical infrastructure also increased. In consequence, the headcount poverty ratio declined from 71 per cent in 1973–74 to 40 per cent in 2000, with the 1990s witnessing a more rapid decrease. Human development indicators also improved: the decline in infant mortality in Bangladesh, from 95 (per 100 live births) in 1972 to 30 (per 100 live births) in 2000, was among the fastest in the developing world. In spite of these achievements, the basic development challenges remain, requiring continued government intervention to ensure income diversification and improvements in living standards in the future.

Sri Lanka

Sri Lanka has been often hailed for its high achievements in human development, despite limited growth. Since independence in 1948, Sri Lanka has registered an average annual growth rate of 3–4 per cent while achieving adult literacy rates of 92 per cent and life expectancy of more than 72 years by 2001. Nevertheless, Sri Lanka remains a low-income, food-deficit country with more than 34 per cent of households below the poverty line in 2001.

In the 1970s, poverty alleviation strategies included free or highly subsidized social and economic services, such as education and health care. There were food subsidies as well as producer subsidies intended to help agricultural producers. These created the highest levels of human development in the region. However, since 1977, Sri Lanka pursued an economic liberalization policy. This was initially accompanied by increased public investments financed by external borrowing, but this could not be sustained (Dunham and Jayasuriya, 2000). From the mid-1980s, the government pursued a macroeconomic stabilization course along with mounting military expenditures in the civil war. From the early 1990s, deregulation and liberalization policies were supposed to bring about economic growth, but success here has been confined to the garments industry. Employment has stagnated, real wages have fallen throughout the 1977–2001 period, and most of the population continues to be employed in commodity production activities as farmers or production workers, and possesses limited education. Remittances (largely from women workers abroad) have been crucial in stabilizing the balance of payments and providing buffer incomes for the poor.

Nepal

A landlocked least-developed country, Nepal has major infrastructure gaps and much of it is physically inaccessible. There are marked regional differences between the Terai, hill and mountain areas, with decreasing infrastructure and increasing poverty as one moves up. Agriculture dominates the production structure, and rural people dominate the poor. The problems of backwardness are compounded by a long and relatively open border with India, which creates a high degree of dependence upon the Indian economy and also a vulnerability to India's macroeconomic policies. The extremely adverse material conditions have generated extreme political movements, which in turn have generated much political instability.

There was a systematic decline in public investment from around 25 per cent of GDP in the late 1970s, to 18 per cent in the late 1980s, to only 11 per cent in the 1990s. This was associated with low and falling rates of private investment. Neoliberal reforms from the early 1990s have been associated with stagnation in agricultural incomes. The manufacturing sector has focused on exports, of garments, carpets and pashminas; there has been little development of manufacturing for the home market. Import penetration has prevented the development of infant industries and led to closure of small units and low employment generation in manufacturing. Financial sector reforms have been associated with reduced flow of institutional credit to agriculture and small enterprises, and fewer possibilities of subsidies for providing micro-credit to the poor through government channels. Open unemployment rates are low, but rates of underemployment are very high, estimated to be around 43 per cent of the labour force. However, recent political changes in Nepal give rise to guarded optimism about the future, and the possibility of economic strategies that will focus more on livelihoods of the poor and productive employment generation in a diversified economy.

Conclusion

Dirigisme outside India was always more vulnerable, since the countries concerned were small in size. Paradoxically however, unlike in Latin America, no effort was made during the entire *dirigiste* period for closer economic integration between the South Asian countries. On the contrary, competition between the South Asian countries, such as between Sri Lanka and India in tea, and between Pakistan and India in jute, was intense and yielded suboptimal outcomes, compared to what cooperation even after an initial phase of competition could have achieved.

South Asia's quest for economic development has been accompanied by a parallel quest for building modern nation-states encompassing multiple religious, linguistic, ethnic and regional groups. The difficulties of this

process, exemplified by Tamil secessionism in Sri Lanka, by the break-up of erstwhile Pakistan, by the periodic eruptions of secessionist movements in India, and by a host of conflicts dotting the entire region at any time, have been compounded by the consequences of the development trajectories pursued. In a *dirigiste* regime the distribution of the 'gains' from development across the proto-bourgeoisies belonging to different ethnic, linguistic and such other groups is determined by the nature of the state, by who has greater control over the state apparatus. In a neoliberal regime there exists a spontaneous tendency towards increasing divergence starting from certain initial differences. The resulting discontent under neoliberalism allows scope for self-serving intervention by powerful elements of metropolitan capital backed by metropolitan states, which creates a tendency towards a fracturing of the nation. The growing inequalities become particularly significant in this context: since any growing hiatus necessarily has some regional or linguistic or religious dimension, it can become a means of fragmentation of the nation. *Dirigisme* in South Asia did not always succeed in keeping these contradictions in check. Neoliberalism is further accentuating these contradictions. The development of humane societies in this region still has a long way to go.

Note

1. The 'poor', officially, are those with consumption below the official poverty line. This is determined by applying price-index adjustment to the observed cost of that 1973–74 consumption basket at which a certain number of calories were just accessed.

References

- Bhagwati, Jagdish (1993), *India in Transition: Freeing the Economy*, Oxford: Clarendon Press.
- Chandrasekhar, C.P. and Jayati Ghosh (2004), *The Market that Failed: Neoliberal Economic Reforms in India*, 2nd edn, New Delhi: Leftword Books.
- Dreze, Jean and Amartya Sen (1994), *India: Economic Development and Social Opportunity*, Delhi: Oxford University Press.
- Dunham, David and Sisira Jayasuriya (2000), 'Equity, Growth and Insurrection: Liberalization and the Welfare Debate in Contemporary Sri Lanka', *Oxford Development Studies*, 28 (1): 97–110.
- Hussain, Akmal (2004), 'Institutions, Economic Structure and Poverty in Pakistan', *South Asia Economic Journal*, 5 (1): 69–102.
- Joshi, Vijay and I.M.D. Little (1996), *India's Economic Reforms 1991–2001*, Oxford: Clarendon Press.
- Mahmud, W. (ed.) (2000), *Adjustment and Beyond: The Reform Experience in South Asia*, International Economic Association, London: Macmillan.
- Muqtada, M., A.M. Singh and M.A. Rashid (2002), *Economic and Social Challenges of Globalization in Bangladesh: Policy Perspectives*, Bangladesh: The University Press, Dhaka.
- Patnaik, Prabhat (ed.) (1994), *Macroeconomics*, New Delhi: Oxford University Press.

70 The East Asian newly industrializing countries

*Jomo K.S.*¹

The image of rapid growth throughout East Asia following Japan has implied a more coherent region than actually exists, and a corresponding tendency to see economic progress in the region as similar in origin and nature, leading to regional terms such as the 'Far East', 'Asia-Pacific' and 'Pacific Asia' preceding others such as 'East Asia', 'yen bloc', 'flying geese', 'tigers', 'mini-dragons' and so on. On the other hand, rather amazingly, the World Bank's (1993) influential *East Asian Miracle* volume hardly considered the significance of geography or location. For the Bank, it did not seem to matter that the eight high-performing Asian economies (HPAEs) – incredibly, with the benefit of hindsight, excluding China – were all from the same part of the world and were not only contiguous but often also linked by trade, finance and investment relations.

The next section starts with a critical review of the World Bank's (1993) influential study of the East Asian economic miracle. The following section emphasizes the diversity of East Asian experiences and the significance of recognizing this diversity for drawing appropriate lessons. The chapter then considers implications of the East Asian financial crises of 1997–98, before the final section addresses some issues in drawing lessons from the East Asian experience.

Interpretations of the 'East Asian miracle'

There are at least three major competing explanations for the rapid growth and industrialization of several East Asian economies, often referred to as the 'East Asian miracle'. The dominant version in the 1970s and 1980s presented sustained rapid growth in the East Asian region as essentially due to unfettered market forces (Little et al., 1970). The obvious policy implication was to liberalize, as well as to open up or globalize. In the late 1980s, this was challenged by an almost opposite, *dirigiste* position, which emphasized the leading role of the 'developmental state' (Johnson, 1982; White, 1988; Amsden, 1989; Wade, 1990). The World Bank's 1993 study seemed to offer an intermediate view by acknowledging the role of the state, while emphasizing the importance of the market.

The World Bank's (1993) *East Asian Miracle* identified at least seven types of state interventions. It approved of the first four, deemed functional interventions, namely: (1) ensuring macroeconomic discipline and macroeconomic balances; (2) providing physical and social infrastructure; (3) providing good governance more generally; and (4) raising savings and thus investment rates. Functional interventions were said to compensate for market failures, and hence, were deemed acceptable, if not desirable, and less distortive of markets. Three types of market-distortive strategic interventions were considered in the areas of finance: directed (that is, subsidized) credit, trade policy and industrial policy.

The interventions were not just market-conforming, but also played important proactive roles which have been more than simply market-augmenting, as suggested by the World Bank's analysis. Of the more controversial strategic interventions, the Bank conceded that financial interventions had been important and successful in East Asia, particularly in North-East Asia, that is, in Japan, Korea and Taiwan, but considered other strategic interventions to be failures. However, the Bank maintained that other developing-country governments were not capable of successfully pursuing the types of policies North-East Asians had successfully implemented because their state capabilities were almost unique and virtually non-replicable.

The Bank volume's evaluation of the role and record of Japan's Ministry of International Trade and Industry (MITI) and its counterparts elsewhere in the region is more predictable, arguing that government interventions have been trade-distortive and generally unsuccessful in East Asia, although export promotion in particular comes in for much praise. The Bank disapproves import protection and fails to see the connection with export promotion. The Japanese, South Korean and Taiwanese governments implemented import substituting industrialization (ISI) policies from the 1950s, but also pursued export-promotion soon after to ensure that their industries quickly become internationally competitive by requiring a rapid transition from import substitution to export orientation, that is, 'effective protection conditional on export promotion' (EPconEP). EPconEP is quite different from just import substitution or export promotion for example in export-processing zones (EPZs). Both import substitution and export promotion (for example subsidies) are trade distortions. It is also misleading to suggest that EPconEP is tantamount to 'simulated free trade' because the 'bad' import-protection distortion is negated by the 'good' export-promotion distortion. EPconEP allows industries and firms to enjoy rents from import protection to develop new industrial and technological capabilities in order to become internationally competitive in terms of both cost and quality.

Infant industries have often been provided with effective protection conditional on export promotion, which has had the effect of forcing firms and industries quickly to become internationally competitive. By giving firms protection for certain periods, depending on the product, and by also requiring that they begin exporting certain shares of output within specified periods, discipline was imposed on the firms in return for the temporary trade protection they enjoyed. Such policies forced firms to reduce their unit production costs as quickly as possible, for example by trying to achieve greater economies of scale and accelerating progress up learning curves. Requiring exports has also meant that producers have had to achieve international quality standards quickly, which imposed pressures to progress technologically in terms of product design and quality as well as technological processes. With strict discipline imposed, but also some flexibility in enforcement, many firms managed rapidly to achieve international competitiveness.

Regional diversity

Some important differences among the East Asian miracle economies suggest that not all East Asian economies have been proceeding inexorably in the same direction in a similar manner. The East Asian experiences are far from constituting a single model. Some of the major differences in East Asia are themselves very instructive. In the case of the role of foreign direct investment (FDI), tremendous contrasts are found. In the case of Singapore, FDI has constituted about a quarter of gross domestic capital formation and about 15 per cent in Malaysia. At the other end of the spectrum, in the case of Japan and Korea, the percentage has long been below 2 per cent. Some of the other countries fall between these two extremes, with few near the mean for developing countries of around 5 to 6 per cent. Those most successful in developing industrial capacities and capabilities in East Asia – namely Japan, South Korea and Taiwan – have hardly depended on FDI.

FDI's far greater importance in South-East Asia has been due to a variety of reasons, which have not been entirely economic. One reason for the major role of FDI in Singapore and Malaysia is political. After Singapore seceded from Malaysia in 1965, the regime decided to attract FDI in massive quantities to ensure its own survival, so that the major foreign powers would quickly develop a stake in the survival of the Singapore regime. Subsequently, this FDI preference has been justified in terms of improving access to foreign markets and technology.

Malaysia has long had ethnic rivalries and an ethnic affirmative action policy. Some policy-makers tried to limit ethnic Chinese control of the economy by encouraging foreign direct investment. Again, one finds a

political motivation for the important role of FDI in Malaysia. In other words, political considerations have been a very important reason for attracting, even privileging foreign investment in Singapore and Malaysia.

There has also been considerable diversity in the role and performance of public investments, including state-owned enterprises (SOEs), in East Asia. In South Korea, Japan and, of course, Hong Kong, SOEs are hardly important today, but historically, state-owned enterprises were important in Japan before World War II, and in South Korea until more recently. Conversely, however, one finds that state-owned enterprises have been extremely important in Singapore and Taiwan. Again, this is partly explained by political factors, but there are also economic considerations. And very importantly, the performance of these SOEs has also been quite impressive.

Industrial and technology policies in East Asia have also been quite diverse. One extreme, of course, is Hong Kong, where there has been relatively little industrial policy, although more than most opponents of industrial policy care to admit, especially in recent years. It is far more detailed and sophisticated in Japan and Korea at the other end of the spectrum. In Korea, industrial policy is largely oriented towards large firms, whereas in Taiwan, much more emphasis is given to medium-sized and relatively smaller enterprises.

Industrial policies in the region have also had different orientations, emphases and instruments. For example, trade policy has been very important in almost all economies in the region except Hong Kong and Singapore, while financial policy has been important in all the countries, including Singapore, but again, with the exception of Hong Kong before the 1998 crisis. Since Hong Kong's reversion to China in mid-1997, there have been many indications of greater government interventions in the territory, presumably in line with its new status and China's envisaged role for the deindustrialized financial centre.

The World Bank recommended that the rest of the developing world emulate South-East Asia, not North-East Asia because of important differences between them. These misleading claims require us to recognize the far more impressive achievements and superior economic performance of the first-tier East Asian newly industrialized economies (NIEs) (including Singapore), compared to the second-tier South-East Asian NIEs. The World Bank (1993) argued that the South-East Asian high-performing economies were the preferable model for emulation by other countries seeking late development.

According to Yoshihara (1988), the South-East Asian economies have been characterized by 'ersatz capitalism' because of the compromised and inferior role of their states, their maltreatment of ethnic Chinese and their failure to develop better technological capabilities. Jomo et al. (1997)

criticized the World Bank's claims, suggesting various problems associated with the growth experiences of the South-East Asian economies praised by the Bank. In any case, the currency and financial crises of 1997–98 radically transformed international opinion about the East Asian models, with earlier praise quickly transformed into condemnation (Jomo, 1998).

The Malaysia, Indonesia and Thailand (MIT) economies as second-tier or second-generation East Asian newly industrializing countries (NICs) share some common characteristics with Singapore, which is also in the region. However, they are not only far less advanced in developmental terms, but also quite different from the city-state's heavy reliance on trade and financial services besides manufacturing. Essentially, the MIT economies have had somewhat different, even ersatz developmental states and industrial policies, compared to the first-generation East Asian newly industrialized economies (NICs). Although Singapore too has pursued industrial policy, it has used fewer trade policy instruments and has been far more reliant on foreign direct investment compared to the other East Asian NIEs. Though Singapore, like Hong Kong, has eschewed trade policy instruments, it has used state-owned enterprises (SOEs) – usually referred to as government-linked corporations (GLCs) in the island republic – more than any other East Asian economy, and perhaps any other economy in the world in the early twenty-first century.

Most importantly, the South-East Asian high-growth economies (including Singapore) have relied much more heavily on foreign direct investment (FDI) to develop most of their internationally competitive industrial capabilities. Trade policy instruments in the region have been less well formulated and implemented, with rather mixed consequences, but have nonetheless been part of the region's industrial policy story. Generally, government interventions in the region have been influenced by a variety of considerations besides economic development and late industrialization. Consequently, industrial policy has also varied in nature, quality and effectiveness. Yet, the economies in the region would not have achieved as much as they did without selective government interventions, including industrial policy.

Growth performance has been superior in North–East Asia over the long term despite the much greater resource wealth of South–East Asia. Over the period studied by the Bank, that is, from the 1960s until the early 1990s, growth in the former averaged about 8 per cent, compared to about 6 per cent for the latter. A 2 per cent difference, compounded over a period of a quarter-century or more, adds up to a lot. Very importantly, except in Hong Kong (due to immigration from China) and perhaps Singapore, population growth has been much lower in the former compared to the latter. Immigration into Hong Kong and Singapore involves a very high

proportion of the labour force, thus raising the average labour utilization rate. Political factors have also ensured far more equitable distribution of economic welfare than would otherwise have been the case in the first-tier NIEs, whereas such considerations have been less influential in the second-tier South–East Asian NICs despite Malaysia’s ethnic ‘social contract’ and Indonesia’s rural development efforts to achieve political stability.

Improvements in per capita income and economic welfare have been much more significant in North–East Asia, compared to South–East Asia (with the exception of Singapore), despite the greater resource wealth of the latter. Income inequalities have also been far less in North–East Asia, although there is some evidence of rapid recent increases in inequality. In other words, what South–East Asia has achieved has been less impressive in some critical ways. Drawing from this contrast, some people now argue that resource wealth is not a blessing, but a curse, in so far as it may have postponed the imperative to industrialize.

The North–East Asian NIEs have generally had much more sophisticated and effective industrial policy compared to South–East Asia’s NICs. This accounts, in no small way, for the very important differences in industrial and technological capabilities between North–East Asia and South–East Asia. Also, South–East Asian industrialization is still driven by FDI, whereas North–East Asian industrialization is primarily an indigenous phenomenon.

Japan and the first-generation NIEs began to industrialize in the very specific economic and political conditions of the post-World War II Golden Age and Cold War. North–East Asia grew rapidly in the immediate post-war period under a ‘security umbrella’ provided by the Americans, including aid and privileges no longer available to others. Besides providing generous aid, the Americans were anxious for them to ‘succeed’ economically in order to be showcased as attractive alternatives to their neighbours under communist rule or influence. Hence, the Americans were quite happy to tolerate trade, finance, investment, intellectual property and other policies violating neoliberal economic norms that they are now strongly opposed to. Such conditions are simply no longer available to others, and hence, their experiences are more difficult to emulate. To discourage other developing countries from trying to emulate the first-generation East Asian NIEs, it is now often argued that their state capabilities are culturally unique and impossible for others to emulate.

The Guomindang government of Taiwan was the same regime driven out of mainland China by the communists because of its incompetence and corruption. One could say the same of the Rhee regime in Korea in the 1950s. Japan has hardly been scandal-free in recent years, and most observers trace recent abuses to the nature of Japan’s post-World War II

political economy. The superior policy-making and implementation capabilities of the North–East Asian decision-makers was, at least until the 1997–98 financial crisis, widely acknowledged, but this in itself does not prove that policy-makers were thoroughly competent and incorruptible.

Some also claim that East Asia cannot be emulated owing to its very different initial conditions. Such differences are real, but often exaggerated. There is no doubt that Japan as well as the first-tier East Asian NIEs have also been distinguished by much higher levels of educational achievement. However, the level of literacy in Korea in 1950 was lower than the literacy rate in contemporary Ethiopia, which has one of the lowest literacy rates in Africa today. The educational achievements of contemporary South Koreans reflect the tremendous human resource investments in East Asia in the post-World War II period.

Some fortuitous circumstances must also be considered. Japan, South Korea and Taiwan all implemented relatively virtuous American-sponsored land reforms soon after the end of World War II. There was also significant redistribution of other non-land assets in Japan, most notably, of the pre-war and war-time *zaibatsu* industrial conglomerates. Much of the motivation for such redistributive reforms was, of course, anti-communist, that is, to undermine and minimize support for the communists by those desiring asset redistribution.

In contrast to South–East Asia, more egalitarian asset redistribution in Japan, South Korea, Taiwan and China have also been important. The Americans were not uninfluenced by the left, partly because of the nature of the wartime anti-Axis alliance and the nature of the most influential scholarship available. During the post-World War II American occupation of Japan, it was widely presumed that the *zaibatsu* ‘military industrial complex’ had been responsible for the militarization of pre-war Japan. The American occupation forcibly broke up family control of the *zaibatsu* to create a unique, corporatist ‘stakeholder’ economy. Assets were sold to employees, and to local communities, thus developing worker and community stakes in the companies.

Hence, the peculiarly Japanese economy was created by deliberately redistributive policies with unique outcomes. The ‘human relations’ school of industrial relations influenced the post-World War II development of guaranteed life-long employment and the seniority wage system, which have effectively strengthened a strong employee commitment to the fate of their firm.

From miracle to debacle

Although East Asian economic performance before 1998 was debated, the East Asian debacle of 1997–98 was not anticipated, partly because it was

not principally due to a failure of the real economy despite various recognized economic weaknesses (Jomo et al., 1997). The financial crisis from mid-1997 was precipitated by an eventually successful currency attack on the Thai baht, overvalued after the strengthening of the US dollar after mid-1995. The crisis was greatly exacerbated by herd-like panicky withdrawals from the entire East Asian region, inducing currency and stock market collapses (Jomo, 1998).² Those who control financial assets usually enjoy disproportionate policy influence in most contemporary economies, especially in 'emerging markets'. The greater role of foreign capital in South-East Asia subordinated domestic industrial capital in the region, allowing finance capital, both domestic and foreign, to become more influential in the region, thus rendering it more economically vulnerable (Jomo, 1998).

Finance capital in the region had developed complex symbiotic relations with politically influential rentiers, dubbed 'cronies' in the aftermath of the 1997–98 crisis. Although threatened by the full implications of international financial liberalization, East Asian financial interests were quick to identify and secure new possibilities for capturing rents from arbitrage as well as other opportunities offered by gradual international financial integration. Thus, foreign dominance of South-East Asian industrialization facilitated the ascendance of financial interests and politically influential rentiers.

Such increasingly influential alliances were primarily responsible for promoting financial liberalization in the region, both externally and internally. However, insofar as the interests of domestic financial capital did not entirely coincide with international finance capital, international financial liberalization was necessarily partial. The processes were necessarily uneven, reflecting the variety of interests involved and their varying strengths in different parts of the region. Such flows were also desired to finance current account deficits in both countries, principally due to service account deficits (mainly for imported financial services as well as investment income payments abroad) and growing imports for consumption and output of non-tradeables, mainly in the property (real estate) sector. There is little evidence that such capital inflows contributed significantly to accelerating economic growth, especially of the tradeable sectors. Instead, they often contributed to asset price bubbles, consumption binges and 'over-investments'.

Challenges

There are important lessons to be drawn from East Asia, but clearly, there is no single model as such. For other reasons as well, it does not make much sense to emulate any particular economy in East Asia. Most other

developing countries would find it impossible to do so even if they wanted to. Nevertheless, some important lessons can be drawn from the East Asian experiences. Such lessons are best drawn from careful analysis rather than broad-brushed generalizations about a rather diverse region.

Economic liberalization, including globalization, since the 1980s has fundamentally changed the environment and conditions for selective industrial policy and, hence, for aspiring developmental states. Most importantly, economic liberalization – at both national and international levels – has seriously constrained the scope for government policy interventions, especially selective industrial promotion efforts. This is true of both the international and domestic policy environments, where policy conditionalities and World Trade Organization (WTO) membership obligations have radically reduced the scope for national economic policy initiatives.

The mid-1980s onwards has seen widespread, sweeping and rapid opening up of trade, investment, finance and other flows. Very often, such liberalization has been externally imposed by the Bretton Woods institutions as part of conditions imposed to secure access to emergency credit during the debt crisis of the 1980s, and more recently, in the wake of the financial crises since the mid-1990s. Various policy packages for (price) stabilization in the short term or for structural adjustment in the medium term have involved such conditionalities. The new intellectual and policy environment from the 1980s – the so-called ‘Washington Consensus’ – has promoted such policy reforms.

Such policy changes as well as limited government capabilities have meant little preparation in terms of a proactive strategy to anticipate and cope with the new international competition. Few industrial policy instruments of the past are viable or feasible options today, including many tools used successfully in post-World War II East Asia. Many, if not most of, the main industrial policy tools still available are already intensively used by most advanced industrial economies.

These policies are probably necessary, but certainly not sufficient for stimulating and sustaining economic growth and structural change for developing countries’ ‘catch-up’. Special policies are urgently needed to prevent such economies – already at a historical disadvantage in various respects – from falling further behind, if not to begin to close the gap with the industrially more developed economies of the North as well as the industrial economies that have emerged in recent decades, that is, during the last third of the twentieth century.

As a region, East Asia has led other developing-country regions in terms of economic performance, growth rate, increased exports as well as technological progress. Lall (2003) notes the great divergence between East

Asian countries with and without selective industrial policy, and finds the latter (mainly in South–East Asia) far more economically vulnerable.

Industrial development in the new circumstances clearly requires international competitiveness, and such competitiveness is increasingly defined in manufacturing, related services and institutions, and not simply in terms of wage costs or exchange rate competitiveness, as important as these may be. Inability to compete effectively implies being bypassed, and likely, stagnation at the lower end of the technological and income ladder.

Appropriate industrial policy will require selective interventions as well as effective coordination among firms, clusters and factor markets, which should presumably be consistent with a clear and coherent ‘vision’ of the future as well as the ‘road-map’ towards policy goals (Lall, 2003). For this purpose, there are still many useful lessons to be drawn from the varied experiences of the more successful East Asian NIEs and China, as well as the more modest and flawed achievements of the South–East Asian NICs.

Notes

1. This chapter is based on considerable earlier work. I am grateful to all who have provided me with critical feedback on this work. Needless to say, however, no one else bears responsibility for this version.
2. After the East Asian crisis, even the IMF seemed to back off from its previous advocacy of financial market liberalization.

References

- Amsden, Alice (1989), *Asia's Next Giant*, New York: Oxford University Press.
- Johnson, Chalmers (1982), *MITI and the Japanese Miracle*, Stanford, CA: Stanford University Press.
- Jomo K.S. (ed.) (1998), *Tigers in Trouble: Financial Governance, Liberalization and Crises in East Asia*, London: Zed Books.
- Jomo K.S. with Chen Yun Chung, Brian C. Folk, Irfan ul-Haque, Pasuk Phongpaichit, Batara Simatupang and Mayuri Tateishi (1997), *Southeast Asia's Misunderstood Miracle: Industrial Policy and Economic Development in Thailand, Malaysia and Indonesia*. Boulder, CO: Westview.
- Lall, Sanjaya (2003), ‘Reinventing Industrial Strategy: The Role of Government Policy in Building Industrial Competitiveness’, paper presented to the Intergovernmental Group on Monetary Affairs and Development (G-24), Geneva, September.
- Little, Ian, Tibor Scitovsky and Maurice Scott (1970), *Industry and Trade in Some Developing Countries*, New York: Oxford University Press.
- Wade, Robert (1990), *Governing the Market*, Princeton, NJ: Princeton University Press.
- White, Gordon (1988), *Developmental States in East Asia*, London: Macmillan.
- World Bank (1993), *The East Asian Miracle: Economic Growth and Public Policy*, New York: Oxford University Press.
- Yoshihara, Kunio (1988), *The Rise of Ersatz Capitalism in Southeast Asia*, Singapore: Oxford University Press.

71 Transition economies: lessons for development

Ruud Knaack and Henk Jager

Introduction

‘Transition’ is a designation that is widely used for the transformation of state socialist countries, characterized by state-controlled economies and political dictatorship, into democratic market economies. Transition started to receive intensive attention when a number of countries freed themselves in 1989 from either particularly the Soviet control (think of Poland, Hungary, Czechoslovakia, East Germany and the Baltic states) or mainly domestic tyranny (such as Romania and Albania). Transition was extended in 1992 to the former Soviet Union (FSU) states, also dubbed the Commonwealth of Independent States (CIS), after the disintegration of the Soviet Union in December 1991.

We also will consider countries such as China and Vietnam as transition economies, though there is a reason to exclude them. In China and Vietnam political power remained in the hands of the leaders of the communist parties. In contrast, the member countries of the FSU and the Central and Eastern European (CEE) countries have undergone a dramatically rapid democratization process – with some exceptions, such as Belarus, Turkmenistan and Uzbekistan. If a high speed of political turmoil were a determining feature of transition, countries such as China and Vietnam would stay out. The differences between the outcomes in, on the one hand, the countries in CEE and of the FSU and, on the other hand, China are striking. From the countries of the first group, only three countries were able to recover from the ‘transition depression’ of gross domestic product (GDP) after ten years. However, China did not experience a decline in GDP in any year from 1978 onwards – China’s transition period – while it had a high and stable annual GDP growth figure in the order of 8 percent on average. It is, of course, quite interesting and informative to analyze the reasons for such economic divergences. For that goal, we will not employ a rapid political transformation as a distinctive feature of transition, so that we will focus on economic transition.

At the start of transition in Europe, around 1990, knowledge about the preferred materialization of the transition process was partly borrowed from the experience obtained from the process of economic growth in

developing countries. Now, more than 15 years later, there is ample information about the transition process to consider the reverse question: What are the lessons of this transition from practice for developing countries that are eager to give additional spurs to growth? The search for these lessons is the aim of this chapter. The chapter consists of three parts. It starts with a short history of political developments, mainly in Russia, the motor of state socialism after World War II. The next part describes the stylized facts of the transition process in the countries in CEE and of the FSU, and China as well. The following part discusses the lessons for developing countries that can be drawn from the transition process.

A short political history

After the October Revolution of 1917, the Bolsheviks found all political power in an internationally isolated and underdeveloped country rested in their hands. They considered it their main task to industrialize the country at full speed, using internally raised investment funds. It was for this purpose that the Soviet planning system was created (Knaack, 1996).

The Soviet Union experienced a 'golden age' in the 1950s. The country grew rapidly, propelled by increases of capital, labor and raw materials. Economic growth directly benefited consumers as their diet and housing improved apace (Schroeder, 1992). Space flights and Nobel prizes symbolized the achievements of Soviet science.

However, from the 1960s onwards, the Soviet economy settled on a slower growth path. According to Ellman and Kontorovich (1992, pp. 10–12), there are three explanations for this slowdown of economic growth. First, a loss of control of the economy. If an economy becomes more complex, coordination from above becomes increasingly difficult. Second, a reduction in the growth rates of both the volume and the productivity of production factors. One could not, for example, increase the participation rate of women any more, and also stocks of natural resources were running out. Third, a weakening of the 'entrepreneurial spirit'. In a command economy, pressures from above provide the main source of dynamics in an economy. Brezhnev's policy of 'stability of the cadres' represented a codification of the process of slackening the pressures from above.

Declining growth rates alone cannot explain the collapse of the Soviet Union at the end of the 1980s. According to Kornai (1992), the command system was still able to guarantee the population a decent way of life. What caused the crisis was its weak economic performance relative to that of the USA and other Organisation for Economic Co-operation and Development (OECD) countries. The dynamic efficiency argument was the *raison d'être* of the Soviet Union. Consequently, the relatively poor growth figures of the 1980s threatened the political legitimacy of the whole system.

Economic reform under Gorbachev must be understood as part of his effort to revive and modernize the Soviet economy. The results of the reform process were disappointing, mainly because Gorbachev's policies were not feasible (Hewitt et al., 1987). But his policies also had unintended consequences. This happened especially for his glasnost policy intended to unmask bureaucrats sabotaging the reform process.¹

On 1 November 1989, the Berlin Wall fell. From that moment on in CEE one country after another unlinked itself from the Soviet dominance and started a process of conversion from a centrally governed economy to a market economy. On 30 June 1990, the two parts of Germany were officially reunited and the German economic, monetary and social union was created. In Czechoslovakia, the Velvet Revolution ended the ruling of the Communist Party in November 1989. The new government of Prime Minister Klaus introduced a series of measures as from 1 January 1991, aimed at the integration of the Czechoslovakian economy into the world economy. The Soviet Union was also contaminated with this spirit of the time. After the breakdown of its economic system the new rulers strove for a rapid transition towards a new system characterized by market relations, private ownership, and a liberal democracy. After the abortive *coup d'état* of August 1991, the Soviet Union disintegrated and Yeltsin became President of the new Republic of Russia. On 2 January 1992, the Gaidar administration introduced a number of market reforms, which inflicted an enormous shock to the Russian economy.

Stylized facts of the transition process

Central and Eastern Europe and the former Soviet Union

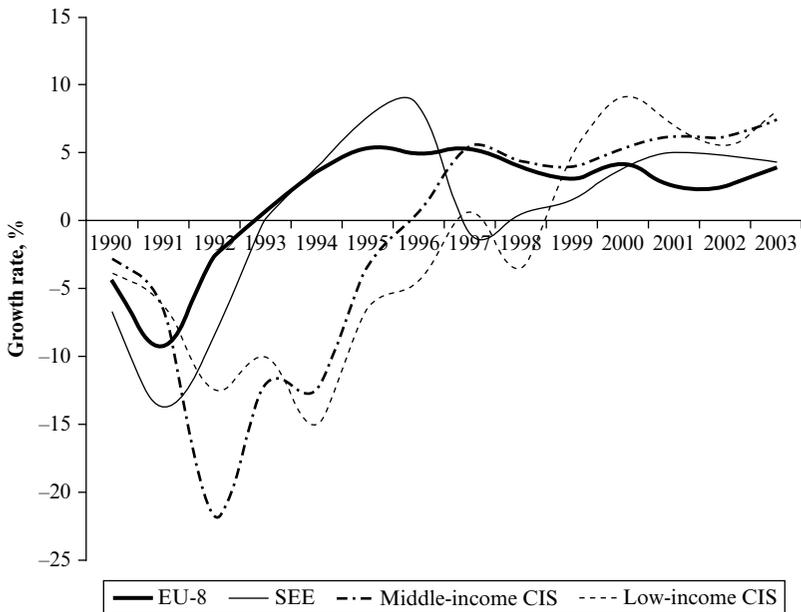
The processes of change in the countries of CEE were not based on a blueprint showing how a formerly communist country ought to be restructured as a capitalist country. According to the prevailing opinion, that was not necessary. Due to the German unification, East Germany took the West German legislation and rules over at one stroke and Eastern German enterprises were privatized at a quick pace. Other countries had the possibility to copy a large part of the existing and tried and tested recipe as applied earlier by, for example, Spain at its accession to the European Union and developing countries at their integration into the world economy. It was the recipe developed by the International Monetary Fund (IMF) and the World Bank (Taylor, 1993) and is often dubbed the Washington Consensus.

The Washington Consensus comprises the 'Holy Trinity' of stabilization, liberalization and privatization (Rutland, 1999). This Consensus initially was advice for a reform agenda for the Latin American countries at the end of the 1980s to adapt their policies and institutions.² Applied to the former

communist countries at the time, this trinity would read as follows. 'Stabilization' refers to the need to reduce inflation, both open and suppressed, to create a currency with a stable external value, and establish balance-of-payments equilibrium. To achieve these goals, a standard package of fiscal and monetary measures was recommended, together with price liberalization so as to eliminate suppressed inflation. 'Liberalization' means the freeing of enterprises and individuals of the old planning institutions. The initial expectation was that the disappearance of these planning institutions would create space for the new market institutions to arise spontaneously. 'Privatization' was considered crucial for both political and economic reasons. Politically it would create powerful interest groups with a stake in transition and, for that reason, strongly opposed to the old system. In addition, it would secure active support from the West. It was thought, moreover, that only privately owned enterprises could operate successfully in a market economy. The collapse of the communist regimes in CEE after 1989 and the dissolution of the Soviet Union in 1991, spread optimism about the chances of rapid economic growth.

The transition process appears to have several characteristics, both intended and unintended (Campos and Coricelli, 2002; Ellman, 2005; Knaack and Jager, 2007). First, all countries of the former Eastern bloc faced a dramatic fall in output. The individual country patterns of recession and recovery cover the years 1989–97. They have largely been of the L-curve shape (sharp fall, followed by slow recovery), rather than the initially hoped-for J-curve type (small fall, followed by fast growth). Figure 71.1 displays the growth rates of real GDP in the years of transition for the distinct country groups. The strongly negative GDP growth rates for each group in the first years of transition indicates that real GDP levels should exhibit L-curves. The fall of investment expenditures, especially for inventories and housing, was even larger than the fall in real GDP. Defense expenditures on equipment and materials have declined sharply as a proportion of GDP, especially in the countries of the FSU. Consequently, during the period of falling output, private consumption has declined only moderately relative to production in most transition economies.

Second, a large inter-country variation in the adjustment patterns is observable, with large differences in both depth and length of the L-curve. According to Figure 71.1, for the whole group of CEE countries the average growth rate already became positive again in 1993, after a depth of –9 percent for the sub-group EU8 (the first eight CEE countries that joined the European Union, that is, Poland, Hungary, the Czech Republic, Slovakia, Slovenia and the three Baltic States) and –14 percent for the countries in the sub-group SEE (or South-Eastern Europe), both in 1991. The CIS, however, only experienced for the first time since 1989 positive economic growth in



Source: World Bank (2005, p. 17).

Figure 71.1 GDP growth rates for four regions in CEE and the FSU, 1990–2003

1997. The low-income CIS countries had their depth in economic growth in 1992, with a decline in GDP of over 20 percent, and the middle-income CIS countries in 1994 with a negative growth rate of 15 percent. The variation in the cumulative fall of real GDP per country in the beginning of the transition process has been large. It ranges from moderate for Poland (6 percent, in two consecutive years of output decline) to high for Russia (40 percent, in seven years of consecutive years of output decline) and extremely high (over 60 percent fall) for Armenia (63 percent, in four consecutive years), Moldova (63 percent, seven years) and Georgia (78 percent, five years).

Third, like output, the level of foreign trade in transition economies has followed a pattern of decline and recovery. In the countries of Eastern Europe, foreign trade declined by 62 percent in the period 1990–93 and then rose to 71 percent of 1990's export level in 1998. As for trade redirection, the share of the Western countries in the export of the CEE countries has increased tremendously. It nearly doubled from 35.8 percent in 1992–93 to 67.5 percent in 1998–99 (World Bank, 2002, p. 7). For the CIS countries this share did not change at all: it remained at 28 percent. On the other hand, the

former CMEA trade collapsed.³ Appropriate exchange rate changes to promote competitiveness proved to be a useful help in redirecting and promoting foreign trade. For example, the Czech Republic was successful in redirecting its trade to the West by devaluating its currency by 50 percent. In contrast, East Germany could not benefit from devaluation at all. Instead, it suffered from a managed real revaluation. Given the one-to-one substitution of the West mark for the East mark, in practice the competitiveness of East Germany badly worsened, creating mass unemployment.

Fourth, the sectoral composition of GDP in current prices changed during transition in favor of services and away from manufacturing. Largely this was due to the relatively larger contraction of manufacturing during recessions. However, productivity gains in the manufacturing sector and, therefore, declining relative prices of manufacturing products during economic recoveries also contributed to this sectoral change.

Fifth, during the transition the participation rates in the labor markets changed rapidly. Under communism, the participation rates were high, in particular for women. During the transition there was a significant increase of flows out of the labor force, especially for women and older persons near to retirement age. At the same time, officially registered unemployment rates increased much faster in Eastern Europe relative to the FSU countries. Labor also moved in great numbers from the state sector to the private sector. In contrast, geographical labor mobility remained very low.

Sixth, the production fall in the formal sector led to sharp increases of the official unemployment rates. Across the transition countries, the employment rates differed widely. The outcome for a specific country strongly depended on whether or not the loss of employment in the old enterprises was compensated by the creation of jobs in new (*de novo*) enterprises. In transition countries where restoring sustained growth has proved relatively illusive, new enterprises account for a low share of employment and value added – both between 10 and 20 percent (World Bank, 2002, p. xxv). According to the World Bank, in practice the transition process proves only to get momentum when the share of medium- and small-scale enterprises in the national employment is more than 40 percent. The CEE countries reached this percentage in 1996. In this respect, the countries of the FSU stayed behind. In these countries the growth of small-scale enterprises stagnated completely; here the share of small enterprises did not rise above 20 percent. Both the government and the Mafia seem to be responsible for that, as they viewed the new enterprises as cash cows instead of centers of new economic initiatives that have to be stimulated.

Seventh, a major result of the transition has been the emergence of a large informal sector. Partly, this concerns new enterprises, which wish to escape the attention of bureaucrats and tax authorities. Partly, it concerns

enterprises that already existed under the old system. These enterprises undertake substantial volumes of activity 'off the books' and make extensive use of barter trade. In addition, there has been a widespread informalization of the labor market. This takes the form of, for instance, unilateral determination of wages and conditions by the management of the firms, regardless of laws about these matters. Other examples are employers that disregard paying wages on time, and employees that do not receive maternity leave, are not protected from dangerous working conditions and are not represented by trade unions.

Eighth, a dramatic result of the transition has been a huge increase in relative and absolute poverty, as well as income inequality. According to the World Bank (2002, p. 8), between 1990 and 1998 the population living on less than US\$1 a day in the CEE and CIS regions increased from 1.5 to 5.1 percent. It was the only area in the world for which this share increased in that period (in the poorest area worldwide, that is, sub-Saharan Africa, the share remained stable at 47 percent). The norm of US\$1 is in effect not appropriate for the transition region as the costs of living are higher there than in other regions. Think, for example, of the cost of heating. Adapting the poverty line for region-specific costs, the World Bank estimates that between 1987–88 and 1993–95, the share of the poor in the total population in the FSU and CEE rose from 3 percent to about 25 percent, and in number of persons from 7 million to 89 million. This means that the transition process pushed more than one-fifth of the population below the poverty line. A group of the population that was particularly badly affected was children. Pensioners also suffered heavily. As a survival strategy, the population in all countries involved put much time and energy into cultivating food in the many allotments. For example, in 1996 this line of food production was estimated to yield 43 percent of Russia's total food production. The deterioration of the living conditions led also to demographic changes. There has been a striking increase in mortality, concentrated among adult men in Russia and Ukraine. This increased the gender gap in life expectancy (in Russia women have a life expectancy that is 13 years longer than men). There was also a dramatic decline in birth rate and an increase in emigration. The combined effects of the current trends in mortality, birth rate and migration are that the population in many countries in CEE and the FSU declines.

Poverty increased not just because of the fall in output, but also due to greater income inequality in all European and FSU transition countries. Table 71.1 contains these developments over time in the years 1987–98, expressed through the Gini coefficients. Without any exception, the countries included in the table had a higher Gini coefficient, and thus a more unequal income distribution, at the end of that period. The CIS exhibited,

Table 71.1 *Income inequality during transition*

Countries	Gini coefficient of income per capita		
	1987–90	1993–94	1996–98
<i>CSB</i>	0.23	0.29	0.33
Bulgaria	0.23	0.38	0.41
Croatia	0.36	–	0.35
Czech Republic	0.19	0.23	0.25
Estonia	0.24	0.35	0.37
Hungary	0.21	0.23	0.25
Latvia	0.24	0.31	0.32
Lithuania	0.23	0.37	0.34
Poland	0.28	0.28	0.33
Romania	0.23	0.29	0.30
Slovenia	0.22	0.25	0.30
<i>CIS</i> ^a	0.28	0.36	0.46
Armenia	0.27	–	0.61
Belarus	0.23	0.28	0.26
Georgia	0.29	–	0.43
Kazakhstan	0.30	0.33	0.35
Kyrgyz Republic	0.31	0.55	0.47
Moldova	0.27	–	0.42
Russian Federation	0.26	0.48	0.47
Tajikistan	0.28	–	0.47
Turkmenistan	0.28	0.36	0.45
Ukraine	0.24	–	0.47

Notes:

— Not available.

a. Median of countries with data.

CSB is the acronym of Central and South-East European countries and the Baltic States.

Source: World Bank, 2002, p. 9.

on average, a much larger increase than the CSB (Central and South-East European countries and the Baltic States). The Gini coefficient of Hungary hardly increased, whereas Armenia showed the largest increase. The picture of changes in the Gini coefficient in the years after 1998 (until 2003) is divergent: for example, Poland and Romania experienced further increases in the coefficient, for Hungary there is no change, while Russia and Armenia exhibit substantial declines in the coefficient (see, World Bank, 2005, p. 15).

Ninth, growth of crime and the widespread criminalization of society has been a significant feature of transition. This has been particularly

marked in the FSU. In some countries, kleptocracy has been an important part of the political system. In these countries, there are also close links between the criminal, political and business worlds. This is connected with the inability of the state to perform even its night watchman functions. Hence, the supply of property protection and rule enforcement are privatized, that is, taken over by criminal organizations.

Tenth, during transition the old institutions collapsed, creating an institutional vacuum (Schmieding, 1993; Knaack, 1996). As the transition experience has advanced, it has become clearer that the disparity between the rapid progress in liberalization and privatization and the slow development of institutions that support markets and private enterprises directly and negatively affects overall economic performance.

China

The reform processes in the countries of Eastern Europe and the former Soviet Union can be better evaluated when we compare them with the reform process in China. It is interesting to compare, for example, Russia and China, because there are so many similarities between them as economies in transition and yet the outcomes of the transition process were so different. As said before, the fall in GDP in Russia in the period 1991–94 was more than 40 percent, while in the period 1978–2002 the annual growth rate in China was 9.4 percent. During this period in China, the per capita income of rural and urban population, measured at constant prices, increased 5.3 and 4.7 times, respectively (Lin, 2004). How can we explain these differences?

There are many similarities between Russia and China as economies in transition (Buck et al., 2000). These include the enormous geographical scale, abundance of cheap labor, and large potential markets to attract foreigners. In addition, both countries inherited similar economic and political ideologies from their Stalinist and Maoist periods, with a common emphasis on Party control, high defense budgets, large industrial enterprises, heavy bureaucratic and tariff protection against manufactured imports, and subsidized public services.

Of course, there were important differences too. The main difference was the economic structure. On the eve of the start of the transition process (in Russia in 1991 and in China in 1978) China was largely an underdeveloped and rural country with 71 percent of the workforce active in agriculture, despite heavy industrialization efforts in the 1950s and 1960s. Russia was an industrialized country with only 13 percent of the population working in agriculture. China's financial system was also relatively underdeveloped.

To some extent, these differences in initial conditions and structural characteristics can explain the divergence of transition policies. Since

agriculture was the biggest sector, accounting for 37 percent of output and 71 percent of employment, agriculture was the biggest Chinese economic problem in 1978. It explains why the transition process started in agriculture and gradually spread to industry. When China started its transition in 1978, initially the government did not question the feasibility of the old system. Its attempt was simply to improve the system by giving agents in collective farms and state enterprises some autonomy so that a closer link between personal rewards and individual efforts could be established. Agricultural prices were raised substantially, mandatory quotas for sown area and output eliminated, and compulsory procurement quotas reduced, with the sale of above-quota output on free markets and increased possibilities for so-called sideline activities. Finally, the commune system gave place to the family responsibility system, accounting for 94 percent of the peasant households in 1984. In response to these reforms, per capita grain production rose by one-sixth between 1978 and 1985, the per capita output of cotton rose by two-thirds, and that of oilseeds, sugar and tobacco doubled (Balassa, 1987, p. 411).

With respect to industry, the Chinese government adopted a dual-track policy, liberalizing the entries to the labor-intensive sectors, but also creating conditions to address the viability issue of those firms in the capital-intensive high-priority sectors. The enterprises that were most able to take advantage of the entry option were collective enterprises sponsored by local governments, particularly in rural areas. These township and village enterprises had started already in the Maoist period, but expanded rapidly after 1978. They already produced 33 percent of total industrial output in 1991. Private firms then produced 11 percent, implying that the share of the state sector reduced to a mere 56 percent of total production (Naughton, 1994, Table 1). These developments in industry had two effects. The intensified competition resulted in the disappearance of monopoly rents and the occurrence of a division of labor between the state and non-state sectors. The state sector increasingly consists of large firms in heavy industry, while the non-state sector consists of medium-sized and small firms in light industry.

The success of the reforms in agriculture and industry had a positive effect on the macroeconomic stability of the economy. Formal rationing was in place for more than 20 consumer goods in 1978, accounting for more than 50 percent of consumer expenditures. Shortages of even the basic commodities were common. Shortages of consumer goods disappeared quickly, even though price controls remained in place in China. These policies had two effects: there was a significant increase in living standards and, therefore, an increase of popular support for the reform measures. Due to high saving rates of the Chinese population, there was no collapse of

investment in China, but a dramatic change in the way these investments were financed. Savings by government and by state enterprises has dropped sharply, while savings by private business and households has increased sharply. By the late 1980s, households were saving 15 percent of their income, compared with 2 percent before 1978.

The dynamics of the non-state enterprises exerted a heavy pressure on the state enterprises and triggered a state policy of managerial reforms. These reforms had some success. For example, during the 1980s the output of the state enterprises grew 7.7 percent annually – though other ownership forms of production exhibited even more rapid growth rates. Total factor productivity also grew in the state sector, although less than in the non-state sector. As a result, the state sector is still less competitive than the non-state sector. The relatively low export levels in the state sector compared with the export levels of comparable firms in the West are a clue to that (Buck et al., 2000, p. 393). The continued government subsidization and protectionism of the state enterprises through low-interest loans and monopolistic practices is both cause and consequence of this disappointing competitiveness. The government hopes to eliminate these distortions in the near future.

China also put into effect a policy of opening up to the outside world, but the scope of the opening to the outside world was only expanded gradually. The first step was to set up special economic zones in the south near Hong Kong and Macao, as laboratories for market reforms. Enterprises in these zones had the right to retain most of the foreign exchange they earned and were more flexible in firing workers. Furthermore, foreign investors in these zones could repatriate profits and own land. After the successful performance of these zones, a number of coastal cities were opened, followed by economic areas along some rivers, such as the Yangtze River, and along the borders, and finally capital cities of the various provinces were turned into open cities. In the first years, it was especially the ‘overseas Chinese’, including Taiwanese, who made use of the new possibilities. Later China could welcome a huge influx of foreign direct investment (FDI), also compared, for example, with Russia. In the period 1989–95, Russia attracted FDI to the amount of \$3.9 billion, or \$1.1 per capita, in 1995. In the same period, China received \$121.7 billion of FDI, equivalent to \$18.2 per capita (Buck et al., 2000, p. 384). Virtually all FDI was in the form of joint ventures. Joint ventures were made possible, although the government limited foreign ownership and control of businesses. In the late 1970s, foreign involvement was limited to 35 percent of any venture. This was raised to 49 percent in 1985, while from 1988 onwards majority foreign ownership is allowed, though still subject to state approval. The joint ventures were crucial for the success of the Chinese reform process, because the foreign

companies delivered the technological knowledge for the Chinese partners to produce goods that could compete on the world market.

An important difference between China and the CEE and the FSU is that in China the Communist Party kept control, while in the other transition countries the power of the Communist Party dissolved in favor of a democratic political system. The main argument of the Chinese leadership is that the continued rule of the Communist Party guaranteed social stability, which is conducive for the success of the economic reform. Leaders occupied with competing for power would create uncertainty about whether or not the economic reforms would be continued. This is not to say that there were no political reforms in China. Actually, economic reform often contains some elements of political reform. In the case of China, there was devolution of power from the center to the provinces. This empowerment of the regions has created what is now the major driving force behind economic reform (Woo, 1994, p. 289).

Lessons to be learned

For most economists and politicians the depth and duration of the depression which accompanied the transition process in CEE and the FSU came as a surprise. The initial idea was that the removal of the overwhelming apparatus of political control over economic activity could only imply additional prosperity in the medium term. The previous system was characterized by a myriad of distortions, and the removal of most of them would lead to a vigorous impulse to output. This optimism was not a monopoly of neoclassical economists. Well-known heterodox economists, like Janos Kornai, also held this view (Campos and Coricelli, 2002). This raises the question of how to explain this anomaly.

Economic stabilization

As mentioned before, in order to suppress open and hidden inflation, most transition countries adopted a standard package of restrictive fiscal and monetary policy (sometimes supported by exchange rate and income policy). Experience has shown that, despite widespread initial skepticism, such packages – if persisted in – are successful in reducing macroeconomic instability (Ellman, 2005). However, the time for the package to work may be far longer than anticipated. For example, Poland implemented the stabilization package at the beginning of 1990, but only nine years later inflation fell below 10 percent. Moreover, this restrictive macroeconomic policy may contribute to a sharp decline in output and welfare, as happened in Russia after the unsuccessful attempt at shock therapy by Prime Minister Gaidar. As a result, Kornai (1994) argued – writing on the Hungarian situation where inflation was about 20 percent – that growth must be the main

economic objective when the danger of accelerating inflation does not exist anymore. Stiglitz (1998) strengthens this view, summarizing the evidence that only high – and not moderate – inflation is costly: When countries cross the threshold of 40 percent annual inflation in an upward direction, they fall into a high-inflation, low-growth trap. Below that level, there is little evidence that inflation is costly. Recent research even suggests that low levels of inflation may improve economic performance relative to what it would have been with zero inflation.

Optimal sequencing

Economic stabilization is a prominent part of the discussion on the optimal order of reforms. The transition process of the 1990s has intensified and widened this discussion, which previously was concentrated on the reform process in developing countries. It concerns reform on three levels of aggregation. On the highest level of aggregation, it is about the optimal order over time of economic liberalization, economic stabilization, privatization and the creation of supporting institutions, necessary for a smooth working of the markets. On a lower level of aggregation, the optimal-sequencing discussion focuses the order over time of the different parts of economic liberalization, in combination with economic stabilization. Here one distinguishes the liberalization of domestic goods and labor markets, international trade, domestic financial markets, international capital flows, and the foreign exchange market. On the lowest level of aggregation, the focal point is the optimal order of domestic sectoral reform: agriculture prior to industrialization, or perhaps the other way around?

A very useful empirical analysis of the transition lessons of the optimal order of the highest aggregation level is Beyer (2001). His data set consists for each of 14 CEE countries of the months in which they switched to the new regime, the majority of their prices were liberalized, their most substantial attempt for stabilization was undertaken, and a new or adjusted constitution was adopted (as an indicator of a country's new institutional system). By using groups of countries with a similar order of reform it appears that the sequence over time of stabilization, privatization and liberalization is significantly the best sequence in terms of GDP level six years after the system switch. Beyer dubs this sequence the *graduality approach*. Slovenia and Hungary have followed this adjustment path. These two countries started the reform process with constitution-building. If instead liberalization took place together with stabilization at the start of the reform process, Beyer considers the adjustment process would be a *'big bang'* approach. The countries that belong to this group (Albania, Bulgaria, the Czech Republic, Poland and Slovakia) display a worse development of GDP relative to the gradualists, but a significantly better outcome than the

transition countries that liberalize and/or privatize before stabilization. Early privatization appears to work badly, as Belarus, Romania, Russia and the Ukraine witness. However, Estonia and Lithuania are counter-examples.

The previous paragraph gives the important clue for optimal sequencing at the lower level of aggregation: that stabilization should in any case not come after liberalization. This gives a strong support to the earlier literature on optimal sequencing for developing countries, which concluded that stabilization should be carried out first, followed by liberalization. The standard outcome of that literature with respect to the optimal sequence within liberalization is: domestic goods and labor markets, international trade simultaneously with unifying the exchange rate and realizing the equilibrium value of that rate, domestic financial markets, and finally, freeing international capital flows from border restrictions.

The transition gives some support to this optimal order, though only of an anecdotal character. The anecdotes that follow have to do with the space for the exchange rate to find its equilibrium value in time. In former East Germany, the unification with West Germany implied a 'big bang' liberalization and the introduction of a unified, but highly overvalued currency from former East Germany's viewpoint. The politically motivated choice of a one-to-one conversion of the East mark into the West mark resulted in a huge deterioration of competitiveness of former East Germany and a concomitant disaster with respect to its GDP, creating mass unemployment. This outcome gives support to the earlier-mentioned optimal liberalization order, which requires that the exchange rate is able to reach its equilibrium value in the process of trade opening.

The collapse of the CMEA trade soon after the start of transition and the resulting loss of jobs in the big state enterprises had to be counterbalanced by an increase of exports to mainly the West and the creation of new jobs by small and medium-sized enterprises. The Czech Republic, for example, was successful in both respects. The large devaluation of its currency resulted in a strong swing of its foreign trade to the West and the process of 'small' privatization contributed to strong employment growth in the private sector. Obviously, the Czech Republic also profited from its geographical position and the possibilities of the tourist industry, especially in Prague.

Like East Germany, Russia has opted for, ultimately, a less successful road. As with the Czech Republic, it devaluated its currency in 1992, but made the mistake to choose a more or less fixed exchange rate of the rouble against the dollar under conditions of high internal inflation. The cause was that Russia had already liberalized international capital movements before the economy was sufficiently stabilized. The potential instability was the lax policy stance on fiscal deficits. Large capital inflows initially allowed

the government to finance a continuing fiscal deficit at relatively low interest rates. This can be considered the so-called good equilibrium (see Gros and Steinherr, 2004, p. 243). However, in the meantime, for Russia the debt-to-GDP ratio increased and so did the country's debt service burden. This development gradually undermined the country's credibility. Reinforcing simultaneous developments were a growing overvaluation of the rouble, in response to the large capital inflow, and inflationary pressure, due to capital inflow as long as the central bank stabilized the exchange rate. Both reinforcing developments usually worsen the country's current account over time – in a gradual, but inevitable way. The real appreciation of the rouble did not lead so much to a deterioration of the trade balance, given the strong export potential of the gas and oil reserves. But this appreciation resulted in the crowding-out of the Russian industrial production in line with 'Dutch disease' features. Industry became more and more expensive and lost its possibilities to export. The industrial loss of sales became still more severe because the real rouble appreciation stimulated the Russian consumers to opt for cheaper foreign consumer goods. In this state the country was ripe for a shift in expectations leading to the so-called bad equilibrium (see Gros and Steinherr, 2004, p. 243). Given the openness of the capital account, Russia was exposed to sudden withdrawals of foreign capital. This fear became reality during the summer of 1998, after which the rouble collapsed and a severe economic crisis occurred.

With respect to the third level of aggregation and the concomitant optimal sequencing of sectoral reform, the different reform roads of China and Russia are informative. From the success of the economic reform process in China, some economists concluded that the Chinese road of agriculture first was also applicable to Russia. By starting the reform in industry, Russia was unable to gain the credibility that probably would have come if it had started the reform process in agriculture. This position is debatable. The situation in China and Russia was quite different. In China, the agricultural sector was the biggest sector, accounting for 37 percent of the output and 71 percent of employment. In Russia, the agricultural sector employed only 13 percent of the labor force and generated 18 percent of gross national product (GNP). Moreover, relative to China the Russian agriculture was much more mechanized. It is much easier to assign property rights to the individual plots that farmers have been working on than to assign property rights to the capital equipment that workers have been using jointly.

However, the argument gains weight when we include in agriculture the activities of the townships and village enterprises. In the 1990s they already produced 33 percent of industrial output. Including the private sector, the enormous dynamism of the non-state sector had a positive effect on the

supply of consumer goods and the living standards of the population, and therefore created popular support for the reform measures. A rapid growth of the non-state sector can also absorb the unemployed in the state sector. This happened also in the Czech Republic. In the period 1989–1995, about 25 percent of the workers in the state sector left that sector. They could easily find new jobs in the new private sector, especially in the new service sector. Especially in the Prague area, the new private sector showed amazing growth. A recent report of the World Bank (2002) confirms the importance of the stimulation of the non-state sector in an early phase of the reform process. According to this report, a key for economic growth in transition countries is the shift from capital-intensive to labour-intensive enterprises. The last group consists overwhelmingly in small enterprises (with a maximum of 50 workers). According to the World Bank, the transition gets momentum when the share of medium- and small-scale enterprises in the national employment is more than 40 percent. China fulfils this condition, as well as to a lesser extent some CEE countries.

Compared with China and the Czech Republic, the position of the small and medium-sized enterprises in Russia is delicate. In the years 1995–97 employment in Russian businesses with a maximum of 50 employees fell officially by 50 percent. Disappointing productivity growth cannot be an explanation. On the contrary, these firms often had good performance (Commander et al., 1996, Chapter 8). The true explanation is the ‘grasping hand’ of the Russian government and the Mafia. They compelled small enterprises to pay highly unpredictable taxes and regular payments to their ‘protectors’. This explains why the transparency of government behavior and a reduction of risks in the business environment are important determinants of the success of the reform process.

Institutions

After the fall of the Berlin Wall in 1989, in principle all CEE countries followed a liberalization process directed at the breaking down of the planning systems. It was expected that markets would arise spontaneously as soon as the old planning bureaucrats disappeared. In other words, the policy-makers expected that a process of ‘organic growth’ would create the political and economic institutions necessary for the smooth functioning of a market economy. Obviously they trusted that the fundamental propensities of human nature to ‘truck, barter and exchange one thing for another’, as postulated by Adam Smith, were not foregone during the decades of communist rule (Knaack, 1999, p. 357).

However, they did not take into account the fact that proper functioning markets require an institutional infrastructure and that it takes a lot of time before the new institutional system and the persons who have to work in

those markets are adapted to the new circumstances. A clearly delineated system of property rights; a regulatory apparatus curbing the worst forms of fraud, anti-competitive behavior and moral hazard; a moderately cohesive society exhibiting trust and social cooperation, the rule of law and clean government – these are the social arrangements that economists usually take for granted, but which were absent in the transition economies (Rodrik, 2000). Not only must new institutions be created, but they must also prove their value during a time-consuming process of trial and error. Each economic transition process is fundamentally an incremental process, during which the country constantly experiments with new forms and finally keeps that form which is satisfactory. In this way, the existing institutional structure actually improves.

The abolition of the old planning system in one stroke without the construction of new institutions of a market economy has irrevocably led to an institutional vacuum. That vacuum has many forms (Knaack, 1999, p. 363). The old rules lose their value, but the enterprises have not yet learned how to behave in the new situation. Further, the information structure of the old system disappeared, while the new market signals were not yet fully developed. For the enterprises it was difficult to find new customers, and when they finally succeeded it was difficult to assess their creditworthiness. As a result, the enterprises operate in an environment characterized by an extreme uncertainty.

It is obvious that the institutional vacuum must be filled. In the CEE countries and the countries of the FSU this happened in different ways. In the case of East Germany, the country took over in one stroke the institutions of West Germany. The Czech Republic profited heavily from the neighborhood of the large markets of Germany and Austria, and the country also learned quickly from the international trade relations. Moreover, from 1995 onwards the Czech Republic gradually adopted the *acquis communautaire*, the legal structure of the European Union. It must be stressed that imported blueprints do not do their work instantaneously. The main reason is that blueprints are highly incomplete. Much of the knowledge to operate with the blueprints has not been written down and has to be learned. However, blueprints give a direction for knowledge acquirement.

Russia did not have these advantages. For a big country, it has a surprisingly huge international trade. However, nearly all export is energy and raw materials. Actually, these are the features of a small and open developing country. With regard to the possibility of the import of institutions it only had to fulfil the requirements of the IMF when it borrowed some money. Consequently, much more than the other small CEE countries, Russia had to fill the institutional vacuum on its own terms. Given the fact that the

creation of new institutions is a time-consuming process, one can understand that it fell back on its old routines and that, given the weakness of the state, organizations such as the Mafia also filled the vacuum.

Compared with the CEE and FSU countries, the problem of the institutional vacuum in China was less acute. First, we have to remember that after the reforms of 1978 the overwhelming majority of the economic relations in China was still shaped and guided by the official planning system. Only in a very gradual way was there a shift from the planning system to more market relations. Consequently, firms and individuals had enough time to adapt to the new circumstances and to learn the rules of a market economy. Second, insofar as the enterprises in the economic zones had to obey immediately the rules of the market, they could profit from the experiences, knowledge and economic networks of the so-called 'overseas Chinese', businessmen especially from Hong Kong, who invested heavily in the zones. Third, sales in the big cities of agricultural surpluses and light industry products of the village enterprises need relatively little organization and structure. Permission from the local authorities to sell on a street corner or a square is sometimes enough. This also explains why, for example, in the Czech Republic it took so little time to start small enterprises in the service sector.

Political reform

There is an intensive debate about the relationship between political regime type and economic performance. Based on the experience in a handful of economies in East and South-East Asia, which (until recently at least) registered the world's highest growth rates, under authoritarian regimes, one could conclude that economic development requires a strong hand from above. To embark on self-sustained growth, deep economic reforms are often needed, which cannot be undertaken in the messy pull and push of fragile democratic politics. The main argument is that economic reform necessarily imposes costs on some segments of society, and that political openness would provide the losers with the opportunity to form coalitions to stop the reform. An example of this occurred in 1992 in Russia when the apparatchik Chernomyrdin replaced the reformist Gaidar. The first new acts were to squeeze the thousands of small shops that had appeared since January 1992 and to extend cheap credits to the industries under the Ministry of Oil and Gas that Chernomyrdin had headed (Woo, 1994, p. 288). A strong and committed leadership can also push economic reforms against the interests of some interest groups. For example, Buck et al. (2000) describe that the Chinese Communist Party stimulated joint ventures with foreign partners against possible dissent of insiders of state enterprises excluded from the deals. The central authorities stimulated

foreign investors to select from the state enterprises the best physical and human assets to form joint enterprises, usually geographically separated from the unreformed parts of the state enterprise left behind. Normally, the incumbent manager and workers repel any outside investor, unless they are prepared to give employment guarantees for all branches of the enterprise, including those involved with social provisions.

This line of thought met a lot of criticism. Rodrik (2000), for example acknowledges that in effect the Asian countries have prospered under authoritarianism, but that many more have seen their economies deteriorate – think of Zaire, Uganda or Haiti. Moreover, some of the most successful economic reforms in South America were implemented under newly elected democratic governments – witness the stabilizations in Bolivia (1985), Argentina (1991) and Brazil (1994). Moreover, the transitions in the democratic European countries were more successful than the transitions in the authoritarian FSU countries, like Belarus, Turkmenistan and Uzbekistan. It is obvious that we cannot subtract from these examples a clear-cut answer about the relationship of political and economic reform. But it is evident that the reform process is helped when a strong government is able and willing to create the necessary market institutions and resist the interest, especially of the insiders in the state enterprises.

Conclusions

After the fall of the Berlin Wall in 1989, one country after another in Central and Eastern Europe freed itself from Soviet domination and started a transition process from a centrally planned economy into a market economy. In this transition process, they followed the recipe from the IMF and the World Bank, developed earlier for developing countries, mainly in Latin America. In all transition countries in Central and Eastern Europe, the results were rather disappointing. In 1999, only three of all these transition countries surpassed the 1989 national income levels. The new countries that belonged to the former Soviet empire underwent an even more severe income fall. The national income reductions were much more profound than initially expected. When we compare these figures with China's experience, the difference is striking. China started its reform process in 1978, and for many years had double-digit positive growth figures without any intermediate fall.

The length and depth of the recession in most countries can be explained by the fact that the reform process was based on an incomplete theory about the functioning of a market economy. The policy-makers recognized too late the precondition for the functioning of a market economy, namely an institutional infrastructure, and the dynamics of the reform process, namely that it takes time before the new institutional infrastructure and the

persons who have to work in it are adapted to each other and the new situation. The collapse of the old planning institutions placed the enterprises in an extremely uncertain situation, in which it was difficult to find new customers and to decipher how trustworthy they were. From this perspective the length and depth of the depression depended on the time it cost to build new institutions, for example the new private property rights, and the time for the market players to adapt to them.

Our study also reveals that the negative aspects of transition can be compensated for, and in the Chinese case even more than fully compensated for if the authorities allow structural flexibility. This takes two forms. First, the speed with which the *de novo* enterprises can expand is important for the success of the transition process. This change is observable in the Czech Republic and especially in China. According to the World Bank the transition gets momentum if the share of medium-sized and small-scale enterprises in national employment is more than 40 percent. This condition is fulfilled in China. In Russia, to the contrary, the *de novo* enterprises were unable to expand. The government could not protect the new enterprises against the negative practices of the Mafia and the already existing big enterprises. Behaviour of the government itself, such as an erratic tax burden, also was counterproductive. Second, the loss of the export market that the COMECON was before the regime switch had to be offset by an increase of exports to mainly the West. The Czech Republic was particularly successful in this respect. The strong devaluation of its currency resulted in a strong swing of foreign trade to the West. China's export possibilities to the West also profited strongly from an undervalued domestic currency, the yuan. In contrast, Russia did badly in the 1990s due to an overvalued rouble, leading to a crowding-out of the traditional industry.

Besides these institutional lessons, which are also applicable to developing countries, the transition process in Eastern European countries and the newly independent countries that arose from the former Soviet Union also produced some lessons about the order of reform measures. Institutional adjustments and economic stabilization, both early in the reform process, prove to have positive effects on a rapid restoration of the pre-transition national income level. Late stabilization, in contrast, is devastating in this respect. A late adjustment over time of the exchange rate system towards more flexible – and thus less rigid, disequilibrium – exchange rates appeared to be economically costly in the transition countries: witness the negative experience of former East Germany and Russia, and the positive experience of the Czech Republic. This is an implicit support for the optimal sequence of liberalization steps as developed before the transition experience of the 1990s.

Summing up, the success of a transition process, and thus a development policy in general, not only depends on the building of a viable market

sector. It also depends on the existence of a strong government that is able and willing to create the necessary market institutions, fight the vested interests, and formulate an economic policy that aims at an immediate and thorough economic stabilization. Privatization and full liberalization can come later. As soon as domestic markets function, a rapid opening of international trade and stimulus of the international trade relations, including the introduction of a unified exchange rate which closely approaches its equilibrium value, should be focal points of economic policy.

Notes

1. For example, latent nationalism was fuelled by new publications about the Chernobyl catastrophe and the contents of the Molotov–Ribbentrop pact.
2. See Williamson (2003) for an elaboration on the reform agenda, consisting of ten reforms, that John Williamson himself in 1989 labelled the ‘Washington Consensus’.
3. CMEA is the group of countries that belonged to the Council for Mutual Economic Assistance. The latter was the body that was supposed to govern trade among Soviet-bloc nations.

References

- Balassa, B. (1987), ‘China’s Economic Reforms in a Comparative Perspective’, *Journal of Comparative Economics*, 11: 410–26.
- Beyer, J. (2001), ‘Beyond the Gradualism–Big Bang Dichotomy: The Sequencing of Reforms and its Impact on GDP’, in J. Beyer, J. Wielgohs and H. Wiesenthal (eds), *Successful Transitions*, Baden-Baden: Nomos, pp. 23–39.
- Buck, T., I. Filatotchev, P. Nolan and M. Wright (2000), ‘Different Path to Economic Reform in Russia and China: Causes and Consequences’, *Journal of World Business*, 35: 379–400.
- Campos, N. and F. Corricelli (2002), ‘Growth in Transition: What We Know, What We Don’t, and What We Should’, *Journal of Economic Literature*, 40: 793–836.
- Commander, S., Q. Fan and M. Schaffer (1996), *Enterprise Restructuring and Economic Policy in Russia*, Washington, DC: World Bank.
- Ellman, M. (2005), ‘Transition: Intended and Unintended Processes’, *Comparative Economic Studies*, 47: 595–614.
- Ellman, M. and V. Kontorovich (1992), ‘Overview’, in M. Ellman and V. Kontorovich (eds), *The Disintegration of the Soviet Economic System*, London and New York: Routledge, pp. 1–39.
- Gros, D. and A. Steinherr (2004), *Economic Transition in Central and Eastern Europe; Planting the Seeds*, Cambridge: Cambridge University Press.
- Hewitt, E.A., B. Roberts and J. Vanous (1987), ‘On the Feasibility of the Key Targets in the Soviet Twelfth Five Year Plan (1986–90)’, in Report of the Joint Economic Committee of the Congress, *Gorbachev’s Economic Plans*, Washington, DC.
- Knaack, R.K. (1996), ‘The Collapse of the Russian Economy: An Institutional Explanation’, in B. Dellago and L. Mittone (eds), *Economic Institutions, Markets and Competition*, Cheltenham, UK and Brookfield, USA: Edward Elgar, pp. 252–74.
- Knaack, R.K. (1999), ‘De ineenstorting van de Russische economie’, *Maandschrift Economie*, 63: 356–73.
- Knaack, R.K. and H. Jager (2007), ‘On the Relevance of Institutional Economics for International Economics’, in J. Groenewegen (ed.), *Teaching Pluralism in Economics*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 200–225.
- Kornai, J. (1992), *The Socialist System*, Oxford: Clarendon Press.
- Kornai, J. (1994), ‘Lasting Growth as the Top Priority’, Harvard Institute of Economic Research, Discussion Paper No. 1697.

- Lin, J.Y. (2004), 'The Role of Government in China's Transition to a Market Economy', paper for the conference A Liberal Agenda for the New Century: A Global Perspective, Moscow, 8–9 April.
- Naughton, B. (1994), 'What is Distinctive to China's Economic Transition? State Enterprise Reform and Overall System Transformation', *Journal of Comparative Economics*, 18: 470–90.
- Rodrik, D. (2000), 'Institutions for High-Quality Growth: What They Are and How to Acquire Them', NBER Working Paper 7540.
- Rutland, P. (1999), 'Mission Impossible? The IMF and the Market Transition in Russia', *Review of International Studies*, 25 (Fall): 183–200.
- Schmieding, H. (1993), 'From Plan to Market: On the Nature of the Transformation Crisis', *Weltwirtschaftliches Archiv*, 129: 216–53.
- Schroeder, G.E. (1992), 'Soviet Consumption in the 1980s: A Tale of Woe', in M. Ellman and V. Kontorovich (eds), *The Disintegration of the Soviet Economic System*, London and New York: Routledge, pp. 86–105.
- Stiglitz, J. (1998), 'More Instruments and Broader Goals: Moving Towards the Post Washington Consensus', *WIDER Annual Lectures*, 2, United Nations University, Helsinki.
- Taylor, L. (1993), 'Stabilization, Adjustment and Reform', in L. Taylor (ed.), *The Rocky Road to Reform*, Boston, MA: MIT Press, pp. 39–94.
- Williamson, J. (2003), 'From Reform Agenda to Damaged Brand Name: A Short History of the Washington Consensus and Suggestions for What to Do Next', *Finance and Development*, September: 10–13.
- Woo, W.T. (1994), 'The Art of Reforming Centrally Planned Economies, Comparing China, Poland and Russia', *Journal of Comparative Economics*, 18: 276–308.
- World Bank (2002), *Transition – The First Ten Years: Analysis and Lessons for Eastern Europe and the Former Soviet Union*, Washington, DC: World Bank.
- World Bank (2005), *Growth, Poverty, and Inequality in Eastern Europe and the Former Soviet Union*, Washington, DC: World Bank.

Index

- '1-2-3' model, *see* dependent economy model
- Acemoglu, D. 266–70, 385–6
- Afghanistan 101
- Africa 24, 458–9, 463
poverty in 220
skilled worker migration from 122
- Africa, sub-Saharan 103, 125, 152, 459, 463, 490–516
brain drain from 119
child labor in 254
children's health in 250–51
democratization in 512
demographic change in 495, 507–8
development strategies 514–16
economic structure 491–5
gender gaps in education in 237
geographical disadvantages 505–7
growth experience 490–98
interpretations of poor growth performance 498–508
poverty in 220, 490, 508
- African Development Bank 157
- aggregate demand 557–8, 560–61
- agricultural productivity
and gender inequality 239
- agriculture 37, 397–8, 400
in China 588, 593
effects of war on 430
foreign aid and 108
gender issues and 228, 229, 230, 231
in India 557–8, 562
in Russia 593
in South Asia 556
- AIDS 154, 221, 224, 225, 251, 383, 435, 507–8
- Aitken, B. 142
- AK model of growth 4
- Amsden, A. 168
- Arab Free Trade Agreement 527
- Asia 73, 75, 76, 78
gender inequality in 236
- Asian Development Bank 157
- Asian financial crisis 72, 74–6, 84–5, 463, *see also* financial crisis in East Asia.
- asymmetric information 342
in international capital markets 88
in technology markets 138, 140
international migration and 124–5
- balance of payments 151, 451–2
of India 563–4
- Baland, J-M. 394–406
- Balasubramanyam, V. 22–3, 48–58, 143–4
- Bangladesh 556, 565–6
- Bank for International Settlements 66
- bank lending
international 71, 79–88
- banks 79–88, 89–91, 385, 453, 525
- Bardhan, P. K. 381–93
- Barro, R. J. 22, 195, 413, 420
- Bates, R. 270, 498–9
- Bauer, P. 98
- Bebbington, A. 414
- behavioral economics 389
- Benabou, R. 201
- Berg, E. 498–9
- Beyer, J. 591
- Bhagwati, J. 5–6, 51, 170
- Birdsall, N. 193–212, 461
- Blecker, R. 170
- Bloch, H. 24, 26
- Bloom, D. 126
- bonds
GDP-indexed 65
local currency-denominated 65–6
- Borensztein, E. 144
- Borjas, G. 125
- Boserup, E. 225–6
- Brady plan 481
- brain drain, 119–22, 132–3, 171
- Branstetter, L. 145
- Braunstein, E. 242–3
- Brazil 72, 154, 156
ethnic/racial inequality in 264

- Bretton Woods institutions 150–53, 159
- Bruton, H. 30–47
- buffer stocks 27
- bureaucracy 52
- Burgstaller, A. 170
- Burundi 265
- Canada
- ethnic/racial inequality in 264–5
 - capabilities 312, 418, 420
 - capital account liberalization 531
 - in India 563
 - capital accumulation 32, 35, 166–7, 349–50, 458–9
 - capital flows *see* international capital flows
 - capital markets 382, 384, 525
 - capital mobility 295
 - capital requirements 68
 - Caribbean
 - skilled worker migration from 121
 - CEPAL *see* ECLAC
 - centrally planned economies, lessons from 282–4
 - Chang, H-J. 367
 - Chen, S. 219
 - child labor 253–4, 255–6, 259–60
 - children
 - and development 250–61
 - effect of female education on 238
 - Chin, J. 173
 - China 54, 78, 165, 457, 463, 536–50, 579, 587–90
 - comparison to India 537–8
 - FDI in 51–3
 - gender issues in 234–5
 - growth experience 536–42
 - income distribution in 189–90, 541
 - investment levels 539
 - legal system in 387–8
 - population control in 538–9
 - poverty in 220, 536, 541
 - private sector participation 543–9
 - state owned enterprises in 542, 544–5, 548
 - total factor productivity 539–41
 - Chiswick, B. 124
 - civil war 265, 423–41, 511 *see also* war class 390–91
 - classical economists 165
 - Coase theorem 342, 385
 - coefficient of variation 182–3
 - collective action 398–9
 - Collier, P. 270, 426, 427, 498, 501–2
 - Colombia 427
 - colonies 31, 167, 267–9, 529
 - commodity agreements 158
 - Commonwealth of Independent States (CIS) 579 *see also* transition economies
 - comparative advantage 16, 21–2, 31, 153, 167
 - Compensatory Financing Facility (CFF) 151
 - Computable General Equilibrium (CGE) Models 347–8, 354
 - conditionality
 - for foreign aid 111–13
 - IMF 151–2
 - contagion, financial 60, 81, 88
 - contract enforcement 203, 411
 - contracts, informal 388
 - convergence, of income across countries 169, 172
 - corporate governance 387
 - corruption 199–200, 369–77, 386
 - costs of 371–2
 - in East Asia 574–5
 - foreign aid and 107, 113
 - and growth 290, 372
 - levels and types of 370–71
 - measurement of 369–70
 - policies towards 372–6
 - and size of government 374
 - cost-benefit analysis 32, 351–3
 - Cote d'Ivoire 425
 - credit markets 198–9, 384, 389
 - crime 203
 - in transition economies 586–7
 - cross-country growth regressions 10–12
 - culture 407–22
 - definition of 409
 - cycles, effect of international capital flows on 90–94
 - Dahi, O. 522–35
 - Damon, A. 250–61
 - Darity, W.A. 262–77
 - data problems 215–16, 217–18, 229–30
 - international migration and 120

- De Soto, H. 411
- Deardorff, A. 173
- debt crisis 88, 151, 454
- debt relief 101
- deflation 297
- Deininger, K. 208
- Demir, F. 522–35
- democracy 408, 412, 531–2
and war 427–8
- dependency theory 167, 478
- dependent economy model 322–4,
331–2
- Development Assistant Committee
(DAC) 99
- discrimination
against women 231
economic 197
ethnic/racial 264
- distortions 5–6
market 32–3, 37–8
policy-induced 6
- divergence, in growth 455–6
- division of labor,
international 287
- Doha development round 151, 153,
154, 155
- dollarization 298
- Douglas, M. 417
- dual economy 36
- Dutch disease 43, 322, 333, 523,
526
foreign aid and 108
remittances and 132
- Dutt, A. 22, 163–76
- East Asia 48, 203–7, 459, 569–78
asset redistribution in 575
diversity of experiences 571–5
foreign direct investment in 571–2
industrial policy in 572–3
interpretations of growth experience
569
lessons from 576–7
trade policy in 572–3
- Easterlin, R. 194
- Easterly, W. 98, 201–2, 264, 269, 270,
411
- Economic Commission for Latin
America, (ECLAC) 158, 474–6,
478
- education 199, 209, 217
of children 251–3, 256–8
effects of war on 435
inequality of 206
in Latin America 484–5
in MENA 528–9
policy 224, 438
policy, for children 259
primary 207
in South Asia 556
university 206
- efficiency 55, 342–3
of private firms and state-owned
enterprises compared 361–4,
366–7
- Elson, D. 227–49
- enclaves 460–61
- Engel's law 20
- Engerman, S. 201, 385
- England 381
- entitlements, individual 430–32,
434–5
- entrepreneurs 359
- environment 173–4
- environment, and war 428, 429
- environmental Kuznets curve 173
- environmental planning models 349
- equity 342–3, 360
- Ethiopia 24
- ethnic diversity 270–75
- ethnic fractionalization 504
- ethnic/racial conflict 264–6, 275
- ethnicity 262–77, 423–4, 531
- ethno-linguistic fractionalization
(ELF) index 270–74
- Europe, Central and Eastern 579, 581
see also transition economies
- evolutionary economics 405–6
- exchange rate 206, 243, 245, 526
devaluation of 335
as a nominal anchor 297–8
policy 40, 42, 293, 295–6, 526
stability 296
targets 301, 336
- exchange rate regime
fixed 296–8
floating 296–8
intermediate 296–8
- export processing zones
in East Asia 570

- export promotion 10, 21, 50–51, 144
 - in East Asia 570
- export-oriented industrialization
 - women and 232
- exports 33, 35, 39, 40, 54
 - effects of war on 430
 - and gender inequality 239
 - of India 561
 - manufacturing 449–50, 459
 - technology transfer and 137
- externalities 130, 167, 168, 169, 342, 384, 395, 396, 399, 400–401, 447
 - from skilled emigration 132–3
- factor price equalization 168
- families,
 - bargaining power of women in 232, 235
 - migration decision by 126
- famines 418
- Fel'dman model 350
- female education
 - effect on children 238
 - effect on fertility 237
 - effect on growth 236–7
 - effect on population growth 237
- fertility 237
- finance 388
- financial crisis 71, 72, 74–5, 454, 531
 - in East Asia 62, 575–6
 - in Mexico 72, 74–6, 82–3
 - in Russia 62
- financial liberalization 88, 89, 90, 91–4, 452–3, 458, 463
 - in East Asia 576
- financial markets 452–3
 - women and 240–41
- financial volatility 60, 71, 89–90
 - effects on growth 59–60
- Findlay, R. 166
- fiscal deficit 312–13
 - relation with growth 313
 - in Latin America 477, 481–2
 - in Sub-Saharan Africa 510
- fiscal policy 91, 305–17
 - characteristics in developing countries 305–11
 - and gender equality 241–3, 244
 - in India 558
 - and privatization 360–61
 - procyclical 311
 - in South Asia 555, 557
- fisheries 402–3
- Fishlow, A. 469–89
- FitzGerald, V. 319–40
- food security 438
- foreign aid 98–118, 430, 436–8
 - amounts of 100
 - definition of 99
 - donor motives for 105
 - donors of 101, 104
 - foreign policy and 104
 - recipients 101–3
 - tying of 104–5, 110
 - types of 99, 110
- foreign debt 170
- foreign direct investment 48–58, 61, 64, 170, 172, 448, 451
 - in China 589
 - determinants of 49–53
 - in East Asia 571–2
 - effect of 53–6
 - effects of war on 433
 - export oriented 143
 - and gender equality 244
 - import substituting 143
 - in Latin America 476–7
 - tax incentives for 54
 - and technology transfers 138–46
- foreign exchange 54, 106, 447–8
 - effects of war on 430
- Foster-Greer-Thorbecke index 215
- fragility, economic 77–8
- France, legal system in 387
- Friedman, B. 410
- Friedman, M. 98, 390
- Fukuyama, F. 413–14
- gap models 106, 350
- gender,
 - and development 227–49
 - distribution, effect of growth and development on 227–36
 - distribution, effect on growth and development 236–41
 - inequality, in India 563
- General Agreement on Trade and Tariffs (GATT) 150
- genocide 265–6
- Ghosh, J. 554–68

- Gibson, B. 341–55
 Gini coefficient 179, 182–3, 195–7, 208, 463
 Glass, A. 137–49, 173
 Glewwe, P. 250–61
 global warming 174
 globalization 71, 119, 234, 288, 445–6, 473, 485–6, 577
 Goulet, D. 417–19
 governance 134, 388
 and foreign aid 113
 and growth 290
 government 39, 343, 436
 failure 201, 286, 343
 policy intervention 39
 revenue 359
 foreign aid and 107–8
 in transition economies 596
 government expenditure 209
 effects of war on 433–4
 reform 316
 Gray, C.W. 369–80
 greed and grievance, and war 426, 429
 Griffith-Jones, S. 59–70
 Grilli, E. 27
 gross domestic product
 -indexed bonds 65
 Grossman, G. 173
 growth 23–4, 30, 33, 37, 88, 165, 203–7, 447, 450, 454–6,
 effect of brain drain on 133–4
 effect of ethnic diversity on 270
 effect of foreign aid on 105–10
 effect of income distribution on 194–201
 effect of international capital flows on 90–94
 effect on ethnic/racial inequality 262
 effect on gender inequality 235–6
 effect on well-being of children 254–8
 effect of war on 432
 in India 560–61
 in MENA 524, 529–30
 models 349–50
 in South Asia 555
 of transition economies 582
 relation with life expectancy 223–4
 relation with literacy 223–4
 relation with poverty 223–4
- Hagen, E. 412
 Hamilton, A. 168
 happiness 194, 207–8
 Harrison, A. 142
 Harrison, L. 411, 412, 416
 Harrod-Domar model 32, 36, 106, 321
 Hatton, T. 125–6
 Heal, G. 342, 349
 health 199, 217–18, 383
 of children 250–51, 256–8
 effects of war on 435
 foreign aid and 106, 107
 impact of brain drain 132
 in MENA 528–9
 policy 224, 257, 438
 policy, for children 258
 in South Asia 556–7
 in transition economies 585
 Heckscher, E. 31
 Heckscher-Ohlin-Samuelson model 129–30, 167, 168
 Heinz, J. 242–3
 Helpman, E. 173
 HIV 221, 224, 225, 435, 507–8
 Hoksbergen, R. 407–22
 Homer-Dixon, T. 428
 Hong Kong 463
 horizontal inequality, *see* inequality, horizontal
 Huang, Y. 52–3, 536–53
 human capital 55–6, 133, 144, 171, 270
 effect of war on 432
 in MENA 528
 Human Development Index 218, 418
 for Bangladesh 566
 ethnic/racial inequality in 263–4
 Hymer, S. 49–50
- identity 409, 416
 immigration policy 127
 imperfect information 389
 imperialism 167
 import substitution 10, 21, 32, 33, 36, 42, 43, 50–51, 144, 288, 523, 524–5
 in East Asia 570
 in India 558
 in Latin America 477–8
 income, measurement of 215–16
 income distribution 193–212, 364, 398,
 see also income inequality

- effect of migration on 129
- international migration and 124
- income inequality
 - aggregation problems in measuring 181–5
 - in China 541
 - comparisons across countries 184–5
 - comparisons over time 185
 - determinants of 185–8
 - in different regions 183–4
 - and economic growth 188–90
 - in Latin America 484–5
 - measurement of 179–91
 - policies towards 190
- increasing returns to scale 130, 167, 168, 169, 171, 386, 403–4, 446–7
- India 156, 203, 415, 557–64
 - FDI in 51–3
 - gender issues in 230, 234–5
 - information technology sector in 133, 561, 563
 - international migration from 120
 - legal system in 387–8
 - uneven pattern of poverty reduction in 221
- Indonesia 139, 425
- industrial policy 168, 460
 - and gender equality 244
- industrialization 188, 475–6
- industrial sector, gender issues and 228
- industry,
 - in India 558
- inefficiency 33, 395, 396
- inequality 194–201, 399, 557, *see also*
 - equity, income distribution
 - between women 236
 - in distribution of education 199–200
 - effect on growth 194–201
 - ethnic/racial 262–4
 - horizontal 208–9, 425, 426, 428
 - in India 562
 - international 164–5, 456–7
 - international migration and 124
 - in landownership 199
 - in MENA 531
 - North–South 30
 - in South Asia 555–6
 - in transition economies 585–6
 - vertical 427
- inflation 295, 301, 306
 - effects of war on 433
 - in Latin America 477, 481
 - in Sub-Saharan Africa 510
- inflation targeting 294, 298–9, 325–30
 - and central bank independence 300
 - experience in developing countries 299–301
 - prerequisites 299–300
- informal employment 232, 240
- informal sector 37, 433–4, 437
 - in transition economies 584–5
- information technology 133
 - in India 561, 563
- infrastructure, effects of war on 432
- innovation 134, 172, 385
- input–output models 344–5
- institutional economics 381, 385
 - new 394
- institutions 41–2, 194, 201, 202, 266–9, 385–6
 - dysfunctional 385–6
 - in transition economies 587, 594–6
- insurance markets 198, 389
- intellectual property rights 138, 145–6, 172, 382–4
- Inter-American Development Bank 157
- International Bank for Reconstruction and Development (IBRD) 104, 150–52, 157, *see also* World Bank
- international capital flows 55, 59–68, 71–118, 169–70, 328, 448–9, 452–3, 463, *see also* foreign direct investment
 - boom–bust cycles of 59–61, 64
 - effects of war on 430, 432–3
 - main characteristics of 60–64
 - policies towards 64–8
 - regulation of 66–7, 77, 88, 94, 156, 295
 - to India 564
- international capital markets 88
- International Development Agency (IDA) 151
- International Development Association (IDA) 104
- international institutions 150–62
- international migration 119–36, 171
 - effects of 128–33
 - growth in 120
 - returns to 123–4

- International Monetary Fund (IMF)
 53, 66, 111–12, 114, 151, 153,
 155–7, 159, 243, 288, 319, 530–31,
 581
 Basic Financial Programming
 Framework 320–21
 international production chains 451,
 460
 international trade 3–13, 31–48, 167–9,
 287, 448
 effects on income distribution 7
 effects on growth 4, 8–12
 effects of war on 433
 gains from 4–8, 16, 167
 liberalization 11, 38–9, 172, 449–52,
 482, 510–11
 liberalization, and gender equality
 243
 mainstream view of 3–4
 in MENA 526–8
 policy 21–2, 30–47, 168, 225, 460
 of transition economies 583–4
 International Trade Organization 150
 investment 35, 36, 197, 451, 458
 effects of war on 432
 foreign aid and 106, 108
 public 458–9
 Iraq 101
 Islam 533
- Jager, H. 579–600
 Jha, R. 305–18
 joint ventures 144
 Jomo K. S. 569–78
- Kaldor, N. 207
 Kaminsky, G. 71–97
 Keen, D. 426
 Kerala 222, 224
 Keynes, J. M. 27, 150
 Kindleberger, C. P. 54
 Knaack, R. 579–600
 knowledge,
 codified 36
 tacit 41, 43, 44
 Korea, South 21, 39, 40
 Kornai, J. 580
 Korten, D. 416
 Kozul-Wright, R. 445–68
 Krugman, P. 168, 171–2
- Kuran, T. 533
 Kuznets curve 188–90
 Kuznets, S. 188, 193
- labor markets 123–4
 effects of war on 430
 in MENA 528–9
 labor, skilled 55
 Lall, S. 56
 Lamb, J. 420
 land 394, 397, 402
 markets, women and 231
 reform 201, 386, 557
 rights, women and 230–31
 titles 382
 Landes, D. 413
 Latin America 38, 203–7, 459, 469–86
 capital flows to 72, 73, 75, 76, 78
 gender inequality in 236
 gender issues in 230
 growth experience 470–83
 law
 and development 381–93
 civil 387
 commercial 388
 common 387
 contract 381
 rule of 381–2, 386, 391
 learning by doing 167, 168, 172, 384
 Leontief model 345
 Leontief, W. 344
 Levine, R. 264, 270
 Lewis, W. A. 36, 166, 227–8, 275, 286
 liberalization 445, 581
 in India 560–61
 in MENA 531–2
 in South Asia 545–55
 licensing, technology 137, 139, 144
 life expectancy 219–24
 linear programming models 346–7
 linkages 188–9
 liquidity 77
 List, F. 160
 literacy rate 219–20, 223–4
 loan-pushing 170
 lock-in 167
 Lorenz curve 182, 184, 463
 Lucas, R. 171
 lump sum transfers 6–7
 Lutz, M. 23–4

- macroeconomic policy 225
 - and gender distribution 241–9
- Mahalanobis model 351
- Mahalonobis, P. C. 561–2
- Malaysia 54, 78–9
- Mali 24
- Malloch, T. 414
- Mansfield, E. 140
- market 41, 342, 354, 394, 401, 408–9, 446–8
 - failures 198–9, 285–6, 342
 - imperfections, capital 126–7
 - role in development 281–91
 - structure 20
 - and technological progress 282
 - in transition economies 594
- Marshall Plan 150, 157–8
- Masson, P. 293–304
- Mattoo, A. 153, 154, 155
- McCleary, R. 413, 420
- McClelland, D. 411–12
- McNamara, R. 151
- Mexico 72, 400
 - financial crisis 72, 74–6, 82–3
 - migration from 120, 130
- microfinance 240
- middle class 201–2
- Middle East and North Africa (MENA) 522–35
- migration, determinants of 122
- Milanovic, B. 165
- military expenditure 434, 530
- Millennium Development Goals 157, 237
- Mishra, P. 130
- modernization theory 407, 410, 416
- monetary policy 91, 293–303
 - and development 293–4
 - and domestic nominal anchors 298–9
 - and gender equality 241–3
 - independence 294–5
 - see also* stabilization policy
- Montalvo, J. 273–4
- moral hazard, in technology markets 140
- Moran, T. 143
- Morgan, W. 27
- Morley, S. 179–92
- mortality, child 107
- mortality, infant 250–51, 255–6, 425, 435
- mortality rate
 - adult 435
 - child (under 5) 219–20, 250–52, 255–7
- multinational firms, *see* transnational corporations
- mutual funds 72–9, 95
- Myrdal, G. 417
- natural disasters 105
- natural resources 394, 395, 396, 397, 400–402, 404, 450, 461, 526
 - war and 426, 428
- Ndulu, B. 490–521
- need-for-achievement 411
- neoclassical approach 167, 168
- Nepal 425, 567
- new growth theory 22, 167, 168, 171, 350
- New International Economic Order (NIEO) 159
- newly-industrialized countries (NICs) 174, 460
- Nigeria 101
 - ethnic/racial inequality in 264
- non-government organizations (NGOs) 99, 416, 434, 437
- Norback, P.-J. 139
- norms, social 202, 389
- North Africa 552–3
- North America 201
- North, D. 371
- North–South
 - inequality 164–5
 - models 21, 145, 165–7
 - relations 160, 163–76
 - trade 167–9
- Nurkse, R. 286
- nutrition
 - effects of war on 434–5
 - in South Asia 556
- nutritional status, of children 257
- O’Connell, S. 490–521
- official development assistance 99
- Ohlin, B. 31
- oil 523, 526
- Okediji, T. 274

- Olson, M. 399, 411
 open access 394, 395, 398
 openness 38–9, 450, 460
 growth and 462
 Organization for Petroleum Exporting Countries (OPEC) 158–9
 outsourcing 561
 outward-oriented development 40, 524
- Pack, H. 141
 Pakistan 564–5
 patents 383, 391
 Patnaik, P. 554–68
 pharmaceuticals 383
 planning 341–55, 594–5, 598
 economy-wide 342–4
 in India 561–2
 models 341, 344–50
 see also centrally planned economies
 Platteau, J.-P. 394–406
 polarization 273
 policies
 and effects of foreign aid 108
 to promote child welfare 257–60
 towards economies in conflict 437–8
 policy 343, *see also under* education, fiscal, health, industrial, monetary and trade policy
 planning models and 353–4
 political constraints, on privatization 365–6
 political economy
 of India 560
 of migration policy 127
 of South Asia 564
 political instability 200, 529
 political reform, in transition
 economies 596–7
 political voice 217–18
 politically relevant ethnic groups index 271–2
 pollution 173
 population 237, 428
 portfolio flows, international 71
 Posner, D. 271–2
 poverty 213–26, 235, 417–18
 absolute 216
 in China 536, 541
 foreign aid for reducing 104
 gap 214
 in India 563
 line 214–17, 568
 measurement of 214–15
 in MENA 531
 one dollar a day 216–17, 221
 rate 219
 relative 216
 in sub-Saharan Africa 490, 508
 targeted programs 325
 in transition economies 585
 two dollars a day 217, 221–2
 and violent conflict 427
 war and 428
 Prebisch, R. 18, 21, 27, 158, 475, 478
 Prebisch-Singer thesis, 18, 20–22, 25–6, 27, 158
 primary commodities,
 exporters of 25
 income inelasticity of 20
 price inelasticity of 20
 principal agent problem, foreign aid and 111, 113–14
 Pritchett, L. 165
 privatization 208, 356–9, 364–7, 400, 401, 403, 533, 581
 in Latin America 482
 reasons for 360–64
 in transition economies 592
 productivity 446
 growth 34, 141, 459
 project appraisal 351–2
 property rights 38, 201, 202, 267, 381, 389, 395, 401, 404, 405, 411
 property, common 389, 396, 400, 401
 protection
 conditional on export promotion 570–71
 import 38, 153, 201
 import in OECD countries 153
 psychological factors 410–12
 public goods 282, 399
 global 174
 public investment 206
 public sector 524
 purchasing power parity 216
 Putnam, R. 202, 264, 269, 413–14
 Putterman, L. 281–92
- Quah, D. 164
 Quesnay, F. 344

- race 262–77
 Radelet, S. 98–118
 Raffler, K. 150–62
 Ramachandran, V. 140
 Randomized experiments 258–9
 Ravallion, M. 220
 Rayment, P. 445
 reforms,
 in China 589–90
 in MENA 530–32
 sequencing of, in transition economies 591–4
 regional planning models 350
 Reinhart, C. 88
 religion 412–14, 417, 420, 423–4
 remittances,
 migrant 119, 127, 130–32, 523
 migrant, to South Asia 563–4, 565
 rent-seeking 33, 201, 202, 264–5, 525, 532
 Research and Development 55, 138, 382–3
 reverse engineering, 55, 137
 Reynal-Querol, M. 273–4
 Ricardo, D. 15–17, 21, 31
 risk
 indicators 77–8
 political 529
 Rivera-Batiz, F. 119–36
 Rodríguez, F. 3–15
 Rodrik, D. 152, 200–201, 382
 Romeo, A. 140
 Rostow, W.W. 286
 Roy model 124
 Russia 594
 financial crisis 62, 74–6
 Rwanda 265

 Saavedra-Rivano, N. 170
 Sachs, J. 98
 Saggi, K. 137–49, 173
 Sala-i-Martin, X. 164
 Samuelson, P. 168
 sanctions, economic 438, 529
 Sapsford, D. 16–29
 Sarkar, P. 19
 saving 197
 foreign aid and 106, 107, 108
 private 324
 relation with growth 313
 school attendance 217
 school enrollment
 gender gap in 229
 primary 217, 219, 251–3, 435
 secondary 253, 255
 Schumpeter, J. 407–8
 Seguino, S. 236, 238
 self-employment 230, 240
 Sen, A. 312, 356–68, 415, 418–20
 services, in India 563
 shadow price 346, 351–2
 Sierra Leone 225
 Singapore 54, 463
 Singer, H. 18–21, 23–4, 27, 170
 Sjaastad, L. 122
 skilled workers 119, 120–22, 124, 130, 133, 168, 171
 slavery 167, 267, 270
 Smarzynska, B. 140
 Smith, A. 16, 165, 207, 381, 394
 Social Accounting Matrices 347
 social capital 202–3, 413–15
 effects of war on 430
 social conflict 200
 social diversity index 274
 social inclusion 217–18
 Sokoloff, K. 201, 385
 Solow growth model 43
 South Africa 154
 South America 201
 South Asia 459, 554–68
 child labor in 254
 children's health in 251
 poverty in 200
 South East Asia 460
 Soviet Union 341, 580–81
 speculative attack 297, 302
 spillovers
 technology 55–6, 141–3, 145, 452, 462
 wage 142
 spiritual capital 414
 Spraos, J. 27
 Squire, L. 207, 208, 213–26
 Sri Lanka 222, 224, 566
 Srinivasan, T. N. 5–6
 stabilization 285, 581
 effect on women's employment 233
 new Keynesian approach to 330

- policy 319–37
- in transition economies 590–91
- Stark, O. 126
- state 427
 - capture 386
 - as coordinator of development 287
 - developmental 208, 385, 462, 524
 - in India 557, 562
 - intervention 390
 - in MENA 531–3
 - role in development 281–91
- state-owned enterprises 356, 359–60, 524–5
 - in China 542, 544–5, 548, 589
 - in East Asia 572–3
 - reasons for establishing 359–60
- Stern, N. 98, 152
- Stewart, F. 423–41
- Stiglitz, J. 98
- stock markets 77, 89, 90, 95, 365, 384
 - effect of financial liberalization on, 81
- Stolper-Samuelson theorem 7
- Streeten, P. 56–7
- structural adjustment 112, 152, 288, 319–37, 356, 458
 - effect on women's employment 233
 - and income distribution 335–6
 - in MENA 530–32
- structuralist CGE models 347–8
- subjective well-being 194, 207–9
- Subramanian, A. 153, 154, 155
- subsidies
 - on FDI 50, 54
 - in MENA 532
 - to Northern farmers 154
- surplus labor 286, 330

- Taiwan 39, 40
- Tanzania 203
- tariffs, import 32, 50
- tax buoyancy 310
- tax-GDP ratio 306
- tax reform 313–16
- Taylor, J. E. 126
- Taylor, L. 166
- teacher absenteeism 259
- technological capability 172
- technological change 20–21, 56, 134, 403, 447–8
- technology 53, 385
 - diffusion 145
 - transfer 36, 53–5, 137–49, 171–2, 384, 449, 462
 - transfer, foreign aid and 106
- terms of trade
 - agriculture-industry 559
 - between primary products and manufactured goods 17–20
 - international 16–29, 20–21, 451, 461
 - net barter 17, 26
 - North-South 18–19, 31, 170
 - volatility of 23–4
- Thompson, E. P. 391
- Throsby, D. 411, 418
- total factor productivity 141, 142, 384
 - growth of 33, 206
- township and village enterprises 588, 593–4
- trade, *see* international trade
- trade, free 39
- Trade-related Intellectual Property Rights (TRIPs) 154, 383
- Trade-related Investment Measures (TRIMs) 154
- transaction costs 389, 394, 397, 401–3, 405
- transition economies 579–600
 - capital flows to 73, 76–8
- transnational corporations 48, 53, 138–40, 146, 170, 173, 391, 451
- Triplett, R. E. 262–77
- trust 202
- Turkey 524–5, 530–31

- underemployment 233, 242
- unemployment 233, 242, 459
 - in MENA 528
 - in South Asia 556
 - in transition economies 584
- uneven development 22, 166–8, 169, 171
- United Nations 157
- United Nations Conference on Trade and Development (UNCTAD) 158–9, 453
- United Nations Development Programme (UNDP) 158, 218, 263

- unpaid work, women and 230, 231, 232, 233, 234
- urban bias 499
- Uruguay round 154, 155, 172
- USA 154
 - legal system in 387
 - migration to 130
- Value at Risk models 60–61
- values 409, 412–13, 418, 419
- Venezuela, technology spillovers in 142
- Vietnam, well-being of children in 257
- violence 203
- violent conflict 423–41
 - causes of 423–9
 - consequences of 429–37
 - in South Asia 567–8
- Voluntary Export Restrictions 154
- von Neumann model 345
- wage inequality, gender 232
- Walrasian CGE models 347
- war 225, 423–41, 529–30, *see also*
 - violent conflict
 - foreign aid and 107
- Washington Consensus 37, 39–40, 42, 44, 483–4, 577, 581–2
- water, in MENA 526
- Weber, M. 386, 407
- Weil, D. 413
- Weingast, B. 381
- Wilber, C. 407–22
- Williams, E. 266–70
- Williamson, J., 44
- Williamson, J. G. 125–6
- Williamson, O. 415–16
- women, missing 234–5
- Women-in-Development literature 228, 229, 231
- Wood, A. 168
- working time, for men and women 234
- World Bank 27, 38, 104, 109, 111–14, 158, 216, 243, 319–21, 414, 457–8, 530, *see also* International Bank for Reconstruction and Development (IBRD)
 - interpretation of East Asian miracles 569–70, 572
 - Revised Minimum Standard Model 321–5
- World Food Program 158
- World Health Organization 383
- World Trade Organization (WTO) 11, 48, 150, 153–6, 158, 159, 172–3, 383
- Yang, M. 27