DEBRE MARKOS UNIVERSITY

COLLAGE OF NATURAL COMPUTATIONAL SCIENCE

DEPARTMENTMENT OF STATISTICS

TIME SERIES ANALYSIS [STAT-2042]

LECTURES NOTES

2nd Year Statistics program students

Course instructor(s) Contact email/ for any communication purpose:

1. Amare Wubishet : amarewubishet21@gmail.com
2. Luel Mekonnen : amleul117@gmail.com

Chapter one 1
1 Introduction 1
1.1 Definition and terminology 1
Definition of some terms in time series 3
1.2 Objective of time series analysis 4
1.3 Significance of Time Series 4
1.4 Component of time series 5
1.5 Decomposition Models of Time Series 6
1.6 Editing of Time Series Data 7
Chapter two 9
2 Test of randomness 9
2.1 Introduction of randomness 9
2.2 Statistical test of randomness 9
2.2.1 Turning Points Test 9
2.2.2 Difference Sign test 11
2.2.3 Phase length Test 12
2.2.4 Rank Test 14
Chapter Three 16
3 Estimation of Trend Component 16
3.1 Introduction 16
3.2 Constant means model and its estimations 16
3.2.1 Freehand method 17
3.2.2 Method of Semi Average 17
3.2.3 Least Square Method 19
3.2.4 Moving Averages method 19
3.2.4.1 Centered Moving Average Method 21
3.2.5 Exponential smoothing method 23
3.3 Linear Trend and Its Estimation 26
3.3.1 Least Square Method 26
3.3.2 Linear Trend estimation using moving average method 29
3.3.3 Linear Trend estimation using exponential smoothing method 31
3.4 Nonlinear trend and its Estimation 34
CHAPTER FOUR 37
4 Estimation of Seasonal Component 37

4.1 Introduction 37
4.2 The simple Average Method 38
4.3 Link relative method 40
4.4 The Ratio to Moving Average Methods 43
4.5 Ratio to Trend Method 47
4.6 Uses of Dummy Variables 50
4.7 Smoothing Methods for Seasonal data (series) 52
4.7.1 Deseasonalization of Data 52
4.8 Estimation of Cyclical and Irregular Components 53
4.8.1 Estimation of irregular components for additive and multiplicative model 53
CHAPTER FIVE 56
5 Introduction to Box-Jenkins Models 56
5.1 Introduction 56
5.2 The concept of Stationarity 57
5.2.1 Nonstationarity 58
5.2.2 Differencing 58
5.2.3 Differencing for seasonal data 58
5.2.4 Backward shift operator (Lag operator) 59
5.3 ARIMA models 60
5.3.1 Moving average process 61
5.3.2 Autoregressive process 62
5.3.3 Autoregressive Moving average process 63
5.4 Methodological tools for model identification 63
5.4.1 Autocorrelation function, ... 63
5.4.2 Partial autocorrelation function 65
5.5 Stages of Box-Jenkins methodology 67
5.5.1 Model Identification/selection 67
5.5.2 Parameter estimation 69
5.5.3 Model Diagnostic Checking 70
CHAPTER SIX 75
6 MODEL IDENTIFICATION AND ESTIMATION 75
6.1 Introduction 75
6.2 Autoregressive (AR) process 76
6.2.1 Special cases for Autoregressive Model/Process 76

6.2.1.1 $AR(1)$ process 76
6.2.1.2 $AR(2)$ process 79
6.2.1.3 The General pth Order Autoregressive $AR(p)$ Process 85
6.3 Moving Average (MA) process 85
6.3.1 Special Cases of Moving Average (MA) process 86
6.3.1.1 $MA(1)$ Process 86
6.3.1.2 $MA(2)$ Process 88
6.3.1.3 The General qth Order Moving Average $MA(q)$ Process 92
6.4 Autoregressive moving average models: 93
6.4.1 Special case of $ARMA(p, q)$ process 94
6.4.1.1 The $ARMA(1, 1)$ process 94
6.5 Autoregressive integrated moving average model 98
6.6 Seasonal ARIMA 99
6.7 Estimation of parameters 102
6.7.1 Parameter Estimation of AR Process 103
6.7.2 Parameter Estimation of MA Process 103
6.8 Diagnostic checking 104
6.8.1 Autocorrelation Check (Studying residuals) 104
6.8.2 An Approximate Chi-squareTest 104
6.8.3 The Ljung-Box Test 105
6.8.4 Over fitting and Parameter Redundancy 107
CHAPTER SEVEN 109
7 FORECASTING 109
7.1 Introduction 109
7.2 Needs and use of forecasting 110
7.3 Some Forecasting Methods 110
7.3.1 Average Methods 110
7.3.2 Exponential Smoothing Methods 112
7.3.3 Box-Jenkins Method 114
7.4 The Accuracy of Forecasting Methods 119

Chapter one

1 Introduction

1.1 Definition and terminology

Statistics is the way of to get information from data. Information is a set of data corresponding to a specific aspect of knowledge combined in an organized way.

What is time series?

A time series is defined as a set of quantitative observations arranged in chronological. Time Series Analysis is defined as the analysis of data organized across units of time and Time series analysis is a systematic approach to deal with mathematical and statistical questions posed by time correlation. The main features of many time series are trends and seasonal variations. But, another important feature of most time series is that observations close together in time tend to be correlated (serially dependent). Most often, the measurements are made at regular time intervals. Granger and Newbold (1986) describe a time series as a sequence of observations ordered by a time parameter.

A time series is a sequence of observations on a variable measured at successive points in time or over successive periods of time. The measurements may be taken every hour, day, week, month, or year, or at any other regular interval. A time series is a series of observations, y_{it}; $i = 1, 2, \dots, n$ & $t = 1,2,3, \dots, T$ Made sequentially over time. Here i indexes the different measurements made at each time point t; n is the number of variables being observed and T is the number of observations made. If n is equal to one then the time series is referred to as univariate (Chatfield, 1989), and if it is greater than one the time series is referred to as multivariate (Hannan, 1970).

- According to the role of time data can be classified into cross-section and time series data.

A time series data: is a collection of observation generated by sequentially through time. The special features of a time series are the data order with respect to time. Under this successive observations are expected to be dependent. The dependency from one time period t to another will be expected in making reliable forecast. The ordered of observation is denoted by subscript t i.e. X_t. $X_t \rightarrow$ the t observation at time t. the preceding observation is denoted by X_{t-1} and the next observation is denoted by X_{t+1}.

Example:

1. Industrial production of goods and services in Ethiopia over the last 40 years.

2. Number of traffic accidents in Ethiopia over the last 35 years.
3. Number of graduate student from Ethiopia University in over the last 10 years.

Cross-section data: is a collection of observation generated from different individual or groups at single point in time. It is a data that are observed on different observational unit at the same point in time. The order of the observation does not play any role.

Example:

1. Number of graduate students from Ethiopia University in a given year. i.e. AAU, DMU. HWU, BDU, UOG…, JU in 2010.
2. Health service delivery by health institutions in a given year.
3. Industrial production in a given year by different industry.

Example: data collected from different localities in the production of wheat at time is cross-section data; but collecting information with the same locality through time is time series data.

- Depending upon the information obtained time series data can be categorized into two types.
  - Discrete time series data: is only defined for a discrete set of points in time. These time series does not contain any information the time between these points.

    Example: Interest rate, Yields, Volume of sales and etc.
  - Continuous time series data: is typically recorded data either at fixed time interval or after a certain change in the Y values has taken place the drelogical parameters are often continuously recorded. The value function f is defined $f(t \quad 1, t \quad 2) = Y$.

    Example: Daily temperature reading in a given place, Monthly rainfall Industrial production (monthly, quarterly, yearly)

Application of Time series data arise in a variety of fields. Here are just a few examples.

- In **business**, we observe daily stock prices, weekly interest rates, quarterly sales, monthly supply figures, annual earnings, etc.
- In **agriculture**, we observe annual yields (e.g., crop production), daily crop prices, annual herd sizes, etc.
- In **engineering**, we observe electric signals, voltage measurements, etc.
- In **natural sciences**, we observe chemical yields, turbulence in ocean waves, earth tectonic plate positions, etc.

- In **medicine**, we observe EKG measurements on patients, drug concentrations, blood pressure readings, etc.
- In **epidemiology**, we observe the number of flu cases per day, the number of health-care clinic visits per week, annual tuberculosis counts, etc.
- In **meteorology**, we observe daily high temperatures, annual rainfall, hourly wind speeds, earthquake frequency, etc.
- In **social sciences**, we observe annual birth and death rates, accident frequencies, crime rates, school enrollments, etc.

Definition of some terms in time series

- **Univariate time series:** are those where only one variable is measured over time.
- **Deterministic time series:** if the future values of a time series are exactly determined by some mathematical function such as $Z_t = \cos(2\pi f t)$.
- **Nondeterministic time series**: if the future value can be determine only in terms of a probability distribution. It also known as statistical time series.
- **Multiple time series** are those, where more than one variable are measured simultaneously.
- **Autocovariance**: Measure of the degree of linear association/dependence among different time points.
- **Correlation coefficient**: Measure of linear dependence of pair of observation.
- **Autocorrelation:** Autocorrelation refers to the correlation of a time series with its own past and future values. Autocorrelation is sometimes called "serial correlation", which refers to the correlation between members of a series of numbers arranged in time.
- **Partial Autocorrelation:** It is the correlation between variables $x_t$and x_{t+2} after removing the influence x_{t+1}. Here the interdependence is removed from each of x_t and x_{t+2}.
- **Stationary:** Time series with constant mean and variance are called stationary time series. A time series Y_t is stationary if its probability distribution does not change over time, that is if the joint distribution of $(Y_{s+1}, Y_{s+2}, \dots, Y_{s+T})$ does not depend on S; otherwise, Y_t is said to be nonstationary. A pair of time series X_t and Y_t are said to be jointly stationary if the joint distribution of $\left((X_{s+1}, Y_{s+1}), (X_{s+2}, Y_{s+2}), \dots, (X_{s+T}, Y_{s+T})\right)$ does not depend on S. Stationarity requires the future to be like the past, at least in probabilistic sense.
- **Deseasonalization:** The process of eliminating seasonal fluctuations or deseasonalization of data consists of dividing each value in the original series by the corresponding value of the seasonal index

1.2 Objective of time series analysis

The objective of time series analysis is to discover a pattern in the historical data or time series and then extrapolate the pattern into the future; the forecast is based solely on past values of the variable and/or on past forecast errors

1. **Description**:- In time series presentation, it is the first step in the analysis using the time series plot because a series of ups and downs. A **time plot** is a two dimensional plot of time series data. The time plot will shows the important features of the series such as trend, seasonality, outliers, discontinuities and then to obtain simple descriptive measures of the main properties of the series and to help in formulating a sensible model.

2. **Explanation**:- it explains the variation in the series when there are two or more variables. It may be possible to the variation in time series to explain the variation in other series. A **univariate** model for a given variable is based only on past values of that variable, while a **multivariate** model for a given variable may be based, not only on past values of that variable, but also on present and past values of other (predictor) variables. In the latter case, the variation in one series may help to explain the variation in another series. For example, how see level is affected by temperature and pressure.
3. **Prediction**: - Developing a model and predicting (forecasting) the possible change in the future values using time series data. Given an observed time series, one may want to predict the future values of the series. This is an important task in sales forecasting, and in the analysis of economic and industrial time series. The term 'prediction' and 'forecasting' interchangeably, but note that some authors do not. For example, Brown (1963) uses 'prediction' to describe subjective methods and 'forecasting' to describe objective methods.

4. **Control**:- Time series are sometimes collected or analyzed so as to improve control over some physical or economic system. Using a structural model, as in (2), we may seek to control a system either by generating warning signals of future unfortunate events or by examine what happen if we alter either the inputs to the system or its parameters.

1.3 Significance of Time Series

The analysis of time series is of great significance not only to the economists and business man but also to the scientists, astronomists, geologists, sociologists, biologists, research worker etc. Time series analysis is of great significance in decision-making for the following reasons.

- **It helps in the understanding of past behavior of the data**: By observing data over a period of time, one can easily understand what changes have taken place in the past. Such analysis will be extremely helpful in predicting the future behavior.
- **It helps in planning future operations**: If the regularity of occurrence of any feature over a sufficient long period could be clearly established then, within limits prediction of probable future variations would become possible.
- **It helps in evaluating current accomplishments**: The actual performance can be compared with the expected performance and the cause of variation analyzed. For example, if expected sales for 2006/7 were 20000 colored TV sets and the actual sales were only 19000 one can investigate the cause for the shortfall in achievement.
- **It helps to facilitates comparison**: Different time series are often compared and important conclusions drawn there from.

1.4 Component of time series

The values of a time series may be affected by the number of movements or fluctuations, which are its characteristics. The types of movements characterizing a time series are called components of time series or elements of a time series. A time series is assumed to consist of a mixture of some or all of the following four components:

- Trend Component
- Seasonal Component
- Cyclical Component
- Irregular Component

The first three components are deterministic which are called "Signals", while the last component is a random variable, which is called "Noise". In order to understand and measure these components and to undertake any time series analysis it initially needs removing the component effects from the data (decomposition). After the effects are measured, making a forecast involves putting back the components on forecast estimates (re-composition).

1. **Trend (T_t):** This is the long term upward or downward movement of the series due to factors that influence the mean of the series. This long-term trend (growth or decay) is typically modeled as a linear, nonlinear, quadratic or exponential function. The trend may also be defined as a slowly changing non-random component of a time series.

It is the long term movement or Secular Trend of the long run direction of the time series. By trend we mean the general tendency of the data to increase or decrease during a long period of time. This is true of most of series of Business and Economic Statistics. For example an upward tendency would be seen in data pertaining to population, agricultural production, currency in circulation etc., while, a downward tendency will be noticed in data of birth rate, death rate etc.

2. **Seasonal Variation (S_t)**: Seasonal variations are the periodic and regular movement in a time series with period less than one year. For example demand of umbrella in the rainy season, demand of worm clothe in the winter, demand of cold drinks in the summer etc. The factor that causes seasonal variations is
    i. Climate and weather conditions
    ii. Customs, traditions and habits etc
3. **Cyclical variations (C_t)**: The oscillatory movements in a time series with period of oscillation more than one year are termed as cyclic fluctuations. One complete period is called a 'cycle'. The cyclical movements in a time series are generally attributed to the so called business cycle. There are four well-defined periods or phase in the business cycle namely prosperity, recession (decline), depression and recovery.
4. **Irregular Component(I_t)**: It is also called Erratic, Accidental or Random Variations. The three variations trend, seasonal and cyclical variations are called as regular variations, but almost all the time series including the regular variation contain another variation called as random variation. This type of fluctuations occurs in random way or irregular ways which are unforeseen, unpredictable and due to some irregular circumstances which are beyond the control of human being such as earth quakes, wars, floods, famines, lockouts, etc. These factors affect the time series in the irregular ways. These irregular variations are not so significant like other fluctuations.

1.5 Decomposition Models of Time Series

Decomposition methods are among the oldest approaches to time series analysis models. It is a crude yet practical way of decomposition the original data (including cyclic pattern in the trend itself) is to go for a seasonal decomposition either assuming multiplicative or additive models.

Breaking down the time series data into its component parts is called decomposition. The decomposition model assumes that the time series data are affected by four factors: the general trend in the time series data, general economic cycles, seasonality, and irregular or random occurrences. Decomposition begins by assuming that the time series data can be represented by a trend, seasonal, cyclic and irregular components which are combined in either a multiplicative or additive fashion. The combined time series models can be decomposed into its components, such as Seasonality (S_t), Trend (T_t), Cycling (C_t) and Irregularity (I_t) portions.

The Different types of time series models are:

a) **Additive Model**: $Y_t = T_t + S_t + C_t + T_t$
b) **Multiplicative model** : $Y_t = T_t \quad S_t \quad C_t \quad T_t$
c) **Mixed model**: $Y_t = T_t \quad S_t \quad C_t + I_t$– multiplicative model but it has an additive irregular component.

Where Y_t = observation at period of time t

T_t =Trend component at period t.

S_t = Cyclical trend component at period t.

I_t =Irregular component at time t.

NB: The multiplicative form of model occurs mostly in practice

The logarithms transformation multiplicative model gives additive model. We get.

$\log Y_t = \log T_t + \log S_t + \log C_t + \log I_t$.

So that the logarithm of the original data can be modeled by a pure additive model.

Therefore the additive model is appropriate if the magnitude (amplitude) of the seasonal variation does not vary with the level of the series, while the multiplicative version is more appropriate if the amplitude of the seasonal fluctuations increases or decreases with the average level of the time series.

1.6 Editing of Time Series Data

The process of checking through data is often called cleaning the data, or data editing. It is an essential precursor to attempts at modeling data. Data cleaning could include **modifying outliers,** identifying and correcting obvious **errors** and filling in (or imputing) any **missing observations**. This can sometimes be done using fairly crude devices, such as down weighting outliers to the next most extreme value or replacing missing values with an appropriate mean value.

It is necessary to make certain adjustments in the available data. Some important adjustments are:

- **Time Variation**: When data are available on monthly basis, the effect of time variation needs to be adjusted because all months of the year do not have the same number of days. This adjustment of time variation is done by dividing each monthly total by daily average, it is then multiplied by $\frac{365}{12}$ which is the average number of days in a month.
- **Population changes**: Adjustment for population change becomes necessary when a variable is affected by change in population. If we are studying National Income figures such adjustment is necessary. In this case, adjustment is to divide the income by the number of persons concerned. Then we can have per capita income figures.
- **Price changes**: Adjustment for price changes becomes necessary wherever we have real value changes. Current values are to be deflated by the ratio of current prices to base year prices.
- **Comparability**: In order to have valid conclusion the data which are being analyzed should be comparable. When we are dealing with the analysis of time series it involves the data relating to past which must be homogeneous and comparable. When we are dealing with the analysis of time series it involves the data relating to past which must be homogenous and comparable. Therefore, effects should be there to make the data as homogeneous and comparable as possible.

Example: The sale of commodity in January is 6000 units and sale in February is 5000 units. Here we can't make comparison, since January contains 31 days and February contains 28 days: yet we need to adjust these data to make comparison. In order to make the comparisons we have two possibilities:

Possibility I (changing the Sale of February in terms of January):

That is February data- $5000 \quad \frac{31}{28} = 5535.71$, since February has 28 days and this is compared with 6000 units.

Possibility II (changing the Sale of January in terms of February):

January data $6000 \quad \frac{28}{31} = 5419.36$, since January has 31 days and this is then compared with 5000 units.

Therefore, comparison between these two periods could be done after adjustment.

Chapter two

2 Test of randomness

2.1 Introduction of randomness

Randomness implies different meanings in different contexts. Usually a random sample from a population means that we have observed a set of identically and independently distributed random variables from a population with a specified probability model.

The simplest time series is a random model, in which the observations vary around a constant mean, have a constant variance, and are probabilistically independent. In other words, a random time series has not time series pattern. Observations do not trend upwards or downwards, the variance does not increase over time, the observations do not tend to be bigger in some periods than in other periods.

2.2 Statistical test of randomness

2.2.1 Turning Points Test

It is a type of test based on counting the number of turning points. Meaning the number of times there is a local maximum or minimum in the series. A local maximum is defined to be any observation Y_t such that $Y_t > Y_{t-1}$ and also$Y_t > Y_{t+1}$. A converse definition applies to local minimum. If the series really is random, one can work out the expected number of turning points and compare it with the observed value. Count the number of peaks or troughs in the time series plot. A peak is a value greater than its two neighbors. Similarly, a trough is a value less than its two neighbors. The two (peak and trough) together are known as Turning Points.

Now, define a counting variable C, where

$$C_i = \begin{cases} 1, & \text{if } \mathrm{Y_i < Y_{i+1} > Y_{i+2}} \text{ or } \mathrm{Y_i > Y_{i+1} < Y_{i+2}} \\ 0, & ot\ \ er\, wise, \end{cases}$$

Therefore, the number of turning points p in the series is given by $P = \sum_{\mathrm{i=1}}^{\mathrm{N-2}} \mathrm{C_i}$ and then the probability of finding a turning points in N consecutive values is $E(P) = E(C_i) = \frac{2(N-2)}{3}$ and $Var(P) = \frac{16N-29}{90}$.

Test of procedure

1. $\boldsymbol{H_o}: Y_t, t = 1,2,3, \ldots\ldots, n\,, independent\ identically\ distributed\ (test\ of\ random)\ \boldsymbol{Vs}$ $\boldsymbol{H_o}: \boldsymbol{not\ H_o}.\boldsymbol{where}\ Y_t = observation\ \ at\ t$
2. The level of significance ($\alpha = 0.05$)

3. Let p is the number of turning point for the set of observation

   $E(p) = \frac{2(n-2)}{3}$, where n is the number of observation.

   $Var(P) = \frac{16N-29}{90}, W \;\; ere \; p \sim N\big(E(p), Var(P)\big)$

4. Test of statistics, $Z = \frac{p-E(P)}{\sqrt{Var(P)}} \sim N(0,1)$

5. Critical value at $Z\alpha/_2$

6. Decision Rule; Reject $\boldsymbol{H_o}$ if $|Z_{cal}| > Z\alpha/_2$, that means the time series is not independently identically distributed. If $|Z_{cal}| < Z\alpha/_2$, accept $\boldsymbol{H_o}$ this indicates the time series independent identically distributed

NB: Based on the decision rule of reject the null hypothesis if P is not in the interval of:

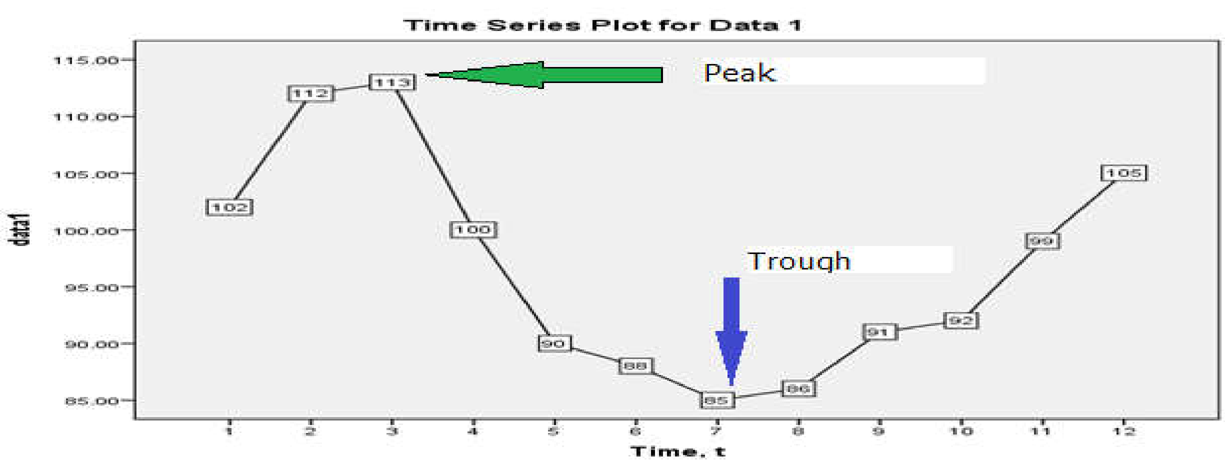
$$\frac{2(N \quad 2)}{3} \pm Z\alpha/_2 \sqrt{\frac{16N \quad 29}{90}}$$

Example-1: consider the following series at time t (in year) and test randomness of the series using turning points.

Time, t	1	2	3	4	5	6	7	8	9	10	11	12
Data 1	102	112	113	100	90	88	85	86	91	92	99	105
Data 2	102	112	88	95	75	103	98	106	98	82	87	92

In order to apply test of randomness for the series by using turning point test first plot the series.

Solutions for data1:



Then from the plot P = 2 and the CI = $\frac{2(12-2)}{3} \pm 2^* \sqrt{\frac{16 \;\; 12-29}{90}}$ = (3.98, 9.36) and 2 is not in the interval and hence we can conclude that the series is not random.

Solutions for data2:

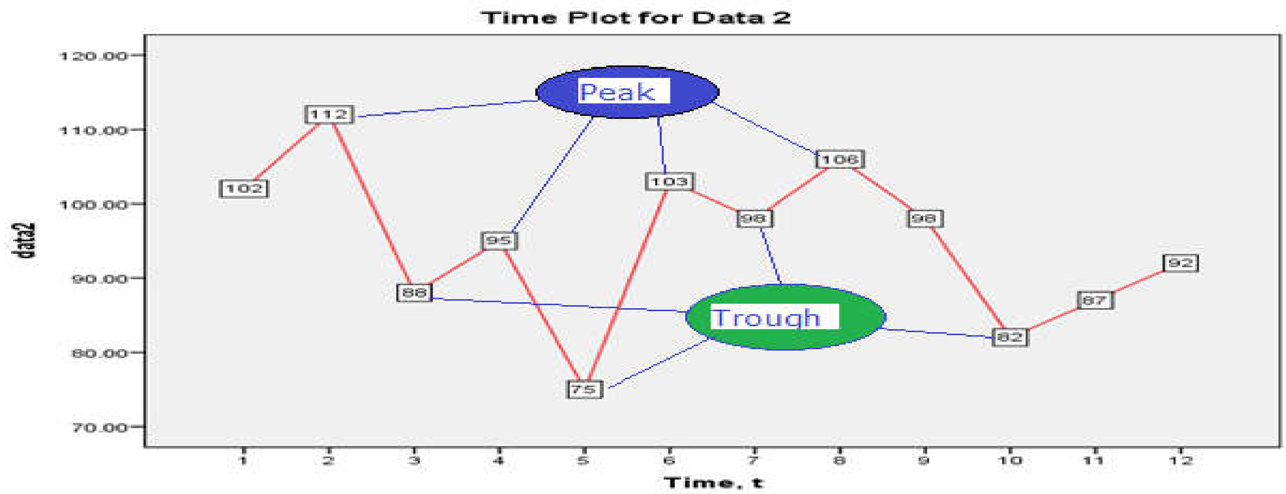


Figure 2.2: Counting turning points

Then from the plot P = 8 and the CI = $\frac{2(12-2)}{3} \pm 2*\sqrt{\frac{16\ 12-29}{90}}$ = (3.98, 9.36) and 8 is in the interval and hence we can conclude that the series is random.

2.2.2 Difference Sign test

This test consists of counting the number of positive first difference of the series, that is to say, the number of points where the series increases (we shall ignore points where there is neither increase nor decrease) with a series of N terms we have N-1 differences. Let us define a variable$Y_i = \begin{cases} 1, & \text{if } Y_{i+1} > Y_i \\ 0, & \text{if } Y_{i+1} < Y_i \end{cases}$, where $i = 1, 2, 3, \dots, N \quad 1$. Then the number of points of increase, say W, is given as $W = \sum_{i=1}^{N-1} Y_i$ and assume it's distribution is normal with mean $E(W) = \frac{(N-1)}{2}$ and variance $Var(W) = \frac{(N+1)}{12}$ as n becomes infinity. This is because $E(W) = \sum_{i=1}^{N-1} Y_i\, p(yi) = \frac{1}{2}\sum_{i=1}^{N-1} Y_i = \frac{(N-1)}{2}andVar(W) = \sum_{i=1}^{N-1} Y_i \left(p(yi) \quad E(W)\right)^2 = \frac{(N+1)}{12}$.

Test of procedure

1. $\boldsymbol{H_o}: Y_t, t = 1,2,3, \dots\dots, n$, *independent identically distributed* (*test of random*) ***Vs***
   $\boldsymbol{H_o}$: ***not*** $\boldsymbol{H_o}$***.where*** $Y_t = observation\ \ at\ t$
2. The level of significance ($\alpha = 0.05$)
3. Let W is the number of positive integers for the set of observation
   $E(w) = \frac{N+1}{2}$, where N is the number of observation.
   $Var(P) = \frac{N-1}{12}, W\ \ ere\ W \sim N\big(E(W), Var(W)\big)$
4. Test of statistics, $Z = \frac{W-E(W)}{\sqrt{Var(W)}} \sim N(0,1)$

5. Critical value at $Z\alpha/_2$
6. Decision Rule; Reject $\boldsymbol{H_o}$ if $|Z_{cal}| > Z\alpha/_2$, that means the time series is not independently identically distributed. If $|Z_{cal}| < Z\alpha/_2$, accept $\boldsymbol{H_o}$ this indicates the time series independent and identically distributed.

NB: Based on the decision rule of reject the null hypothesis if W is not in the interval of:

$$\frac{N+1}{2} \pm Z\alpha/_2 \sqrt{\frac{N \quad 1}{12}}$$

***Example*-2.** Consider the following series and test the randomness of it by applying difference sign test.**[Use 5% SL]**

Time, t	1	2	3	4	5	6	7	8	9	10	11	12
Data 3	35	46	51	46	48	51	46	42	41	43	61	55

Solution: in order to test the randomness of the series first we should find the difference of the series and obtain the number of increasing points.

Time, t	1	2	3	4	5	6	7	8	9	10	11	12
Data 3	35	46	51	46	48	51	46	42	41	43	61	55
differnce	-	11	5	-5	2	3	-5	-4	-1	2	18	-6
Y_i	-	1	1	0	1	1	0	0	0	1	1	0

The $w = \sum_{i=1}^{N-1} Y_i = 6, E(w) = \frac{12-1}{2} = 5.5, Var(w) = \frac{12+1}{12} = 1.083, Zcal = \frac{6-5.5}{\sqrt{1.083}} = 0.45$

and $Z_{0.05/2} = 1.96$. Hence, Z_{cal} is less than Z_{crit} and we retain H_0 at 5% significance level and conclude that the series is statistically random.

***Exercise*-3**: Consider Data 1 and Data2 in the **Above Example-2** and test the randomness of the series by applying difference sign test.

2.2.3 Phase length Test

The phae length test is based on the length of interval between the turing points. A phase is an interval between two turning points (peak and trough or trough and peak). To define a phase of length d from the series we require $d + 3$ points, i. e., a phase of length 1 will requires 4 points, a phase of length 2 will requires 5 points, and the like.

The randomeness hypothesis is tested by comparing the obseved frequence with the expected values. Since the length of the phases are not independent, a slight modification in the chisquuare test is necessary. It is recommended that a three way classification

$d = 1,\ d = 2$ and $d \geq 3$ be tested with 2.5 degree of freedom for the estimated chisquare ≥ 6.3. for the smaller values $\frac{6}{7}\chi^2$(estimated chisquatre) can be tested with two degree of freedom

Consider the $d + 3$ values arranged in increasing order of magnitude. Then the probability of a phase either rising or falling is $\frac{2(d^2+3d+1)}{(d+3)!}$. Now in a series of length n there are $n \quad d \quad 2$ possible phase of length d and the expected number of phases of length is:

$$E(d) = \frac{2(n \quad d \quad 2)(d^2 + 3d + 1)}{(d + 3)!}$$

The phase length test compares the deserved number with the expected number through the χ^2 statistic with a slight modification on the decision rule.

Step1: classify the observed and expected counts of phase length in to three categories as $d = 1, d = 2\ and\ d \geq 3.$

Step2: calculate χ^2 as: $\chi^2 = \sum_{i-1}^{n-1} \frac{[d-E(d)]^2}{E(d)}$

Step3: if $\chi^2_{cal} \geq$ 6.3, compare it with $\chi^2_{\alpha,2.5}$ otherwise compare $\frac{6}{7}\chi^2_{cal}$ with $\chi^2_{\alpha,2}$.

Hypothesis; H_o:Series is random $\boldsymbol{Vs}$ H_1Series is not random

Decision: reject H_o if $c\chi^2_{cal} > \chi^2_{\alpha,df}$, where $c = 1$ or 6/7 and df 2.5 or 2.

Example-4*:* consider data 2 above Example-1 and test the randomness of it by using phase length test [Use 5% level of significance].

Solution*:*

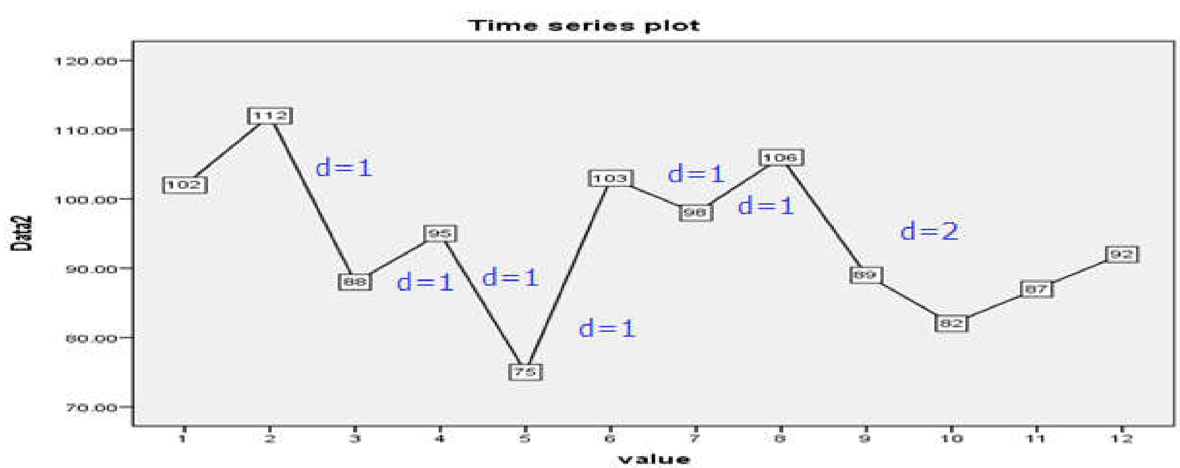


Figure 2.3: Identifying phase length

D	#d	E(d) = $\frac{2(n-d-2)(d^2+3d+1)}{(d+3)!}$
1	6	3.75
2	2	1.467
≥ 3	0	3.333

$$\chi^2 = \sum_{i=1}^{N-1} \frac{[\#d \quad E(d)]^2}{E(d)} = \frac{(6 \quad 3.75)^2}{3.75} + \frac{(1 \quad 1.467)^2}{1.467} + \frac{(0 \quad 0.3694)^2}{0.3694}$$

$$= 1.35 + 0.149 + 3.333 = 1.868 < 6.3$$

and we use $\chi^2_{0.05,2} = 5.991$ and $\frac{6}{7}\chi^2_{cal} = \frac{6}{7} \quad 1.868 = 1.601$.

Decision: since $\frac{6}{7}\chi^2_{cal} < \chi^2_{0.05,2}$ $(i.e., 1.601 < 5.991)$ we do not reject H_o at 5% level of significance and conclude that the series is random.

2.2.4 Rank Test

Given a set of series in which having Trend pattern. From the series let us count the number of cases in which each observation is greater than the previous observation(s), i.e.$Y_t > Y_{t-1}, Y_{t-2}, \ldots, Y_1$. Now let each count be M_t and add up to $M = \sum M_t$, then calculate Kendall's correlation coefficient, $r = \frac{4M}{n(n-1)} \quad 1, \; 1 \leq r \leq 1$.

Hypothesis: series is random (H_0) versus series is not random (H_1). Use $\boldsymbol{r}$ as a test statistic by assuming that it's distribution is normal with mean 0 and variance $\frac{2(2n+5)}{9n(n-1)}$.

Decision: reject H_0 if $|r| > \sqrt{Var(r)} = \sqrt{\frac{2(2n+5)}{9n(n-1)}}$ otherwise retain H_0.

Example5: consider the following series and test the randomness of the series by using rank test.

Time, t	1	2	3	4	5	6	7	8	9	10	11	12
Data 4	10	9	11	10	12	13	12	13	14	12	15	12

Solution:

Time, t	1	2	3	4	5	6	7	8	9	10	11	12	**Total**
Data 4	10	9	11	10	12	13	12	13	14	12	15	12	
M_t	0	0	2	1	4	5	4	6	8	4	10	4	**48**

Therefore, $r = \frac{4M}{n(n-1)} \quad 1 = \frac{4 \; 48}{12(12-1)} \quad 1 = \frac{192}{132} \quad 1 = 0.454545 = 0.45$

and $\sqrt{Var(r)} = \sqrt{\frac{58}{1188}} = 0.22$.

Therefore, since $|r| > \sqrt{\frac{2(2n+5)}{9n(n-1)}}$ (i. e., 0.45 > 0.22) we reject H0 and conclude that the series is not random.

Exercise:

1. In a certain time series there are 56 values with 35 turning points. Then, test the randomness of the series using turning points method. (Use $\alpha = 0.05$).
2. Consider Q1 and let the series has 34 phases with a phase-length of 1, 2 and ≥ 3 are 23, 7 and 4 respectively. So, what is your decision about the randomness of the series if you apply phase-length test? (Use $\alpha = 0.05$)
3. Test the randomness of the series by using difference sign test to distinguish the randomness of the series having 73 observations and $w = 35$. (Use $\alpha = 0.05$)
4. Test the randomness of the series by using rank test to distinguish the randomness of the series having 73 observations and the sum of positive differences are 39.
5. Annual U.S. Lumber Production Consider the annual U.S. lumbers production from 1947 through 1976. The data were obtained from U.S. Department of Commerce Survey of Current Business. The 30 observations are listed in Table below.

35404	36762	32901	38902	37515
37462	36742	36356	37858	38629
32901	33385	32926	32926	32019
33178	34171	35697	35697	35710
34449	36124	34548	34548	36693
38044	38658	32087	32087	37153

Test the randomness of the above data using

a. Rank test
b. Difference sign test
c. Turning point test
d. Phase length test

6. Consider the following data and test whether the time series is random or not. Repeat question number (a-b)

Year	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983
Index	157	161	149	142	125	129	135	127	130	129
Year	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993
Index	138	132	136	152	171	169	215	258	219	255

Chapter Three

3 Estimation of Trend Component

3.1 Introduction

Trend estimation deals with the characterization of the underlying, or long–run, evolution of a time series. In order to measure trend, we are to eliminate seasonal, cyclical and the irregular components from the time series data. For short series it is assumed that the trend and the cyclical components are inextricably mixed up. Hence, either we assumed that the seasonal components have already eliminated out of the series or even if these are presents we are to find methods that smooth out the seasonal effects while determining the trend value. There are different methods of estimations of trend value.

i) Free-hand method

ii) Method of semi-averages

iii) Least squares method

iv) Moving average method

v) Exponential smoothing method

A few time series methods such as freehand curves and moving averages simply describe the given data values, while other methods such as semi-average and least squares help to identify a trend equation to describe the given data values.

3.2 Constant means model and its estimations

The constant mean model can estimated by using Free-hand Method, Method of semi-averages, Least squares method, moving average, Exponential smoothing method.

This model assumes that the data vary randomly around the mean which is a time series generated by a constant mean and random error. $i.e\ X_t = \mu + \varepsilon_t$

Where ε_t = random error of time series and μ = Constant over all mean

Assumption:

- $E(\varepsilon_t) = 0$
- $E(\varepsilon_t, \varepsilon_{t-j}) = 0, for\ j \neq 0$
- $E(X_t) = \mu$
- $V(X_t) = \sigma^2$

3.2.1 Freehand method

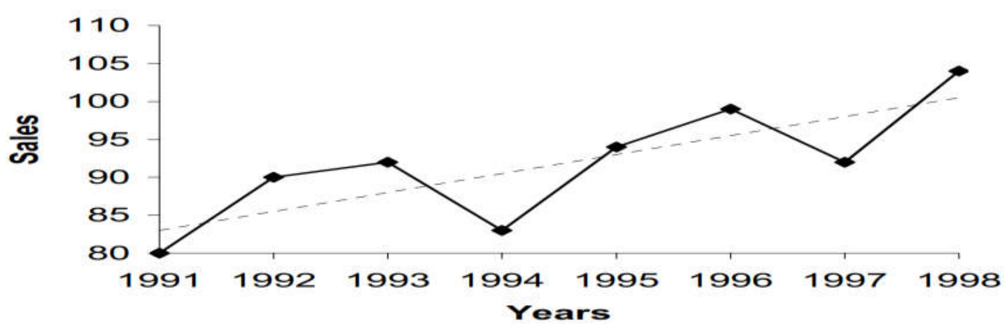
A freehand curve drawn smoothly through the data values is often an easy and, perhaps, adequate representation of the data. The forecast can be obtained simply by extending the trend line. A trend line fitted by the freehand method should conform to the following conditions:

- The trend line should be smooth- a straight line or mix of long gradual curves.
- The sum of the vertical deviations of the observations above the trend line should equal the sum of the vertical deviations of the observations below the trend line.
- The sum of squares of the vertical deviations of the observations from the trend line should be as small as possible.
- The trend line should bisect the cycles so that area above the trend line should be equal to the area below the trend line, not only for the entire series but as much as possible for each full cycle.

Example: Fit a trend line to the following data by using the freehand method.

Year	1991	1992	1993	1994	1995	1996	1997	1998
Sales turnover	80	90	92	83	94	99	92	104

Solution: Presents the graph turnover from 1991 to 1998.



Limitations of freehand method

- This method is highly subjective because the trend line depends on personal judgement and therefore what happens to be a good-fit for one individual may not be so for another.
- The trend line drawn cannot have much value if it is used as a basis for predictions.
- It is very time-consuming to construct a freehand trend if a careful and conscientious job is to be done.

3.2.2 Method of Semi Average

The semi-average method permits us to estimate the slope and intercept of the trend the quite easily if a linear function will adequately described the data. The procedure is simply to divide the data into two parts and compute their respective arithmetic means. These two points are plotted corresponding to their midpoint of the class interval covered by the respective part and then these points are joined by a straight line, which is the required trend line.

- The arithmetic mean of the first part is the intercept value.
- The slope is determined by the ratio of the difference in the arithmetic mean of the number of years between them, that is, the change per unit time.

The resultant is a time series of the form: $\boldsymbol{\hat{y} = a + bx}$. The $\hat{y}$ is the calculated trend value and *a* and *b* are the intercept and slope values respectively.

Example: Fit a trend line to the following data by the method of semi-average and forecast the sales for the year 2002.

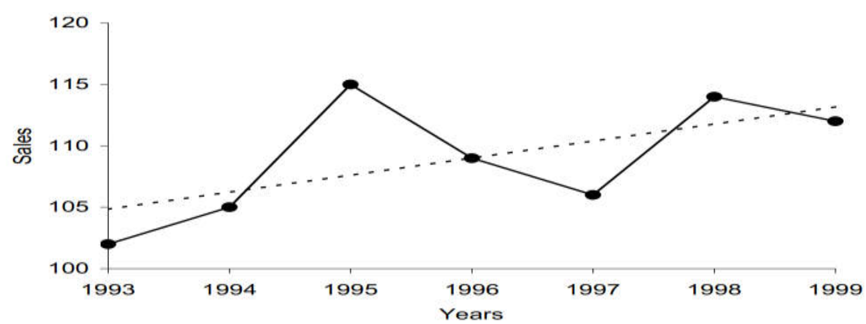
Year	1993	1994	1995	1996	1997	1998	1999
Sales	102	105	114	110	108	116	112

Solution: Since numbers of years are odd in number, therefore divide the data into equal parts (A and B) of 3 years ignoring the middle year (1996). The average of part A and B is

$$\bar{x}_A = \frac{102 + 105 + 114}{3} = \frac{321}{3} = 107$$

$$\bar{x}_B = \frac{108 + 116 + 112}{3} = \frac{336}{3} = 112$$

Part A is centered upon 1994 and part B on 1998. Plot points 107 and 112 against their middle years, 1994 and 1998. By joining these points, we obtain the required trend line as shown below. The line can be extended and be used for prediction.



To calculate the time-series $\boldsymbol{\hat{y} = a + bx}$, we need

$$slope\ b = \frac{\Delta y}{\Delta x} = \frac{c\ \ ange\ in\ sales}{c\ \ ange\ in\ year} = \frac{112 \quad 107}{1998 \quad 1994} = \frac{5}{4} = 1.25$$

Intercept $a = 107\ units\ \ and\ t\ \ e\ trend\ line\ equation\ is$have $\hat{y} = 107 + 1.25x$

Since 2002 is 8 year distant from the origin (1994), therefore we have

$$\hat{y} = 107 + 1.25(8) = 117$$

3.2.3 Least Square Method

This method also called *Ordinary Least Square (OLS)* method: Suppose that all observations from the origin of time through the current period; say $X_1, X_2, X_3, \dots, X_T$ are available. The least squares method criterion is to choose μ so as to minimize sum of the squares of error (SSE); i. e. $SSE = \sum_{t=1}^{T}(X_t \quad \mu)^2$. From $\frac{dSSE}{d\hat{\mu}} = 0$ one can obtain $\mu = \frac{1}{T}\sum X_t$. Therefore, μ is the estimates of trend pattern in the series.

The least squares method criterion is finding the average of all observation in which the sum square of value is minimum.

Suppose $X_T = \mu + \varepsilon_t \qquad \varepsilon_t = X_T \quad \mu$

$$SSE = \sum_{t=1}^{T} \varepsilon_t^2 = \sum_{t=1}^{T}(X_t \quad \mu)^2 \rightarrow s \;\; olud\ be\ minimum$$

$$\frac{dSSE}{d\mu} = \quad 2\sum_{t=1}^{T}(X_t \quad \mu) = 0 \qquad \sum_{t=1}^{T} X_t \quad \sum_{t=1}^{T} \mu \;= 0$$

$$\sum_{t=1}^{T} X_t \quad \sum_{t=1}^{T} \mu \;= 0 \qquad \sum_{t=1}^{T} X_t = T\mu \qquad \mu = \frac{\sum_{t=1}^{T} X_t}{T}$$

Example: Estimate μ for the following time series data by using the least square method.

Year(t)	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999
Value(X_t)	14	15	10	14	17	12	15	11	12	18

Solution: $T = 10$ and $\mu = \frac{\sum_{t=1}^{T} X_t}{T} = \frac{14+15+10+14+17+12+15+11+12+18}{10} = \frac{138}{10} = 13.8$

Exercise: Plot the two series on the some coordinate plane and observe this behavior.

3.2.4 Moving Averages method

- In OLS method, the arithmetic mean includes all past observations of the series with equal weights $\frac{1}{T}$.
- But the unknown parameters μ can change slowly with time and it is reasonable to give more weight to the most recent observations to forecast the future value.
- It assumes that the observations nearby in time are likely to be chose values.
- Thus averaging them provides a reasonable estimate of the trend value at line.
- $\hat{T}_t = M_t = kMA = \frac{1}{k}(X_{t-m} + X_{t-m+1} + \dots + X_t + X_{t+1} + \quad + X_{t+m})$, where $m = \frac{k-1}{2}$ and X_t is the midpoint in the range of k

$$\mu = \frac{\sum_{j=1}^{M} X_{t-j}}{M}, M_t = period\ of\ moving$$

$m = \frac{k-1}{2}$, k is odd and it is a number of data to be include.

Example: Estimating μ Using 3-yearly and 5-year moving averages, determine the trend and short-term-error.

Year (t)	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999
Values(X_t)	14	15	10	14	17	12	15	11	12	18

Solution:

t	Year	Values(X_t)	3-MA	5-MA
1	1990	14	___	____
2	1991	15	13	____
3	1992	10	13	14
4	1993	14	13.67	13.6
5	1994	17	14.33	13.6
6	1995	12	14.67	13.8
7	1996	15	12.67	13.4
8	1997	11	12.67	13.6
9	1998	12	13.67	___
10	1999	18	___	___

$$T = 10 \ and \ M = 3$$

$$M_3 = \frac{X_1 + X_2 + X_3}{3} = \frac{14 + 15 + 10}{3} = 13$$

$$M_4 = \frac{X_2 + X_3 + X_4}{3} = \frac{15 + 10 + 14}{3} = 13$$

$$M_5 = \frac{X_3+X_4+X_5}{3} = \frac{10+14+17}{3} = 13.67$$

$$M_6 = \frac{X_4+X_5+X_6}{3} = \frac{14+17+12}{3} = 14.33$$

$$M_7 = \frac{X_5+X_6+X_7}{3} = \frac{17+12+15}{3} = 14.67$$

$$M_8 = \frac{X_6+X_7+X_8}{3} = \frac{12+15+11}{3} = 12.67$$

$$M_9 = \frac{X_7+X_8+X_9}{3} = \frac{15+11+12}{3} = 12.67$$

$$M_{10} = \frac{X_8+X_9+X_{10}}{3} = \frac{11+12+18}{3} = 13.67$$

$$T = 10 \ and \ M = 5$$

$$M_5 = \frac{X_1 + X_2 + X_3 + X_4 + X_5}{5} = \frac{14 + 15 + 10 + 14 + 17}{5} = 14$$

$$M_6 = \frac{X_2 + X_3 + X_4 + X_5 + X_6}{5} = \frac{15 + 10 + 14 + 17 + 12}{5} = 13.6$$

$$M_7 = \frac{X_3 + X_4 + X_5 + X_6 + X_7}{5} = \frac{10 + 14 + 17 + 12 + 15}{5} = 13.6$$

$$M_8 = \frac{X_4 + X_5 + X_6 + X_7 + X_8 +}{5} = \frac{14 + 17 + 12 + 15 + 11}{5} = 13.8$$

$$M_9 = \frac{X_5 + X_6 + X_7 + X_8 + X_9}{5} = \frac{17 + 12 + 15 + 11 + 12}{5} = 13.4$$

$$M_{10} = \frac{X_6 + X_7 + X_8 + X_9 + X_{10}}{5} = \frac{12 + 15 + 11 + 12 + 18}{5} = 13.6$$

Exercise: Draw the plot of the estimated 3-Period MA, 5MA and compare with Actual value.

Assignment: Discuss briefly Weighted moving averages method of estimation trend component.

3.2.4.1 Centered Moving Average Method

When the period selected for the moving average consists of numbers of$(3, 5,\ 7, 9, 11\ and\ etc)$ there is no problem of centering it. The average obtained is written against into middle of the length of MA period.

But when the period selected for the MA consists of even numbers $(4, 6, 8,\ 10\ and\ etc)$ of years, months or weeks over this difficultly will average the two MAs smoothers.

Formula: $2\quad kMA = \frac{1}{2}(M'_{kMA} + M''_{kMA})$

$$= \frac{1}{2k}([Y_{t-m} + Y_{t-m+1} + \ ... + Y_{2m}] + [Y_{t-m+1} + Y_{t-m+2} ... + Y_{2m} + Y_{2m+1}])$$

$$= \frac{1}{2k}(Y_{t-m} + 2Y_{t-m+1} + 2Y_{t-m+2}\ ... + 2Y_{t-2m} + 2Y_{2m} + Y_{2m+1}])$$

$$= \frac{1}{2k}Y_{t-m} + \frac{Y_{t-m+1} + Y_{t-m+2} ... + Y_{2m}}{k} + \frac{1}{2k}Y_{2m+1}$$

Where m = $\frac{k}{2}$ and k is even number

$$i.e.\quad CMA = \frac{1}{2}\{T_{2.5} + T_{3.5}\}$$

$$T_{2.5} = \frac{1}{4}(X_1 + X_2 + X_3 + X_4) \quad and \quad T_{3.5} = \frac{1}{4}(X_2 + X_3 + X_4 + X_5)$$

$$CMA = \frac{1}{2}\{T_{2.5} + T_{3.5}\} = \frac{1}{2}\left\{\frac{1}{4}(X_1 + X_2 + X_3 + X_4) + \frac{1}{4}(X_2 + X_3 + X_4 + X_5)\right\}$$

$$= \frac{1}{2}\left\{\frac{1}{4}(X_1 + 2X_2 + 2X_3 + 2X_4 + X_5)\right\}$$

$$= \frac{1}{8}\{X_1 + 2X_2 + 2X_3 + 2X_4 + X_5\}$$

t	Year	Values(X_t)	4-MA	CMA
1	1990	14	___	____
2	1991	15	___	____
3	1992	10	13.25	13.625
4	1993	14	14	13.625
5	1994	17	13.25	13.875
6	1995	12	14.5	14.125
7	1996	15	13.75	13.125
8	1997	11	12.5	13.25
9	1998	12	14	___
10	1999	18	___	___

Then

	T	1	2	3	4	5	6	7	8	9	10
	Y_t	14	15	10	14	17	12	15	11	12	18
	3MA	-	13	13	13.7	14.33	14.67	12.67	12.67	13.67	-
	5MA	-	-	14	13.6	13.6	13.8	13.4	13.6	-	-
$\hat{T}_t$	2*4MA	-	-	13.63	13.6	13.875	14.125	13.13	13.25	-	-

Therefore, the estimated trend, $\hat{T}_t$, is either 3MA or 5MA or 2*4MA and the best one is an estimate that have minimum mean square error (MSE) among 3MA, 5MA or 2*4MA.

N.B.: Moving average cannot be calculated for some values at the beginning and at the end of the series. However, we can make an adjustment for those values by taking the average of two values in the previous column or previous series.

Exercise: Why we use centered moving average method than 4-period moving average method?

Example: consider the previous example and adjust the missing values in all kMA.

Solution: $for\ 3MA \quad 14.5 = \frac{14+15}{2},\ \ 15 = \frac{12+18}{2}$

$for\ 5MA \quad 13.75 = \frac{14.5+13}{2}, 13.00 = \frac{13+13}{2}, 13.17 = \frac{12.67+13.67}{2}, 14.335 = \frac{13.67+15}{2}$

$for\ CMA \quad 13.375 = \frac{13.75+13}{2}, 13.500 = \frac{13+14}{2}, 13.385 = \frac{13.6+13.17}{2},\ 13.75 = \frac{13.17+14.34}{2}$

	T	1	2	3	4	5	6	7	8	9	10
	Y_t	14	15	10	14	17	12	15	11	12	18
	3MA	14.5	13	13	13.67	14.33	14.67	12.67	12.67	14	15
$\hat{T}_t$	5MA	13.75	13	14	13.6	13.6	13.8	13.4	13.6	13	14.34
	2*4MA	13.375	13.5	13.625	13.63	13.875	14.13	13.13	13.25	13	13.75

Example: Assume a four-yearly cycle and calculate the trend by the method of moving average from the following data relating to the production of tea in India.

Year	1987	1988	19989	1990	1991	1992	1993	1994	1995	1996
Production	464	515	518	467	502	540	557	571	586	612

Solution: The first 4-year moving average is:

$$MA_{2(3)} = \frac{464+515+518+467}{4} = \frac{1964}{4} = 491$$

$$MA_{3(4)} = \frac{515+518+467+502}{4} = \frac{2002}{4} = 500.5$$

$$MA_{4(5)} = \frac{518+467+502+540}{4} = \frac{2027}{4} = 506.5$$

$$MA_{5(6)} = \frac{467+502+540+557}{4} = \frac{2066}{4} = 516.5$$

$$MA_{6(7)} = \frac{502 + 540 + 557 + 571}{4} = \frac{2170}{4} = 542.5$$
$$MA_{7(8)} = \frac{540 + 557 + 571 + 586}{4} = \frac{2254}{4} = 563.5$$
$$MA_{8(9)} = \frac{557 + 571 + 586 + 612}{4} = \frac{22326}{4} = 581.5$$

Year(t)	Production (X_t)	4-Year Moving Total	4-Year Moving Average	4-Year Moving Average Centered
1987	464	____	____	____
1988	515	____	____	____
		1964	491	____
1989	518			495.75
		2002	500.50	
1990	467			503.62
		2027	506.5	
1991	502			511.62
		2066	516.5	
1992	540			529.5
		2170	542.5	
1993	557			553
		2254	563.5	
1994	571			572
		2326	581.5	____
1995	586	____	____	____
1996	612	____	____	____

3.2.5 Exponential smoothing method

Exponential smoothing is a type of moving-average forecasting technique which weighs past data in an exponential manner so that the most recent data carries more weight in the moving average. Simple exponential smoothing makes no explicit adjustment for trend effects where as adjusted exponential smoothing does take trend effect into account.

This method involves re-estimating the model parameters each period's in order to incorporate the most recent period's observations. Assume that at the end of the period T, we have available the estimate of μ made at the end of previous period, μ_{T-1}, and the current periods actual observation X_t. A reasonable way to obtain the new estimate is to modify the old by some fraction of the forecast error resulting from using the old estimate to forecast X_t in the current period.

The forecast error, $i.e.\ \varepsilon_T = X_T \quad \mu_{T-1}$

Now α if the desired fraction, the new estimate becomes

$$\mu_T = \mu_{T-1} + \alpha(X_T - \mu_{T-1}) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

Let $\mu_T = S_T$ and rewrite the equation (1) as

$$S_T = S_{T-1} + \alpha(X_T - S_{T-1}) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2)$$

Rearranging equation (2) terms we have

$$S_T = \alpha X_T + (1 - \alpha)S_{T-1} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3)$$

The process given by Equation 3 is known as simple exponential smoothing

In effect each smoothed value before the weighted average of the pervious observation where the weights decreased exponentially based on the value α, $0 \leq \alpha \leq 1$.

If $\alpha = 1$, then the pervious observation are ignored entirely.

If $\alpha = 0$, the current observation are ignored entirely.

➔ Simple exponential smoothing requires a starting value, S_0

- If historical data are available, one can use a simple average of the most recent N observation as S_0 (that is to use $S_0 = M_0$)
- If there is no reliable past data available, then some subjective prediction must be made.

Example: Consider the following data estimate μ of constant mean model using the simple exponential smoothing with $\alpha = 0.1$

Year(t)	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999
Value(X_t)	14	15	10	14	17	12	15	11	12	18

Solution: $S_T = \alpha X_T + (1 - \alpha)S_{T-1}$

Let S_0 *be* the average of the last four observations.

$$S_0 = \frac{15 + 11 + 12 + 18}{4} = 14$$

$$S_1 = \alpha X_1 + (1 - \alpha)S_0 = 0.1 \times 14 + 0.9 \times 14 = 14$$

$$S_2 = \alpha X_2 + (1 - \alpha)S_1 = 0.1 \times 15 + 0.9 \times 14 = 14.1$$

$$S_3 = \alpha X_3 + (1 - \alpha)S_2 = 0.1 \times 10 + 0.9 \times 14.1 = 13.69$$

$$S_4 = \alpha X_4 + (1 - \alpha)S_3 = 0.1 \times 14 + 0.9 \times 13.69 = 13.721$$

$$S_{10} = \alpha X_{10} + (1 - \alpha)S_9 = 0.1 \times 18 + 0.9 \times 13.516 = 13.964$$

Year(t)	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999
Value(X_t)	14	15	10	14	17	12	15	11	12	18
S_T	14	14.1	13.69	13.721	14.0489	13.843	13.9587	13.663	13.516	13.964

Exercise: plot the two series on the some coordinate plane and observe this behavior.

Properties of simple exponential smoothing

Consider that $S_T = \alpha X_T + (1 \quad \alpha)S_{T-1} \dots \dots (1)$

$$S_{T-1} = \alpha X_{T-1} + (1 \quad \alpha)S_{T-2} \dots \dots (2)$$

Substitute equation 2 in equation 1 and we get

$$S_T = \alpha X_T + (1 \quad \alpha)[\alpha X_{T-1} + (1 \quad \alpha)S_{T-2}]$$

$$= \alpha X_T + \alpha(1 \quad \alpha)X_{T-1} + (1 \quad \alpha)^2 S_{T-2} \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots (3)$$

$$S_{T-2} = \alpha X_{T-2} + (1 \quad \alpha)S_{T-3} \dots \dots \dots \dots \dots \dots (4)$$

Substitute equation 4 in equation 3 and we get

$$S_T = \alpha X_T + \alpha(1 \quad \alpha)X_{T-1} + (1 \quad \alpha)^2[\alpha X_{T-2} + (1 \quad \alpha)S_{T-3}]$$

$$= \alpha X_T + \alpha(1 \quad \alpha)X_{T-1} + (1 \quad \alpha)^2 X_{T-2} + (1 \quad \alpha)^3 S_{T-3}$$

Upon successive substitutions S_{T-t}, we will ultimately get

$$S_T = \alpha X_T + \alpha(1 \quad \alpha)X_{T-1} + \alpha(1 \quad \alpha)^2 X_{T-2} + \quad + \alpha(1 \quad \alpha)^{T-1}X_1 + (1 \quad \alpha)^T S_0$$

$$= \alpha \sum_{t=0}^{T-1}(1 \quad \alpha)^t X_{T-t} + (1 \quad \alpha)^T S_0$$

For T sufficiently large so that $(1 \quad \alpha)^T S_0$ is close to zero, the exponential smoothing process gives us unbiased estimator for μ. If the average statistics of the samples of equal size give to the population parameter, then that statistic is unbiased estimator. That is

$$E(S_T) = E(\alpha \textstyle\sum_{t=0}^{T-1}(1 \quad \alpha)^t X_{T-t}) = \alpha \sum_{t=0}^{T-1}(1 \quad \alpha)^t E(X_{T-t})$$

$$= \alpha \sum_{t=0}^{T-1}(1 \quad \alpha)^t E(\mu + \varepsilon_t)\, , from\ t\ \ e\ constant\ mean\ model\ X_{T-t} = \mu + \varepsilon_t$$

$$= \alpha \sum_{t=0}^{T-1}(1 \quad \alpha)^t \mu\, , T \to \infty \qquad E(S_T) = \mu, w\ \ ic\ \ \ is\ unbiased\ estimator$$

Generally: The speed at which the older responses are dampened (smoothed) is a function of the value of α. When α is close to 1, dampening is quick and when α is close to 0, dampening is slow. This is illustrated in the table below:

Smoothing constant	Most recent period α	$\alpha(1 - \alpha)^1$	$\alpha(1 - \alpha)^2$	$\alpha(1 - \alpha)^3$	$\alpha(1 - \alpha)^4$
$\alpha = 0.9$	0.9	0.09	0.009	0.0009	0.00009
0.5	0.5	0.25	0.125	0.0625	0.03125
0.1	0.1	0.09	0.081	0.0729	0.06561

Example: Consider the following data set consisting 12 observations taken over time and estimate Trend component at time t by assuming $\hat{X}_0 = 71$ and $\alpha = 0.1$ and 0.5. Which α is appropriate? Why?

T	1	2	3	4	5	6	7	8	9	10	11	12
X_t	71	70	69	68	64	65	72	78	75	75	75	70

Solution:

T	1	2	3	4	5	6	7	8	9	10	11	12	**MSE**
X_t	71	70	69	68	64	65	72	78	75	75	75	70	
α = 0.1	71	70.9	70.71	70.44	69.8	69.32	69.58	70.43	70.88	71.29	71.67	71.5	
α = 0.5	71	70.5	69.75	68.88	66.44	65.72	68.86	73.43	74.21	74.61	74.8	72.4	
Error (0.1)	0	-0.9	-1.71	-2.44	-5.8	-4.32	2.42	7.57	4.12	3.71	3.33	-1.5	**14.1**
Error (0.5)	0	-0.5	-0.75	-0.88	-2.44	-0.72	3.14	4.57	0.79	0.39	0.2	-2.4	**3.78**

Estimation of trend using α =0.5 is better than that of α =0.1. Because the mean square error for α =0.5 is smaller than α =0.1.

3.3 Linear Trend and Its Estimation

A time series that exhibits a trend is a **non-stationary** time series. Modeling and forecasting of such a time series is greatly simplified if we can eliminate the trend. One way to do this is to fit a **regression model** describing the trend component to the data and then subtracting it out of the original observations, leaving a set of residuals that are free of trend. The trend models that are usually considered are the linear trend, in which the mean of X_t is expected to change linearly with time as in $E(X_t) = b_0 + b_1 t$ or as a quadratic function of time $E(X_t) = b_0 + b_1 t + b_1 t^2$ or even possibly as an exponential function of time such as $E(X_t) = b_0 \varepsilon^{b_1 t}$. But here in this section we will consider linear trend and estimate based on three different methods.

3.3.1 Least Square Method

Assume that there are T periods of the data available, say $X_1, X_2, X_3, \dots, X_T$. Denote the estimator of $\hat{b}_o$and $\hat{b}_1$ respectively. The fitted model is $\hat{X}_t = \hat{b}_o + \hat{b}_1 t$ and the random error is obtained by $\varepsilon_t = X_t \quad \hat{X}_t$

$$SSE = \sum_{t=1}^{T} \left(X_t \quad \hat{X}_t \right)^2 = \sum_{t=1}^{T} \left(X_t \quad \hat{b}_o \quad \hat{b}_1 t \right)^2$$

To be the minimum the SSE it is necessary that b_0 and b_1 satisfy the conditions

$$\frac{\partial SSE}{\partial \hat{b}_o} = 0 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.(1)$$

$$\frac{\partial SSE}{\partial \hat{b}_1} = 0 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(2)$$

From equation (1) we get

$$2\sum_{t=1}^{T}\left(X_t \quad \hat{\beta}_o \quad \hat{\beta}_1 t\right) = 0 \qquad \sum_{t=1}^{T}(X_t) = \sum_{t=1}^{T}\left(\hat{\beta}_o + \hat{\beta}_1 t\right)$$

$$\sum_{t=1}^{T}(X_t) = \sum_{t=1}^{T}\widehat{\beta}_o + \sum_{t=1}^{T}\widehat{\beta}_1 t$$

$$\sum_{t=1}^{T}(X_t) = T\widehat{\beta}_o + \widehat{\beta}_1 \sum_{t=1}^{T} t$$

$$\sum_{t=1}^{T}(X_t) = T\widehat{\beta}_o + \widehat{\beta}_1 \frac{T(T \quad 1)}{2} \quad since \sum_{t=1}^{T} t = \frac{T(T+1)}{2}$$

$$T\widehat{\beta}_o = \sum_{t=1}^{T}(X_t) \quad \widehat{\beta}_1 \frac{T(T+1)}{2}$$

$$\hat{\beta}_o = \frac{1}{T}\sum_{t=1}^{T}(X_t) \quad \hat{\beta}_1 \frac{(T+1)}{2} \ldots\ldots\ldots\ldots\ldots.\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(3)$$

From equation (2) we get

$$2\sum_{t=1}^{T}\left(X_t \quad \hat{\beta}_o \quad \hat{\beta}_1 t\right)(t) = 0 \qquad 2\sum_{t=1}^{T}\left(t\, X_t \quad t\hat{\beta}_o \quad \hat{\beta}_1 t^2\right) = 0$$

$$\sum_{t=1}^{T} t\, X_t \quad \widehat{\beta}_o \sum_{t=1}^{T} t \quad \widehat{\beta}_1 \sum_{t=1}^{T} t^2 = 0$$

$$we\ know\ t\ \ at \sum_{t=1}^{T} t = \frac{T(T+1)}{2} \quad and \qquad \sum_{t=1}^{T} t^2 = \frac{T(T+1)(2T+1)}{6}$$

$$\sum_{t=1}^{T} t\, X_t \quad \widehat{\beta}_o \frac{T(T+1)}{2} \quad \widehat{\beta}_1 \frac{T(T+1)(2T+1)}{6} = 0$$

$$\hat{\beta}_1 \frac{T(T+1)(2T+1)}{6} = \sum_{t=1}^{T} t\, X_t \quad \hat{\beta}_o \frac{T(T+1)}{2} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.(4)$$

Substitute equation (3) in to equation (4) as:

$$\hat{\beta}_1 \frac{T(T+1)(2T+1)}{6} = \sum_{t=1}^{T} t\, X_t \quad \frac{T(T+1)}{2}\left(\frac{1}{T}\sum_{t=1}^{T}(X_t) \quad \hat{\beta}_1 \frac{(T+1)}{2}\right)$$

$$\hat{\beta}_1 \frac{T(T+1)(2T+1)}{6} = \sum_{t=1}^{T} t\, X_t \quad \frac{(T+1)}{2}\sum_{t=1}^{T} X_t + \hat{\beta}_1 \frac{(T+1)}{2} x \frac{T(T+1)}{2}$$

$$= \sum_{t=1}^{T} t\, X_t \quad \frac{(T+1)}{2} \sum_{t=1}^{T} X_t + \hat{\beta}_1 \frac{T(T+1)^2}{4}$$

$$\hat{\beta}_1 \frac{T(T+1)(2T+1)}{6} \quad \hat{\beta}_1 \frac{T(T+1)^2}{4} = \sum_{t=1}^{T} t\, X_t \quad \frac{(T+1)}{2} \sum_{t=1}^{T} X_t$$

$$\hat{\beta}_1 \left(\frac{2T(T+1)(2T+1)}{12} \quad \frac{3T(T+1)^2}{12} \right) = \sum_{t=1}^{T} t\, X_t \quad \frac{(T+1)}{2} \sum_{t=1}^{T} X_t$$

$$\hat{\beta}_1 \frac{T(T^2 \quad 1)}{12} = \sum_{t=1}^{T} t\, X_t \quad \frac{(T+1)}{2} \sum_{t=1}^{T} X_t$$

Therefore

$\hat{\beta}_1 = \frac{12}{T(T^2-1)} \sum_{t=1}^{T} t\, X_t \quad \frac{6}{T(T-1)} \sum_{t=1}^{T} X_t$ $\qquad \hat{\beta}_o = \frac{2(2T+1)}{T(T \quad 1)} \sum_{t=1}^{T} (X_t) \quad \frac{6}{T(T-1)} \sum_{t=1}^{T} t\, X_t$

Therefore, $\hat{T}_t = \hat{\beta}_0 + \hat{\beta}_1 t$ is the estimated of linear trend by least square method.

The magnitude of $\hat{\beta}_1$ indicates the trend (or average rate of change) and its sign indicates the direction of the trend. $\hat{\beta}_0$ indicates the value at time $t = 0$.

Example: assume linearity and estimate the trend pattern from the following series by least square method.

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep
Price	3	6	2	10	7	9	14	12	18

Solution:

Month(t)	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Total
Price(X_t)	3	6	2	10	7	9	14	12	18	81
tX_t	3	12	6	40	35	54	98	96	162	506
t^2	1	4	9	16	25	36	49	64	81	285

Therefore,

$$\hat{\beta}_0 = \frac{2(2T+1)}{T(T \quad 1)} \sum_{t=1}^{T=9} X_t \quad \frac{6}{T(T \quad 1)} \sum_{t=1}^{T=9} tX_t = \frac{2(2 \quad 9+1)}{9(9 \quad 1)} \; 81 \quad \frac{6}{9(9 \quad 1)} \; 506 = 0.58$$

And

$$\hat{\beta}_1 = \frac{12}{T(T^2 \quad 1)} \sum_{t=1}^{T=9} tX_t \quad \frac{6}{T(T \quad 1)} \sum_{t=1}^{T=9} X_t = \frac{12}{9(9 \quad 1)^2} 506 \quad \frac{6}{9(9 \quad 1)} 81 = 1.68$$

The fitted model is: $\hat{X}_t = \hat{T}_t = \hat{\beta}_O + \hat{\beta}_1 t = 0.58 + 1.68t$.

Now estimate the error term as:

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep
Price	3	6	2	10	7	9	14	12	18
$\hat{T}_t$	2.26	3.94	5.62	7.30	8.98	10.66	12.34	14.02	15.70
Error	0.74	2.06	-3.62	2.70	-1.98	-1.66	1.66	-2.02	2.30

$$MSE = \frac{1}{T}SSE = 4.89$$

Exercise: Plot the two series on the some coordinate plane and observe these behavior.

3.3.2 Linear Trend estimation using moving average method

Neither the mean of all data nor the moving average of the most recent moving values is able to cope with a significant ***trend***. There exists a variation on the moving average procedure that often does a better job of handling trend. It is called ***Double Moving averages*** for a linear Trend process. It calculates a second moving average from the original moving average, using the same value for moving. As soon as both single and double moving averages are available. Therefore, estimation of trend in a time series which has a linear relationship with t is as follows.

$$Recall\ t\ \ at: \qquad M_T = \frac{X_T + X_{T\ \ 1} + X_{T\ \ 2} + X_{T\ \ 3} + \quad + X_{T\ \ N+1}}{N}$$

$$E(M_T) = E\left(\frac{X_T + X_{T\ \ 1} + X_{T\ \ 2} + X_{T\ \ 3} + X_{T\ \ 4} + X_{T\ \ 5} + \quad + X_{T\ \ N+1}}{N}\right)$$

$$= \frac{E(X_T) + E(X_{T-1}) + E(X_{T-2}) + E(X_{T-3}) + \quad + E(X_{T-N+1})}{N}$$

$$= \frac{1}{N}(b_o + b_1T + b_o + b_1(T \quad 1) + b_o + b_1(T \quad 2) + \quad + b_o + b_1(T \quad N+1))$$

$$Since \quad X_T = b_o + b_1T + \varepsilon_t$$

$$= \frac{1}{N}(b_o + b_1T + b_o + b_1T \quad b_1 + b_o + b_1T \quad 2b_1 + \quad + b_o + b_1T \quad (N \quad 1)b_1)$$

$$= \frac{1}{N}\left(Nb_o + Nb_1T \quad \frac{N(N \quad 1)b_1}{2}\right), \qquad Since\ \frac{(N \quad 1)(N \quad 1+1)}{2} = \frac{N(N \quad 1)}{2}$$

$$\therefore \quad E(M_T) = b_o + b_1T \quad \frac{(N \quad 1)b_1}{2}$$

$$E(M_T) = E(X_T) \quad \frac{(N \quad 1)\hat{b}_1}{2} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots..(1)$$

This last equation tells us that M_T lag behind the observations at time T by an amount equal to $\frac{(N-1)\hat{b}_1}{2}$ and hence it is biased. Thus, M_T is a biased estimator.

Consider a double Moving average, $M_T^{(2)}$

$$M_T^{(2)} = \frac{M_T + M_{T\ \ 1} + M_{T\ \ 2} + M_{T\ \ 3} + \quad + M_{T\ \ N+1}}{N}$$

$$E(M_T^{(2)}) = E\left(\frac{M_T + M_{T-1} + M_{T-2} + M_{T-3} + \cdots + M_{T-N+1}}{N}\right)$$

$$= \frac{E(M_T) + E(M_{T-1}) + E(M_{T-2}) + E(M_{T-3}) + \cdots + E(M_{T-N+1})}{N}$$

$$= \frac{1}{N}\left(b_o + b_1 T - \frac{(N-1)b_1}{2} + b_o + b_1(T-1) - \frac{(N-1)b_1}{2} + \cdots + b_o + b_1(T-N+1) - \frac{(N-1)b_1}{2}\right)$$

$$= \frac{1}{N}\left(Nb_o + Nb_1 T - \frac{N(N-1)b_1}{2} - \frac{N(N-1)b_1}{2}\right)$$

$$= \frac{1}{N}(Nb_o + Nb_1 T - N(N-1)b_1)$$

$$\therefore\ E(M_T^{(2)}) = b_o + b_1 T - (N-1)b_1$$

$$E(M_T^{(2)}) = E(X_T) - (N-1)b_1 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots .(2)$$

Solving the equation 1 and 2 the estimator of M_T and $M_T^{(2)}$ simultaneously

$$\begin{cases} M_T = E(X_T) - \frac{(N-1)b_1}{2} \\ M_T^{(2)} = E(X_T) - (N-1)b_1 \end{cases}$$

$$M_T - M_T^{(2)} = (N-1)b_1 - \frac{(N-1)b_1}{2}$$

$$2\left(M_T - M_T^{(2)}\right) = b_1[2(N-1) - (N-1)]$$

$$\hat{b}_1 = \frac{2}{N-1}\left(M_T - M_T^{(2)}\right)$$

$$\hat{b}_o = 2M_T - M_T^{(2)} - \hat{b}_1 T$$

Hence, $\hat{X}_T = \hat{b}_o + \hat{b}_1 T = 2M_T - M_T^{(2)}$.This is the linear moving average model.

Example: Using the above example for the price series, estimate the trend using the linear MA procedure with $N = 3$ (the number of smoother)

Using simple moving average Method

$$M_3 = \frac{X_1 + X_2 + X_3}{3} = \frac{3 + 6 + 2}{3} = 3.67$$

$$M_4 = \frac{X_2 + X_3 + X_4}{3} = \frac{6 + 2 + 10}{3} = 6$$

$$M_5 = \frac{X_3 + X_4 + X_5}{3} = \frac{2 + 10 + 7}{3} = 6.33$$

$$M_6 = \frac{X_4 + X_5 + X_6}{3} = \frac{10 + 7 + 9}{3} = 8.67$$

$$M_7 = \frac{X_5 + X_6 + X_7}{3} = \frac{7 + 9 + 14}{3} = 10$$

$$M_8 = \frac{X_6 + X_7 + X_8}{3} = \frac{9 + 14 + 12}{3} = 11.67$$

$$M_9 = \frac{X_7 + X_8 + X_9}{3} = \frac{14 + 12 + 18}{3} = 14.67$$

Using double moving average method

$M_4^{(2)} = \frac{M_3+M_4+M_5}{3} = \frac{3.67+6+6.33}{3} = 5.33$

$M_5^{(2)} = \frac{M_4+M_5+M_6}{3} = \frac{6+6.33+8.67}{3} = 7$

$M_6^{(2)} = \frac{M_5+M_6+M_7}{3} = \frac{6.33+8.67+10}{3} = 8.33$

$M_7^{(2)} = \frac{M_6+M_7+M_8}{3} = \frac{8.67+10+11.67}{3} = 10.11$

$M_7^{(2)} = \frac{M_6+M_7+M_8}{3} = \frac{10+11.67+14.67}{3} = 12.11$

Month	Period	Price(X_T)	M_T	$M_T^{(2)}$	$\hat{X}_T = 2M_T \quad M_T^{(2)}$
Jan	1	3	-	-	-
Feb	2	6	3.67	-	-
March	3	2	6.00	5.33	6.67
Apr	4	10	6.33	7.00	5.66
May	5	7	8.67	8.33	9.01
June	6	9	10.00	10.11	9.89
July	7	14	11.67	12.11	11.23
Aug	8	12	14.67	-	-
Sep	9	18	-	-	-

3.3.3 Linear Trend estimation using exponential smoothing method

As we previously observed, single smoothing does not excel in following the data when there is a trend. This situation can be improved by the introduction of a second equation with a second constant, β, which must be chosen in conjunction with α.

In simple exponential smoothing we have,

$$S_T = \alpha X_T + (1 \quad \alpha)S_{T-1}$$

$$= \alpha \sum_{t=0}^{T-1} (1 \quad \alpha)^t X_{T-t} + (1 \quad \alpha)^T S_0$$

$$let\ \beta = 1 \quad \alpha\ t\ \ en\ we\ \ \ ave$$

$$S_T = \alpha \sum_{t=0}^{T-1} \beta^t X_{T-t} + \beta^T S_0, Taking\ expected\ value, we\ get$$

$$E(S_T) = \alpha \sum_{t=0}^{T-1} \beta^t E(X_{T-t}) + \beta^T S_0$$

$$= \alpha \sum_{t=0}^{T-1} \beta^t \left(b_o + b_1(T \quad t)\right) + \beta^T S_0$$

$$As\ T \to \infty\ and\ \beta^T \to 0\ we\ obtain,\ E(S_T) = \alpha \sum_{t=0}^{T-1} \beta^t \left(b_o + b_1(T \quad t)\right)$$

$$E(S_T) = \alpha(b_o + b_1 T)\sum_{t=0}^{\infty} \beta^t \quad \alpha b_1 \sum_{t=0}^{\infty} t\beta^t$$

$$= \alpha(b_o + b_1 T)\left(\frac{1}{1 \quad \beta}\right) \quad \alpha b_1 \frac{\beta}{(1 \quad \beta)^2}$$

$$= b_o + b_1 T \quad \frac{\beta}{\alpha} b_1,\ \ since\ \alpha = 1 \quad \beta$$

$$= E(X_T) \quad \frac{\beta}{\alpha} b_1$$

$$= E(X_T) \quad \frac{1 \quad \alpha}{\alpha} b_1 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots . (1)$$

Using geometric series

$$\sum_{k=0}^{\infty} x^k = 1 + x^2 + x^3 + \quad = \frac{1}{1 \quad x}, for\ |x| < 1$$

$$\sum_{k=0}^{\infty} kx^k = 1 + 2x^2 + 3x^3 + \quad = \frac{x}{(1 \quad x)^2}, for\ |x| < 1$$

Now suppose we apply the exponential smoothing operator to the output data of exponential smoothing. The result is

$$S_T^{(2)} = \alpha S_T + (1 \quad \alpha) S_{T-1}^{(2)}$$

By applying the same procedure and using similar argument we can show that,

$$S_T^{(2)} = \alpha S_T + (1 \quad \alpha) S_{T-1}^{(2)}$$

$$= \alpha S_T + \alpha(1 \quad \alpha) S_{T-1} + (1 \quad \alpha)^2 S_{T-2}^{(2)}$$

$$= \alpha \sum_{t=0}^{T-1} (1 \quad \alpha)^t S_{T-t} + (1 \quad \alpha)^T S_0^{(2)}, T \to \infty, (1 \quad \alpha)^T \to 0$$

Using both sides expectation, then we can get

$$E\left(S_T^{(2)}\right) = \alpha \sum_{t=0}^{T-1} (1 \quad \alpha)^t E(S_{T-t}), \qquad let\ (1 \quad \alpha) = \beta$$

$$= \alpha \sum_{t=0}^{T-1} \beta^t \left(b_o + b_1(T \quad t) \quad \frac{\beta}{\alpha} b_1\right)$$

$$= \alpha\left(b_o + b_1 T \quad \frac{\beta}{\alpha} b_1\right) \sum_{t=0}^{T-1} \beta^t \quad \alpha b_1 \sum_{t=0}^{T-1} t\beta^t$$

$$= \alpha\left(b_o + b_1 T \quad \frac{\beta}{\alpha} b_1\right) \frac{1}{1 \quad \beta} \quad \alpha b_1 \frac{\beta}{(1 \quad \beta)^2}$$

$$= b_o + b_1 T \quad \frac{\beta}{\alpha} b_1 \quad \frac{\beta}{\alpha} b_1$$

$$T\ \ erefore \quad E\left(S_T^{(2)}\right) = E(S_T) \quad \frac{\beta}{\alpha} b_1$$

$$= E(X_T) \quad \frac{\beta}{\alpha} b_1 \quad \frac{\beta}{\alpha} b_1$$

$$= E(X_T) \quad 2\frac{\beta}{\alpha}b_1 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots..(2)$$

$E(X_T) \quad 2\frac{\beta}{\alpha}b_1$ From equation $E(S_T)$ and $E\left(S_T^{(2)}\right)$ one can get

$$b_1 = \frac{\alpha}{\beta}\left(E(S_T) \quad E\left(S_T^{(2)}\right)\right)$$

$b_1 = \frac{\alpha}{\beta}\left(E(S_T) \quad E\left(S_T^{(2)}\right)\right)$ Logically b_1 may be estimated from $\hat{b}_1(T) = \frac{\alpha}{\beta}(S_T \quad S_T^{(2)})$ and

$$\hat{b}_o(T) = \hat{X}_T \quad T\hat{b}_1(T)$$

The expected value of X at the end of period T may be obtained from the equation for $E(S_T)$ and $E\left(S_T^{(2)}\right)$ as

$$E(X_T) = E(S_T) + \frac{\alpha}{\beta} \quad \frac{\beta}{\alpha}\left(E(S_T) \quad E\left(S_T^{(2)}\right)\right)$$
$$= 2E(S_T) \quad E\left(S_T^{(2)}\right)$$

Again $\hat{X}_T = 2S_T \quad S_T^{(2)} \Rightarrow$ Linear exponential smoothing method

This procedure may be referred to as the Brown's one parameter linear exponential smoothing.

At time T, we estimate the intercept as

$$\hat{b}_0(T) = \hat{X}_T \quad T\hat{b}_1(T) \quad = 2S_T \quad S_T^{(2)} \quad T\frac{\alpha}{\beta}\left(S_T \quad S_T^{(2)}\right)$$

The initial values S_0 and $S_0^{(2)}$ are obtained from estimates of the coefficients b_0 and b_1 which may be developed through simple linear regression analysis from historical data, but in the absence of historical data estimates $\hat{b}_1$ and $\hat{b}_0$ are subjectively determined.

If the initial estimates $\hat{b}_1(0)$ and $\hat{b}_0(0)$ given above may be solved with $T = 0$ that is

$$S_0 = \hat{b}_0(0) \quad \frac{\alpha}{\beta}\hat{b}_1(0) \;\; and \;\; S_0^{(2)} = \hat{b}_0(0) \quad 2\frac{\alpha}{\beta}\hat{b}_1(0)$$

Example: Apply the Brown's method with $\alpha = 0.2$ to estimate the linear trend for the price series given above.

Solutions: from the OLS method we obtained $\hat{X}_T = 0.58 + 1.68T$,

The initial value may be taken to be $\hat{b}_o(0) = 0.58$ and $\hat{b}_1(0) = 1.68$

$$S_0 = 0.58 \quad \frac{0.8}{0.2} \quad 1.68 = \quad 6.14 \quad and \quad S_0^{(2)} = 0.58 \quad 2\frac{0.8}{0.2} \quad 1.68 = \quad 12.86$$

$Based\ on\ t\ \ ese\ intial\ result\ S_1 = \alpha X_1 + (1 \quad \alpha)S_0$

$$S_1 = \alpha X_1 + (1 \quad \alpha)S_0 = 0.2 \quad 3 + 0.8 \quad (\quad 1.68) = \quad 4.31$$

$$S_2 = \alpha X_2 + (1 \quad \alpha)S_1 = 0.2 \quad 6 + 0.8 \quad (\quad 4.31) = \quad 2.25$$

$$S_9 = \alpha X_9 + (1 \quad \alpha)S_8 = 0.2 \quad 6 + 0.8 \quad (7.09) = 9.27$$

$Likewise \quad S_1^{(2)} = \alpha S_1 + (1 \quad \alpha)S_0^{(2)}$

$$S_1^{(2)} = \alpha S_1 + (1 \quad \alpha)S_0^{(2)} = 0.2 \quad (\quad 4.31) + 0.8(\quad 12.86) = \quad 11.15$$

$$S_2^{(2)} = \alpha S_2 + (1 \quad \alpha)S_1^{(2)} = 0.2 \quad (\quad 2.25) + 0.8(\quad 11.25) = \quad 9.37$$

$$S_9^{(2)} = \alpha S_9 + (1 \quad \alpha)S_8^{(2)} = 0.2 \quad 9.27 + 0.8 \quad 0.71 = 2.42$$

$T\ \ us \quad \hat{X}_T = 2S_T \quad S_T^{(2)}$

$$\hat{X}_1 = 2S_1 \quad S_1^{(2)} = 2 \quad (\quad 4.31) \quad (\quad 11.15) = 2.53$$

$$\hat{X}_1 = 2S_1 \quad S_1^{(2)} = 2 \quad (\quad 2.25) — 9.37 = 4.87$$

$$\hat{X}_9 = 2S_9 \quad S_9^{(2)} = 2 \quad 9.27 \quad 2.24 = 16.12$$

Summary:

Month	Period	Price(X_t)	S_T	$S_T^{(2)}$	$\hat{X}_t = 2S_T \quad S_T^{(2)}$
Jan	1	3	-4.31	-11.15	2.53
Feb	2	6	-2.25	-9.37	4.87
march	3	2	-0.72	-7.64	6.20
Apr	4	10	1.42	-5.83	8.67
May	5	7	2.54	-4.16	9.24
June	6	9	3.83	-2.56	10.22
July	7	14	5.86	-0.88	12.60
Aug	8	12	0.27	0.71	13.47
Sep	9	18	9.27	2.42	16.12

3.4 Nonlinear trend and its Estimation

A linear trend equation is used when the data are increasing (or decreasing) by equal amounts but a nonlinear trend equation is used when the data are increasing (or decreasing) by increasing amounts over time. When data increase (or decrease) by equal percents or proportions plot will show curvilinear pattern. If the trend is not linear but rather the increases tend to be a constant percent, the X_t values are converted to logarithms, and a least squares equation is determined using the log transformation.

Many times the line which draw by "Least Square Method" is not prove 'Line of best fit' because it is not present actual long term trend So we distributed Time Series in sub- part and make following equation:-

$$\boldsymbol{X_t = a + bt + ct^2}$$

- Quadratic trend: $X_t = b_0 + b_1 t + b_2 t^2 + \varepsilon_t$ Polynomial trend of order 2
- Polynomial trend of order k: $X_t = b_0 + b_1 t + b_2 t^2 + \quad + b_k t^k + \varepsilon_t$ Polynomial of degree K.
- Simple Exponential trend: This trend indicates the annual growth rate. The equation of logistic trend is $X_t = ab^t$

By applying logarithm transformation we will have: $\log X_t = \log a + t \log b$

If this equation is increase up to second degree then it is **Parabola of second degree** and if it is increase up to third degree then it **Parabola of third degree**. There are three constant $a, b, c, d, \ldots$ Its are calculated by following three equations:-

Parabolic second degree

$$\sum x = Ta + b\sum t + c\sum t^2 \quad , \qquad \sum tx = a\sum t + b\sum t^2 + c\sum t^3$$

$$\sum t^2 x = a\sum t^2 + b\sum t^3 + c\sum t^4$$

If we take the deviation from 'Mean year' then the all three equation are presented like this:

$$\sum X = Ta + c\sum t^2, \quad \sum tX = b\sum t^2 \quad , \qquad \sum t^2 X = a\sum t^2 + c\sum t^4$$

Example: draw a parabola of second degree from the following data

Year	1992	1993	1994	1995	1996
Production	5	7	4	9	10

Solution

Year	Production (X)	Dev. from middle year(t)	tY	t^2	t^2Y	t^3	t^4	Trend value $X_t = a + bt + ct^2$
1992	5	-2	-10	4	20	-8	16	5.7
1993	7	-1	-7	1	7	-1	1	5.6
1994	4	0	0	0	0	0	0	6.3
1995	9	1	9	1	9	1	1	8
1996	10	2	20	4	40	8	16	10.5
Total	35	0	12	10	76	0	34	

We take deviation from middle year so the equations are as below

$$\sum x = Ta + c\sum t^2 \qquad \sum tx = b\sum t^2 \qquad \sum t^2x = a\sum t^2 + c\sum t^4$$

Now we put the value of $\sum t, \sum x,\ \sum tx,\ \sum t^2,\ \sum t^2x,\ \sum t^3,\ \sum t^4,\ T$

$$35 = 5a + 10c \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (i)$$
$$12 = 10b \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (ii)$$
$$76 = 10a + 34c \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (iii)$$

From equation (ii) we get $b = \frac{12}{10} = 1.2$

Equation (ii) is multiply by 2 and subtracted from (iii):

$$10a + 34c = 76 \ldots\ldots\ldots\ldots\ldots (iv)$$
$$10a + 20c = 70 \ldots\ldots\ldots\ldots\ldots (v)$$
$$14c = 6 \ or\ c = \frac{6}{14} = 0.43$$

Now we put the value of c in equation (i)

$$5a + 10\,(0.43) = 35$$
$$5a = 35 \quad 4.3 \qquad 5a = 30.7$$
$$a = 6.14$$

Now after putting the value of $a, b\ and\ c$, Parabola of second degree is made that is:

$$X_t = 6.34 + 1.2t + 0.43t^2$$

<u>Parabola of Third degree</u>:-There are four constant $a, b, c\ and\ d$ which are calculated by following equation. The main equation is $\boldsymbol{X_t = a + bt + ct^2 + dt^3}$. There are also four normal equations.

$$\sum x = Ta + b\sum X + c\sum t^2 + d\sum t^3$$
$$\sum tx = a\sum t + b\sum t^2 + c\sum t^3 + d\sum t^4$$
$$\sum t^2x = a\sum t^2 + b\sum t^3 + c\sum t^4 + d\sum t^5$$
$$\sum t^2x = a\sum t^3 + b\sum t^4 + c\sum t^5 + d\sum t^6$$

Exercise: Using the above example data and draw a parabola of third degree.

CHAPTER FOUR

4 Estimation of Seasonal Component

4.1 Introduction

The seasonal component refers to the variation in the time series that occurs within one year. These movements are more or less consistent from year to year in terms of placement in time, direction and magnitude. Seasonal fluctuations can occur for many reasons, including natural conditions like the weather, administrative factors like the beginning and end of school holidays as well as various social/cultural/religious traditions (for example Christmas holidays). The term seasonal effect is often used to describe the effects of these factors. Some time series are strongly influenced by these effects, while other series are not affected at all.

Economic time series are often affected by events which recur each year at roughly the same time. These time series are said to be influenced by "seasonal effects". For example, major household purchases undertaken prior to the Christmas holiday result in a seasonal effect whereby retail sales increase considerably from October to November and also from November to December. Similarly, the existence of widespread holidays in July contributes to a drop in production from June to July. The magnitude of seasonal fluctuations often complicates the interpretation and analysis of many statistics.

The method employs three steps in the computation of the index

1. Computing an annual moving average of the series being studied
2. Dividing the values of the original data through by the corresponding moving average values and
3. Averaging and adjusting the ratios computed in step 2 to obtain the seasonal index.

Seasonal indexes (SI) are used to adjust a time series for seasonal variation (i.e. remove the seasonal effects on the series) and to aid in short-term forecasting. According to our model, we may remove the effects of seasonal variations from a time series by dividing the original data for the series through by the seasonal index. In symbols we have

$$\frac{Y}{SI} = \frac{TxSxCI}{S} = TxCI$$

- Sales, purchasing, customs of goods and services at New Year.
- Rainfall amount

- Sales of stationary materials monthly
- Temperature records daily.

There are many methods of estimation of seasonal component some of the well known are:

1. Simple average methods
2. Link relative methods
3. Ratio to moving average method
4. Ratio to trend methods
5. Use of dummy variables

4.2 The simple Average Method

This is the easiest and the simplest method of studying seasonal variations. This method is used to the time series variable consists of only the seasonal and random component. The effect of taking average of data corresponding to the same period (say first quarter of each year) is to eliminate the effect of random component and thus, the resulting average consists of only seasonal component. These averages are converted into seasonal indices.

In this method the data for each weekly, monthly, or quarterly observation are expressed as percentages of the average for the year. The percentages for the corresponding weekly, monthly, or quarterly observations of different years are then averaged using either mean or median. If the mean is used it is best to avoid extreme values which may occurs. For monthly data the resulting 12 percentages gives the seasonal index. If the sum of these 12 percentages is not equal to 1200 (mean is not 100) there should be an adjustment by multiplying by a suitable factor.

Steps:-

1. Compute the simple average for each year (the data may involve weekly, monthly, or quarterly observations)
2. Divide each of the given weekly, monthly or quarterly values by the corresponding yearly averages to express the result as a percentage
3. Sort out these percentages by weeks, months or quarters
4. Find the mean percentage of each week, month or quarter
5. Adjust them (the means obtained in step (4) if there mean is not 100 $\left(i.e\ \frac{X}{4} \neq 100\right)$

$$\frac{I \quad 400}{X}, \quad w\ \ ere\ X = t\ \ e\ sum\ of\ seasonal\ index$$

$$I = t\ \ e\ sseasonal\ index\ for\ eac\ \ \ season.$$

Example: Calculate the seasonal indexes using the Simples Average Method for the following data showing the amount of money in million dollars spent on passenger's traveler in country.

Year	Quarter			
	I	II	III	Iv
2005	71	89	106	78
2006	71	90	108	79
2007	73	91	111	81
2009	76	97	122	89

Solution

1. Complete the simple average for each all the year.

Year	Quarter				Yearly average of the data
	I	II	III	IV	
2005	71	89	106	78	$\frac{71+89+106+78}{4}=\frac{344}{4}=86$
2006	71	90	108	79	$\frac{71+90+108+79}{4}=\frac{348}{4}=87$
2007	73	91	111	81	$\frac{73+91+111+81}{4}=\frac{356}{4}=89$
2009	76	97	122	89	$\frac{76+97+122+89}{4}=\frac{384}{4}=96$

2. Divide each of the given quarterly values by the corresponding yearly averages and express the result as a percentage

Year	Quarter			
	I	II	III	IV
2005	$\frac{71}{86}\ 100=82.56$	$\frac{89}{86}\ 100=103.49$	$\frac{106}{86}\ 100123.26$	$\frac{78}{86}\ 100=90.7$
2006	$\frac{71}{87}\ 100=81.61$	$\frac{90}{87}\ 100=103.45$	$\frac{108}{87}\ 100=124.14$	$\frac{79}{87}\ 100=90.8$
2007	$\frac{73}{89}\ 100=82.02$	$\frac{91}{89}\ 100=102.25$	$\frac{111}{89}\ 100=124.72$	$\frac{81}{89}\ 100=91.01$
2009	$\frac{76}{96}\ 100=79.17$	$\frac{97}{96}\ 100=101.04$	$\frac{122}{96}\ 100==127.08$	$\frac{89}{96}\ 100=92.71$

3. Sort out these percentages by quarters and Find the mean percentage of each quarter.

Year	Quarter			
	I	II	III	IV
2005	$\frac{71}{86}\ 100 = 82.56$	$\frac{89}{86}\ 100 = 103.49$	$\frac{106}{86}\ 100123.26$	$\frac{78}{86}\ 100 = 90.7$
2006	$\frac{71}{87}\ 100 = 81.61$	$\frac{90}{87}\ 100 = 103.45$	$\frac{108}{87}\ 100 = 124.14$	$\frac{79}{87}\ 100 = 90.8$
2007	$\frac{73}{89}\ 100 = 82.02$	$\frac{91}{89}\ 100 = 102.25$	$\frac{111}{89}\ 100 = 124.72$	$\frac{81}{89}\ 100 = 91.01$
2009	$\frac{76}{96}\ 100 = 79.17$	$\frac{97}{96}\ 100 = 101.04$	$\frac{122}{96}\ 100 == 127.08$	$\frac{89}{96}\ 100 = 92.71$
Total	325.36	410.23	499.20	365.22
Mean(SI)	81.34	102.56	124.80	91.30

The seasonal indexes are 81.34 for quarter $I, 102.56$ for quarter $II, 124.80$ for quarter III and 91.30 for quarter IV.

Note: Obviously, in this data set each seasonal indices add up to 400 (as if each quarter had 100). What would be done if they sum up to $> or < 400$?

We manipulate each indexes by $\frac{I\ 400}{X}$ where I represent the seasonal index and x represents the sum of the indices.

For instance

- If the sum of the indices is 410 then we will adjust by $\frac{81.4\ 400}{410}$
- If the sum of the indices is 395 then we will adjust by $\frac{81.4\ 400}{395}$
- and so on for what ever X is.

4.3 Link relative method

This method is based on the assumption that the trend is linear and cyclic variations are uniform pattern. The link relatives are percentages of the current period (quarter or month) are compared with the previous period. With the computation of the link relative and their average, the effect of cyclic and the random component are minimized. Further, the trend gets eliminated in this process of adjustment of chain relatives. This method is also known as the Pearson's method. In this method the data for each weekly, monthly, or quarterly observation are expressed as percentages of the data for the previous month. These percentages are called linked relative, since they linked each weekly, monthly, or quarterly observations to the preceding one. Then, an appropriate average of the link relatives for corresponding weekly, monthly, or quarterly is then taken. The following steps are used for Link Relative method.

1. Express each of the weekly, Monthly or quarterly values or a percentage of the value for the previous week, month or quarter (These are called link relatives)

$$Link\ Relative\ for\ any\ period(LR_t) = \frac{Current\ periods\ figure}{Previous\ periods\ figure} x100 == \frac{Y_t}{Y_{t\ \ 1}} x100$$

2. Sort out the link relatives by weeks, months, or quarters and take an appropriate average of them for each week, month or quarter.

   It is calculated as $\overline{LR} = \frac{\sum LR}{S}, w\ \ ere\ S\ is\ t\ \ e\ number\ of\ obsevation\ in\ eac\ \ \ season$

3. Convert these averages into a series of chain relatives by setting the value to the first week, January or the first quarter as 100%, and carrying the process to include the first unit of the next period.

$$CR = \frac{average\ LR\ for\ t\ \ e\ current\ period\ \ \ C\ \ ain\ relative\ of\ t\ \ e\ pervious\ period}{100}$$

$$CR_i = \frac{LR_i\ \ \ CR_{i-1}}{100}, i = 1,2,3, \dots, L, \qquad w\ \ ere\ \ L\ is\ t\ \ e\ lengt\ \ \ of\ season$$

4. A discrepancy due to the trend increment $(+ve\ or - ve)$ exists between the chain relatives for the 1^{st}January or the first quarter and that for the next January or next quarter. Adjust the chain relatives for the effect of trend by subtracting the correction factor $\frac{k}{n}d$ from $(k+1)^{th}$ chain relative respectively

   Where $k\ =\ 1,2, \dots \dots .11$ for monthly and $k\ =\ 1,2,3$ for quarterly data,

   n is number of season

$$d = \frac{1}{100}\begin{pmatrix} average\ Link\ Relative\ for\ t\ \ e\ First\ Seaseon \\ C\ \ ain\ Relative\ for\ t\ \ e\ t\ \ e\ Last\ season \end{pmatrix} \quad 100$$

$$= \frac{Average\ LR\ for\ I(1\ \ \ \ mont\ \ )\ \ \ \ CR\ for\ IV(12\ mont\ \ )}{100} \quad 100$$

5. Reduce the adjusted chain relatives to the same level as the first season.

Example: Obtain the seasonal indices for the travel expense data using link relative method.

Solution: step1- computes Link Relatives

Year	Quarter	Y_t	$LR = \frac{Y_t}{Y_{t-1}} \quad 100$
	1	71	-
	2	89	125.4
	3	106	119.1
2005	4	78	73.6
	1	71	91.0
	2	90	126.8
	3	108	120.0
2006	4	79	73.1
	1	73	92.4
	2	91	124.7
	3	111	122.0
2007	4	81	73.0
	1	76	93.8
	2	97	127.6
	3	122	125.8
2008	4	89	73.0

Step 2: Sort out the link relatives by quarters and take an appropriate average of them for each quarter

Year	Quarter			
	I	II	III	IV
2005	-	125.4	119.1	73.6
2006	91.0	126.8	120	73.1
2007	92.4	124.7	122	73
2008	93.8	127.6	125.8	73
Total	277.2	504.5	486.9	292.7
LR Average	92.4	126.125	121.7	73.2

Step 3: Calculation of chain relatives for each quarter.

- Chain relative for $Quarter \quad I = 100$
- Chain relative for $Quarter \quad II = \frac{126.1 \;\; 100}{100} = 126.1$
- Chain relative for $Quarter \quad I = \frac{121.7 \;\; 126.1}{100} = 153.5$
- Chain relative for $Quarter \quad I = \frac{73.2 \;\; 153.5}{100} = 112.4$

LR Average	92.1	126.1	121.7	73.2
CR	100	126.1	153.5	112.4

Step 4- calculation of adjusted chain relatives and adjusted seasonal indices

To calculate the adjusted chain relatives first we find the correction factor by using this formula $\frac{92.4 \quad 112.4}{100} = 103.8576 \qquad 103.9$ *and* then the correction factor is $103.9 \quad 100 = 3.9$

Since the difference is positive, so, we subtract $\frac{1}{4}$ of 3.9, $\frac{2}{4}$ of 3.9 and $\frac{3}{4}$ of 3.9 from the second, the third and the fourth quarter of chain relative respectively in order to obtain the mean of adjusted chain relatives. Here the mean value of adjusted chain relative is 121.4 and hence adjust it to get adjusted seasonal indices of the series

$$\text{Adjusted } CR\ for\ II = 126.13 \quad \frac{1}{4} \quad 3.9 = 126.13 \quad 0.975 = 125.155$$

$$\text{Adjusted } CR\ for\ III = 153.52 \quad \frac{2}{4} \quad 3.9 = 153.5 \quad 1.95 = 151.55$$

$$\text{Adjusted } CR\ for\ IV = 112.34 \quad \frac{3}{4} \quad 3.9 = 112.3 \quad 2.925 = 109.375$$

Step-5: Reduce the adjusted chain relatives to the same level as the 1^{st} quarter.

The sum of the adjusted chain relative is 486.085. Thus, the sum of the seasonal indices are exceeds 400, so it needs to adjustment;

- $Adjusted\ Seasonal\ Index\ \ for\ I = \frac{100\ x\ 400}{486.085} = 82.29$
- $Adjusted\ Seasonal\ Index\ for\ II = \frac{125.155\ x\ 400}{486.085} = 102.99$
- $Adjusted\ Season\ Index\ for\ III = \frac{151.52\ x\ 400}{486.085} = 124.71$
- $Adjusted\ Seasonal\ Index\ for\ IV = \frac{109.79\ x\ 400}{486.085} = 90.01$

Quarter	I	II	III	IV	Total	Mean
Adjusted CR	100	125.155	151.55	109.38	486.085	121.52
Adjusted S.I.	82.29	102.99	124.71	90.01	400	100.0

Therefore, the seasonal indices for quarter $I, II, III\ and\ IV$ are $82.29, 102.99, 124.71\ and\ 90.01$ respectively. These shows that the effect of the season in quarter I and IV is decreased by 17.71% and 9.99% from the mean while the effect of the season in quarter II and III is increased by 2.99% and 24.71% from the mean respectively.

4.4 The Ratio to Moving Average Methods

The ratio to moving average is the most commonly used method of measuring seasonal variations. This method assumes the presence of all the four components of a time series. The ratio to moving average method is also known as percentage of moving average method. In this method a 12 month moving average is computed, since results thus obtained fall between successive months instead of in the middle of the month as for the original data, we compute a 2 month moving average of this 12 month moving average. The result is often called a centered 12 month moving average.

Assume the multiplicative model $Y_t = T_t \, x \, S_t x \, C_t \, x \, I_t$

A centered 12 month moving average of Y serves to eliminate seasonal and irregular movements S and I and is thus equivalent to values given by TC. Then division of the original data by TC yields SI. The subsequent averages over corresponding months serve to eliminate the irregularity I and thus result in a suitable index S

The steps necessary for determining seasonal variations by this method are

Steps:

1. Smooth the time series or Complete the 12 month or 4 quarter centered Moving Average
2. Divide the Original data for each month or quarter by the corresponding centered moving average &express each result as a percentage.
3. Construct a table containing these % ages by months or quarters and find the monthly or quarterly averages using either the mean or the median.
4. Adjust them if they don't average to 100 using the normalizing constant $\frac{400}{X}$ or $\frac{4}{x}$, where X is the sum of all the monthly or quarterly averages.

Generally:

- If we divide the original data by CMA $\left(\frac{Y_t}{CMA}\right)$ then the method is called Ratio to Moving Average Method for multiplicative model.
- If we subtract CMA from original Y_t CMA then the method is called Ratio to Moving Average Method for additive model.

Example: Let us calculate the seasonal index by the ratio-to-moving average method from the following data: Assume that the model is multiplicative model

Year	Quarter			
	I	II	III	IV
2006	75	60	54	59
2007	86	65	63	80
2008	90	72	66	85
2009	100	78	72	93

Solution: Then calculation of moving average and ratio-to-moving averages are shown in the table below.

Step1: Smoothing the time series data using Center Moving Average method

$$4CMA = \frac{1}{2}\left(\frac{x_1 + x_2 + x_3 + x_4}{4} + \frac{x_2 + x_3 + x_4 + x_5}{4}\right) = \frac{1}{8}(x_1 + 2x_2 + 2x_3 + 2x_4 + x_5)$$

Step 2

Year	Quarter	Y_t	$4 \quad CMA$	$\frac{Y_t}{4 \quad CMA}$
2006	I	75	___	___
	II	60	___	___
	III	54	63.375	85.21
	IV	59	65.375	90.25
2007	I	86	67.125	128.12
	II	65	70.875	91.71
	III	63	74	85.14
	IV	80	75.375	106.14
2008	I	90	76.625	117.46
	II	72	77.625	92.75
	III	66	79.5	83.02
	IV	85	81.5	104.29
2009	I	100	83	120.48
	II	78	84.75	92.04
	III	72	___	____
	IV	93	___	____

$Step$ 3: Construct a table containing these % ages by quarters and find the quarterly averages using either the mean or the median.

Year	Quarter			
	I	II	III	IV
1998	____	_____	85.21	90.25
1999	128.12	91.71	85.14	106.14
2000	117.46	92.75	83.02	104.29
2001	120.48	92.04	_____	_____
Total	336.06	276.5	253.37	300.68
Mean	122.02	92.17	84.46	100.23

Step-4: Adjust them if they don't average to 100

Quarter	I	II	III	IV	Total
Mean	122.02	92.17	84.46	100.23	398.88

$$Adjusted\ SI\ \ for\ \ I = \frac{122.02 \quad 400}{398.88} = 122.36$$

$$Adjusted\ SI\ \ for\ \ II = \frac{92.17 \quad 400}{398.88} = 92.43$$

$$Adjusted\ SI\ \ for\ \ III = \frac{84.46 \quad 400}{398.88} = 84.70$$

$$Adjusted\ SI\ \ for\ \ IV = \frac{100.23 \quad 400}{398.88} = 100.51$$

Therefore, the seasonal indices for quarter I, II, III and IV are $122.36, 92.43, 84.70$ and 100.51 respectively. These shows that the effect of the season in quarter II and III is decreased by 7.57% and 15.3% from the mean while the effect of the season in quarter I and IV is increased by 22.36% and 0.51% from the mean respectively.

Estimation of Seasonal Component for the Additional Model $Y_t = T_t + S_t + C_t + I_t$

Steps:

1. Compute 12 month or 4 quarter centered moving average
2. Subtract the moving average from the actual data
3. Construct the table containing these differences by months or quarter the mean or the median.
4. Adjust them if they don't total zero by the addition or subtraction of a correction factor.

$Correction\ factor(Normalizing\ constant) = Grant\ average\ wit\ \ sign\ reversed.$

$Adjusts\ Seasonal\ Index = Average\ for\ t\ \ e\ mont\ \ or\ quarter + Adjusted\ factor$

Example: Find the seasonal Indexes for the Travel express data by using the moving average method assuming the addition model.

Solution

1. Step1: Compute 4 quarter centered moving average

Step 2

Year	Quarter	Y_t	4 CMA	Y_t CMA
2005	I	71	___	___
	II	89	___	___
	III	106	86	20
	IV	78	86.1	-8.1
2006	I	71	86.5	-15.5
	II	90	86.9	3.1
	III	108	87.2	20.8
	IV	79	87.6	-8.6
2007	I	73	88.1	-15.1
	II	91	88.8	2.2
	III	111	89.4	21.6
	IV	81	90.5	-9.5
2008	I	76	92.6	-16.6
	II	97	95	2
	III	122	___	____
	IV	89	___	____

Step 3: Construct the table containing these differences by quarter the mean.

Year	Quarter			
	I	II	III	IV
2005	____	_____	20	-8.1
2006	-15.5	3.1	20.8	-8.6
2007	-15.1	2.2	21.6	-9.5
2008	-16.6	2	_____	_____
Total	-47.2	7.3	62.4	-26.2
Mean	-15.67	2.43	20.8	-8.73

Here the grand mean can be calculated as follows

$$Total = \quad 15.67 + 2.43 + 20.8 \quad 8.73 = \quad 1.17 \neq 0$$

$$Grand\ mean = \frac{1.17}{4} = \quad 0.2923 \approx \quad 0.3$$

$$Correction\ factor (normalizing\ constant) = Grant\ average\ wit \quad sign\ reversed = 0.3$$

$$Adjusts\ Seasonal\ Index = Average\ for\ t \ \ e\ quarter + Adjusted\ correction\ factor$$

$$Adjusts\ Seasonal\ Index\ \ for\ quarter\ I = \quad 15.7 + 0.3 = \quad 15.4$$

$$Adjusts\ Seasonal\ Index\ \ for\ quarter\ II = 2.4 + 0.3 = 2.7$$

$$Adjusts\ Seasonal\ Index\ \ for\ quarter\ III = 20.8 + 0.3 = 21.1$$

$$Adjusts\ Seasonal\ Index\ \ for\ quarter\ IV = \quad 8.7 + 0.3 = \quad 8.4$$

Interpretation: The average Travel Expense for quarter II is 2.7 Million dollars above the average quarter i.e. the average Travel Expense of Months April, May June of years 2005, 2006, 2007 and 2008 is about 2.7 million Dollars.

4.5 Ratio to Trend Method

This method is also known as percentage to trend method. It is used when the cyclic variations are absent from the data, i.e. the time series variable Y_t consists of trend, seasonal and random components. This method provides seasonal indices free from trend and is an improved version of the simple average method as it assumes that seasonal variation for a given period is a constant fraction of the trend. If we divide the original data by CMA $\left(\frac{Y_t}{CMA}\right)$ then the method is called Ratio to Moving Average Method for multiplicative model. But if the division is done by $(\hat{b}_0 + \hat{b}_1 t)$ $\left(\frac{Y_t}{\hat{b}_0+\hat{b}_1 t}\right)$ then the method is called Ration to Trend Method.

Assume the multiplicative model $Y_t = T_t\ x\ S_t x\ C_t\ x\ I_t$

$$\frac{Y_t}{\hat{b}_0 + \hat{b}_1 t} = \frac{Y_t}{T_t} = S_t x\ C_t\ x\ I_t \qquad Ratio\ to\ Trend\ Met \ \ od$$

Steps:

1. Obtain the trend values by the method of least squares for each period by establishing the trend by fitting a straight line or second degree parabola or a polynomial.
2. Divide each original value (Y_t) by the corresponding trend value and multiply it by 100 to express the results as percentage.
3. Obtain the seasonal indices free from the cyclic and irregular variations, we find average (mean or median) of ratio to trend values (or percentages values) for each season for any number of years.
4. Adjust them if they do not average to 100.

Example: use ratio-to-trend method and estimate the seasonal index for the following series of sales of a certain firm.

Year	Quarter			
	I	II	III	IV
1950	30	40	36	34
1951	34	52	50	44
1952	40	58	54	48
1953	54	76	68	62
1954	88	92	78	82

Solution: For determining the seasonal variation by ratio to trend method, we first calculate the quarterly trend by considering the average quarterly sales for each year. This eliminates the seasonal effect on the trend value. First estimate trend by least square method by giving dummy variables for the average values of the year. Assuming that the given quarterly data correspond to the middle of the quarter, we calculate the trend values as when $x = 0$ which corresponds to January 1, 1952.

1. Obtain the trend values by the method of least squares for each period by establishing the trend by fitting a straight line or second degree parabola or a polynomial.

Year	Total	Average	t	Yt	t^2	$\hat{Y}_t = \hat{b}_0 + \hat{b}_1 t$
1950	140	35	-2	-70	4	32
1951	180	45	-1	-45	1	44
1952	200	50	0	0	0	56
1953	260	65	1	65	4	68
1954	340	85	2	170	82	80
Total		280	0	120	10	

Using the above table data and the least square method of estimation the linear trend parameter

$$\hat{b}_0 = \bar{Y} \quad \hat{b}_1\bar{t} = 56 \; since \sum t = 0 \; and \; \bar{Y} = \frac{\sum Y \; average}{T} = \frac{280}{5}$$

$$\hat{b}_1 = \frac{\sum Yt}{\sum t^2} = \frac{120}{10} = 12$$

$$\hat{Y}_t = \hat{b}_0 + \hat{b}_1 t = 56 + 12t \qquad is \; t \;\; e \; trend \; line$$

The estimated trend at January 1, 1952 is $56 \; + 12 \quad 0 = 56$. This is because the trend increment is 12 in a year, but the trend value of quarter III of year 1952 is $56 + \frac{3}{2} = 57.5$ and the trend value of quarter IV of year 1952 is $57.5 + 3 = 60.5$ and soon. Similarly, the trend value of quarter II of year 1952 is $56 \quad \frac{3}{2} = 54.5$ and for quarter I is $54.5 \quad 3 = 51.5$ and soon. Calculating in this manner we have the quarterly trend values as follows.

Generally for 1951 at January for $Q_3 = 33.5 + \frac{3}{2} = 33.5$, $Q_4 = 33.5 + 3 = 36.5$

$$Q_2 = 32 \quad \frac{3}{2} = 30.5 \qquad Q_1 = 30.5 \quad 3 = 27.5$$

Year	Quarter			
	I	II	III	IV
1950	27.5	30.5	33.5	36.5
1951	39.5	42.5	45.5	48.5
1952	51.5	54.5	57.5	60.5
1953	63.5	66.5	69.5	72.5
1954	75.5	78.5	81.5	84.5

2. Dividing the actual values by the corresponding trend estimates and expressing the result in percentages $\frac{Y_t}{\hat{T}_t} \quad 100$. Due to this we obtain the following results.

Year	Quarter			
	I	II	III	IV
1950	109.1	131.1	107.5	93.1
1951	86.1	122.4	109.9	90.7
1952	77.7	106.4	93.9	79.3
1953	85.0	114.3	97.8	85.5
1954	106.0	117.2	105.5	97.0

3. Obtain the average for each month or quarter for all the years to remove the seasonal effect and adjust these if they do not average to 100.

The average values of estimated trend for each quarter are given below in the table.

For instance $92.8 = \frac{109.1+86.1+77.7+85+106}{5}$ is the estimated season value for quarter I.

Average of averages for season estimate is 100.8 and now seasonal indices for various quarters are obtained by expressing the average percentages of the quarters as the percentages of the overall average is 100.8, i.e., the adjusted seasonal index for quarter I is (92.8/100.8)*100=92.1 and for quarter II is $\frac{118.3}{100.8}$ $100 = 117.4$ and the like.

Quarter	I	II	III	IV	Mean
Average	92.8	118.3	102.9	89.1	100.8
Adjusted S.I.	92.1	117.4	102.1	88.4	100.0

The seasonal indices for quarter I, II, III and IV are $92.1,\ 117.4,\ 102.1$ and 88.4 respectively. Thus the effect of the season in quarter I and IV are decreased by 7.9% and 11.6% from the overall mean (expected mean which is 100) while in quarter II and III are increased by 17.4% and 2.1% from expected mean (100), respectively.

4.6 Uses of Dummy Variables

Seasonally adjusted time series are obtained by removing the seasonal component from the data. Statistician may implement a seasonal adjustment procedure using The Simple Averages Method, Link Relatives Method, Ratio-to-Moving Average Method and Ratio-to-Trend Method and report the deseaonalized time series. Another method for removing the seasonal factor is by the use of dummy variables. The seasonal dummy variables can be created with the number of periods in the seasonal cycle (4 for quarterly data and 12 for monthly data).

- Dummy (binary) categorical variables are used to classify numerical observation by non numerical categories.
- The seasonal adjustment can be estimated by including among the explanatory variables called dummy variables, say Q_1, Q_2, Q_3 quarterly data.
- The quarterly regression model

$$Y_t = b_0 + b_1 t + \beta_1 D_1 + \beta_2 D_2 + \beta_3 D_3 + \varepsilon_t$$

$$w\ \ ere \qquad D_1 = \begin{cases} 1, & if\ t\ \ e\ data\ is\ from\ \ t\ \ e\ 1^{st}\ quarter \\ 0 & \end{cases}$$

$$D_2 = \begin{cases} 1, & if\ t\ \ e\ data\ is\ from\ t\ \ e\ 2^{nd} quarter \\ 0 & \end{cases}$$

$$D_3 = \begin{cases} 1, & if\ t\ \ e\ data\ is\ from\ t\ \ e\ 3^{rd}\ quarter \\ 0 & \end{cases}$$

- The monthly regression model

$$Y_t = b_0 + b_1 t + \beta_1 M_1 + \beta_2 M_2 + \beta_3 M_3 + \quad + \beta_{11} M_{11} + \varepsilon_t$$

$$w\ \ ere \qquad M_1 = \begin{cases} 1, & if\ t\ \ e\ data\ is\ from\ t\ \ e\ 1^{st} mont \\ 0 & \end{cases}$$

$$M_2 = \begin{cases} 1, & if\ t\ \ e\ data\ is\ from\ t\ \ e\ 2^{nd}\ mont \\ 0 & \end{cases}$$

$$M_1 = \begin{cases} 1, & if\ t\ \ e\ data\ is\ from\ t\ \ e\ 1^{st} mont \\ 0 & \end{cases}$$

Generally; for Quarterly data the dummy variable has the form:

		Quarter I	Quarter II	Quarter III	Quarter IV
Year 1	Quarter I	1	0	0	0
	II	0	1	0	0
	III	0	0	1	0
	IV	0	0	0	1
Year 2	Quarter I	1	0	0	0
	II	0	1	0	0
	III	0	0	1	0
	IV	0	0	0	1
Year 3	Quarter I	1	0	0	0
	II	0	1	0	0
	III	0	0	1	0
	IV	0	0	0	1

etc

Example: Use the traveler expense data fit the appropriate model using use of dummy variable

Year	Quarter			
	I	II	III	Iv
2005	71	89	106	78
2006	71	90	108	79
2007	73	91	111	81
2009	76	97	122	89

Solution: either use SPSS, STATA or R-software, we obtain the following estimated model results

1. If we assuming the actual series is in the form of $Y_t = b_0 + b_1 t + \varepsilon_t$

$$\hat{Y}_t = 80.75 + 1.029t \;\; wit \;\; R^2 = 0.099$$

2. If we assuming the actual series is in the form of $Y_t = b_0 + \beta_1 Q_1 + \beta_2 Q_2 + \beta_3 Q_3 + \varepsilon_t$

$$\hat{Y}_t = 81.7 \quad 9Q_1 + 10Q_2 + 30Q_3 \;\; wit \;\; R^2 = 0.922$$

3. If we assuming the actual series is in the form of

$$Y_t = b_0 + b_1 t + \beta_1 Q_1 + \beta_2 Q_2 + \beta_3 Q_3 + \varepsilon_t$$

$$\hat{Y}_t = 73.45 + 0.8t \quad 6.6Q_1 + 11.6Q_2 + 30.8Q_3 \; wit \;\; R^2 = 0.979$$

- $T \;\; erefore \; based \; on \; t \;\; eir \; R^2 \; values \; model \; 3 \; is \; better.$

4.7 Smoothing Methods for Seasonal data (series)

Smoothing is defined as the removing of outliers/removing the estimated trend (detrend)

Assume the model is additive $Y_t = T_t + S_t + C_t + I_t$ to estimate $\hat{Y}_t = \hat{S}_t + \varepsilon_t$

1. Estimate the trend using $CMA = \hat{T}_t$
2. Detrend (remove the trend from the model) $D_t = Y_t \quad CMA \approx S_t + \varepsilon_t$

Assume the model is multiplicative $Y_t = T_t \, x \, S_t x \, C_t \, x \, I_t$ to estimate $\hat{Y}_t = \hat{S}_t \quad \varepsilon_t$

1. Estimate the trend using $CMA = \hat{T}_t$
2. Detrend (remove the trend from the model) $D_t = \frac{Y_t}{CMA} \approx S_t \quad \varepsilon_t$

4.7.1 Deseasonalization of Data

- Deseasonalization defined as removal of seasonal component from the time series data by subtracting and dividing each value in the original series by the corresponding value of the seasonal index for additive and multiplicative model respectively.
- For additive model Deseasonalization the observation as

$$D_{S_t} = Y_t \quad \hat{S}_t = \hat{T}_t + \varepsilon_t$$

- For multiplicative model Deseasonalization the observation as

$$D_{S_t} = \frac{Y_t}{\hat{S}_t} = \hat{T}_t \quad \varepsilon_t$$

Purpose of seasonal index

Computing and studying purpose of the seasonal movement with the objective of:

1. Avoiding or minimizing the consequence of the seasonal effect
2. To smooth out seasonal fluctuations
3. To take advantage of them (to move benefit out of them)

Use of seasonal index

1. To adjust original data for seasonality.
   - This process yields deseasenalized data.
2. For economic forecasting and managerial control

Exercise: briefly discus about the use and purpose of the seasonal indices?

4.8 Estimation of Cyclical and Irregular Components

The analysis of cyclical and irregular influences on data is useful for describing past variations but because of their unpredictable nature; their value in forecasting is very limited. If we ignore the Cycle and Irregular components, since by definition they cannot be predicted, the forecasting model will become $\hat{Y} = TxS$.

There is no complete agreement on the way to best measure cycles or on the causes of cycles. The most generally accepted and used time series model specifies that values of the time series (Y) are the product of all the components i.e. $Y = T\ x\ C\ x\ S\ x\ I$ and since the separation of cycle and irregular movements is difficult and the effect of an irregular factor on the series is similar to that of a cycle the model may be rewritten as $Y = T \times S \times CI$ where the cycle and irregular are combined as a single component. The cycle irregular component can be calculated by dividing the original data through by means of trend and seasonal, thereby eliminating them and leaving only the cycle irregular. In symbols

$$\frac{Y}{TxS} = \frac{TxSxCI}{TxS} = CI$$

There are many methods of estimation of cyclic component. Some of the well known are:

- Residual Method
- References cycle analysis method
- Direct Method
- Harmonic Analysis Method

Exercise: briefly discus about the estimation methods of cyclic component?

4.8.1 Estimation of irregular components for additive and multiplicative model

Assume the model is additive $Y_t = T_t + S_t + C_t + I_t$

1. Estimate the trend using $CMA = \hat{T}_t$
2. Detrend (remove the trend from the model) $D_t = Y_t \quad CMA \approx S_t + \varepsilon_t$
3. Estimate the seasonal component

$$\hat{S}_t = \sum_{i=1}^{L} \frac{D_{t+i}}{L}, \qquad L\ lengt\ \ of\ t\ \ e\ period$$

4. Deseasonalization the observation

$$D_{S_t} = Y_t \quad \hat{S}_t$$

5. Fit the trend curve to estimate the trend $\hat{T}_t$
6. Estimate the irregular component

$$I_t = Y_t \quad \hat{S}_t \quad \hat{T}_t$$

7. In case the cyclic component exists we can estimate it by using $\hat{C}_t = CMA_t$
8. Finally $\quad I_t = Y_t \quad \hat{S}_t \quad \hat{T}_t \quad \hat{C}_t$

Assume the model is Multiplicative $Y_t = T_t \, x \, S_t x \, C_t \, x \, I_t$

1. Estimate the trend using $CMA = \hat{T}_t$
2. Detrend (remove the trend from the model) $D_t = \frac{Y_t}{CMA} \approx S_t \quad \varepsilon_t$
3. Estimate the seasonal component using normalizing the average

$$\hat{S}_t = \frac{\sum_{i=1}^{L} D_{t+1}}{L} = L\,, \;\; L\, lengt \;\; of\; t \;\; e\, period$$

4. Deseasonalization the observation

$$D_{S_t} = \frac{Y_t}{\hat{S}_t}$$

5. Fitting the trend curve to estimate the trend $\hat{T}_t$
6. Estimate the irregular component

$$I_t = \frac{Y_t}{\hat{S}_t \quad \hat{T}_t}$$

7. In case the cyclic component exists we can estimate it by using $\hat{C}_t = CMA_t$
8. Finally

$$I_t = \frac{Y_t}{\hat{S}_t \quad \hat{T}_t \quad \hat{C}_t}$$

Example: Consider the following series and estimate the irregular component using additive model procedures above.

Year	Quarter			
	I	II	III	IV
1992	30	135	96	188
1993	51	156	115	209
1994	70	175	136	228
1995	98	196	175	249
1996	111	215	176	270

Solution: The overall steps for estimation of irregular component are given below in the following table.

$$D_t = Y_t \quad CMA \quad and \quad \hat{S}_t = \sum_{i=1}^{L} \frac{D_{t+i}}{L}$$

$$D_{S_t} = Y_t \quad \hat{S}_t$$

$$I_t = Y_t \quad \hat{S}_t \quad \hat{T}_t \quad \hat{C}_t$$

Year	Quarter	$Col(1)$	$Col(2)$	$Col(3)$	$Col(4)$	$Col(5)$	$Col(6)$	$Col(7)$	$Col(8)$
		T	Y_t	CMA	Y_t CMA	$\hat{S}_t$	D_{S_t}	$\hat{T}_t$	$\hat{I}_t$
1992	I	1	30			-74.3	104.3	104.52	-0.22
	II	2	135			23.7	111.3	109.72	1.58
	III	3	96	114.875	-18.875	-16.2	112.2	114.93	-2.73
	IV	4	188	120.125	67.875	66.8	121.2	120.13	1.07
1993	I	5	51	125.125	-74.125	-74.3	125.3	125.33	-0.03
	II	6	156	130.125	25.875	23.7	132.3	130.54	1.76
	III	7	115	135.125	-20.125	-16.2	131.2	135.74	-4.54
	IV	8	209	139.875	69.125	66.8	142.2	140.94	1.26
1994	I	9	70	144.875	-74.875	-74.3	144.3	146.15	-1.85
	II	10	175	149.875	25.125	23.7	151.3	151.35	-0.05
	III	11	136	155.75	-19.75	-16.2	152.2	156.55	-4.35
	IV	12	228	161.875	66.125	66.8	161.2	161.75	-0.55
1995	I	13	98	169.375	-71.375	-74.3	172.3	166.96	5.34
	II	14	196	176.875	19.125	23.7	172.3	172.16	0.14
	III	15	175	181.125	-6.125	-16.2	191.2	177.36	13.84
	IV	16	249	185.125	63.875	66.8	182.2	182.57	-0.37
1996	I	17	111	187.625	-76.625	-74.3	185.3	187.77	-2.47
	II	18	215	190.375	24.625	23.7	191.3	192.97	-1.67
	III	19	176			-16.2	192.2	198.18	-5.98
	IV	20	270			66.8	203.2	203.38	-0.18

Note: Column (5) is obtained by adjusting the mean of each quarter from column (4) as follows.

Quarter	I	II	III	IV	Mean
Seasonal Mean=Adjusted S.I.	-74.3	23.7	-16.2	65.9	0.0

For instance:

- $seasonal\ index\ for\ \ I = \quad 74.25 = \frac{-74.125\ -74.875\ -71.375-76.625}{4}$
- $seasonal\ index\ for\ \ II = 23.6875 = \frac{25.875+\ 25.125+19.125+24.625}{4}$
- $seasonal\ index\ for\ \ III = \quad 16.21875 = \frac{-18.875-20.125-19.75-6.125}{4}$
- $seasonal\ index\ for\ \ IV = 66.75\ = \frac{67.875+69.125+66..125+63..875}{4}$

Column (7) is obtained by fitting column (6) with column (1) i.e. $T_t = 99.318 + 5.203 \quad t$

Column 8 can be obtained as $col(2) \quad col(5) \quad col(7) = \hat{I}_t = Y_t \quad \hat{S}_t \quad \hat{T}_t$

CHAPTER FIVE

5 Introduction to Box-Jenkins Models

5.1 Introduction

The Box-Jenkins methodology refers to a set of procedures for identifying and estimating time series models within the class of autoregressive integrated moving average (ARIMA) models. ARIMA models are regression models that use lagged values of the dependent variable and/or random disturbance term as explanatory variables. These models rely heavily on the autocorrelation pattern in the data. The three basic ARIMA models for a stationary time series Y_t: Models generalize regression but "explanatory" variables are past values of the series itself and unobservable random disturbances. ARIMA models exploit information embedded in the autocorrelation pattern of the data. Estimation is based on maximum likelihood; not least squares. This method applies to both non-seasonal and seasonal data. In this chapter, we will only deal with non-seasonal data.

Let $Y_1, Y_2, \dots, Y_n$ denote a series of values for a time series of interest. These values are observable Let $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ denote a series of random disturbances. These values are unobservable.

- The three basic Box-Jenkins models for Y_t are:

1. Autoregressive model of order p $[AR(p)]$

   $Y_t = \delta + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \quad + \phi_p Y_{t-p} + \varepsilon_t$ (i.e., Y_t depends on its p previous values)

2. Moving average model of order q $(MA(q))$:

   $Y_t = \delta + \varepsilon_t \quad \theta_1 \varepsilon_{t-1} \quad \theta_2 \varepsilon_{t-2} \qquad \theta_q \varepsilon_{t-q}$ (i.e.,Y_t depends on its q previous random error terms)

3. Autoregressive moving average model of orders p and q ($ARMA(p; q)$):

   $Y_t = \delta + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \quad + \phi_p X_{t-p} + \varepsilon_t \quad \theta_1 \varepsilon_{t-1} \quad \theta_2 \varepsilon_{t-2} \qquad \theta_q \varepsilon_{t-q}$ (i.e., Y_t depends on its p previous values and q previous random error terms)

   - Obviously, AR and MA models are special cases of ARMA.
   - ε_t is typically assumed to be "white noise"; i.e., it is identically and independently distributed with a common mean 0 and common variance σ^2 across all observations.
   - We write $\varepsilon_t \sim iid(0, \sigma^2)$.
   - The white noise assumption rules out possibilities of serial autocorrelation and heteroscedasticity in the disturbances.

5.2 The concept of Stationarity

Stationarity is a fundamental property underlying almost all time series statistical models. Stationarity of data is a fundamental requirement underlying ARIMA and most other techniques involving time series data. Time series is said to be stationary if its mean and variance are constant over time and the value of the covariance between the two periods depends only on the distance or gap or lag between the two time periods and not the actual time at which the covariance is computed. Stationarity means that there is no growth or decline in the data. That data must be roughly horizontal along the time axis. In other words the data fluctuates around a constant mean, independent of time, and the variance of the fluctuation remains essentially constant over time.

A time series Y_t is said to be strictly stationary if the joint distribution of $\{Y_t, Y_2, \dots, Y_n\}$ is the same as that of $\{Y_{1+k}, Y_{2+k}, \dots, Y_{n+k}\}$. That is, when we shift through time the behaviour of the random variables as characterized by the density function stays the same.

A time series Y_t is said to be weakly stationary if it satisfies all of the following conditions:

1. $E(Y_t) = \mu_y$ for all t
2. $Var(Y_t) = E(Y_t - \mu_y)^2 = \sigma_y^2$ for all t
3. $Cov(Y_t, Y_{t-k}) = \gamma_k$ for all t

- A time series Y_t is said to weakly stationary if
  - its mean is the same at every period,
  - its variance is the same at every period, and
  - Its autocovariance with respect to a particular lag is the same at every period.
- A time series Y_t *is said to be* white noise series is stationary if:
  1. $E(\varepsilon_t) = 0$ for all t
  2. $Var(\varepsilon_t) = E(\varepsilon_t - 0)^2 = \sigma^2$ for all t
  3. $Cov(\varepsilon_t, \varepsilon_{t-k}) = 0$ for all t and $k \neq 0$
- A time series Y_t is said to be non stationary or random walk is
  1. $E(Y_t) = t\mu$ for all t
  2. $Var(Y_t) = E\left(Y_t - \mu_y\right)^2 = t\sigma_\varepsilon^2$ for all t
  3. $Cov(Y_t, Y_{t-k}) = (t - k)\sigma_\varepsilon^2$ for all t

5.2.1 Nonstationarity

A non-stationary time series will have a time-varying mean or a time-varying variance or both. Random walk model (RWM): a classical example of nonstationary time series. The term random walk is often compared with a drunkard's walk. Leaving a bar, the drunkard moves a random distance $\boldsymbol{\varepsilon_t}$ at time $\boldsymbol{t}$, and continuing to walk indefinitely, will eventually drift farther and farther away from the bar. The same is said about stock prices. Today's stock price is equal to yesterday's stock price plus a random shock.

5.2.2 Differencing

We know the problems associated with non-stationary time series, the practical question is what to do? To avoid the spurious regression problem that may arise from regressing a non-stationary time series on one or more non-stationary time series, we have to transform non-stationary time series to make them stationary. This procedure is known as differencing. Differencing is a technique commonly used to transform a time series from a non-stationary to stationary by subtracting each data Y_t in a series from its predecess or its lagged values Y_{t-1}. Hence, differencing turns out to be a useful filtering procedure in the study of non-stationary time series. The set of observations $Y_t^{'}s$ that correspond to the initial time period (t) when the measurement was taken is described as a series at level.

One of the most mechanisms removing non stationary time series is differencing. We define the differencing series as the change between the observations in the original series.

$$\left.\begin{aligned} \nabla x_t &= x_t \quad x_{t-1} \\ \nabla x_{t-1} &= x_{t-1} \quad x_{t-2} \\ \nabla x_{t-2} &= x_{t-2} \quad x_{t-3} \end{aligned}\right\} 1^{\text{st}} \text{ order differencing}$$

$$\left.\begin{aligned} \nabla^2 x_t &= x_t \quad 2x_{t-1} + x_{t-2} \qquad i.e\;\; \nabla^2 x_t = \nabla x_t \quad \nabla x_{t-1} \\ \nabla^2 x_{t-1} &= x_{t-1} \quad 2x_{t-2} + x_{t-3} \qquad i.e\;\; \nabla^2 x_{t-1} = \nabla x_{t-1} \quad \nabla x_{t-2} \end{aligned}\right\} 2^{\text{nd}} \text{ order differencing}$$

The first and second order differencing is used to removing the trend effect from the series.

5.2.3 Differencing for seasonal data

A seasonal difference is the difference between an observation and the corresponding observation from the previous year. It can be defined as: $\nabla_s x_t = x_t \quad x_{t-s}$

$$\nabla_4 x_t = x_t \quad x_{t-4},\; 1^{\text{st}} \text{ order differencing for quarter data}$$

$$\nabla_{12} x_t = x_t \quad x_{t-12},\; 1^{\text{st}} \text{ order differencing for monthly data}$$

Second order differencing for seasonal data

$$For\ quarterly\ data \quad \nabla_4^2 x_t = \nabla_4 x_t \quad \nabla_4 x_{t-4}$$

$$= x_t \quad x_{t-4} \quad (x_{t-4} \quad x_{t-8})$$

$$= x_t \quad 2x_{t-4} + x_{t-8}$$

$$For\ montly\ data \quad \nabla_{12}^2 x_t = \nabla_{12} x_t \quad \nabla_{12} x_{t-12}$$

$$= x_t \quad x_{t-12} \quad (x_{t-12} \quad x_{t-24})$$

$$= x_t \quad 2x_{t-12} + x_{t-24}$$

When both seasonal and trend difference are applied together it makes no different which is done first.

$$\nabla_{12} \nabla x_t = \nabla_{12}(x_t \quad x_{t-1}) = (x_t \quad x_{t-1}) \quad (x_{t-12} \quad x_{t-13})$$

$$= (x_t \quad x_{t-12}) \quad (x_{t-1} \quad x_{t-13})$$

$$= x_t \quad x_{t-1} \quad x_{t-12} + x_{t-13}$$

$$= \nabla \nabla_{12} x_t \ \ (commultative)$$

5.2.4 Backward shift operator (Lag operator)

Lag (L) or backshift operator (B) operates on an element of a time series to produce the previous element as current element. Defined by

$$BX_t = X_{t-1}$$

$$B^2 X_t = X_{t-2}$$

$$B^p X_t = X_{t-p}$$

- The relationship between differencing and backward shift (lag) operator are defined as follows:

$$B^2 X_t = B(BX_t) = BX_{t-1} = X_{t-2}$$

$$B^3 X_t = B(B^2 X_t) = BX_{t-2} = X_{t-3}$$

$$\nabla x_t = x_t \quad x_{t-1} = \nabla x_t = x_t \quad Bx_t = (1 \quad B)x_t$$

$$\nabla^2 x_t = x_t \quad 2x_{t-1} + x_{t-2} = x_t \quad 2Bx_t + B^2 x_t$$

$$= (1 \quad 2B + B^2)x_t = (1 \quad B)^2 x_t$$

- Seasonal difference for "S" **span** period to lag(backward shift) operator

$$\nabla_s x_t = x_t \quad x_{t-s} = (1 \quad B^s)x_t$$

$$\nabla_s \nabla x_t = \nabla_s(x_t \quad x_{t-1}) = \nabla_s(1 \quad B)x_t = (1 \quad B)\nabla_s x_t$$

$$= (1 \quad B)(1 \quad B^s)x_t$$

$$= (1 \quad B \quad B^s + B^{s+1})x_t$$

Example: consider the following time series data on certain company applying the backward differencing operator find $\nabla x_t, \nabla^2 x_t, \nabla^3 x_t, \nabla^4 x_t$

Time	1	2	3	4	5	6	7	8	9	10
x_t	8	12	15	19	25	30	34	40	45	51

Solution

Time	x_t	∇x_t	$\nabla^2 x_t$	$\nabla^3 x_t$	$\nabla^4 x_t$
1.	8	___	___	___	___
2.	12	4	___	___	___
3.	15	3	-1	___	___
4.	19	4	1	2	___
5.	25	6	2	1	-1
6.	30	5	-1	-3	-2
7.	34	4	-1	0	3
8.	40	6	2	3	3
9.	45	5	-1	-3	-6
10.	51	6	1	2	5

5.3 ARIMA models

Autoregressive Integrated Moving Average models were popularized by George Box and Gwilym Jenkins in the early 1970s. ARIMA models are a class of linear models that is capable of representing stationary as well as non-stationary time series and ARIMA models do not involve independent variables in their construction. They make use of the information in the series itself to generate forecasts. ARIMA models are regression models that use past values of the series itself and unobservable random disturbances as explanatory variables.

ARIMA is an acronym that stands for Auto-Regressive Integrated Moving Average. Specifically,

- Autoregression(AR). A model that uses the dependent relationship between an observation and some number of lagged observations.
- Integrated (I). The use of differencing of raw observations in order to make the time series stationary.
- Moving Average (MA). A model that uses the dependency between an observation and a residual error from a moving average model applied to lagged observations.

Note that **AR** and **MA** are two widely used linear models that work on stationary time series, and **I** is a preprocessing procedure to stationarize time series if need.

A standard notation is used of $ARIMA(p, d, q)$ where the parameters are substituted with integer values to quickly indicate the specific ARIMA model being used.

- **p** The number of lag observations included in the model, also called the lag order.
- **d** The number of times that the raw observations are differenced, also called the degree of differencing.
- **q** The size of the moving average window, also called the order of moving average.

5.3.1 Moving average process

The moving-average (MA) model is a common approach for modeling univariate time series. The moving-average model specifies that the output variable depends linearly on the **current** and various **past values** of a stochastic term.

Suppose that ε_t be a white noise (purely random) process with mean zero and finite variance σ_ε^2 , and let the process $X_t, t = 0, \pm1, \pm2, \ldots.$ be defined by

$$X_t = \mu + \varepsilon_t \quad \theta_1\varepsilon_{t-1} \quad \theta_2\varepsilon_{t-2} \qquad\quad \theta_q\varepsilon_{t-q}$$

Where $\theta_j, j = 1, 2, 3, \ldots, q$ are real constant with $\theta_q \neq 0$, then the process X_t is said to be Moving Average process with order q, which is usually shortened to $MA(q)$.

A $MA(q)$ process can be described by the following stochastic difference equation,

$$X_t = \mu + \varepsilon_t \quad \theta_1\varepsilon_{t-1} \quad \theta_2\varepsilon_{t-2} \qquad\quad \theta_q\varepsilon_{t-q}$$

With $\theta_q \neq 0$ where ε_t is again a pure random process with zero mean and variance σ^2. Using the lag operator we can also write:

$$X_t \quad \mu = \varepsilon_t \quad \theta_1 L\varepsilon_t \quad \theta_2 L^2\varepsilon_t \qquad\quad \theta_q L^q\varepsilon_t$$

$$X_t \quad \mu = (1 \quad \theta_1 L \quad \theta_2 L^2 \qquad\quad \theta_q L^p)\varepsilon_t$$

$$= \theta(L)\varepsilon_t, \qquad w\ ere\ \theta(L) = 1 \quad \theta_1 L \quad \theta_2 L^2 \qquad\quad \theta_q L^p$$

In the moving average process no problems of convergence, and every finite MA(q) process is stationary, no matter what values are used for θ_j,j=1, 2, 3, …, q

Example: Let X_t be the $MA(2)$ model is given by $X_t = \varepsilon_t \quad \theta_1\varepsilon_{t-1} \quad \theta_2\varepsilon_{t-2}$

For this model $\gamma_0 = \sigma_\varepsilon^2(1 + \theta_1^2 + \theta_2^2) = \sigma_x^2$

$$\rho_k = \begin{cases} \frac{\theta_1 + \theta_1\theta_2}{1+\theta_1^2+\theta_2^2}, & k = 1 \\ \frac{\theta_2}{1+\theta_1^2+\theta_2^2}, & k = 2 \\ 0, & k \geq 3 \end{cases}$$

5.3.2 Autoregressive process

An autoregressive (AR) model is a representation of a type of random process. The autoregressive model specifies that the output variable depends linearly on its own **previous values** and on a **stochastic term**.

Suppose ε_t be a white noise (purely random) process with mean zero and finite variance σ_ε^2 , and let the process $X_t, t = 0, \pm1, \pm2, \ldots$. be defined by

$$X_t = \mu + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \quad + \phi_p X_{t-p} + \varepsilon_t$$

Where $\phi_j, j = 1, 2, 3, \ldots, p$ are constant numbers with $\phi_p \neq 0$, then the process X_t is said to be Autoregressive process with order p, which is usually shortened to $AR(p)$.

Assumption of Autoregressive process:

1. $E(\varepsilon_t) = 0$ for all t
2. $Var(\varepsilon_t) = E(\varepsilon_t, \varepsilon_m) = \begin{cases} \sigma^2, if\ t = m \\ 0, if\ t \neq m \end{cases}$
3. $Cov(\varepsilon_t, \varepsilon_{t-k}) = \gamma_k$ for all t and $k \neq 0$

- An $AR(p)$ process can be described by the following stochastic difference equation,

$$X_t = \mu + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \quad + \phi_p X_{t-p} + \varepsilon_t$$

With $\phi_p \neq 0$ where ε_t is again a pure random process with zero mean and variance σ^2. Using the lag operator we can also write:

$$X_t \quad \phi_1 X_{t-1} \quad \phi_2 X_{t-2} \qquad \phi_p X_{t-p} = \mu + \varepsilon_t$$

$$(1 \quad \phi_1 L \quad \phi_2 L^2 \quad \phi_3 L^3 \qquad \phi_P L^P) X_t = \mu + \varepsilon_t$$

The AR(p) process in is stationary if the stability conditions are satisfied:

1. if the characteristic equation $\lambda^P \quad \phi_1 \lambda^{P-1} \quad \phi_2 \lambda^{P-2} \quad \phi_3 \lambda^{P-3} \qquad \phi_P = 0$ only has roots with absolute values smaller than one
2. If the solutions of the lag polynomial $1 \quad \phi_1 L \quad \phi_2 L^2 \quad \phi_3 L^3 \qquad \phi_P L^P = 0$ only have roots with absolute values larger than one

5.3.3 Autoregressive Moving average process

Autoregressive moving-average ($ARMA$) models provide a parsimonious description of a stationary stochastic process in terms of two polynomials, one for the **autoregressive** and the second for the **moving average**. The model consists of two parts, an autoregressive (AR) part and a moving average (MA) part. The AR part involves regressing the variable on its own past values. The MA part involves modeling the error term as a linear combination of error terms occurring contemporaneously and at various times in the past.

BOX AND JENKINS (1970) where the first, who developed a systematic methodology for identifying and fitting a combination of the AR and MA processes, which were originally investigated by Yule. An ARMA model consists according to his name of two components: the weighted sum of past values (autoregressive component) and the weighted sum of past errors (moving average component). Formally, an ARMA model of order (p, q) can be formulated as follows:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \quad + \phi_p X_{t-p} + \varepsilon_t \quad \theta_1 \varepsilon_{t-1} \quad \theta_2 \varepsilon_{t-2} \qquad \theta_q \varepsilon_{t-q}$$

According to the stationarity condition of both processes, the ARMA process is stationary, if the roots of the polynomial $\phi(Z) = 0$ lie outside unit circle. This is the only condition, as every MA(q) process is stationary. In contrast, an ARMA process is called invertible, if the roots of $\theta(z) = 0$ lie outside the unit circle, this is the only condition as every AR(p) process is invertible.

5.4 Methodological tools for model identification

- The main objective of model identification is to identify an appropriate subclass of model from the general ARIMA family models.
- The most commonly used statistical tools for model identification are:
    - Autocorrelation function (ACF)
    - Partial Autocorrelation function (PACF)

5.4.1 Autocorrelation function,

Autocorrelation refers to the correlation of a time series with its own past and future values. It is sometimes called "serial correlation", which refers to the correlation between members of a series of numbers arranged in time. The degree of dependency in a time series is determined by the magnitude of the autocorrelations that can vary between -1 and 1 with a value of 0 indicating no relationship.

$$r_k = \frac{\sum_{t=1}^{n-k}(x_t \quad \bar{x})(x_{t-k} \quad \bar{x})}{\sum_{t=1}^{n}(x_t \quad \bar{x})^2}$$

Where r_k is the autocorrelation coefficient at lag k

Given N Pairs of observations on two variables X and Y. The correlation coefficient is defined as:

$$\rho_{xy} = \frac{cov(x,y)}{\sigma_y \sigma_x} = \frac{\sigma_{xy}}{\sigma_y \sigma_x} = \frac{\sum(x_i \quad \bar{x})(y_i \quad \bar{y})}{\sqrt{\sum(x_i \quad \bar{x})^2 \sum(y_i \quad \bar{y})^2}}$$

Give N observations $X_1, x_2, x_3, \dots, x_N$ on a descriptive time – series.

We can form $(N \quad 1)$ pairs of observations namely $(x_1, x_2), (x_2, x_3), (x_3, x_4) \dots, (x_{N-1}, x_N)$

$$r_1 = \frac{\sum_{t=1}^{N-1}(x_t \quad \bar{x}_1)(x_{t-1} \quad \bar{x}_1)}{\sum_{t=1}^{n}(x_t \quad \bar{x})^2}, w\ ere\ \bar{x}_1 = \frac{\sum_{t=1}^{N-1} x_t}{N \quad 1}\ and\ \bar{X} = \frac{\sum_{t=1}^{N} x_t}{N}$$

In a similar way, we can define the correlation between observations a distance k apart, which are given by

$$r_k = \frac{\sum_{t=1}^{N-k}(x_t \quad \bar{x})(x_{t-k} \quad \bar{x})}{\sum_{t=1}^{N}(x_t \quad \bar{x})^2}$$

Example: Consider the following set of data: {23: 32 32: 33 32: 88 28: 98 33: 16 26: 33
29: 88 32: 69 18: 98 21: 23 26: 66 29: 89}

- The general formula of autocorrelation function is defined as

$$\gamma_k = \frac{\sum_{t=k+1}^{n}(y_t \quad \bar{y})(y_{t-k} \quad \bar{y})}{\sum_{t=1}^{n}(y_t \quad \bar{y})^2}, w\ ere\ \bar{y} = \frac{\sum_{t=1}^{n} y_t}{n}\quad t\ en$$

$$\bar{y} = \frac{\sum_{t=1}^{n} y_t}{n} = \frac{7.801}{12} = 0.65 \quad and \quad \sum_{t=1}^{n}(y_t \quad \bar{y})^2 = 2.931$$

A. Calculate the lag-one sample autocorrelation of the time series.

$$\gamma_1 = \frac{\sum_{t=2}^{n}(y_t \quad \bar{y})(y_{t-1} \quad \bar{y})}{\sum_{t=1}^{n}(y_t \quad \bar{y})^2} = \frac{0.2896}{2.931} = \quad 0.0988$$

B. Calculate the lag-two sample autocorrelation of the time series.

$$\gamma_2 = \frac{\sum_{t=3}^{n}(y_t \quad \bar{y})(y_{t-2} \quad \bar{y})}{\sum_{t=1}^{n}(y_t \quad \bar{y})^2} = \frac{0.93415}{2.931} = \quad 0.3187$$

C. Calculate the lag-three sample autocorrelation of the time series.

$$\gamma_3 = \frac{\sum_{t=4}^{n}(y_t \quad \bar{y})(y_{t-3} \quad \bar{y})}{\sum_{t=1}^{n}(y_t \quad \bar{y})^2} = \frac{0.29136}{2.931} = 0.0994$$

A graph displaying the γ_k versus the lag k is called the autocorrelation function (ACF)

Properties of γ_k

i) γ_k is unitless

ii) $|\gamma_k| \leq 1$that is $1 \leq \gamma_k \leq 1$

iii) $\gamma_k = \gamma_k\ as\ \gamma_{xy} = \gamma_{yx}$

iv) For a random time series $\gamma_k \sim N(0, \frac{1}{N})$

Exercise:

1. The first ten sample autocorrelation coefficients of 400 random numbers are
   $r_1 = 0.02, r_2 = 0.05, r_3 = \ 0.09, r_4 = 0.08, r_5 = \ 0.02, r_6 = 0.00, r_7 = 0.12, r_8 = 0.06, r_9 = 0.02, r_{10} = \ 0.08$. Is there any evidence of non-randomness?

Solution: Here $= 400$, $\pm\frac{2}{\sqrt{N}} = \pm\frac{2}{\sqrt{400}} = \pm 0.1$. Thus r_7 is just significant but unless there is some physical at lag 7. There is no real evidence of non-randomness, as one expects 1 in 20 values of r_k to be significant when the data are random.

2. Suppose that the correlogram of a time series of 100 observations has
   $r_1 = 0.31, r_2 = 0.37, r_3 = \ 0.05, r_4 = 0.06, r_5 = \ 0.21, r_6 = 0.11, r_7 = 0.08, r_8 = 0.05, r_9 = 0.12, r_{10} = \ 0.01$. Suggest an ARMA model which may be appropriate?

Solution: Here $N = 100$, r_1, r_2 and $r_5$values outside $\pm\frac{2}{\sqrt{N}} = \pm\frac{2}{\sqrt{100}} = \pm 0.2$ are significant. A second order moving average process has an autocorrelation function for this type.

3. The following data shows product sales of certain company in successive week's periods over 2010.

Time	1	2	3	4	5	6	7	8	9	10	11	12
x_t	111	170	243	178	248	202	163	139	120	96	95	53

A. Plot the time series and determine its basic features

B. Compute the autocorrelation function and plot it hence interpret the characteristics of the series.

5.4.2 Partial autocorrelation function

It is the correlation between variables x_t and x_{t+2} after removing the influence x_{t+1}. Here the interdependence is removed from each of x_t and x_{t+2}. Suppose there was a significant autocorrelation between x_t and x_{t+1}. Then there will also be a significant correlation between x_t and x_{t+2}. since they also one time unit apart. Consequently, there will be a correlation between x_{t+1} and x_{t+2} because both are related to x_{t+1}. So to measure the real correlation between x_t and

x_{t+2} we need to take out the effects of the intervening value x_{t+1}. This is what partial autocorrelation does. The partial autocorrelations at lag $1, 2, \ldots$, make up the partial autocorrelation function (PACF)

The partial autocorrelations function at lag k can be denoted as ϕ_{kk} and defined as

$$\phi_{kk} = \frac{Cov(x_t, x_{t+k}/\, x_{t+1},\, x_{t+2},\, x_{t+3}, \ldots,\, x_{t+k-1})}{Var(x_t/\, x_{t+1},\, x_{t+2},\, x_{t+3}, \ldots,\, x_{t+k-1})}$$

The plot of ϕ_{kk} against lag k gives partial autocorrelation function ($PACF$).

If the autocorrelation matrix for stationary time series length k is given matrix (KxK)

$$E(\varepsilon_t, \varepsilon_{t-i}) = E(\varepsilon_{t-i}, \varepsilon_t), i = 1, 2, 3, \ldots, k \quad 1$$

$$\begin{matrix} 1 & \rho_1 & \rho_2 & \ldots & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \ldots & \rho_{k-2} \\ \rho_2 & \rho_1 & 1 & \ldots & \rho_{k-3} \\ \\ \rho_{k-1} & \rho_{k-2} & & & 1 \end{matrix} \quad for\ stationary\ time\ series\ data$$

The autocorrelation function can be denoted in matrix as

$$\Omega = \begin{matrix} 1 & \rho_1 & \rho_2 & \ldots & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \ldots & \rho_{k-2} \\ \rho_2 & \rho_1 & 1 & \ldots & \rho_{k-3} \\ \\ \rho_{k-1} & \rho_{k-2} & & & 1 \end{matrix}$$

$$\rho_k = \frac{\gamma_k}{\gamma_0}, \quad ere \;\; \rho_0 = \frac{\gamma_0}{\gamma_0} = 1\,, \rho_{kk-1} = \frac{\gamma_{k-1}}{\gamma_0} \; and \; \rho_3 = \frac{\gamma_3}{\gamma_0}$$

Then ϕ_{kk} can be computed as

$$\phi_{kk} = \frac{|\Omega_k|}{|\Omega_k|} \; w \;\; ere\ \Omega_k\ is\ \Omega_k\ wit \;\;\; last\ column\ replaced\ by \begin{bmatrix} \rho_1 \\ \rho_2 \\ \\ \rho_k \end{bmatrix}$$

$$\phi_{22} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} = \frac{\rho_2 \quad \rho_1^2}{1 \quad \rho_1^2}$$

Example: compute ρ_1, ρ_2, ρ_3 and ϕ_{22} for the following series

Time	1	2	3	4	5	6	7	8	9	10
x_t	47	64	23	71	38	64	55	41	59	48

Solution: $\bar{x} = \frac{\sum_{t=1}^{10} x_t}{n} = \frac{47+64+23+ \quad +59+48}{10} = \frac{510}{10} = 51$

Time	1	2	3	4	5	6	7	8	9	10
x_t	47	64	23	71	38	64	55	41	59	48
$x_t \quad \bar{x}$	-4	13	-28	20	-13	13	4	-10	8	-3

$$\gamma_0 = \sum_{t=1}^{n} (x_t \quad \bar{x})^2 = 16 + 169 + \quad + 100 + 64 + 9 = 1896$$

$$\gamma_1 = \sum_{t=1}^{n-1} (x_t \quad \bar{x})(x_{t+1} \quad \bar{x}) = (\quad 4)13 + 13(\quad 28) + \quad + 8(\quad 3) = \quad 1497$$

$$\gamma_2 = \sum_{t=1}^{n-2} (x_t \quad \bar{x})(x_{t+2} \quad \bar{x}) = 876$$

Therefore $\rho_0 = \frac{\gamma_0}{\gamma_0} = 1,\ \rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{-1497}{1896} = \quad 0.790,\ \rho_2 = \frac{\gamma_2}{\gamma_0} = \frac{876}{1896} = 0.462$

$$\phi_{22} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} = \frac{\begin{vmatrix} 1 & 0.790 \\ 0.790 & 0.462 \end{vmatrix}}{\begin{vmatrix} 1 & 0.790 \\ 0.790 & 1 \end{vmatrix}} = \frac{0.462 \quad (\quad 0.790)^2}{1 \quad (\quad 0.790)^2} = \quad 0.429$$

Lag k	1	2	3	4	5	6	7	8
ACF	-0.790	0.462	-0.164	-0.123	0.253	-0.227	0.120	-0.037
PACF	-0.790	-0.429	0.061	-0.293	-0.242	0.014	-0.111	-0.264

5.5 Stages of Box-Jenkins methodology

There are three stages in the Box-Jenkins approach to time series analysis and forecasting are identification, estimation and diagnostic checking, and the forecasts themselves.

5.5.1 Model Identification/selection

Model identification can represent the primary goal of the analysis, especially if a researcher is trying to identify the basic underlying process represented in a time series data set, and perhaps link this process with important theoretical underpinnings.

Plot the time series/ACF and examine whether it is stationary. If not, try some transformation and or differencing, until the data seems stationary. Compare ACF and PACF of the time series data and identify the ARIMA model to be used. To judge the significance of autocorrelation and partial autocorrelation, the corresponding sample values may be compared with $\pm\frac{2}{\sqrt{n}}$. Use the principle of parsimony (that is, in its simplest form and it has the fewest parameters).

Parsimony is property of a model where the specification with the fewest parameters capable of capturing the dynamics of a time series is preferred to other representations equally capable of capturing the same dynamics.

1. Transform data using natural log transformation which was found the most appropriate.
2. Removing trend component by using the first order differencing.

3. Removing the seasonal variation by using the first order seasonal differencing.
4. Model identification by plotting ACF and PACF of monthly observations

Generally the behavior of ACF and PACF for ARIMA model is given as follows

Model	ACF	PACF
Autoregressive with order p $X_t = \mu + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \quad + \phi_p X_{t-p} + \varepsilon_t$	Dies down (tails off)	Cuts off after lag p
Moving average with order q $Y_t = \mu + \varepsilon_t \quad \theta_1 \varepsilon_{t-1} \quad \theta_2 \varepsilon_{t-2} \qquad \theta_q \varepsilon_{t-q}$	Cuts off after lag p	Dies down (tails off)
Autoregressive moving average model of orders p and q (ARMA(p,q)): $Y_t = \delta + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \quad + \phi_p X_{t-p} + \varepsilon_t \quad \theta_1 \varepsilon_{t-1} \quad \theta_2 \varepsilon_{t-2} \qquad \theta_q \varepsilon_{t-q}$	Dies down (tails off)	Dies down (tails off)

Key points

- ☞ Cut off means from high value decreasing at once. Tail off means decreasing slowly

Behaviour of Autocorrelation and partial Autocorrelation functions		
Model	ACF	PACF
First order Autoregressive model $X_t = \mu + \phi_1 X_{t-1} + \varepsilon_t$	Dies down in a damped exponential fashion; specifically $\rho_k = \phi_1^k \ for\ k \geq 1$	Cuts off after lag 1
First order Autoregressive model $X_t = \mu + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \varepsilon_t$	Dies down according to a mixture of damped exponentials and/or damped sine waves; specifically: $\rho_1 = \frac{\phi_1}{1-\phi_2}$ $\rho_2 = \phi_2 + \frac{\phi_1^2}{1 \quad \phi_2}$ $\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2}\ for\ k \geq 3$	Cuts off after lag 2
Moving average with order 1 $Y_t = \mu + \varepsilon_t \quad \theta_1 \varepsilon_{t-1}$	Cuts off after lag 1;specifically $\rho_2 = \frac{\theta_1}{1+\theta_1^2}$ $\rho_3 = 0\ , for\ k \geq 2$	Dies down in a fashion dominated by damped exponential decay
Moving average with order 2 $Y_t = \mu + \varepsilon_t \quad \theta_1 \varepsilon_{t-1} \quad \theta_2 \varepsilon_{t-2}$	Cut off after lag 2; specifically $\rho_1 = \frac{\theta_1 + \theta_1 \theta_2}{1+\theta_1^2+\theta_2^2}$ $\rho_2 = \frac{\theta_2}{1+\theta_1^2+\theta_2^2}$ $\rho_3 = 0\ , for\ k \geq 3$	Dies down according to a mixture of damped exponentials and/or damped sine waves
Autoregressive moving average model (*ARMA*(*1,1*)): $Y_t = \delta + \phi_1 X_{t-1} + \varepsilon_t \quad \theta_1 \varepsilon_{t-1}$	Dies down in a damped exponential fashion; specifically: $\rho_1 = \frac{(1 \quad \phi_1 \theta_1)(\phi_1 \quad \theta_1)}{1+\theta_1^2 \quad 2\phi_1 \theta_1}$ $\rho_k = \phi_1 \rho_{k-1}, for\ k \geq 2$	Dies down in a fashion dominated by damped exponential decay

5.5.2 Parameter estimation

After identification of one or more models, Estimate parameters in potential models and Select best model using suitable criterion. It is a way that finding the values of the model coefficient $(\phi_1, \phi_2, \phi_3, \dots, \phi_p \ and \ \ \theta_1, \ \theta_2, \theta_3, \dots, \theta_q)$. There are various methods of estimation that can be used, the most well-known:

- least squares estimation
- Yule-Walker estimation
- Maximum likelihood estimation.

Maximum likelihood estimation of ARMA models

ARMA models are typically estimated using maximum likelihood (ML) estimation assuming that the errors are normal, using either conditional maximum likelihood, where the likelihood of x_t given $x_{t-1}, x_{t-2}, \dots$ is used, or exact maximum likelihood where the joint distribution of $[x_1, x_2, \dots, x_{t-1}, x_t\ ]$ is used.

Conditional Maximum Likelihood

Conditional maximum likelihood uses the distribution of of x_t given $x_{t-1}, x_{t-2}, \dots$ to estimate the parameters of an ARMA. The data are assumed to be conditionally normal, and so the likelihood is

$$f(\,x_t \,|\, x_{t\ \ 1}, x_{t\ \ 2}, \dots, \phi, \theta, \sigma^2) = (2\pi\sigma^2)^{-1/2} exp\left(\frac{-\varepsilon_t^2}{2\sigma^2}\right)$$

$$= (2\pi\sigma^2)^{-1/2} exp\left(\frac{\left(x_t \quad \phi_0 \quad \sum_{i=1}^{p} \phi_i x_{t-i} \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} \quad \phi_0\right)^2}{2\sigma^2}\right)$$

Since the ε_t series is assumed to be a white noise process, the joint likelihood is simply the product of the individual likelihoods,

$$f(\,x_t \,|\, x_{t\ \ 1}, x_{t\ \ 2}, \dots, \phi, \theta, \sigma^2) = \prod_{t=1}^{T} (2\pi\sigma^2)^{-1/2} exp\left(\frac{\varepsilon_t^2}{2\sigma^2}\right)$$

The conditional log likelihood is defined as

$$L(\phi, \theta, \sigma^2, x_t \,|\, x_{t\ \ 1}, x_{t\ \ 2}, \dots,) = \prod_{t=1}^{T} (2\pi\sigma^2)^{-1/2} exp\left(\frac{\varepsilon_t^2}{2\sigma^2}\right)$$

$$= \quad \frac{1}{2}\left(\sum_{t=1}^{T} ln2\pi + \ln\sigma^2 + \frac{\varepsilon_t^2}{2\sigma^2}\right)$$

Using first condition for the mean parameters from a normal log-likelihood does not depend on σ^2 and that given the parameters in the mean equation, the maximum likelihood estimate of the variance is

$$\hat{\sigma}^2 = T^{-1} \sum_{t=1}^{T} \left(x_t \quad \phi_0 \quad \sum_{i=1}^{p} \phi_i x_{t-i} \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}\right)^2$$

$$= T^{-1} \sum_{t=1}^{T} \varepsilon_t^2$$

Estimation using conditional maximum likelihood is equivalent to least squares, although unlike linear regression the objective is nonlinear due to the moving average terms and so a nonlinear maximization algorithm is required. If the model does not include moving average terms ($q = 0$), then the conditional maximum likelihood estimates of an $AR(P\ )$ are identical the least squares estimates from the regression

$$x_t = \phi_0 + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \phi_3 x_{t-3} + \quad + \phi_p x_{t-p} + \varepsilon_t$$

$$T\ \ erefore\ \hat{\sigma}^2 = T^{-1} \sum_{t=1}^{T} \left(x_t \quad \hat{\phi}_0 \quad \sum_{i=1}^{p} \hat{\phi}_i x_{t-i} \sum_{j=1}^{q} \hat{\theta}_j \varepsilon_{t-j} \right)^2$$

Example: The following parameter estimates were computed for the $AR(2)$ model based on the differenced data using maximum likelihood estimation method.

Source	Parameter estimate	Standard error	95% confidence interval
Intercept	0.0050	0.0119	
AR(1)	0.4064	0.0149	(0.4884, 0.3243)
AR(2)	0.1649	0.0419	(0.2469, 0.0829)

Number of Observations: 558

Degrees of Freedom: 558 - 3 = 555

Residual Standard Deviation: 0.4423

Both AR parameters are significant since the confidence intervals do not contain zero. The model for the differenced data Y_t is an $AR(2)$ model:

$$Y_t = \quad 0.0050 \quad 0.4064 Y_{t-1} \quad 0.1649 Y_{t-2}$$

5.5.3 Model Diagnostic Checking

In the model-building process, if an ARIMA (p, d, q) model is chosen (based on the ACFs and PACFs), some checks on the model adequacy are required. A residual analysis is usually based on the fact that the residuals of an adequate model should be approximately white noise. Therefore, checking the significance of the residual autocorrelations and comparing with approximate two standard error bounds, i.e.,$\pm\frac{2}{\sqrt{n}}$ are need.

After estimating the model parameters, the diagnostic checking is applied to see if the model is adequate or not. Model checking is an obligatory activity to investigate validity and reliability of all inference procedures made by ARIMA before one is going to use these models to forecast future patterns of series. Therefore a range of diagnostic tests is available for checking the model assumptions and properties formally. Therefore the following statistical tests are used:

1. t-tests for coefficient significance
2. Residual portmanteau test:
3. AIC and BIC for model selection

Residual portmanteau test: This test was originally developed by Box and Pierce for testing the residuals from a forecast model. Any good forecast model should have forecast errors which follow a white noise model. If the series is white noise then, the Q statistic has a chi-square distribution with (h-m) degrees of freedom, where m is the number of parameters in the model which has been fitted to the data. The test can easily be applied to raw data, when no model has been fitted, by setting $m = 0$.

The portmanteau test for residual autocorrelation checks the null hypothesis that all residual autocovariances are zero.

$$H_0: \rho_1 = \rho_2 = \quad = \rho_k \ Vs \ \ H_1: \rho_j \neq 0, j = 1,2,3, \dots, k \ at \ leat \ one \ j \neq 0$$

- The Box-Pierce Q statistics is defined as

$$Q = T \sum_{k=1}^{h} r_k^2$$

- Ljung-Box (1978) statistics: Q-statistics is an objective diagnostic measure of white noise for a time series, assessing whether there are patterns in a group of autocorrelations
- The Ljung-Box (modified Box-Pierce) statistic

$$Q \ = T(T+2) \sum_{k=1}^{h} \frac{\gamma_k^2}{T \quad k}$$

Where $\gamma_k^2 =$ the square of the residual autocorrelation coefficients for lags $k = 1,2, \dots, \ $; $T =$ the number of data points in the stationary time series; = the number of autocorrelations used in the summation. Q* has a Chi-square distribution with (m) degrees of freedom.

In general, the data are not white noise if the values of Q or Q* is greater than the value given in a chi square table with α level of significance.

Hypothesis:

- H_0 : ACFs are not significantly different than white noise ACFs (i.e., ACFs = 0).
- H_1 : ACFs are statistically different than white noise ACFs (i.e., ACFs ≠ 0).

Decision rule:

- If $Q \leq$ chi-square; Do not reject Ho, the ACF patterns are white noise.
- If $Q >$ chi-square; Reject Ho, the ACF patterns are not white noise

If a model is rejected at this stage, the model-building cycle has to be repeated. Note: This test only make sense if $k > p + q$

Example: Here is the first 15 ACF value from 40 observations for the white noise is given by below table. Find the box-Pierce Q statistics and the Ljung-Box (modified Box-Pierce) statistics for $= 10$.

Lag	ACF
1.	0.159128
2.	0.12606
3.	0.102384
4.	0.06662
5.	0.08255
6.	0.176468
7.	0.191626
8.	0.05393
9.	0.08712
10.	0.01212
11.	0.05472
12.	0.22745
13.	0.089477
14.	0.017425
15.	0.2004

Solution: The Box-Pierce Q statistics for $= 10$ is

$$Q = T\sum_{k=1}^{h} r_k^2 = 40[(1.59128)^2 + (\quad 0.12606)^2 + (0.102384)^2 + \quad + (\quad 0.01212)^2]$$

$$= 40 \quad 0.141454767 = 5.658191 \quad 5.66$$

- The Ljung-Box (modified Box-Pierce) Q statistic for $= 10$ is

$$Q = T(T+2)\sum_{k=1}^{h} \frac{r_k^2}{T \quad k} = 40(40+2)\left[\frac{(1.59128)^2}{40 \quad 1} + \frac{(\quad 0.12606)^2}{40 \quad 2} + \frac{(0.102384)^2}{40 \quad 3} + \quad + \frac{(\quad 0.01212)^2}{40 \quad 10}\right]$$

$$= 40 \quad 42 \quad 0.06831498 = 114.7692$$

Since the data is not modeled $m = 0$ at $df = 10$. From table chi-square with 10 df, the probability of obtaining a chi-square value is 15.98718 as large as or larger than 5.66 and

114.7692 is at 1% level of significance respectively. The set of 10 γ_k values are not significantly different from zero.

AIC and BIC for model selection: Beginning with the classical approach, we see that comparing the likelihood of different models is of little use because the model with the most parameters will always have greater likelihood.

$$\text{Akaike's Information Criterion:} \qquad AIC = \log \hat{\sigma}_k^2 + \frac{2k}{T}$$

$$\text{Bayesian Schwarz Information Criterion: } BIC = \log \hat{\sigma}_k^2 + \frac{k \log T}{T}$$

Where $\hat{\sigma}_k^2$ is the maximum likelihood estimator for the variance and k is the number of parameters in the model and T is the number of sample size.

Both criteria are likelihood-based and represent a different trade-off between "fit", as measured by the log-likelihood value, and "parsimony", as measured by the number of free parameters, $p + q$. If a constant is included in the model, the number of parameters is increased to $p + q + 1$. Usually, the model with the smallest AIC or BIC values are preferred. While the two criteria differ in their trade-off between fit and parsimony, the BIC criterion can be preferred because it has the property that it will almost surely select the true model.

Example1: We are going to apply the model selection criteria in order to choose a model for the vehicle registration series in Debre Markos town. The table gives the model, the residual variance, the number of parameters and the value of the corresponding selection criterion:

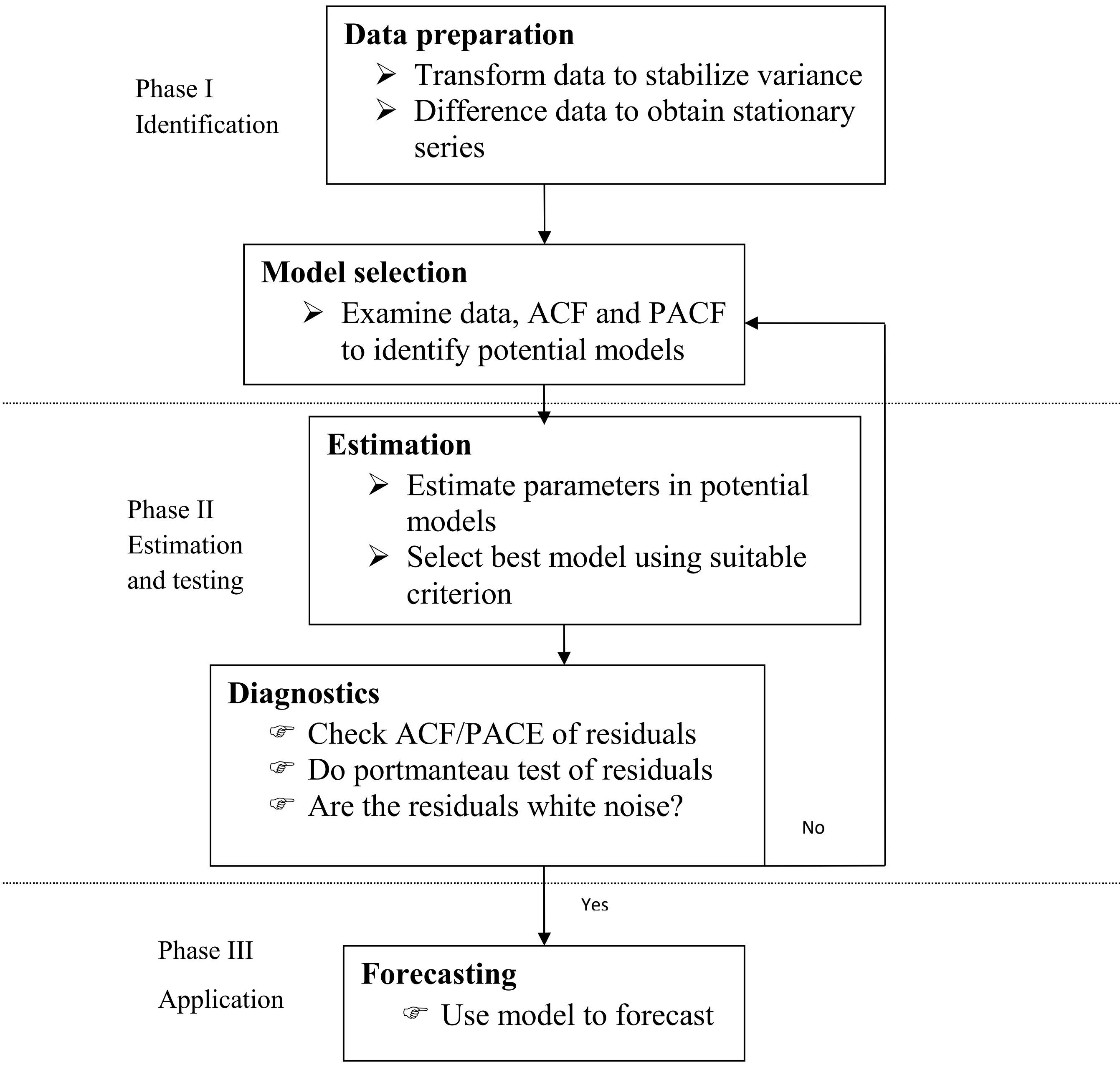
Model	$\hat{\sigma}^2$	T	K	BIC	AIC
$ARIMA(0,1,1)x(0,1,1)_{12}$	0.119^2	466	2	-1.9799	-1.9716
$ARIMA(0,1,1)x(1,1,1)_{12}$	0.116^2	466	3	-2.0017	-1.9892

The best model using both the BIC as well as the AIC is the second one, which obtains the lowest value using both criteria. Based on AIC and BIC values $ARIMA(0,1,1)x(1,1,1)_{12}$ model would be chosen.

Example 2: Select the appropriate model based on the following information

Models	Model selection criteria			
	$t \quad test$	$Q \quad test$	AIC	BIC
$AR(2)$	√	√	1424.66	1437.292
$MA(1)$	√	X	1507.162	1515.583
$ARIMA(2,0,1)$	√ (partially)	√	1425.727	1442.57

Generally stage of Stages of Box-Jenkins methodology



CHAPTER SIX

6 MODEL IDENTIFICATION AND ESTIMATION

6.1 Introduction

Box-Jenkins Methodology is a method for estimating ARIMA models, based on the ACF and PACF as a means of determining the stationarity of the variable in question and the lag lengths of the ARIMA model. Here the ACF and PACF methods for determining the lag length in an ARIMA model are commonly used.

This is the technique for determining the most appropriate $ARMA \text{ or } ARIMA$ model for a given variable. It comprises four stages in all:

1. Identification of the model, this involves selecting the most appropriate lags for the AR and MA parts, as well as determining if the variable requires first-differencing to induce stationarity. The ACF and PACF are used to identify the best model. (Information criteria can also be used)

- To determine the appropriate lag structure in the AR part of the model, the PACF or Partial correlogram is used, where the number of non-zero points of the PACF determine where the AR lags need to be included.
- To determine the MA lag structure, the ACF or correlogram is used, again the non-zero points suggest where the lags should be included.

2. Estimation, this usually involves the use of a least squares estimation process, moment estimation method, maximum likelihood method and Yule-walker method.

3. Diagnostic testing, which usually is the test for autocorrelation. If this part is failed then the process returns to the identification section and begins again, usually by the addition of extra variables.

4. Forecasting, the ARIMA models are particularly useful for forecasting due to the use of lagged variables.

Some of the $ARIMA(p, d, q)$ process are known as:

- $ARIMA(1,0,0) = AR(1)$, first-order autoregressive model
- $ARIMA(0,1,0) = I(1)$, random walk model
- $ARIMA(1,1,0) = ARI(1,1)$, differenced first order autoregressive model

- $ARIMA(0,1,1) = IMA(1,1)$,without constant, simple exponential smoothing
- $ARIMA(0,1,1) = IMA(1,1)$,with constant, simple exponential smoothing with growth
- $ARIMA = IMA(2,1)$ Without constant, linear exponential smoothing
- $ARIMA(1,1,2)$, with constant, dampened-trend linear exponential smoothing

6.2 Autoregressive (AR) process

Autoregressive models are based on the idea that the current value of the series, x_t, can be explained as a function of p past values, $x_{t-1}, x_{t-2}, \dots, x_{t-p}$, where p determines the number of steps into the past needed to forecast the current value.

$$X_t = \mu + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \quad + \phi_p X_{t-p} + \varepsilon_t$$

Where ε_t is a white noise series with mean zero and finite variance σ_ε^2. Using the backward shift or lag operator B (defined as $B^p = X_{t-p}$), the $AR(p)$ can be written as

$$(1 \quad \phi_1 L \quad \phi_2 L^2 \quad \phi_3 L^3 \qquad \phi_P L^P) X_t = \mu + \varepsilon_t$$

The $AR(p)$ process is stationary if the stability conditions are satisfied:

1. if the characteristic equation $\lambda^P \quad \phi_1 \lambda^{P-1} \quad \phi_2 \lambda^{P-2} \quad \phi_3 \lambda^{P-3} \qquad \phi_P = 0$ only has roots with absolute values smaller than one
2. If the solutions of the lag polynomial $1 \quad \phi_1 L \quad \phi_2 L^2 \quad \phi_3 L^3 \qquad \phi_P L^P = 0$ only have roots with absolute values larger than one

6.2.1 Special cases for Autoregressive Model/Process

6.2.1.1 $AR(1)$ process

The simplest example of an AR process is the first-order case given by

$$X_t = \phi_1 X_{t-1} + \varepsilon_t \quad \text{Without drift (constant) value}$$

$$X_t = \delta + \phi_1 X_{t-1} + \varepsilon_t \text{ With drift (constant) value}$$

Assumption $of\ Ar(1)$ process $E(X_{t-k}\varepsilon_t) = 0, k \neq 0$ and $E(\varepsilon_t) = 0$

1. Mean of $AR(1)$ process

Let the $AR(1)$ process can be given as $X_t = \delta + \phi_1 X_{t-1} + \varepsilon_t$, then the mean of $AR(1)$ process of X_t can be defined as follows

$$E(X_t) = E(\delta + \phi_1 X_{t-1} + \varepsilon_t)$$

$$= E(\mu) + \phi_1 E(X_{t-1}) + E(\varepsilon_t)$$

$$= \delta + \phi_1 E(X_{t-1}) + 0$$

- $E(X_t) \quad \phi_1 E(X_{t-1}) = \delta \qquad E(X_t)[1 \quad \phi_1] = \delta$

$$E(X_t) = \frac{\delta}{1 \quad \phi_1}$$

2. **Stationarity condition for $AR(1)$ process**

The $AR(1)$ process is stationary if only if the roots of lag polynomial

$$\phi(L) = 1 \quad \phi_1 L = 0$$

lies outside the unit circle. Which implies that the parameter ϕ_1 and ϕ_2 satisfies the following conditions, $i.e \quad |\phi_1| < 1$

3. **Autocorrelation function for $AR(1)$ process**

To compute the autocorrelations $AR(1)$ process first we obtained autocovariances the process.

$$r_0 = E(X_t X_t) = E((\phi_1 X_{t-1} + \varepsilon_t)^2) = \phi_1^2 r_0 + \sigma_\varepsilon^2 \; = \frac{\sigma_\varepsilon^2}{1 \quad \phi_1^2}$$

$$r_1 = E(X_{t-1} X_t) = E\big(X_{t-1}(\phi_1 X_{t-1} + \varepsilon_t)\big) = \phi_1 r_0$$

$$r_2 = E(X_{t-2} X_t) = E\big(X_{t-2}(\phi_1 X_{t-1} + \varepsilon_t)\big) = \phi_1 r_1 = \phi_1^2 r_0$$

$$r_3 = E(X_{t-3} X_t) = E\big(X_{t-3}(\phi_1 X_{t-1} + \varepsilon_t)\big) = \phi_1 r_2 = \phi_1^3 r_0$$

Therefore: $r_k = E(X_{t-k} X_t) = E\big(X_{t-k}(\phi_1 X_{t-1} + \varepsilon_t)\big) = \phi_1 r_{k-1} = \phi_1^k r_0, \; k > 0$

The autocorrelation can be obtained as follows:

$$\rho_k = \frac{r_k}{r_0} = \frac{\phi_1 r_{k-1}}{r_0} = \frac{\phi_1^k r_0}{r_0} = \phi_1^{|k|}, \qquad k > 0$$

Proof:

Let the $AR(1)$ process is defined by $X_t = \phi_1 X_{t-1} + \varepsilon_t$ is stationary if only if $|\phi_1| < 1$ and ε_t is a white noise. To compute the autocovariances and autocorrelations using backward substitution,

$X_t = \phi_1 X_{t-1} + \varepsilon_t$ must be transformed into a pure MA process by recursive substitution

$$\begin{aligned} X_t &= \phi_1 X_{t-1} + \varepsilon_t \\ &= \phi_1(\phi_1 X_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \\ &= \phi_1^2 X_{t-2} + \phi_1 \varepsilon_{t-1} + \varepsilon_t \\ &= \phi_1^2(\phi_1 X_{t-3} + \varepsilon_{t-2}) + \phi_1 \varepsilon_{t-1} + \varepsilon_t \\ &= \phi_1^3 X_{t-3} + \phi_1^2 \varepsilon_{t-2} + \phi_1 \varepsilon_{t-1} + \varepsilon_t \\ &= \varepsilon_t + \phi_1 \varepsilon_{t-1} + \phi_1^2 \varepsilon_{t-2} + \phi_1^3 \varepsilon_{t-3} \dots \end{aligned}$$

$$= \varepsilon_t + \sum_{i=1}^{\infty} \phi_1^i \varepsilon_{t-i}$$

$$\gamma_0 = Var(X_t) = E(x_t^2) = E([\varepsilon_t + \phi_1\varepsilon_{t-1} + \phi_1^2\varepsilon_{t-2} + \phi_1^3\varepsilon_{t-3} \ldots]^2)$$

$$= E\left[\sum_{i=0}^{\infty} \phi_1^{2i}\, \varepsilon_{t-i}^2 + \sum_{i=0}^{\infty}\sum_{j\neq i}^{\infty} \phi_1^i \phi_1^j \varepsilon_{t-i}\varepsilon_{t-j}\right]$$

$$= \sum_{i=0}^{\infty} \phi_1^{2i}\, E(\varepsilon_{t-i}^2) + \sum_{i=0}^{\infty}\sum_{j\neq i}^{\infty} \phi_1^i \phi_1^j E(\varepsilon_{t-i}\varepsilon_{t-j})$$

$$= \sum_{i=0}^{\infty} \phi_1^{2i}\, \sigma_\varepsilon^2 + \sum_{i=0}^{\infty}\sum_{j\neq i}^{\infty} \phi_1^i \phi_1^j 0$$

$$= \sum_{i=0}^{\infty} \phi_1^{2i}\, \sigma_\varepsilon^2 \ = \sigma_\varepsilon^2 \sum_{i=0}^{\infty} \phi_1^{2i}$$

$$= \frac{\sigma_\varepsilon^2}{1 \quad \phi_1^2} \qquad Since\ |\phi_1| < 1\ t\ \ en \sum_{i=0}^{\infty} \phi_1^{2i} = \frac{1}{1 \quad \phi_1^2}$$

Example1: which of the following $AR(1)$ process is stationary? Why?

$$X_t = 0.3X_{t-1} + \varepsilon_t$$

$$X_t = 1.5X_{t-1} + \varepsilon_t$$

$$X_t = \quad 0.75X_{t-1} + \varepsilon_t$$

$$X_t = \quad 1.2X_{t-1} + \varepsilon_t$$

Example2: consider the $AR(1)$ process is given as

$$X_t = 0.3X_{t-1} + \varepsilon_t$$

Express these models using backshift operator and determine whether the model is stationary and find the equivalent MA representation

Solution

1. Mean of X_t is 0
2. Stationarity of X_t, AR(1) process is stationary because $|\phi_1| = 0.3 < 1$
3. Acovariance and autocorrelation of X_t

- Acovariance function of AR(1) process is defined as

$$r_k = E(X_{t-k}X_t) = E\big(X_{t-k}(\phi_1 X_{t-1} + \varepsilon_t)\big) = \phi_1 r_{k-1} = \phi_1^k r_0$$

$$\therefore r_1 = 0.3r_0, r_2 = 0.09r_0, r_3 = 0.027r_0, \ldots.$$

- Autocorrelation function of AR(1) process is defined as $\rho_k = \phi_1^{|k|}$

$$\rho_1 = 0.3, \rho_2 = \phi_1^2 = 0.09, \qquad \rho_3 = \phi_1^3 = 0.027, ...$$

4. Partial autocorrelatuon of AR(1) process is $\phi_{11} = \rho_1 = 0.3$ and $\phi_{22} = 0, \forall\, k \geq 2$

6.2.1.2 $AR(2)$ process

The second order autoregressive $AR(2)$ process can be defined as

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \varepsilon_t \quad \text{Without drift (constant) value}$$

$$X_t = \delta + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \varepsilon_t \text{ With drift (constant) value}$$

1. **Mean of for $AR(2)$ process**

$$E(X_t) = E(\delta + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \varepsilon_t)$$

$$= E(\delta) + \phi_1 E(X_{t-1}) + \phi_2 E(X_{t-2}) + E(\varepsilon_t)$$

$$= \mu + \phi_1 E(X_{t-1}) + \phi_2 E(X_{t-2}) + 0$$

$$E(X_t) \quad \phi_1 E(X_{t-1}) + \phi_2 E(X_{t-2}) = \delta$$

$$E(X_t)[1 \quad \phi_1 \quad \phi_2] = \delta$$

$$E(X_t) = \frac{\delta}{1 \quad \phi_1 \quad \phi_2}$$

2. **Stationarity condition for $AR(2)$ process**

The $AR(2)$ process is stationary if only if the roots of lag polynomial

$$\phi(L) = 1 \quad \phi_1 L \quad \phi_2 L^2 = 0$$

lies outside the unit circle. Which implies that the parameter ϕ_1 and ϕ_2 satisfies the following conditions.

$$\phi_1 + \phi_2 < 1$$

$$\phi_2 \quad \phi_1 < 1$$

$$1 < \phi_2 < 1$$

3. **Autocorrelation function for $AR(2)$ process**

- The autocovariance and autocorrelation of $AR(2)$

$$\gamma_0 = E(X_t X_t) = E(X_t[\phi_1 X_{t-1} + \phi_2 X_{t-2} + \varepsilon_t])$$

$$= E(\phi_1 X_t X_{t-1} + \phi_2 X_t X_{t-2} + X_t \varepsilon_t)$$

$$= \phi_1 E(X_t X_{t-1}) + \phi_2 E(X_t X_{t-2}) + E(X_t \varepsilon_t)$$

$$= \phi_1\gamma_1 + \phi_2\gamma_2 + \sigma_\varepsilon^2$$

$$= \frac{1 \quad \phi_2}{(1+\phi_2)\left[(1 \quad \phi_2)^2 \quad {\phi_1}^2\right]}\sigma_\varepsilon^2$$

$$\gamma_1 = E(X_{t-1}X_t) = E(X_{t-1}[\phi_1 X_{t-1} + \phi_2 X_{t-2} + \varepsilon_t])$$

$$= \phi_1 E(X_{t-1}X_{t-1}) + \phi_2 E(X_{t-1}X_{t-2}) + E(X_{t-1}\varepsilon_t)$$

$$= \phi_1\gamma_0 + \phi_2\gamma_1$$

$$= \frac{\phi_1}{(1+\phi_2)\left[(1 \quad \phi_2)^2 \quad {\phi_1}^2\right]}\sigma_\varepsilon^2$$

$$\gamma_2 = E(X_{t-2}X_t) = E(X_{t-2}[\phi_1 X_{t-1} + \phi_2 X_{t-2} + \varepsilon_t])$$

$$= \phi_1 E(X_{t-2}X_{t-1}) + \phi_2 E(X_{t-2}X_{t-2}) + E(X_{t-2}\varepsilon_t)$$

$$= \phi_1\gamma_1 + \phi_2\gamma_0$$

$$= \frac{{\phi_1}^2 + \phi_2 \quad {\phi_2}^2}{(1+\phi_2)\left[(1 \quad \phi_2)^2 \quad {\phi_1}^2\right]}\sigma_\varepsilon^2$$

Since $E(X_tX_{t-1}) = E(X_{t-2}X_{t-1}) = \gamma_1$ *and* $E(X_tX_t) = E(X_{t-1}X_{t-1}) = E(X_{t-2}X_{t-2}) = \gamma_0$

- The autocorrelation function satisfies the second order differencing equations as

$$\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2}$$

With satisfying

$$\rho_0 = 1$$

$$\rho_1 = \frac{\phi_1}{1 \quad \phi_2}$$

$$\rho_2 = \frac{\phi_1^2}{1 \quad \phi_2} + \phi_2$$

For $AR(2)$ process the general formula for ACF is

$$\rho_k = \boldsymbol{A}(r_1)^{|k|} + \boldsymbol{B}(r_2)^{|k|}$$

Where **A** and **B** are constants determined by initial conditions that $\rho_0 = 1$ *and* $\rho_1 = \rho_{-1}$. And also that r_1 and r_2 are roots of characteristics polynomial and determined by

$$r_1 = \frac{1}{L_1} \text{ and } r_2 = \frac{1}{L_2}$$

- When the roots are real, the ACF consists of a mixture of dampened exponentials $i.e$ $\phi_1^2 + 4\phi_2 \geq 0$
- When the roots are complex $(\phi_1^2 + 4\phi_2 < 0)$ the $AR(2)$ process displays Pseudo-Periodic behavior. Then the general solution for ACF is

$$\rho_k = d^k(A\cos(ft) + B\sin(ft))$$

with A and B again being arbitrary constants that can be determined by using the initial conditions.

Where $d =$ damping factor, $d = \sqrt{\quad \phi_2}$

f =is the frequency of the oscillation, $f = \arccos\left(\frac{\phi_1}{2\sqrt{-\phi_2}}\right)$

The period of the cycles is $P = \frac{2\pi}{f}$. Processes with conjugate complex roots are well-suited to describe business cycle fluctuations.

4. Partial Autocorrelation function for $AR(2)$ process

The partial correlation between two variables is the correlation that remains if the possible impact of all other random variables has been eliminated. To define the partial autocorrelation coefficient, we use the new notation,

$$x_t = \phi_{k1}x_{t-1} + \phi_{k2}x_{t-2} + \quad + \phi_{kk}x_{t-k} + \varepsilon_t$$

Where ϕ_{ki} is the coefficient of the variable with lag i if the process has order k. (According to the former notation it holds that $\phi_i = \phi_{ki}, i = 1,2, \dots, k.$)

The coefficients ϕ_{kk} are the partial autocorrelation coefficients (of order k), $k = 1,2, \dots$. The partial autocorrelation measures the correlation between x_t and x_{t-k} which remains when the influences of $x_{t-1}, x_{t-2}, \dots, x_{t-k+1}$ on x_t and x_{t-k} have been eliminated.

In order to find the autocorrelations of an $AR(p)$ process, pre-multiply each side by x_{t-k} where $K > 0$ and taking expectation both sides.

$$E(x_{t-k}x_t) = E\left(x_{t-k}(\phi_{k1}x_{t\ 1} + \phi_{k2}x_{t\ 2} + \quad + \phi_{kk}x_{t\ k} + \varepsilon_t)\right)$$

$$r_0 = \phi_{k1}r_1 + \phi_{k2}r_2 + \quad + \phi_{kk}r_k + \sigma_\varepsilon^2$$

$$r_1 = \phi_{k1}r_0 + \phi_{k2}r_1 + \quad + \phi_{kk}r_{k\ 1}$$

$$r_2 = \phi_{k1}r_1 + \phi_{k2}r_0 + \quad + \phi_{kk}r_{k\ 2}$$

$$r_3 = \phi_{k1} r_2 + \phi_{k2} r_1 + \quad + \phi_{kk} r_{k \quad 3}$$

$$r_k = \phi_{k1} r_1 + \phi_{k2} r_0 + \quad + \phi_{kk} r_{k \quad 2}$$

Then the Autocorrelation of the higher order autoregressive process is defined as:

$$\rho_1 = \phi_{k1} + \phi_{k2}\rho_1 + \quad + \phi_{kk}\rho_{k \quad 1}$$
$$\rho_2 = \phi_{k1}\rho_1 + \phi_{k2} + \quad + \phi_{kk}\rho_{k \quad 2}$$
$$\rho_3 = \phi_{k1}\rho_2 + \phi_{k2}\rho_1 + \quad + \phi_{kk}\rho_{k \quad 3}$$

$$\rho_k = \phi_{k1}\rho_1 \quad + \quad \phi_{k2} + \ \dots \ + \phi_{kk}$$

Using Yule-Walk we can derive the partial autocorrelation coefficients ϕ_{kk} from the autocorrelation coefficients

$$\begin{matrix} 1 & \rho_1 & \rho_2 \dots & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 \dots & \rho_{k-2} \\ \rho_2 & \rho_1 & 1 \dots & \rho_{k-3} \\ & & & \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} \dots & 1 \end{matrix} \quad \begin{matrix} \phi_{k1} \\ \phi_{k2} \\ \phi_{k3} \\ \dots \\ \phi_{kk} \end{matrix} = \begin{matrix} \rho_1 \\ \rho_2 \\ \rho_3 \\ \dots \\ \rho_k \end{matrix}, k = 1,2,3,\dots$$

With Cramer's rule we get

$$\phi_{kk} = \frac{\begin{matrix} 1 & \rho_1 & \rho_2 \dots & \rho_1 \\ \rho_1 & 1 & \rho_1 \dots & \rho_2 \\ \rho_2 & \rho_1 & 1 \dots & \rho_3 \\ & & & \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} \dots & \rho_k \end{matrix}}{\begin{matrix} 1 & \rho_1 & \rho_2 \dots & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 \dots & \rho_{k-2} \\ \rho_2 & \rho_1 & 1 \dots & \rho_{k-3} \\ & & & \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} \dots & 1 \end{matrix}}, k = 1,2,3,\dots$$

If the Data Generation Process (DGP) is an $AR(2)$ process, we get for the partial autocorrelation function: $\phi_{11} = \rho_1 = \frac{\phi_1}{1-\phi_2}$

$$\phi_{22} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} = \frac{\rho_2 \quad \rho_1^2}{1 \quad \rho_1^2} = \phi_2$$

$$\phi_{33} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & \rho_3 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}} = 0, k \geq 3$$

$$= \frac{\rho_3 \quad \rho_1\rho_2 \quad \rho_1^2(\rho_3 \quad \rho_1) + \rho_2\rho_1(\rho_2 \quad 1)}{(1 \quad \rho_1^2) \quad \rho_1^2(1 \quad \rho_2) + \rho_2(\rho_1^2 \quad \rho_2)} = 0$$

In general, for ϕ_{kk}, the determinat in the numerator has the same elements as that in the denominator, but with the last column replaced by ρ_k. The quantity ϕ_{kk} regards as a function of the lag k, is called the partial autocorelatuon function.

Example 1: let us consider the $AR(2)$ process of $X_t = 1 + 1.5X_{t-1} \quad 0.56X_{t-2} + \varepsilon_t$ with a variance of error term is **1**. Then check stationarity, find the mean, variance, auto-covariance and autocorrelation of X_t.

Solution:

- To check stationarity for $AR(2)$ use the characteristic equation: $1 \quad 1.5L + 0.56L^2 = 0$. here the roots of $L_1 = 0.8\ and\ L_2 = 0.7$ which is both root values are less than one, then $X_t\ is\ stationry.$
- The expected value of process is defined as, $\mu = \frac{1}{1-1.5-0.56} = 16.667$
- The variance of AR(2) model

$$\gamma_0 = \frac{1 \quad \phi_2}{(1+\phi_2)[(1 \quad \phi_2)^2 \quad {\phi_1}^2]}\sigma_\varepsilon^2 = \frac{1.56}{(0.44)[(1.56)^2 \quad (1.5)^2]} = 19.31$$

$$\gamma_1 = \frac{\phi_1}{(1+\phi_2)[(1 \quad \phi_2)^2 \quad {\phi_1}^2]}\sigma_\varepsilon^2 = \frac{1.5}{(0.44)[(1.56)^2 \quad (1.5)^2]} = 18.568$$

$$\gamma_2 = \frac{{\phi_1}^2 + \phi_2 \quad {\phi_2}^2}{(1+\phi_2)[(1 \quad \phi_2)^2 \quad {\phi_1}^2]}\sigma_\varepsilon^2 = \frac{2.25 \quad 0.56 \quad 0.3136}{(0.44)[(1.56)^2 \quad (1.5)^2]} = 17.038$$

- The Autocorrelation function $\rho_0 = \frac{\gamma_0}{\gamma_0} = 1$

$$\rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{18.568}{19.31} = 0.962$$

$$\rho_2 = \frac{\gamma_2}{\gamma_0} = \frac{17.038}{19.31} = 0.882$$

Example 2: Consider the time series $X_t = \frac{1}{3}x_{t-1} + \frac{2}{9}x_{t-2} + \varepsilon_t$, where$\varepsilon_t$ is WN (0,1)

A. Show that this process is stationary?
B. Find the Yule-Walker equations
C. Find the general expression of ρ_k

Solution

A. We want to Show that $AR(2)$ process is stationary

$$\phi_1 + \phi_2 < 1 \qquad \frac{1}{3} + \frac{2}{9} = \frac{5}{9} < 1$$

$$\phi_2 \quad \phi_1 \ < 1 \qquad \frac{2}{9} \quad \frac{1}{3} = \frac{\quad 1}{9} < 1$$

$$1 < \phi_2 \ < 1 \qquad 1 < \frac{1}{9} < 1$$

$\therefore X_t$ is stationary.

B. Find the Yule-Walker equations

To find the Yule-Walker equation first we find the autocovariance function and autocorrelation function.

$$\gamma_k = \begin{matrix} \frac{1}{3}\gamma_1 + \frac{2}{9}\gamma_2 + \sigma^2, & if\ k = 0 \\ \frac{1}{3}\gamma_0 + \frac{1}{3}\gamma_1, & if\ k = 1 \\ \frac{1}{3}\gamma_1 + \frac{2}{9}\gamma_0, & if\ k = 2 \\ 0, & ot\ \ erwise \end{matrix}$$

$$corr(X_t, X_{t-k}) = \rho_k = \frac{\gamma_k}{\gamma_0} = \frac{1}{3}\rho_{1-k} + \frac{2}{9}\rho_{2-k}$$

$$\rho_k = \begin{matrix} 1, & if\ k = 0 \\ \frac{1}{3}\rho_0 + \frac{2}{9}\rho_1 = \frac{3}{7}, & if\ k = 1 \\ \frac{1}{3}\rho_1 + \frac{2}{9}\rho_0 = \frac{23}{63}, & if\ k = 2 \\ 0, & ot\ \ erwise \end{matrix}$$

$\rho_k = \frac{1}{3}\rho_{1-k} + \frac{2}{9}\rho_{2-k}$ for $k \geq 1$ and Let $\rho = [\rho_1 \ \ \rho_2\ ]'$and $\boldsymbol{a} = [\frac{1}{3} \ \ \frac{2}{9}\ ]'$

$$\rho = \begin{bmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{bmatrix} \begin{bmatrix} 1/3 \\ 2/9 \end{bmatrix} \qquad \rho = \begin{bmatrix} 1 & \frac{3}{7} \\ \frac{3}{7} & 1 \end{bmatrix} \begin{bmatrix} 1/3 \\ 2/9 \end{bmatrix}$$

$\rho = Ra$, this is Yule-Walker equation

C. The general expression of ρ_k can be defined as $\rho_k = A(r_1)^{|k|} + B(r_2)^{|k|}$

Where **A** and **B** are constants determined by initial conditions that $\rho_0 = 1\ and\ \rho_1 = \rho_{-1}$.

And r_1 and r_2 are roots of characteristics polynomial and determined by $r_1 = \frac{1}{L_1}\ and\ r_2 = \frac{1}{L_1}$.

Using lag polynomial $1 \quad \frac{1}{3}L \quad \frac{2}{9}L^2 = 0$

$L_1 = \quad 3\ and\ L_2 = \frac{3}{2}$ then $r_1 = \frac{1}{L_1} = \frac{-1}{3}\ and\ \ r_2 = \frac{1}{L_2} = \frac{2}{3}$

And also from solution B above $\rho_0 = 1\ and\ \rho_1 = \frac{3}{7}$

$$\rho_k = A(r_1)^{|k|} + B(r_2)^{|k|} = A\left(\frac{1}{3}\right)^{|k|} + B\left(\frac{2}{3}\right)^{|k|}$$

$$1 = A + B \qquad A = 1 \quad B \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.$$

$$\frac{3}{7} = \frac{1}{3}A + \frac{2}{3}B \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.$$

Substitute equation * in equation ** then we get

$$\frac{3}{7} = \frac{1}{3}(1 \quad B) + \frac{2}{3}B \qquad B = \frac{16}{21}$$

$$\frac{5}{21}B = \frac{16}{21} \quad t\ \ en\ A = 1 \quad \frac{16}{21} = \frac{5}{21}$$

Therefore $\therefore \rho_k = A(r_1)^{|k|} + B(r_2)^{|k|} = A\left(\frac{-1}{3}\right)^{|k|} + B\left(\frac{2}{3}\right)^{|k|} = \frac{5}{21}\left(\frac{-1}{3}\right)^{|k|} + \frac{16}{21}\left(\frac{2}{3}\right)^{|k|}$

Exercise: Let us consider the $AR(2)$ process of $X_t = 1.4X_{t-1} \quad 0.85X_{t-2} + \varepsilon_t$ with a variance of error term is 1. Then check stationarity, find the mean, variance, auto-covariance and autocorrelation of X_t.

6.2.1.3 The General p^{th} Order Autoregressive $AR(p)$ Process

The general $\mathbf{p^{th}}$ Order Autoregressive AR(p) Process is defined as

$$X_t = \mu + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \quad + \phi_p X_{t-p} + \varepsilon_t$$

$$(1 \quad \phi_1 L \quad \phi_2 L^2 \quad \phi_3 L^3 \qquad \phi_P L^P)x_t = \varepsilon_t$$

The autocovariance function of of the AR(p) process

$$\gamma_k = \phi_1 r_{k-1} + \phi_2 r_{k-2} + \quad + \phi_p r_{k-p}, k \geq 1$$

The autocorrelation function of the AR(R) process

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \quad + \phi_p \rho_{k-p}, k \geq 1$$

The partial autocorrelation function of $AR(p)$ process cuts off after lag p. this is important property enables us to identify whether a given time series is generated by Autoregressive process

6.3 Moving Average (MA) process

A time series, X_t,is said to be a moving average process of order q (abbreviated $MA(q)$) if it is a weighted linear sum of the last q random shocks, that is

$$X_t = \delta + \varepsilon_t \quad \theta_1 \varepsilon_{t-1} \quad \theta_2 \varepsilon_{t-2} \qquad \theta_q \varepsilon_{t-q}$$

With $\theta_q \neq 0$ where ε_t is again a pure random process with zero mean and variance σ^2. Using the lag operator we can also write:

$$X_t \quad \delta = \varepsilon_t \quad \theta_1 L\varepsilon_t \quad \theta_2 L^2\varepsilon_t \qquad \theta_q L^q \varepsilon_t$$

$$X_t \quad \delta = (1 \quad \theta_1 L \quad \theta_2 L^2 \qquad \theta_q L^p)\varepsilon_t$$

$$= \theta(L)\varepsilon_t, \qquad w\ \ ere\ \ \theta(L) = 1 \quad \theta_1 L \quad \theta_2 L^2 \qquad \theta_q L^p$$

<u>Invertibility</u> is a technical requirement stemming from the use of the autocorrelogram and partial autocorrelogram to choose the model, and it plays an important role in achieving unique identification of the MA component of a model.

Stationarity of $MA(q)$ proces: The MA(q) process no problems of convergence and every finite MA(q) process is stationary, no matter what values are used for θ_j, j=1, 2, 3, …, q

Mean of $MA(q)$ process:

- It the $MA(q)$ process is defined as $X_t = \mu + \varepsilon_t \quad \theta_1\varepsilon_{t-1} \quad \theta_2\varepsilon_{t-2} \qquad \theta_q\varepsilon_{t-q}$, then $E(X_t) = \mu$
- It the $MA(q)$ process is defined as $X_t = \varepsilon_t \quad \theta_1\varepsilon_{t-1} \quad \theta_2\varepsilon_{t-2} \qquad \theta_q\varepsilon_{t-q}$, then $E(X_t) = 0$

6.3.1 Special Cases of Moving Average (MA) process

6.3.1.1 $MA(1)$ *Process*

Let x_t be the $MA(1)$ process defined by $\quad X_t = \varepsilon_t \quad \theta_1\varepsilon_{t-1}$

For any values of θ_1, always the $MA(1)$ process is stationary.

1. **Invertibility condition of $MA(1)$ process :**

The first order moving average is invertible if only if the root of the lag polynomial

$$\theta(L) = 1 \quad \theta_1 L = 0$$

is lies outside the unit circle.

The $MA(1)$ process can be inevitable to infinite order autoregressive process $AR(\infty)$

2. **Autocovariance and autocorelation of $MA(1)$ process**

- In order to find r_k and ρ_k for a $MA(1)$ process it is necessary to evaluate the expectation $E(X_{t-k}X_t) = E(X_{t+k}X_t)$

$$r_k = E(X_{t-k}X_t)$$

$$r_0 = E(X_tX_t) = E((\varepsilon_t \quad \theta_1\varepsilon_{t-1})^2) = \sigma_\varepsilon^2 + \theta_1^2\sigma_\varepsilon^2 = \sigma_\varepsilon^2(1+\theta_1^2)$$

$$r_1 = E(X_{t-1}X_t) = E\big((\varepsilon_t \quad \theta_1\varepsilon_{t-1}) \quad (\varepsilon_{t-1} \quad \theta_1\varepsilon_{t-2})\big) = \quad \theta_1\sigma_\varepsilon^2$$

$$r_2 = E(X_{t-2}X_t) = E\big((\varepsilon_t - \theta_1\varepsilon_{t-1})\ (\varepsilon_{t-2} - \theta_1\varepsilon_{t-3})\big) = 0, k \geq 2$$

- Autocorrelation of $MA(1)$ process

$$\rho_k = \frac{r_k}{r_0} = \frac{r_k}{\sigma_\varepsilon^2(1+\theta_1^2)} \qquad \rho_0 = \frac{r_0}{r_0} = \frac{\sigma_\varepsilon^2(1+\theta_1^2)}{\sigma_\varepsilon^2(1+\theta_1^2)} = 1$$

$$\rho_1 = \frac{r_1}{r_0} = \frac{\theta_1\sigma_\varepsilon^2}{\sigma_\varepsilon^2(1+\theta_1^2)} = \frac{\theta_1}{(1+\theta_1^2)}$$

$$\rho_k = 0\,,\ \ for\ all\ k \geq 2$$

3. Partial ACF of $\boldsymbol{MA(1)}$ process

$$\phi_{11} = \rho_1 = \frac{\theta_1}{(1+\theta_1^2)} = \frac{\theta_1(1 - \theta_1^2)}{1 - \theta_1^4}$$

$$\phi_{22} = \frac{\rho_1^2}{1 - \rho_1^2} = \frac{\theta_1^2}{1+\theta_1^2+\theta_1^4} = \frac{\theta_1^2(1 - \theta_1^2)}{1 - \theta_1^6}$$

$$\phi_{33} = \frac{\rho_1^3 - \rho_2\rho_1(2 - \rho_2)}{1 - \rho_2^2 - 2\rho_1^2(1 - \rho_2)} = \frac{\rho_1^3}{1 - 2\rho_1^2} = \frac{\theta_1^2}{1+\theta_1^2+\theta_1^4+\theta_1^6} = \frac{\theta_1^3(1 - \theta_1^2)}{1 - \theta_1^8}$$

Generally the PACF of $\boldsymbol{MA(1)}$ process is defined as

$$\phi_{kk} = \frac{\theta_1^k(1 - \theta_1^2)}{1 - \theta_1^{2(k+1)}} \quad ,k \geq 1$$

Example: Consider the $MA(1)$ models given by

$$X_t = 10 + \varepsilon_t + 0.4\varepsilon_{t-1}$$

Where ε_t is white noise and with variance σ^2

A. Check the stationary and invertible of X_t
B. Calculate the mean and variance of X_t
C. What is the autocovariance function of X_t
D. What is the autocorrelation function (ACF) of X_t

Solution:

A. X_t is stationary because MA models are always weakly stationary.

- A $MA(1)$ process is invertible if and only if the root of the lag polynomial $1 - \theta_1 L = 0$, is larger than one in modulus.

$$1 - \theta_1 L = 0 \qquad 1 + 0.4L = 0$$

$$L = \frac{1}{0.4} = \left| \frac{5}{2} \right| > 1$$

Therefore the roots of the associated lag polynomial $1 \quad \theta_1 L = 0$ are greater than one in absolute value. This indicates that X_t is invertible.

B. Calculate the mean and variance of X_t

- ❖ $E(X_t) = E(10 + \varepsilon_t + 0.4\varepsilon_{t-1}) = 10, \qquad since\ \varepsilon_t \sim WN(0, \sigma^2)$
- ❖ $var(X_t) = \sigma^2(1 + \theta_1^2) = Var(10 + \varepsilon_t + 0.4\varepsilon_{t-1}) \ = \sigma^2(1 + 0.16) = 1.16\sigma^2$

C. What is the autocovariance function of $\{X_t\}$

$$cov(y_t, y_{t-k}) = \gamma_k = E(y_t y_{t-k}) = E[(\varepsilon_t + 0.4\varepsilon_{t-1})y_{t-k}]$$

The autocovariance function X_t for $MA(1)$ process is given as

$$cov(y_t, y_{t-k}) = \gamma_k = E(y_t y_{t-k}) = E[(\varepsilon_t + 0.4\varepsilon_{t-1})y_{t-k}], \text{ then}$$

$$\gamma_k = \begin{cases} \sigma^2(1 + \theta_1^2) = 1.16\sigma^2, & if\ k = 0 \\ \theta_1\sigma^2 = 0.4\sigma^2, & if\ k = 1 \\ 0, & o\ \ erwise \end{cases}$$

D. What is the autocorrelation function (ACF) of X_t

The autocorrelation function $of\ \ X_t$ for $MA(1)$ process is $\rho_k = \frac{\gamma_k}{\gamma_o}$, then:

$$\rho_k = \begin{matrix} 1, & if\ k = 0 \\ \dfrac{\theta_1}{1 + \theta_1^2} = 0.1615, & if\ k = 1 \\ 0, & ot\ \ erwise \end{matrix}$$

6.3.1.2 $MA(2)$ Process

Let x_t be the $MA(2)$ process defined by

$$X_t = \varepsilon_t \quad \theta_1\varepsilon_{t-1} \quad \theta_2\varepsilon_{t-2}$$

For any values of θ_1 and θ_2, always the $MA(2)$ process is stationary.

1. **Invertibility condition of $MA(2)$ process :**

The second order moving average is invertible if only if the root of the lag polynomial

$$\theta(L) = 1 \quad \theta_1 L \quad \theta_2 L^2 = 0$$

is lies outside the unit circle.

2. **Autocovariance and autocorelation of $MA(1)$ process**

- ➢ In order to find r_k and $\rho_k$for a $MA(1)$ process it is necessary to evaluate the expectation $E(X_{t-k}X_t) = E(X_{t+k}X_t)$

- To obtaine

$$r_k = E(X_{t-k}X_t) r_0 = E(X_tX_t) = E((\varepsilon_t \quad \theta_1\varepsilon_{t-1} \quad \theta_2\varepsilon_{t-2})^2) = \sigma_\varepsilon^2 + \theta_1^2\sigma_\varepsilon^2 + \theta_2^2\sigma_\varepsilon^2$$
$$= \sigma_\varepsilon^2(1 + \theta_1^2 + \theta_2^2)$$
$$r_1 = E(X_{t-1}X_t) = E\big((\varepsilon_t \quad \theta_1\varepsilon_{t-1} \quad \theta_2\varepsilon_{t-2}) \quad (\varepsilon_{t-1} \quad \theta_1\varepsilon_{t-2} \quad \theta_2\varepsilon_{t-3})\big)$$
$$= \quad \theta_1\sigma_\varepsilon^2 + \theta_1\theta_2\sigma_\varepsilon^2 = (\quad \theta_1 + \theta_1\theta_2)\sigma_\varepsilon^2$$
$$r_2 = E(X_{t-2}X_t) = E\big((\varepsilon_t \quad \theta_1\varepsilon_{t-1} \quad \theta_2\varepsilon_{t-2}) \quad (\varepsilon_{t-2} \quad \theta_1\varepsilon_{t-3} \quad \theta_2\varepsilon_{t-4})\big)$$
$$= \quad \theta_2\sigma_\varepsilon^2$$
$$r_K = E(X_{t-K}X_t) = 0, k \geq 3$$

- Autocorrelation of $MA(2)$ process

$$\rho_k = \frac{r_k}{r_0} = \frac{r_k}{\sigma_\varepsilon^2(1 + \theta_1^2 + \theta_2^2)} \qquad \rho_0 = \frac{r_0}{r_0} = \frac{\sigma_\varepsilon^2(1 + \theta_1^2 + \theta_2^2)}{\sigma_\varepsilon^2(1 + \theta_1^2 + \theta_2^2)} = 1$$
$$\rho_1 = \frac{r_1}{r_0} = \frac{(\quad \theta_1 + \theta_1\theta_2)\sigma_\varepsilon^2}{\sigma_\varepsilon^2(1 + \theta_1^2 + \theta_2^2)} = \frac{\theta_1 + \theta_1\theta_2}{1 + \theta_1^2 + \theta_2^2}$$
$$\rho_2 = \frac{r_2}{r_0} = \frac{\theta_2\sigma_\varepsilon^2}{\sigma_\varepsilon^2(1 + \theta_1^2 + \theta_2^2)} = \frac{\theta_2}{1 + \theta_1^2 + \theta_2^2}$$
$$\rho_k = 0\,, for\ all\ k \geq 3$$

3. Partial ACF $\boldsymbol{MA}(\mathbf{2})$ **process**

$$\phi_{11} = \rho_1 = \frac{\theta_1(1 \quad \theta_2)}{1 + \theta_1^2 + \theta_2^2}$$
$$\phi_{22} = \frac{\rho_2 \quad \rho_1^2}{1 \quad \rho_1^2}$$
$$\phi_{33} = \frac{\rho_1^3 \quad \rho_2\rho_1(2 \quad \rho_2)}{1 \quad \rho_2^2 \quad 2\rho_1^2(1 \quad \rho_2)}$$

Example 1: Find the ACF of the time series $X_t = \varepsilon_t + \frac{5}{2}\varepsilon_{t-1} \quad \frac{3}{2}\varepsilon_{t-2}$ where $\varepsilon_t \sim wN(0,1)$

Example 2: Find the ACF of the time series $X_t = \varepsilon_t \quad \frac{1}{6}\varepsilon_{t-1} \quad \frac{1}{6}\varepsilon_{t-2}$ where $\varepsilon_t \sim wN(0,9)$.

Example3: Which of the MA models in 1or 2 is invetrible?

Solution: We compute three cases: since the ε_t are uncorrelated, we can ignore any cross-terms of the form $E(\varepsilon_t\varepsilon_k)$ when $t \neq k$ and $\boldsymbol{\varepsilon_t \sim WN(0,1)}$Then

1. $\gamma_0 = E(X_tX_t) = E\left[\left(\varepsilon_t + \frac{5}{2}\varepsilon_{t-1} \quad \frac{3}{2}\varepsilon_{t-2}\right)^2\right] = \sigma_\varepsilon^2 + \frac{25}{4}\sigma_\varepsilon^2 + \frac{9}{4}\sigma_\varepsilon^2 = \frac{19}{2}$

$$\gamma_1 = E(X_tX_{t-1}) = E\left[\left(\varepsilon_t + \frac{5}{2}\varepsilon_{t-1} \quad \frac{3}{2}\varepsilon_{t-2}\right)\left(\varepsilon_{t-1} + \frac{5}{2}\varepsilon_{t-2} \quad \frac{3}{2}\varepsilon_{t-3}\right)\right]$$

$$= E\left(\frac{5}{2}\varepsilon_{t-1}^2\right) \quad E\left(\frac{15}{4}\varepsilon_{t-1}^2\right) = \frac{5}{2}\sigma_\varepsilon^2 \quad \frac{15}{4}\sigma_\varepsilon^2 = \frac{10 \quad 15}{4} = \quad \frac{5}{4}$$

$$\gamma_2 = E(X_tX_{t-2}) = E\left[\left(\varepsilon_t + \frac{5}{2}\varepsilon_{t-1} \quad \frac{3}{2}\varepsilon_{t-2}\right)\left(\varepsilon_{t-2} + \frac{5}{2}\varepsilon_{t-3} \quad \frac{3}{2}\varepsilon_{t-4}\right)\right]$$

$$= E\left(\quad \frac{3}{2}\varepsilon_{t-2}^2\right) = \quad \frac{3}{2}\sigma_\varepsilon^2 = \quad \frac{3}{2}$$

$$\gamma_3 = E(X_tX_{t-3}) = E\left[\left(\varepsilon_t + \frac{5}{2}\varepsilon_{t-1} \quad \frac{3}{2}\varepsilon_{t-2}\right)\left(\varepsilon_{t-3} + \frac{5}{2}\varepsilon_{t-4} \quad \frac{3}{2}\varepsilon_{t-5}\right)\right] = 0$$

$$and\ so\ on$$

$$T\ \ erefore \qquad \rho_k = \begin{cases} 1, & if\ k = 0 \\ \dfrac{\frac{5}{4}}{\frac{19}{2}} = \dfrac{5}{38}, & if\ k = 1 \\ \dfrac{\frac{3}{2}}{\frac{19}{2}} = \dfrac{3}{19}, & if\ \ k = 2 \\ 0, & ot\ \ erwise \end{cases}$$

2. $\gamma_0 = E(X_tX_t) = E\left[\left(\varepsilon_t \quad \frac{1}{6}\varepsilon_{t-1} \quad \frac{1}{6}\varepsilon_{t-2}\right)^2\right] = \sigma_\varepsilon^2 + \frac{1}{36}\sigma_\varepsilon^2 + \frac{1}{36}\sigma_\varepsilon^2 = \left(\frac{19}{18}\right) \quad 9 = \frac{19}{2}$

$$\gamma_1 = E(X_tX_{t-1}) = E\left[\left(\varepsilon_t \quad \frac{1}{6}\varepsilon_{t-1} \quad \frac{1}{6}\varepsilon_{t-2}\right)\left(\varepsilon_{t-1} \quad \frac{1}{6}\varepsilon_{t-2} \quad \frac{1}{6}\varepsilon_{t-3}\right)\right]$$

$$= E\left(\quad \frac{1}{6}\varepsilon_{t-1}^2\right) + E\left(\frac{1}{36}\varepsilon_{t-1}^2\right) = \quad \frac{1}{6}\sigma_\varepsilon^2 + \frac{1}{36}\sigma_\varepsilon^2 = \left(\frac{6+1}{36}\right) \quad 9 = \quad \frac{5}{4}$$

$$\gamma_2 = E(X_tX_{t-2}) = E\left[\left(\varepsilon_t \quad \frac{1}{6}\varepsilon_{t-1} \quad \frac{1}{6}\varepsilon_{t-2}\right)\left(\varepsilon_{t-2} \quad \frac{1}{6}\varepsilon_{t-3} \quad \frac{1}{6}\varepsilon_{t-4}\right)\right]$$

$$= E\left(\quad \frac{1}{6}\varepsilon_{t-2}^2\right) = \quad \frac{1}{6}\sigma_\varepsilon^2 = \left(\quad \frac{1}{6}\right) \quad 9 = \frac{\quad 3}{2}$$

$$\gamma_3 = E(X_tX_{t-3}) = E\left[\left(\varepsilon_t \quad \frac{1}{6}\varepsilon_{t-1} \quad \frac{1}{6}\varepsilon_{t-2}\right)\left(\varepsilon_{t-3} \quad \frac{1}{6}\varepsilon_{t-4} \quad \frac{1}{6}\varepsilon_{t-5}\right)\right] = 0$$

$$and\ so\ on$$

$$T\ \ erefore \qquad \rho_k = \begin{cases} 1, & if\ k=0 \\ \dfrac{\frac{5}{4}}{\frac{19}{2}} = \dfrac{5}{38}, & if\ k=1 \\ \dfrac{\frac{3}{2}}{\frac{19}{2}} = \dfrac{3}{19}, & if\ \ k=2 \\ 0, & ot\ \ erwise \end{cases}$$

3. Let $\theta_1(L) = 1 + \frac{5}{2}L \quad \frac{3}{2}L^2$ and $\theta_b(L) = 1 \quad \frac{1}{6}L \quad \frac{1}{6}L^2$ be the MA lag polynomials of parts (1) and (2) respectively. By using the quadratic formula, the roots of θ_1 are 2 and $\frac{1}{3}$; similarly, the roots of θ_2 are 3 and 2. Then the MA model of **example 1** is not invertible, but the MA model of **example 2** is invertible.

Example 4: A time series satisfies the model x_t,

$$X_t = 10 + \varepsilon_t + 0.4\varepsilon_{t-1} \quad 0.45\epsilon_{t-2}$$

Where ε_t is white noise and with variance σ^2

A. This model can be described as a member of the $ARMA(p,q)$ family. State the values of p and q and verify that X_t is stationary and invertible.

B. Calculate the mean and variance of X_t

C. What is the autocovariance function of X_t

D. What is the autocorrelation function (ACF) of X_t

Solution:

A. This model can be described as a member of the $ARMA(p,q)$ family. State the values of p and q and verify that $x_t$is stationary and invertible.

- Yes , with $p = 0$ and $q = 2,$ then y_t is $ARMA(p,q) = ARMA(0,2) = MA(2)$ process.
- X_t is stationary because MA models are always weakly stationary.
- A $MA(2)$ process is invertible if and only if the root of the lag polynomial $1 \quad \theta_1 L \quad \theta_2 L^2 = 0,$ is larger than one in modulus.

$$1 \quad \theta_1 L \quad \theta_2 L^2 = 0 \qquad 1 + 0.4L \quad 0.45L^2 = 0$$

$$L = \frac{0.4 \pm \sqrt{0.16 + 1.8}}{0.9} = \frac{0.4 \pm \sqrt{1.96}}{0.9} = \frac{0.4 \pm 1.4}{0.9}$$

$$L_1 = \quad 2 \qquad and \qquad L_2 = \frac{10}{9}$$

Therefore all roots of the associated lag polynomial $1 \quad \theta_1 L \quad \theta_2 L^2 = 0$ are greater than one in absolute value. This indicates that X_t is invertible.

B. Calculate the mean and variance of X_t

- ❖ $E(X_t) = E(10 + \varepsilon_t + 0.4\varepsilon_{t-1} \quad 0.45\epsilon_{t-2}) = 10, \quad since\ \varepsilon_t \sim WN(0, \sigma^2)$
- ❖ $var(X_t) = \sigma^2(1 + \theta_1^2 + \theta_2^2) = Var(10 + \varepsilon_t + 0.4\varepsilon_{t-1} \quad 0.45\epsilon_{t-2})$

$$= \sigma^2(1 + 0.16 + 0.2025) = 1.3625\sigma^2$$

C. What is the autocovariance function of $\{X_t\}$

$$cov(y_t, y_{t-k}) = \gamma_k = E(y_t y_{t-k}) = E[(\varepsilon_t + 0.4\varepsilon_{t-1} \quad 0.45\epsilon_{t-2})y_{t-k}]$$

The autocovariance function X_t for $MA(2)$ process is given as

$$cov(y_t, y_{t-k}) = \gamma_k = E(y_t y_{t-k}) = E[(\varepsilon_t + 0.4\varepsilon_{t-1} \quad 0.45\epsilon_{t-2})y_{t-k}], \text{ then}$$

$$\gamma_k = \begin{cases} \sigma^2(1 + \theta_1^2 + \theta_2^2) = 1.3625\sigma^2, & if\ k = 0 \\ \theta_1(1 \quad \theta_2)\sigma^2 = 0.22\sigma^2, & if\ k = 1 \\ \theta_2\sigma^2 = \quad 0.45\sigma^2, & if\ k = 2 \\ 0, & o\ \ erwise \end{cases}$$

D. What is the autocorrelation function (ACF) of X_t

The autocorrelation function $of\ \ X_t$ for $MA(2)$ process is $\rho_k = \frac{\gamma_k}{\gamma_o}$, then:

$$\rho_k = \begin{cases} 1, & if\ k = 0 \\ \dfrac{\theta_1(1 \quad \theta_2)}{1 + \theta_1^2 + \theta_2^2} = 0.1615, & if\ k = 1 \\ \dfrac{\theta_2}{1 + \theta_1^2 + \theta_2^2} = \quad 0.3303, & if\ k = 2 \\ 0, & ot\ \ erwise \end{cases}$$

6.3.1.3 The General q^{th} Order Moving Average $MA(q)$ Process

The general $\mathbf{q^{th}}$ Order Moving Average **MA(q)** process is defined as

$$X_t = \varepsilon_t \quad \theta_1\varepsilon_{t-1} \quad \theta_2\varepsilon_{t-2} \qquad \theta_q\varepsilon_{t-q}$$

$$X_t = (1 \quad \theta_1 L \quad \theta_2 L^2 \qquad \theta_q L^q)\varepsilon_t$$

The autocovariance function of of the MA(q) process

$$\gamma_k = \begin{cases} \sigma_\varepsilon^2 \sum_{i=0}^{q} \theta_i^2, & if\ k = 0 \\ \sigma_\varepsilon^2(\quad \theta_k + \theta_1\theta_{k+1} + \quad + \theta_{q-k}\theta_q), & if\ k = 1,2, \dots, q \\ 0, & k \geq q \end{cases}$$

The autocorrelation function of the MA(q) process

$$\rho_k = \begin{cases} 1\,, & if\ k = 0 \\ \dfrac{\theta_k + \theta_1\theta_{k+1} + \quad + \theta_{q-k}\theta_q}{1 + \theta_1^2 + \theta_2^2 + \quad + \theta_q^2}, & if\ k = 1,2,\dots,q \\ 0, & k \geq q \end{cases}$$

The autocorrelation function of $MA(q)$ process cuts off after lag q. this is important property enables us to identify whether a given time series is generated by moving average process

6.4 Autoregressive moving average models:

Autoregressive moving-average ($ARMA$) models provide a parsimonious description of a stationary stochastic process in terms of two polynomials, one for the **autoregressive** and the second for the **moving average**. The model consists of two parts, an autoregressive (AR) part and a moving average (MA) part. The AR part involves regressing the variable on its own past values. The MA part involves modeling the error term as a linear combination of error terms occurring contemporaneously and at various times in the past.

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \quad + \phi_p X_{t-p} + \varepsilon_t \quad \theta_1\varepsilon_{t-1} \quad \theta_2\varepsilon_{t-2} \qquad \theta_q\varepsilon_{t-q}$$

$$X_t \quad \phi_1 X_{t-1} + \phi_2 X_{t-2} + \quad + \phi_p X_{t-p} = \varepsilon_t \quad \theta_1\varepsilon_{t-1} \quad \theta_2\varepsilon_{t-2} \qquad \theta_q\varepsilon_{t-q}$$

$$\left(1 \quad \phi_1 L + \phi_2 L^2 + \quad + \phi_p L^p\right)X_t = \left(1 \quad \theta_1 L \quad \theta_2 L^2 \qquad \theta_q L^p\right)\varepsilon_t$$

AR(p) process part MA(q) process part

1. **Stationarity Condition of $ARMA$ (p, q) Process**

The stationarity of an ARMA process is related to the AR component in the model and can be checked through the roots of the associated lag polynomial

$$1 \quad \phi_1 L + \phi_2 L^2 + \quad + \phi_p L^p = 0$$

If all the roots of lag polynomial lies outside the unit circle or all the roots of characteristics polynomial less than one in absolute value, then $ARMA(p, q)$ is stationary. This also implies that, under this condition, $ARMA(p, q)$ has an infinite MA representation.

2. **Invertibility Condition of ARMA $(\mathbf{p}, \mathbf{q})$ Process**

Similar to the stationarity condition, the invertibility of an $ARMA$ process is related to the MA component and can be checked through the roots of the associated lag polynomial

$$1 \quad \theta_1 L \quad \theta_2 L^2 \quad \theta_3 L^3 \qquad \theta_q L^p = 0$$

If all the roots of lag polynomial lies outside the unit circle(greater than one in absolute value) or all the roots of characteristics polynomial less than one in absolute value, then $ARMA(p,q)$ is said to be invertible and has an infinite AR representation.

3. ACF and PACF of $ARMA(p,q)$ Process

As in the stationarity and invertibility conditions, the ACF and PACF of an ARMA process are determined by the AR and MA components respectively. The ACF and PACF of an $ARMA(p,q)$ both exhibit exponential decay and/or damped sinusoid patterns, which makes the identification of the order of the $ARMA(p,q)$ model relatively more difficult.

Model	ACF	PACF
$AR(p)$	Dies down (tails off)	Cuts off after lag p
$MA(q)$	Cuts off after lag q	Dies down (tails off)
$ARMA(p,q)$	Dies down (tails off)	Dies down (tails off)

N.B: In this context **Die out** means **tend to zero gradually** and **cutoff** means **disappear or is zero.**

6.4.1 Special case of $ARMA(p,q)$ process

6.4.1.1 *The* $ARMA(1,1)$ *process*

It is the simplest and easiest process of $ARMA(p,q)$ process and it is defined as

$$X_t = \delta + \phi_1 X_{t-1} + \varepsilon_t \quad \theta_1 \varepsilon_{t-1}$$

$$X_t = \phi_1 X_{t-1} + \varepsilon_t \quad \theta_1 \varepsilon_{t-1}$$

1. Stationarity and invertibility condition for $ARMA(1,1)$ process
   - The $ARMA(1,1)$ process is stationary if only if $|\phi_1| < 1$
   - The $ARMA(1,1)$ process is invertible if only if $|\theta_1| < 1$
2. Autocovariance function and ACF of $ARMA(1,1)$ process

$$r_0 = \phi_1 r_1 + \sigma_\varepsilon^2 (1 \quad \theta_1 \phi_1 + \theta_1^2) = \frac{1 + \theta_1^2 \quad 2\theta_1\phi_1}{1 \quad \phi_1^2} \sigma_\varepsilon^2$$

$$r_1 = \phi_1 r_0 \quad \theta_1 \sigma_\varepsilon^2 = \frac{(1 \quad \theta_1\phi_1) \quad (\phi_1 \quad \theta_1)}{1 \quad \phi_1^2} \sigma_\varepsilon^2$$

$$r_k = \phi_1 r_{k-1}, \qquad k \geq 2$$

From the r_k equation we have that

$$\rho_k = \begin{cases} 1\ , & k = 0 \\ \dfrac{(1 \quad \theta_1\phi_1)\quad(\phi_1 \quad \theta_1)}{1+\theta_1^2 \quad 2\theta_1\phi_1}, & k = 1 \\ \phi_1\rho_{k-1} = \phi_1^{k-1}\rho_1\ , & k \geq 2, \end{cases}$$

The ACF decays exponentially from the starting ρ_1, which depends on θ_1 as well as on ϕ_1

3. **Partial Autocorrelation Function:** The partial ACF of mixed $ARMA(1,1)$ process consists of a single initial value $\phi_{11} = \rho_1$. thereafter it behaves like the partial autocorrelation function of a pure $MA(1)$ process and it is dominated by a damped exponential. When θ_1 is positive it is dominated by a smoothly damped exponential which decays from a value of ρ_1, with sign determined by the sign of $(\phi_1 \quad \theta_1)$. Similarly, when θ_1 is negative it is dominated by an exponential which oscillates as it decays from a value of ρ_1, with sign determined by the sign of $(\phi_1 \quad \theta_1)$.

Example:

1. The time series X_t satisfies

$$X_t = 1.8x_{t-1} \quad 0.8x_{t-2} + \varepsilon_t + 0.6\varepsilon_{t-1}$$

   a. Classify X_t as a member of the $ARIMA(p,d,q)$ family, i.e. identify p, d and q.
   b. Investigate Stationarity and Invertibility of X_t

2. Consider the following ARIMA model:

$$X_t = 16 + 0.6x_{t-1} + \varepsilon_t + 0.8\varepsilon_{t-1}$$

$$X_t = 16 \quad 0.7x_{t-1} + \varepsilon_t \quad 0.6\varepsilon_{t-1}$$

   a. Check the stationarity and invertibility of X_t
   b. Find the mean and variance X_t
   c. Compute the ACF of the model

3. A linear time series model whose characteristic polynomials have a common root is said to be redundant since it is equivalent to a simpler model. Show that then model

$$X_t = 2.8x_{t-1} \quad 2.6x_{t-2} + 0.8x_{t-3} + \varepsilon_t \quad 0.4\varepsilon_{t-1} \quad 0.6\varepsilon_{t-2}$$

   is redundant, simplify it and correctly classify it as a member of the $ARIMA\ (p,d,q)$family.

   Solution:

$$X_t = 2.8x_{t-1} \quad 2.6x_{t-2} + 0.8x_{t-3} + \varepsilon_t \quad 0.4\varepsilon_{t-1} \quad 0.6\varepsilon_{t-2}$$

$$X_t \quad 2.8x_{t-1} + 2.6x_{t-2} \quad 0.8x_{t-3} = \varepsilon_t \quad 0.4\varepsilon_{t-1} \quad 0.6\varepsilon_{t-2}$$

$$(1 \quad 2.8L + 2.6L^2 \quad 0.8L^3)X_t = (1 \quad 0.4B \quad 0.6B^2)\varepsilon_t$$

From lag polynomial $1 \quad 2.8L + 2.6L^2 \quad 0.8L^3 = 0$ $AR(3)$process using R-software

- Re(polyroot(c(1,-2.8,2.6,-0.8))) the roots are $L_1 = 1, L_2 = 1, L_3 = 1.25$

From lag polynomial$1 \quad 0.4B \quad 0.6B^2 = 0$ of $MA(2)$ process

$$B = \frac{0.4 \pm \sqrt{0.16 + 2.4}}{1.2} = \frac{0.4 \pm \sqrt{2.56}}{1.2} = \frac{0.4 \pm 1.6}{1.2}$$

$$B_1 = \frac{0.4 + 1.6}{1.2} = \frac{5}{3} \qquad and \qquad B_2 = \frac{0.4 \quad 1.6}{1.2} = 1$$

Therefore the root of associated lag polynomial of $MA(2)$ and $AR(3)$ process are share common roots which are $L_1 = L_2 = B_2 = 1$. It indicates that $X_t$is redundant linear time series model.

- ☐ If the $ARMA(p,q)$ model have share common roots for their lag polynomial roots then we reduced the given $ARMA(p,q)$ in to $ARMA(p \quad 1, q \quad 1)$ by **canceling** the common roots in both sides of the lag polynomial.

$$(1 \quad 2.8L + 2.6L^2 \quad 0.8L^3)X_t = (1 \quad 0.4B \quad 0.6B^2)\varepsilon_t$$

$$\frac{[(1 \quad L_1)(1 \quad L)(1.25 \quad L_3)]X_t}{(1 \quad L_2)} = \frac{[(\frac{5}{3} \quad B_1)(1 \quad B_2)]\varepsilon_t}{(1 \quad L_2)}, Snice\ L = B = lag\ operator$$

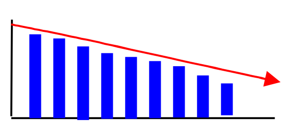
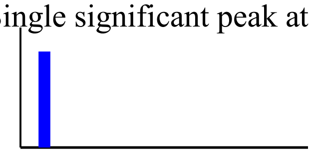
$$(1 \quad L_1)(1.25 \quad L_3)X_t = (\frac{5}{3} \quad B_1)\varepsilon_t$$

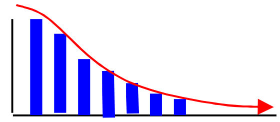
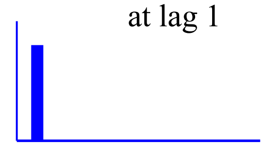
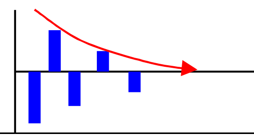
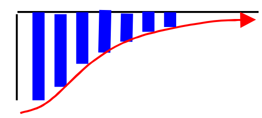
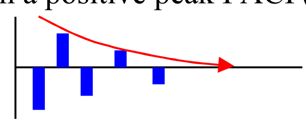
$$(1 \quad L_1)(1.25 \quad L_3)X_t = (\frac{5}{3} \quad B_1)\varepsilon_t \qquad (1 \quad L)(1.25 \quad L)X_t = (\frac{5}{3} \quad B)\varepsilon_t$$

$$(1 \quad 1.8L + 0.8L^2)X_t = (1 + 0.6B)\varepsilon_t$$

$$X_t = 1.8X_{t-1} \quad 0.8X_{t-2} + \varepsilon_t + 0.6\varepsilon_{t-1}$$

$$\therefore T\ \ e\ rduced\ \ model\ is\ ARIMA(2,0,1) = ARMA(2,1)$$

Process(Model)	ACFs	PACFs
ARIMA (0,0,0)	No significant lags	No significant lags
ARIMA (0,1,0)	Linear decline at lag 1, with many lags significant 	Single significant peak at lag 1 

ARIMA (1,0,0) **$1 > \Phi > 0$**	Exponential decline, with first two or many lags significant 	Single significant positive peak at lag 1 
ARIMA (1,0,0) **$-1 < \Phi < 0$**	Alternative exponential decline with a negative peak ACF(1) 	Single significant negative peak at lag 1 
ARIMA (0,0,1) **$1 > \theta > 0$**	Single significant negative peak at lag 1 	Exponential decline of negative value, with first two or many lags significant 
ARIMA (0,0,1) **$-1 < \theta < 0$**	Single significant positive peak at lag 1 	Alternative exponential decline with a positive peak PACF(1) 

We use the standard errors of the sample ACF and PACF to identify non-zero values. Recall that the standard error of the k^{th} sample autocorrelation, r_k is: $se(r_k) = \sqrt{\frac{1+2\sum_{j=1}^{k-1} r_k}{N}}$ and also $se(\hat{\phi}_{kk}) = \sqrt{\frac{1}{N}}$.

Therefore, we would assume that r_k or $\hat{\phi}_{kk}$ to be zero if the absolute value of its estimate is less than twice its standard error. That is $|r_k| \leq 2 \quad se(r_k)$ or $|\hat{\phi}_{kk}| \leq 2 \quad se(\hat{\phi}_{kk})$.

Thus, to test $H_O: \rho_k = 0 \; Vs \; H_1: \rho_k \neq 0$, reject H_O if the value of r_k lies outside the interval and to test $H_O: \phi_{kk} = 0 \; Vs \; H_1: \phi_{kk} \neq 0$, reject H_O if the value of $\hat{\phi}_{kk}$ lies outside the interval

Example1: consider the following ACF and PACF for some time series data with 120 observations and identify statistically the Box-Jenkins model/process.

k	1	2	3	4	5	6	7
r_k	0.709	0.523	0.367	0.281	0.208	0.096	0.132
$se(r_k)$	0.091	0.129	0.146	0.153	0.153	0.153	0.153
$\hat{\phi}_{kk}$	0.709	0.041	-0.037	0.045	-0.007	-0.123	0.204
$se(\hat{\phi}_{kk})$	0.091	0.091	0.091	0.091	0.091	0.091	0.091

As we seen from the table, ACF is tails off and $PACF$ is cuts off at lag 1 and that implies the series follow $AR(1)$ process. That is $X_t = \mu + \phi_1 x_{t-1} + \varepsilon_t$

Example2: consider the following ACF and PACF for a time series with N = 106 and identify the Box-Jenkins model/process.

k	1	2	3	4	5	6	7
r_k	-0.448	0.004	0.052	-0.058	0.094	-0.021	0.012
$se(r_k)$	0.097	0.115	0.115	0.115	0.115	0.115	0.115
$\hat{\phi}_{kk}$	-0.448	-0.423	-0.308	-0.340	-0.224	-0.151	0.009
$se(\hat{\phi}_{kk})$	0.097	0.097	0.097	0.097	0.097	0.097	0.097

As we seen from the table, ACF cuts off at lag 1 while PACF tails off and that implies the series shows $MA(1)$ process is adequate. That is $X_t = \mu + \varepsilon_t \quad \theta_1\varepsilon_{t-1}$.

N.B: for *ARMA* process both ACF and PACF are exponentially decay out ***quickly*** while for *ARIMA* both ACF and PACF are exponentially decay out *slowly*.

6.5 Autoregressive integrated moving average model

ARIMA models (Box-Jenkins models) represent a very powerful and flexible class of models for time series analysis and forecasting

The general $ARIMA(p,d,q)$ model is defined as

$$\left(1 \quad \phi_1 B \quad \phi_2 B^2 \qquad \phi_p B^p\right)\ (1 \quad \mathrm{B})^{\mathrm{d}}\mathrm{x_t} = (1 \quad \theta_1 B \quad \theta_2 B^2 \qquad \theta_q B^q)\varepsilon_\mathrm{t}$$

$AR(p)$ — d differencing — $MA(q)$

$$\left(1 \quad \phi_1 B \quad \phi_2 B^2 \qquad \phi_p B^p\right)\nabla^{\mathrm{d}}\mathrm{x_t} = (1 \quad \theta_1 B \quad \theta_2 B^2 \qquad \theta_q B^q)\varepsilon_\mathrm{t}$$

$$\phi(B)\nabla^{\mathrm{d}}\mathrm{x_t} = \theta(\mathrm{B})\varepsilon_\mathrm{t}$$

Where B *is* backshift operator defined by $\mathrm{B^m x_t = x_{t-m}},\ \mathrm{m} = 0,1,2,\dots,\mathrm{p}$

$\phi(B)$ *is* Autoregressive operator of order p defined as

$$\phi(B) = 1 \quad \phi_1 B \quad \phi_2 B^2 \quad \phi_3 B^3 \qquad \phi_p B^p$$

∇ is the backward differenced operator defined by $\nabla = \mathrm{x_t} \quad \mathrm{x_{t-1}} = (1 \quad \mathrm{B})\mathrm{x_t}$

∇^{d} means the d^{th} difference of the series values $\mathrm{x_t}$

$\theta(B)$ *is* Moving average operator of order q defined as

$$\theta(B) = 1 \quad \theta_1 B \quad \theta_2 B^2 \quad \theta_3 B^3 \qquad \theta_q B^q$$

1. **The $ARIMA(0,1,1)$ process:**

$$\nabla X_t = \varepsilon_t + \theta_1 \varepsilon_{t-1}$$
$$= (1 \quad \theta_1 B)\varepsilon_t$$

Corresponding to $p = 0, d = 1, q = 1, \phi(B) = 1, \theta(B) = 1 \quad \theta_1 B$

2. **The $ARIMA(0, 2, 2)$ process:**

$$\nabla^2 X_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2}$$
$$= (1 \quad \theta_1 B \quad \theta_2 B^2)\varepsilon_t$$

Corresponding to $p = 0, d = 2, q = 2, \phi(B) = 1, \theta(B) = 1 \quad \theta_1 B \quad \theta_2 B^2$

3. **The $ARIMA(1, 1, 1)$ process:**

$$\nabla X_t \quad \phi_1 \nabla X_{t-1} = \varepsilon_t + \theta_1 \varepsilon_{t-1}$$
$$(1 \quad \phi_1 \mathrm{B})\nabla X_t = (1 \quad \theta_1 B)\varepsilon_t$$

Corresponding to $p = 1, d = 1, q = 1, \phi(B) = 1 \quad \phi_1 \mathrm{B}, \theta(B) = 1 \quad \theta_1 B \quad \theta_2 B^2$

6.6 Seasonal ARIMA

Many busines and economic time series contains a seasonal phenomenon that repeats itself after a regular period of time. The smallest time period for this repetitiv phenomenon is called the sseasonal period. For example, the quarterly series of ice-cream sales is high in summer and the series repeats this phenomenon each year,giving a seasonal period of 4. Similarly, monthly auto sales and earning tend to decrease during Auguast and september every year because of the change over new models, and the monthly sales of toys rise every year in the month of December. The seasonal period in these later case is 12. Seasonal phenomenon may stem from factor such as weather,which affects many business and economic activites like tourism and home buliding, cultural events like christmas, which is closely related to sales jewlry,toys, greeting cards and stamps and graduation ceremonies in the summer month,which are directly related to the labour force status in these month.

A seasonal $ARIMA$ model is formed by including additional seasonal terms in the $ARIMA$ models we have seen previous discussion.

The pure seasonal time series is defined as $SARIMA(P, D, Q)_S$

$$\phi_p(B^s)(1 \quad B^s)^D X_t = \theta_0 + \Theta_Q(B^s)\varepsilon_t$$

Where θ_0 is constant

$$\phi_P(B^s) = 1 \quad \phi_1 B^s \quad \phi_2 B^{2s} \quad \phi_3 B^{3s} \qquad \phi_p B^{sP}$$
$$\Theta_Q(B^s) = 1 \quad \Theta_1 B^s \quad \Theta_2 B^{2s} \quad \Theta_3 B^{3s} \qquad \Theta_Q B^{sQ}$$

Example $$SARIMA(0,0,1)_{12} = SMA(1)_{12}$$

$$SARIMA(2,0,1)_{12} = SARMA(2,1)_{12}$$

$$SARIMA(2,1,0)_4 = SARI(2,1)_4$$

- This is a simple seasonal MA model.

$$X_t = \theta_0 + \varepsilon_t \quad \Theta\varepsilon_{t-12}$$

Invertibility condition for seasonal $MA(1)$ process: $|\Theta| < 1$

$$E(X_t) = \theta_0$$

$$Var(x_t) = (1 + \Theta^2)\sigma_\varepsilon^2$$

$$ACF : \rho_{12k} = \begin{cases} \dfrac{\Theta}{1+\Theta^2}, |k| = 12 \\ 0, \quad ot\ erwise \end{cases}$$

- This is a simple seasonal AR model.

$$(1 \quad \Phi B^{12})X_t = \theta_0 + \varepsilon_t$$

Stationarity condition for seasonal $AR(1) process$: $|\Phi| < 1$

Mean of seasonal $AR(1) process$: $E(X_t) = \frac{\theta_0}{1-\Phi}$

Variance of seasonal $AR(1) process$:

$$Var(x_t) = \frac{\sigma_\varepsilon^2}{1 \quad \Phi^2}$$

$$ACF : \rho_{12k} = \Phi^k, k = 0, \pm 1, \pm 2, \ldots$$

When $\Phi = 1$, the series is non-stationary. To test for a unit root, consider seasonal unit root tests.

Multiplicative Seasonal Time Series

A seasonal $ARIMA$ model is formed by including additional seasonal terms in the $ARIMA$ models we have seen previous discussion. A seasonal ARIMA model is classified as an

$$ARIMA(p, d, q)(P, D, Q)_m$$

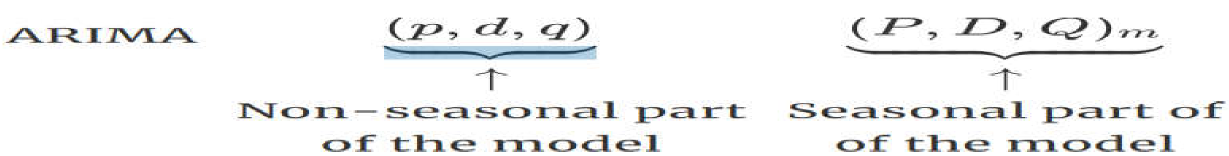
Where (p, d, q) Non-seasonal parts of the model.

$(P, D, Q)_m$ seasonal parts of the model

P =number of seasonal autoregressive (SAR) terms

D =number of seasonal differences

Q =number of seasonal moving average (SMA) terms



Where m = number of observations per year. We use uppercase notation for the seasonal parts of the model, and lowercase notation for the non-seasonal parts of the model.

A special, parsimonious class of seasonal time series models that is commonly used in practice is the multiplicative seasonal model $ARIMA(p,d,q)(P,D,Q)_s$

$$\phi_p(B)\Phi_p(B^s)(1\quad B)^d(1\quad B^s)^D X_t = \theta_0 + \theta_q(B)\Theta_Q(B^s)\varepsilon_t$$

Where all zeros of $\phi(B),\ \Phi(B^s), \theta(B)\ \ and\ \ \Theta(B^s)$ lie outside the unit circle. Of course, there are no common factors between $\phi(B),\ \Phi(B^s)$ and $\theta(B)\ \ and\ \ \Theta(B^s)$

The behavior of sample autocorrelation $(1\quad B^4)(1\quad B)x_t$ is common among seasonal time series. It led to the development of the following special seasonal time series model:

$$(1\quad B^s)(1\quad B)x_t = (1\quad \theta\mathrm{B})(1\quad \Theta B^s)\varepsilon_t$$

Where s is the periodicity of the series, ε_t is a white noise series,$|\ \theta| < 1$ and $|\Theta| < 1$. this model is referred to as the airline model in the literature ;(Box, Jenkins and Reinsel, 1994 chapter 9). It has been found to be widely applicable in modeling seasonal time series. The AR part of the model simply consists of the regular and seasonal difference, where as the MA part involves two parameters. Focusing on the MA part(i.e on the model)

$$w_t = (1\quad \theta\mathrm{B})(1\quad \Theta B^s)\varepsilon_t = \varepsilon_t\quad \theta\varepsilon_t\quad \Theta\varepsilon_{t-s} + \theta\Theta\varepsilon_{t-s-1}$$

Where $w_t = (1\quad B^s)(1\quad B)x_t, s > 1$. It is easy to obtaine that $E(w_t) = 0$

$$Var(w_t) = (1+\theta^2)(1+\Theta^2)\sigma_\varepsilon^2$$

$$Cov(w_t, w_{t-1}) = \quad \theta(1+\Theta^2)\sigma_\varepsilon^2$$

$$Cov(w_t, w_{t-s+1}) = \theta\Theta{\sigma_\varepsilon}^2$$

$$Cov(w_t, w_{t-s}) = \quad \Theta(1+\theta^2)\sigma_\varepsilon^2$$

$$Cov(w_t, w_{t-s-1}) = \theta\Theta\sigma_\varepsilon^2$$

$$Cov(w_t, w_{t-L}) = 0, for\ \ L \neq 0,1,s\quad 1,s,s+1$$

Consequently, the ACF of w_t series is given by

$$\rho_1 = \frac{\theta}{1+\theta^2},\ \ \rho_s = \frac{\Theta}{1+\Theta^2},\ \ \rho_{s-1} = \rho_{s+1} = \rho_s\rho_1 = \frac{\theta\Theta}{(1+\theta^2)(1+\Theta^2)}$$

And $\rho_L = 0$ for $L > 0$ and $L \neq 1, s\quad 1, s, s+1$

For example, if w_t is a quarterly time series, then $s = 4$ and $L > 0$, the ACF of ρ_L is nonzero at logs $1, 3, 4\ and\ 5$ only. It is intresting to compare the prior ACF with those of the MA(1) model

$x_t = (1 \quad \theta \mathrm{B})\varepsilon_t$ and the MA(s) model $z_t = (1 \quad \Theta B^s)\varepsilon_t$. The ACFs of x_t and z_t series are

$$\rho_{1(x)} = \frac{\theta}{1+\theta^2}, \qquad \rho_{l(x)} = 0, l > 1$$

$$\rho_{s(z)} = \frac{\Theta}{1+\Theta^2}, \qquad \rho_{l(z)} = 0, l > 0 \; and \; l \neq s$$

Example: let us consider the $ARIMA(0,1,1)(0,1,1)_{12}$ is given as

$$(1 \quad B)(1 \quad B^{12})X_t = (1 \quad \theta B)(1 \quad \Theta B^{12})\varepsilon_t$$

Find the autocovariance function and autocorrelation function (ACF) for $\theta = 0.4 \; and \; \Theta = 0.6$.

Solution:

The autocovariance of $w_t = (1 \quad \theta B)(1 \quad \Theta B^{12})\varepsilon_t$ can be easily found to be

$$r_0 = (1+\theta^2)(1+\Theta^2)\sigma_\varepsilon^2$$

$$r_1 = \quad \theta(1+\Theta^2)\sigma_\varepsilon^2$$

$$r_{11} = \theta\Theta\sigma_\varepsilon^2$$

$$r_{12} = \quad \Theta(1+\theta^2)\sigma_\varepsilon^2$$

$$r_{13} = \theta\Theta\sigma_\varepsilon^2$$

$$r_j = 0, for \;\; j \neq 0,1, s \quad 1, s, s+1$$

The autocorreletion function

$$\rho_k = \begin{cases} \dfrac{\theta}{1+\theta^2}, & k = 1 \\ \dfrac{\Theta}{1+\Theta^2}, & k = 12 \\ \dfrac{\theta\Theta}{(1+\theta^2)(1+\Theta^2)}, & k = 11,13 \\ 0, & ot\;\; erwise \end{cases}$$

6.7 Estimation of parameters

There are several methods such as methods of moments, maximum likelihood, Yule-walker and least squares that can be employed to estimate the parameters in the tentatively identified model. All of them should produce very similar estimates, but may be more or less efficient for any given model. In general, during the parameter estimation phase a function minimization algorithm is used to maximize the likelihood (probability) of the observed series, given the parameter values.

6.7.1 Parameter Estimation of AR Process

Once the order p is determined, the model parameters $\phi_1, \phi_2, \phi_3, \dots, \phi_p$ are estimate by minimizing $g(\phi_1, \phi_2, \phi_3, \dots, \phi_p)$ over the parameter space. Manually we can estimate the coefficients from Yule-Walker equations as follows.

$$\begin{matrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-2} \\ \rho_2 & \rho_1 & 1 & \cdots & \rho_{k-3} \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & 1 \end{matrix} \quad \begin{matrix} \phi_1 \\ \phi_2 \\ \phi_3 \\ \cdots \\ \phi_k \end{matrix} = \begin{matrix} \rho_1 \\ \rho_2 \\ \rho_3 \\ \cdots \\ \rho_k \end{matrix}, k = 1,2,3, \dots$$

Then the estimates of the coefficients are computed from the sample autocorrelation. This is computed as:

$$\hat{\phi}_k = \frac{r_k \quad \sum_{j=1}^{k\ 1}\left(r_{k\ 1,j}\right) r_{k\ j}}{1 \quad \sum_{j=1}^{k\ 1}(r_{k\ 1}) r_{k\ j}}$$

AR (1) Process: We know that $\rho_k = \phi_1^k$, if r_1 is given, then $r_1 = \hat{\phi}_1$.

***AR*(2) Process:** We know from the Yule-Walker equations that:

$$\rho_{(1)} = \phi_1 + \phi_2 \rho_{(1)}$$

$$\rho_{(2)} = \phi_1 \rho_{(1)} + \phi_2$$

$$\Rightarrow \hat{\phi}_1 = \frac{r_1(1 \quad r_2)}{1 \quad r_1^2}$$

$$\hat{\phi}_2 = \frac{r_2 \quad r_1^2}{1 \quad r_1^2}$$

Example: Given $r_1 = 0.81$ and $r_2 = 0.43$, then we can estimate the parameters and fit the process as follows.

Solution:

$$\hat{\phi}_1 = \frac{r_1(1 \quad r_2)}{1 \quad r_1^2} = \frac{0.81(1 \quad 0.43)}{1 \quad 0.81^2} = 1.32$$

$$\hat{\phi}_2 = \frac{r_2 \quad r_1^2}{1 \quad r_1^2} = \frac{0.43 \quad 0.81^2}{1 \quad 0.81^2} = \quad 0.63$$

Therefore, the model is given below as: $x_t = 1.32x_{t-1} - 0.63x_{t-2} + \varepsilon_t$

6.7.2 Parameter Estimation of MA Process

There is no closed form solution for MA process. But we can compute it iteratively and compute MSEs. Another approach is to rewrite the $MA(q)$ process as $AR(\infty)$ and estimate the AR process with large p-value.

$MA(1)$ Process has the form $X_t = \varepsilon_t \quad \theta_1 \varepsilon_{t-1}$

Recall: $\rho_1 = \frac{-\theta_1}{1+\theta_1^2}$ $\qquad \theta_1 = \rho_1(1+\theta_1^2)$

$$\rho_1\theta_1^2 + \theta_1 + \rho_1$$

$$\hat{\theta}_1 = \frac{1 \pm \sqrt{1 \quad 4\rho_1^2}}{2\rho_1}$$

6.8 Diagnostic checking

Diagnostic checking is the process of checking the adequacy of the fitted model against a range of criteria and possibly returning to the identification stage to respecify the model. After a tentative model has been fit to the data, we must examine its adequacy and, if necessary, suggest potential improvements.

6.8.1 Autocorrelation Check (Studying residuals)

If the specified model is adequate and hence the appropriate orders p and q are identified, then the residuals: $\varepsilon_t = X_t + (\mu + \sum_{i=1}^{p} \hat{\phi}_i X_{t-i} \quad \sum_{i=1}^{q} \hat{\theta}_i \varepsilon_{t-i})$ should behave like white noise. This means that if the model is appropriate, then the residual sample autocorrelation function should have no structure to identify. That is, the autocorrelation should not differ significantly from zero for all lags. This means $|se(r_k)| \leq 2 \quad \sqrt{\frac{1}{N}}$.

Here

$$se(\rho_{k(e)}) = \sqrt{\frac{1+2\sum_{k=1}^{k} \rho_{k(e)}}{n}}, \text{ if no difference in the ARIMA model}$$

$$se(\rho_{k(e)}) = \sqrt{\frac{1+2\sum_{k=1}^{k} \rho_{k(e)}}{n-d}}, \text{ if there is difference in the ARIMA model}$$

But $\rho_{k(e)} = 0$, we assume that the process is white noise

$$se(\rho_{k(e)}) = \sqrt{\frac{1}{n}}, \text{ if no difference in the ARIMA model}$$

$$se(\rho_{k(e)}) = \sqrt{\frac{1}{n-d}}, \text{ if there is difference in the ARIMA model}$$

6.8.2 An Approximate Chi-squareTest

Ruther than considering the $se(r_k)$ terms individually, we may obtain an individual of whether the 1^{st} K residual autocorrelations considered together indicate adequacy of the model. This is done by an appropriate chi-square test. The test statistic is: $Q = (N \quad d)\sum_{k=1}^{K} r_k^2(\varepsilon) \sim \chi_{K-p-q}^2$ and we can decide reject the null hypothesis (Model Adequacy) if Q greater than the chi-square tabulated value with df $(K \quad p \quad q)$ and alpha value.

6.8.3 The Ljung-Box Test

In addition to looking at residual correlations at individual lags, it is useful to have a test that takes into account their magnitudes as a group. For example, it may be that most of the residual autocorrelations are moderate, some even close to their critical values, but, taken together, they seem excessive. Box and Pierce (1970) proposed the statistic

$$Q = T\sum_{k=1}^{h} \rho_k^2(e) \ = T[\rho_1^2(e) + \rho_2^2(e) + \quad + \rho_h^2(e)]$$

By modifying the Q statistic slightly, they defined a test statistic whose null distribution is much closer to chi-square for typical sample sizes. The modified Box-Pierce is **Ljung-Box,** statistic is given by

$$Q \ = T(T+2)\sum_{k=1}^{h} \frac{\rho_k^2(e)}{T \quad k} = T(T+2)\left[\frac{\rho_1^2(e)}{T \quad 1} + \frac{\rho_2^2(e)}{T \quad 2} + \quad + \frac{\rho_h^2(e)}{T}\right]$$

For large n, Q has an approximate chi-square distribution with $K \quad p - q$, degrees of freedom.

NB: If $\frac{(n+2)}{(n-k)} > 1$ for every $k \geq 1$, we have $Q \quad > Q$

We reject the null hypothesis if the Q statistics is greater than the Chi-square distribution with $K \quad p - q$, degrees of freedom, $i.e\ Q > \chi^2_{\alpha,(T-p-q)}$ when H_0 defined as the model is adequately fitted.

Example: From the weekly total number of loan applications in a local branch of a Commercial bank of Ethiopia in Debre Markos for the last two years. It is suspected that there should be some relationship (i.e., autocorrelation) between the number of applications in the current week and the number of loan applications in the previous weeks. Modeling that relationship will help the management to proactively plan for the coming weeks through reliable forecasts $ARIMA(2,1,1)$ is given by

$$\nabla x_t = 0.2121\nabla x_{t-1} + 0.3591\nabla x_{t-2} + \varepsilon_t \quad 0.9474\varepsilon_{t-1}$$

and the first 20 sample autocorrelation function of the residual are summarized as below which was fitted to a series with 104 observations. Test that the model is fitted adequately by using

1. Approximate chi-square statistics
2. Residual autocorrelation
3. Box-Pierce and Ljung-Box statistics method.

Lag	$\hat{\rho}_{k(e)}$	k	$\hat{\rho}_{k(e)}$
1.	0.037	11	-0.106
2.	0.042	12	-0.016
3.	-0.090	13	0.001
4.	-0.076	14	0.132
5.	-0.038	15	-0.073
6.	-0.022	16	0.103
7.	0.102	17	-0.036
8.	-0.064	18	0.120
9.	0.044	19	-0.028
10.	-0.132	20	0.052
Box-Pierce test			
$Q = \chi \quad squared = 12.174, \quad df = 20, \quad p \quad value = 0.91$			
Box-Ljung test			
$Q = \chi \quad squared = 13.973, \quad df = 20, \quad p \quad value = 0.8319$			

Solution

- **Using an Approximate Chi-square Test**

Hypothesis testing: H_0: The model is adequately fitted Vs

H_1: The Model is not fitted adequately

$$\sum_{k=1}^{20} \rho_{k(\mathrm{e})}^2 = 0.001369 + 0.001764 + 0.0081 + \quad + 0.000784 + 0.002704 = 0.1168368$$

Test statistic: $Q = (N \quad d) \sum_{k=1}^{25} \rho_{k(\mathrm{e})}^2 = (N \quad d) \quad 0.1168 = (103 \quad 1) \quad 0.1168 = 12.343$

Decision: Reject H_0 if Q is greater than the critical value $(\chi^2_{0.05,(20-2-1)}) = 27.5871$.

Here $Q = 12.343 < \chi^2_{0.05,(20-2-1)} = 27.5871$, we do not reject H_0

Conclusion: since the calculated value is less than the critical value, there is no enough evidence to say that model is not adequately fitted at 5% significant level.

- **Using Residuals Autocorrelation**

Hypothesis testing: $H_O: \rho_k = 0 \; Vs \; H_1: \rho_k \neq 0$

$$se(\rho_{k(e)}) = \sqrt{\frac{1}{n \quad d}} = \sqrt{\frac{1}{104 \quad 1}} = \sqrt{\frac{1}{103}} = 0.0985$$

Test statistic: The 95% confidence interval is $\pm 2se(\rho_{k(e)}) = \pm 2 \quad 0.0985 = \pm 0.1970$

Decision rule: Reject H_0 if the $\rho_{k(e)}$ value is lies out of the interval $\pm 2se(\rho_{k(e)})$

Conclusion: As we have seen from the above table all standard errors of the residual is inside the confidence interval. Therefore, we can conclude that the residuals are random and the model is fitted adequately.

6.8.4 Over fitting and Parameter Redundancy

The other basic model diagnostic tool is over fitting of the model and redundancy of parameters in the model.

After specifying and fitting what we believe to be an adequate model, we fit a slightly more general model; that is, a model close by that contains the original model as a special case. For example, if an $AR(2)$ model seems appropriate, we might overfit with an $AR(3)$ model. The original $AR(2)$ model would be confirmed if:

1. The estimate of the additional parameter, ϕ_3, is not significantly different from zero, and
2. The estimates for the parameters in common, ϕ_1 and ϕ_2, do not change significantly from their original estimates

Thus the AR and MA characteristic polynomials in the ARMA(2,3) process have a common factor of (1 - *cx*). Even though *Yt* does satisfy the ARMA(2,3) model, clearly the parameters in that model are not unique the constant c is completely arbitrary. We say that we have **parameter redundancy** in the $ARMA(2{,}3)$ model. The implications for fitting and overfitting models are as follows:

1. Specify the original model carefully. If a simple model seems at all promising, check it out before trying a more complicated model.
2. When overfitting, do not increase the orders of both the AR and MA parts of the model simultaneously
3. Extend the model in directions suggested by the analysis of the residuals. For example, if after fitting an MA(1) model, substantial correlation remains at lag 2 in the residuals, try an MA(2), not an ARMA(1,1).

Exercise: The first 25 sample autocorrelations of residuals from an $IMA\ (0, 2, 2)$ process: $\nabla^2 X_t = (1 \quad 0.13B \quad 0.12B^2)\varepsilon_t$ are summarized as below which was fitted to a series with $N\ =\ 226$ observations. Then test the model adequacy by using approximate chi-square statistic.

k	$\rho_{k(e)}$	k	$\rho_{k(e)}$
1	0.03	14	0.022
2	0.002	15	-0.006
3	0.032	16	-0.089
4	0.05	17	0.133
5	-0.078	18	-0.092
6	-0.11	19	-0.005
7	-0.133	20	-0.015
8	-0.033	21	0.007
9	-0.138	22	0.132
10	-0.098	23	0.012
11	-0.129	24	-0.012
12	0.063	25	-0.127
13	-0.084		

Solution

- **Using an Approximate Chi-square Test**

Hypothesis testing:H_0: The model is Adequate Vs H_1: the Model is not Adequate

$$\sum_{k=1}^{25} \rho_{k(\mathrm{e})}^2 = 0.0009 + 000004 + \quad + 0.000144 + 0.016129 = 0.166$$

Test statistic: $Q = (N \quad d)\sum_{k=1}^{25} \rho_{k(\mathrm{e})}^2 = (N \quad d) \quad 0.166 = (226 \quad 2) \quad 0.166 = 37.184.$

Decision: reject H_0 if Q is greater than the critical value $(\chi^2_{(1-0.05),(25-0-2)})\chi^2_{0.05,(23)} = 35.1725$

Conclusion: since the calculated value is greater than the critical value, there is enough evidence that model is not adequately fitted at 5% significant level.

- **Using Residuals Autocorrelation**

Hypothesis testing: $H_O: \rho_k = 0 \; Vs \; H_1: \rho_k \neq 0$

$$se(\rho_{k(e)}) = \sqrt{\frac{1}{n \quad d}} = \sqrt{\frac{1}{226 \quad 2}} = \sqrt{\frac{1}{224}} 0.066815 \approx 0.067$$

Test statistic: The 95% confidence interval is $\pm 2se(\rho_{k(e)}) = \pm 2 \quad 0.067 = \pm 0.134$

Decision rule: reject the null hypothesis if the $\rho_{k(e)}$ value is lies out of the interval $\pm 2se(\rho_{k(e)})$

Conclusion: As we have seen from the above table all standard errors of the residual is inside the confidence interval except the 9^{th} observation. Therefore, we can conclude that the residuals are random and intern to that the model is adequate.

CHAPTER SEVEN

7 FORECASTING

7.1 Introduction

Forecasting is one of the main objectives of univariate and multivariate time series analysis for horizons ≥ 1. Forecasting the future values of an observed time series is an important problem in many areas, including economics, production planning, sales forecasting and stock control. Forecasting problems are often classified as short-term, medium-term, and long-term. Short-term forecasting problems involve predicting events only a few time periods (days, weeks, months) into the future. Medium-term forecasts extend from one to two years into the future, and long-term forecasting problems can extend beyond that by many years.

Forecasting methods may be broadly classified into three groups as follows:

- **Subjective**-forecasts can be made on a subjective basis using judgement, intuition, commercial knowledge or any other relevant information
- **Univariate**- forecasts under this method is done based on the model from a given time series variable, so that $\hat{Y}_{T+k}$ depends only on the values of $Y_T, Y_{Y-1}, Y_{t-2}, \ldots$, possibly augmented by a simple function of time. Methods of this type are sometimes called *naive* or *projection* methods
- **Multivariate**-the forecasts of a given variable depends on the value of two or more other data series called predictors and explanatory variables. For example, sales forecasts may depend on stocks and/or on economic indices. Models of this type are sometimes called causal models.

In general, the choice of method depends on a variety of considerations, including:

- ✓ How the forecast is to be used.
- ✓ The type of time series (e.g. macroeconomic series or sales figures) and its properties (e.g. are trend and seasonality present?). Some series are very regular and hence 'very predictable', but others are not. As always, a time plot of the data is very helpful.
- ✓ How many past observations are available? They are also sometimes called independent variables but this terminology is misleading, as they are typically not independent of each other.
- ✓ The length of the forecasting horizon (lead time). For example, in stock control the lead time for which forecasts are required is the time between ordering an item and its delivery.

- ✓ The number of series to forecast and the cost allowed per series.
- ✓ The skill and experience of the analyst. Analysts should select a method with which they feel 'happy' and for which relevant computer software is available. They should also consider the possibility of trying more than one method.

Definitions of Terms

- ➢ Forecast Period:- the basic unit of time for which forecast are made.
- ➢ Forecast interval:- the frequency with which new forecast are prepared
- ➢ Forecast lead time: the no of periods in the future covered by the forecast
- ➢ Forecast horizon:- the length of time in to the future for which forecast are made

7.2 Needs and use of forecasting

Frequently there is a time lag between awareness of impending event or need and occurrences of that event. This lead time is the main reason for planning & forecasting. If the lead time is long planning or play an important role. Suppose the time series $Y_1, Y_2, \dots, Y_T$ the future value Y_{T+K} made at time 'T' for 'k' steps where k is a lead time.

A wide Varity of methods are available ranging from the most simplified methods, such as the use of the most recent observation as a forecast to highly complex approaches such as econometrics systems of simultaneous equations.

Some areas in which forecasting plays important role are

1. to scheduling existing resources
2. Acquiring additional resources
3. Determine what resources are desired

Based on the above consideration some forecasting methods are:

- ☞ Averaging Methods- the Mean, Simple Moving Average, double Moving Average;
- ☞ Exponential Smoothing- Single and Double Exponential Smoothing;
- ☞ Box-Jenkins Methods.

Whatever forecasting method is used, some sort of forecast monitoring scheme is often advisable, particularly with large numbers of series, to ensure that forecast errors are not systematically positive or negative. We will see these methods one by one in the next sections of this chapter.

7.3 Some Forecasting Methods

7.3.1 Average Methods

A. The Mean

When we forecast using the mean, data must be stationary, and the variance is stable. In other words, the data must have not trend and seasonality. Given the data set covering the start time periods the forecast of the observation in some future period $T+k$ would be $\hat{Y}_{T+k} = \bar{Y} = \frac{\sum_{t=1}^{T} Y_t}{T}$.

B. Simple Moving Averages

Given historical data and a decision to use the most recent N observations for each average, then the forecast for some future period '$T+k$' becomes $\hat{Y}_{T+K} = M_T$

Example: consider the following data and use the three-period simple moving average. Using one-period ahead forecast.

T	1	2	3	4	5	6	7	8	9	10
Y_t	14	15	10	14	17	12	15	11	12	19
$3MA$		13.00	13.00	13.67	14.33	14.67	12.67	12.67	14.00	
$\hat{Y}_{T+k=1}$			13.00	13.00	13.67	14.33	14.67	12.67	12.67	14.00

Solution*:*

T	1	2	3	4	5	6	7	8	9	10
3MA	-	13.00	13.00	13.67	14.33	14.67	12.67	12.67	14.00	-
$\hat{Y}_{T+k=1}$	-	-	13.00	13.00	13.67	14.33	14.67	12.67	12.67	14.00

Therefore $\hat{Y}_{T+k=1} = \hat{Y}_{10+k=1} = \hat{Y}_{10} = 14.00$

C. Double Moving Average

The forecasting for period $T+k$ is obtained by extrapolating the trend k periods in to the future according to $\hat{Y}_{T+K} = \hat{Y}_{T+K} \quad \hat{\beta}$.

Thus, the forecasting equation becomes from the estimation of trend by using double moving average methods.

Recall that $\hat{Y}_T = \hat{\beta}_0 + \hat{\beta}_1 T = 2 \quad M_T^{(1)} \quad M_T^{(2)}$,

$$w \ \ ere \ \hat{\beta}_0 = 2 \quad M_T^{(1)} \quad M_T^{(2)} \quad T\hat{\beta}_{1(0)}$$

$$\hat{\beta}_1 = \frac{2}{N \quad 1}\left(M_T^{(1)} \quad M_T^{(2)}\right)$$

K is the period for moving the series.

Therefore, after some mathematical operation the forecasting formula is given as:

$$\hat{Y}_{T+K} = 2 \quad M_T^{(1)} \quad M_T^{(2)} + K\left(\frac{2}{N-1}\right)\left(M_T^{(1)} \quad M_T^{(2)}\right) = (2 + \frac{2k}{N-1})M_T^{(1)} \quad (1 + k\frac{2}{N-1})M_T^{(2)}.$$

Example: consider the following series and forecast in to the future using double moving average methods in to K=3.

T	1	2	3	4	5	6	7	8	9	10	11	12
Y_t	35	46	51	46	48	51	46	42	41	43	61	55

Solution:

T	1	2	3	4	5	6	7	8	9	10	11	12
3 $M_t^{[1]}$	-	44	47.67	48.33	48.33	48.33	46.33	43	42	48.33	53	-
3 $M_t^{[2]}$	-	-	46.67	48.11	48.33	47.67	45.89	43.78	44.44	47.78	-	-
$\hat{Y}_t$	-	-	48.67	48.55	48.33	48.99	46.77	42.22	39.56	48.88	-	-

$$\hat{Y}_{10+k} = \left(2 + \frac{2k}{3\quad 1}\right) M_{10}^{(1)} \quad \left(1 + \frac{2k}{3\quad 1}\right) M_{10}^{(2)} = 48.88 + 0.55k$$

$$\hat{Y}_{10+1} = 48.88 + 0.55 = 49.43$$

$$\hat{Y}_{10+2} = 48.88 + 0.55 \quad 2 = 49.98$$

$$\hat{Y}_{10+3} = 48.88 + 0.55 \quad 3 = 50.53$$

7.3.2 Exponential Smoothing Methods

A. Single (simple) Exponential Smoothing

Exponential smoothing (ES) is the name given to a general class of forecasting procedures that rely on simple updating equations to calculate forecasts. The most basic form, introduced in this subsection, is called **simple exponential smoothing** (SES), but this should only be used for non-seasonal time series showing no systematic trend. Of course many time series that arise in practice do contain a trend or seasonal pattern, but these effects can be measured and removed to produce a stationary series for which simple ES is appropriate. Alternatively, more complicated versions of ES are available to cope with trend and seasonality. Thus adaptations of exponential smoothing are useful for many types of time series.

Given a non-seasonal time series, say Y_1, Y_2 ,…, Y_n, with no systematic trend, it is natural to forecast $Y_{n+1}$by means of a weighted sum of the past observations. It seems sensible to give more weight to recent observations and less weight too bservations further in the past. An intuitively appealing set of weights are geometric weights, which decrease by a constant ratio for every unit increase in the lag. In order that the weights sum to one, we take$\alpha(1 \quad \alpha)^i$, i=0, 1, 2,…where α is a constant such that 0<α<1. Then the forecasting value is based on exponential smoothing equation expressed as: $\hat{Y}_{T+1} = \alpha Y_T + \alpha(1-\alpha)Y_{T-1} + \alpha(1-\alpha)^2 Y_{T-2} + \ldots$ Strictly speaking, this implies an infinite number of past observations, but in practice there will only be a finite number. Thus it is customarily rewritten in the **recurrence** form as: $\hat{Y}_{T+1} = \alpha Y_T + (1 \quad \alpha)\hat{Y}_{T-1} = \hat{Y}_T$ and it is called simple exponential smoothing. This is because of the assumption of constant mean

model. The adjective 'exponential' arises from the fact that the geometric weights lie on an exponential curve, but the procedure could equally well have been called geometric smoothing.
The above equation sometimes rewritten in the equivalent **error-correction** form

$$\hat{Y}_{T+1} = \alpha\left(Y_T \quad \hat{Y}_{T-1}\right) + \hat{Y}_{T-1} = \hat{Y}_{T+1} = \alpha\varepsilon_T + \hat{Y}_{T-1}$$

where ε_T is the prediction error at time T.
The value of the smoothing constant α depends on the properties of the given time series. Values between 0.1 and 0.3 are commonly used and produce a forecast that depends on a large number of past observations (Chatfield, 2001).

Example: find the forecast for period 11 based on the following series using the single exponential smoothing method with $\alpha = 0.2$ and $\hat{Y}_0 = \bar{Y}$.

T	1	2	3	4	5	6	7	8	9	10
Y_t	14	15	10	14	17	12	15	11	12	19

Solution:

T	1	2	3	4	5	6	7	8	9	10
$\hat{Y}_t$	13.84	14.07	13.26	13.41	14.12	13.70	13.96	13.37	13.09	14.28

Therefore, the forecast value at $T + k$ (i. e. $T + k = 10 + 1 = 11$) is

$$\hat{Y}_{T+1} = \alpha Y_T + (1 \quad \alpha)\hat{Y}_{T-1} = \hat{Y}_T$$

$$\hat{Y}_{11} = 0.2 \quad Y_{10} + (1 \quad 0.2)\hat{Y}_9 = \hat{Y}_{10} = 14.28$$

$$\hat{Y}_{T+1} = \alpha\varepsilon_T + \hat{Y}_{T-1} = \alpha\left(Y_T \quad \hat{Y}_{T-1}\right) + \hat{Y}_{T-1} = 0.2(19 \quad 13.09) + 13.09 = 14.28$$

B. Double Exponential Smoothing

Exponential smoothing may readily be generalized to deal with time series containing trend and seasonal variation. The version for handling a trend with non-seasonal data is usually called Holt's (two-parameter) exponential smoothing, while the version that also copes with seasonal variation is usually referred to as the Holt-Winters (three-parameter) procedure. The general idea is to generalize the equations for SES by introducing trend and seasonal terms, which are also updated by exponential smoothing.
Therefore, to forecast k-periods in to future using double exponential smoothing method, we use the following forecasting equation.

$$\boldsymbol{\hat{Y}_{T+k} = \hat{Y}_T + k \quad b_1}, \textbf{ where } \boldsymbol{\hat{Y}_t = 2 \quad S_T^{(1)} \quad S_T^{(2)}} \text{ and } \boldsymbol{b_1 = \frac{\alpha}{\beta}\left(S_T^{(1)} \quad S_T^{(2)}\right)}$$

$$\boldsymbol{= 2 \quad S_T^{(1)} \quad S_T^{(2)} + k\frac{\alpha}{\beta}\left(S_T^{(1)} \quad S_T^{(2)}\right)}$$

$$= \left(2 + k\frac{\alpha}{\beta}\right) S_T^{(1)} \quad \left(1 + k\frac{\alpha}{\beta}\right) \quad S_T^{(2)}, \qquad where\ \beta = 1 \quad \alpha$$

***Example**:* the number of items of a certain product sold by a department store is given below in the thousands of units: use $\alpha = 0.2$in order to forecast the next few years.

Year	1983	1984	1985	1986	1987	1988	1999	1990	1991	1992	1993	1994
T	1	2	3	4	5	6	7	8	9	10	11	12
Y_t	12	11	14	13	16	15	18	17	20	19	22	21

Solution: when you refer your chapter three of double exponential smoothing methods to estimate trend component, the results are given below.

T	1	2	3	4	5	6	7	8	9	10	11	12
$S_t^{(1)}$	7.54	8.23	9.38	10.10	11.28	12.02	13.22	13.98	15.18	15.94	17.15	17.92
$S_t^{(2)}$	3.57	4.50	5.48	6.40	7.38	8.31	9.29	10.23	11.22	12.16	13.16	14.11

It can be shown that $\hat{Y}_t = 10.26 + 0.96t$ is the least squares expression for the series. Then using

$$S_0^{(1)} = b_0 \quad \frac{\beta}{\alpha} \quad b_1 = 10.26 \quad \frac{0.8}{0.2} \quad 0.96 = 6.42$$

$$S_0^{(2)} = b_0 \quad 2 \quad \frac{\beta}{\alpha} \quad b_1 = 0.26 \quad 2 \quad \frac{0.8}{0.2} \quad 0.96 = 2.58$$

Therefore, $\hat{Y}_{T+k} = (2 + k\frac{\alpha}{\beta})S_T^{(1)} \quad (1 + k\frac{\alpha}{\beta})\ S_T^{(2)}$

$\Rightarrow \hat{Y}_{T+K} = (2 + 0.25k)\ S_T^{(1)} \quad (1 + 0.25k)\ S_T^{(2)}$

$= (2 + 0.25k) \quad 17.92 - (1 + 0.25k) \quad 14.11$

$= 21.74 + 0.95k$

For example, $\hat{Y}_{12+1} = 21.74 + 0.95 \quad 1 = 24.69$

$\hat{Y}_{12+2} = 21.74 + 0.95 \quad 2 = 23.64$, thousands of items and soon.

7.3.3 Box-Jenkins Method

Here we will see the forecasting procedure based on ARIMA models which is usually known as the Box-Jenkins Approach. A forecast is obtained by taking expectation at origin T of the model written at time T+k. A major contribution of Box and Jenkins has been to provide a general strategy for time-series forecasting, which emphasizes the importance of identifying an appropriate model in an iterative way.

Procedures:

- Replace the Y_{t+j} that have not occurred at time T by the forecasts $\widehat{Y}_{T+j}$.
- Replace the ε_{T+j} that have not occurred at time T by zero and then ε_{T-j} that have occurred by the single period forecast error. That is $\varepsilon_{1(T-j)} = Y_{t-j} \quad \widehat{Y}_{T-j}$
- In starting the forecasting process assume that $\varepsilon_{(T-j)} = 0,\ T \quad j \leq 0$

Forecasts with $AR(1)$ Process

For this process, it holds that $X_t = \mu + \phi_1 X_{t-1} + \varepsilon_t$, with $|\phi_2| < 1$.

The optimal k-step-forecast is the conditional mean of X_{T+k}, i.e.

$$E[X_{T+k}] = E[\mu + \phi_1 X_{t+k-1} + \varepsilon_{t+k}]$$
$$= \mu + \hat{\phi}_1 E(X_{t+k-1}) + 0$$

We get the following first order difference equation for the prediction function which can be solved recursively:

$$E(X_{T+1}) = \hat{X}_{T+1} = \mu + \hat{\phi}_1 X_T$$
$$E(X_{T+2}) = \hat{X}_{T+2} = \mu + \hat{\phi}_1 \hat{X}_{T+1}$$
$$E(X_{T+3}) = \hat{X}_{T+3} = \mu + \hat{\phi}_1 \hat{X}_{T+2}$$

In general, $E(X_{T+K}) = \hat{X}_{T+K} = \mu + \hat{\phi}_1 \hat{X}_{T+K-1}, k \geq 1$

Forecasts with $AR(2)$ Process

For this process, it holds that $X_t = \mu + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \varepsilon_t$ with

$$\phi_1 + \phi_2 < 1$$
$$\phi_2 \quad \phi_1 < 1$$
$$|\phi_2| < 1$$

The optimal k-step-forecast is the conditional mean of X_{T+k}, i.e.

$$E[X_{T+k}] = E[\mu + \phi_1 X_{t+k-1} + \phi_2 X_{t+k-2} + \varepsilon_{t+k}]$$
$$= \mu + \hat{\phi}_1 E(X_{t+k-1}) + \hat{\phi}_2 E(X_{t+k-2}) + 0$$

We get the following first order difference equation for the prediction function which can be solved recursively:

$$E(X_{T+1}) = \hat{X}_{T+1} = \mu + \hat{\phi}_1 X_T + \hat{\phi}_2 X_{T-1}$$
$$E(X_{T+2}) = \hat{X}_{T+2} = \mu + \hat{\phi}_1 \hat{X}_{T+1} + \hat{\phi}_2 X_T$$
$$E(X_{T+3}) = \hat{X}_{T+3} = \mu + \hat{\phi}_1 \hat{X}_{T+2} + \hat{\phi}_2 \hat{X}_{T+1}$$

In general, $E(X_{T+K}) = \hat{X}_{T+K} = \mu + \hat{\phi}_1 \hat{X}_{T+K-1} + \hat{\phi}_2 \hat{X}_{T+K-2}, k \geq 3$

Forecasts with $AR(p)$ Process

Starting with the representation: $X_t = \mu + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \phi_3 X_{t-3} + \quad + \phi_p X_{t-p} + \varepsilon_t$

The conditional mean of X_{T+k} is given by:

$$E(X_{T+k}) = \hat{X}_{T+k} = \mu + \hat{\phi}_1 E(X_{t+k-1}) + \hat{\phi}_2 E(X_{t+k-2}) + \quad + \hat{\phi}_p E(X_{t+k-p}) + 0$$

Thus, the above equation can be solved recursively:

$$E(X_{T+1}) = \hat{X}_{T+1} = \mu + \hat{\phi}_1 X_t + \hat{\phi}_2 X_{t-1} + \quad + \hat{\phi}_p X_{t+1-p}$$

$$E(X_{T+2}) = \hat{X}_{T+2} = \mu + \hat{\phi}_1 X_{t+1} + \hat{\phi}_2 X_t + \hat{\phi}_3 X_{t-1} + \quad + \hat{\phi}_p X_{t+1-p}$$

Examples

1. A time series model has fitted to some historical data and yielding $X_t = 25 + 0.34x_{t-1} + \varepsilon_t$ suppose that at time $T = 100$, the observation is $X_{100} = 28$, then
   A. Determine forecasts for periods $101, 102, 103$ etc.
   B. Suppose $X_{101} = 32$, revise your forecasts for periods$102, 103, 104$,..., using period 101 as the new origin of time T.

Solution:

A. Given T = 100, then $X_{100} = 28$ and $\hat{X}_{T+K} = \mu + \hat{\phi}_1 \hat{X}_{T+K-1},\ k \geq 1$

- $\hat{X}_{100+1} = 25 + 0.34\hat{X}_{100+1-1} = 25 + 0.34\hat{X}_{100} = 34.52$
- $\hat{X}_{100+2} = 25 + 0.34\hat{X}_{100+2-1} = 25 + 0.34\hat{X}_{101} = 36.7368$
- $\hat{X}_{100+3} = 25 + 0.34\hat{X}_{100+3-1} = 25 + 0.34\hat{X}_{102} = 37.49051$

B. For $T = 101$, then $X_{101} = 32$

- $\hat{X}_{101+2} = 25 + 0.34\hat{X}_{101+1-1} = 25 + 0.34\hat{X}_{101} = 35.88$
- $\hat{X}_{101+2} = 25 + 0.34\hat{X}_{101+2-1} = 25 + 0.34\hat{X}_{102} = 37.1992$
- $\hat{X}_{101+3} = 25 + 0.34\hat{X}_{101+3-1} = 25 + 0.34\hat{X}_{103} = 37.64773$

2. The following ARIMA model has been fit to a time series:

$$X_t = 25 + 0.8x_{t-1} \quad 0.3x_{t-2} + \varepsilon_t$$

A. Suppose that we are at the end of time period $T = 100$ and we know that $X_{100} = 40$ and $X_{99} = 38$. Determine forecasts for periods 101, 102, 103, from this model at origin 100.

B. Suppose that the observation for time period 101 turns out to be $X_{101} = 35$. Revise your forecasts for periods 102, 103,104 using period 101 as the new origin of time

Solution:

A. Given that $X_{100} = 40$ and $X_{99} = 38$ and $\hat{X}_{T+1} = \mu + \hat{\phi}_1 X_T + \hat{\phi}_2 X_{T-1}$

- $\hat{X}_{101} = 25 + 0.8X_{100} \quad 0.3X_{99} = 25 + 0.8 \quad 40 \quad 0.3 \quad 38 = 45.6$

- $\hat{X}_{102} = 25 + 0.8\hat{X}_{101} \quad 0.3X_{100} = 25 + 0.8 \quad 45.6 \quad 0.3 \quad 40 = 49.48$
- $\hat{X}_{103} = 25 + 0.8\hat{X}_{102} \quad 0.3\hat{X}_{101} = 25 + 0.8 \quad 49.48 \quad 0.3 \quad 45.6 = 50.904$

B. Given that $X_{100} = 40$ and $X_{101} = 35$

- $\hat{X}_{102} = 25 + 0.8\hat{X}_{101} \quad 0.3X_{100} = 25 + 0.8 \quad 35 \quad 0.3 \quad 40 = 41$

$\hat{X}_{103} = 25 + 0.8\hat{X}_{102} \quad 0.3\hat{X}_{101} = 25 + 0.8 \quad 41 \quad 0.3 \quad 35 = 47.3$

Forecasts with $MA(1)$ Process

For this process, it holds that $X_t = \mu + \varepsilon_t \quad \theta_1\varepsilon_{t-1}$with $|\theta_1| < 1$. The conditional mean of X_{t+k} is

$$E(X_{t+k}) = \hat{X}_{t+k} = E\left(\mu + \varepsilon_{t+k} \quad \theta_1\varepsilon_{1,t+k-1}\right) = \mu \quad \hat{\theta}_1\varepsilon_{1,t+k-1}$$

For k = 1, this leads to $\hat{Y}_{T+1} = \mu \quad \hat{\theta}_1\varepsilon_{1,t+k-1}$and for $k \geq 2$, we get $\hat{Y}_{T+1} = \hat{\boldsymbol{\mu}}$ i.e. the unconditional mean is the optimal forecast of Y_{t+k}, k = 2, 3, ..., .

Note that: to be able to perform the one-step forecasts $\hat{Y}_{T+1}$, the unobservable variable $e_{1(T-j}$ has to be expressed as a function of the observable variable Y. To dothis, it must be taken into account that for T≤j, the one-step forecast errorscan be written as: $e_{1(T-j)} = Y_j - \hat{Y}_{T-j}$.

Forecasts with MA(2) Process

For this process, it holds that $X_t = \mu + \varepsilon_t \quad \theta_1\varepsilon_{t-1} \quad \theta_2\varepsilon_{t-2}$ with

$$\theta_1 + \theta_2 < 1$$
$$\theta_2 \quad \theta_1 < 1$$
$$|\theta_2| < 1$$

The conditional mean of $X_{t+k} = \hat{X}_{t+k} = \mu + E(\varepsilon_{t+k}) \quad \hat{\theta}_1 E(\varepsilon_{t+k-1}) + \hat{\theta}_2 E(\varepsilon_{t+k-2})$

$$= \mu \quad \hat{\theta}_1 e_{1,t+k-1} + \hat{\theta}_2 e_{2,t+k-2}$$

This leads; for $k = 1, E(X_{t+1}) = \hat{X}_{t+1} = \mu \quad \hat{\theta}_1 e_{1,T} \quad \hat{\theta}_2 e_{2,T-1}$

For $k = 2$, $\hat{X}_{t+2} = \mu \quad \hat{\theta}_2 e_{2,T}$

For k=3, $\hat{X}_{t+3} = \mu$

Generally $\hat{X}_{t+k} = \mu, k \geq 3$

Similarly, it is possible to show that, after q forecast steps, the optimal forecasts of invertible $MA(q)$ processes, $q > 1$ are equal to the unconditional mean of the process. The forecasts in observable terms are represented similarly to those of the $MA(1)$ process.

Examples

1. The time series model has been fitted to some historical data MA(1) process as:

$X_t = 10 + \varepsilon_t \quad 0.3\varepsilon_{t-1}$ with 200 observations. If the first observation and the last forecast error are given as 19 and   0.45 then find forecasts for period 201, 201, 203,…

2. The time series model has been fitted to some historical data MA(2) process as:

$$X_t = 20 + \varepsilon_t + 0.45\varepsilon_{t-1} \quad 0.35\varepsilon_{t-2}$$

   A. Suppose that at the end of time period $T = 100$ and the observed forecast error for period 100 was 0.5 and for period 99 the observed forecast error was -0.8. Determine forecasts for periods 101, 102, 103, . . . from this model at origin 100

   B. Suppose that the observations for the next four time periods turn out to be $17.5,\ 21.25,\ 18.75$ and 16.75 respectively, then find forecasts for period 102, 103 , …

Solution:

A. Given $\varepsilon_{100} = 0.5 \ and \ \ \varepsilon_{99} = \quad 0.8$

$$\hat{X}_{t+1} = \mu + \hat{\theta}_1 e_{1,T} \quad \hat{\theta}_2 e_{2,\ T-1}$$

- $\hat{X}_{101} = \mu + \hat{\theta}_1 e_{1,100} \quad \hat{\theta}_2 e_{2,\ 99} = 20.505$
- $\hat{X}_{102} = \mu \quad \hat{\theta}_2 e_{2,\ 100} = 20 \quad 0.35 \quad 0.5 = 19.825$
- $\hat{X}_{103} = \mu = 20,$

B. Given $\varepsilon_{100} = 0.5 \ and \ \ \varepsilon_{101} = X_{101} \quad \hat{X}_{101} = \quad 3.755$

- $\hat{X}_{102} = \mu + \hat{\theta}_1 e_{1,101} \quad \hat{\theta}_2 e_{2,\ 100} = 20 + 0.45 \quad 0.5 + .35 \quad 3.75 = 21.5375$
- $\hat{X}_{103} = \mu \quad \hat{\theta}_2 e_{2,\ 101} = 20 \quad 0.35 \quad 0.5 = 19.825$
- $\hat{X}_{103} = \mu = 20,$

Forecasts with $ARMA(p,q)$ Processes

Forecasts for $ARMA(p,q)$ process result from combining the approaches of pure AR and MA processes. Thus, for instance, the one-step a head forecast for a stationary and invertible $ARMA(1,1)$ process as $X_t = \mu + \phi_1 x_{t-1} + \varepsilon_t \quad \theta_1\varepsilon_{t-1}$ is given by:

$$E[X_{T+K}] = E\left[\mu + \hat{\phi}_1 x_{t+K-1} + \varepsilon_{t+K} \quad \hat{\theta}_1 \varepsilon_{t+K-1}\right] = \hat{X}_{T+K} = \mu + \hat{\phi}_1 \hat{X}_{t+K-1} \quad \hat{\theta}_1 \varepsilon_{t+K-1}$$

For $k = 1,\ \hat{X}_{T+1} = \mu + \hat{\phi}_1 \hat{X}_T \quad \hat{\theta}_1 \varepsilon_{1,T}$

For $k = 2,\ \hat{X}_{T+2} = \mu + \hat{\phi}_1 \hat{X}_{T-1}$

For $k = 3,\ \hat{X}_{T+3} = \mu + \hat{\phi}_1 \hat{X}_{T-2}$

Generally, $\hat{X}_{T+k} = \mu + \hat{\phi}_1 \hat{X}_{T+k-1},\ k \geq 3$

Example: The time series model has been fitted to some historical data having 100 observations as $ARMA(1, 1)$ process: $X_t = 0.8x_{t-1} + \varepsilon_t \quad 0.5\varepsilon_{t-1}$. If the last observation and forecast error are given as 91 and   0.54 then find forecasts for period $101, 102, 103, \ldots$

Solution: **Given $X_{100} = 91,\ \varepsilon_{100} = \quad 0.54$ and** $\hat{X}_{T+k} = \mu + \hat{\phi}_1 \hat{X}_{t+K-1} \quad \hat{\theta}_1 \varepsilon_{t+K-1}$

- $\hat{X}_{101} = \mu + \hat{\phi}_1 \hat{X}_{100} \quad \hat{\theta}_1 \varepsilon_{100} = 0.8 \quad 91 \quad 0.5 \quad 0.54 = 73.07$
- $\hat{X}_{102} = \mu + \hat{\phi}_1 \hat{X}_{101} \quad \hat{\theta}_1 \varepsilon_{101} = 0.8 \quad 73.07 \quad 0.5 \quad 0 = 58.456$
- $\hat{X}_{103} = \mu + \hat{\phi}_1 \hat{X}_{102} = 0.8 \quad 58.456 = 46.7648$
- $\hat{X}_{104} = \mu + \hat{\phi}_1 \hat{X}_{102} = 0.8 \quad 46.7648 = 37.41184$

7.4 The Accuracy of Forecasting Methods

The word accuracy refers to the goodness of fit, which intern refers to how well the forecasting model is able to reproduce the data that are already known. To the consumer of forecasts, it is the accuracy of the future forecast that is most important. If Y_t is the actual observation for the period t and $\hat{Y}_t$ is the forecast for the sample period, then the error defined as $e_t = Y_t \quad \hat{Y}_t$.

It is a one-step forecast because it is forecasting one period ahead of the last observation used in the calculation .Therefore ,we describe e_t as a one-step forecast error. It is the difference between the observation Y_t and forecast made using all observations up to but not including Y_t. If there are observations and forecasts for T time periods, then there will be n error terms, and the following standard statistical measures can be defined:

The mean absolute error is often used to avoid this effect i.e.$|e_t| = |Y_t \quad \hat{Y}_t|$. Hence, we can define a measure known as the mean absolute error (MAE) as:

$$MAE = \frac{\sum_{t=1}^{T} |e_t|}{T} = \frac{\sum_{t=1}^{T} |Y_t \quad \hat{Y}_t|}{T}$$

Every forecast error gets the same weight in this measure. The *root mean square error* is often used to give particularly large errors a stronger weight: $RMAE = \sqrt{\frac{\sum_{t=1}^{T} e_t^2}{T}} = \sqrt{\frac{\sum_{t=1}^{T} (Y_t - \hat{Y}_t)^2}{T}}$

Another method is to use the mean squared error (MSE) defined as follows:

$$MSE = \frac{\sum_{t=1}^{T} e_t^2}{T} = \frac{\sum_{t=1}^{T} \left(Y_t \quad \hat{Y}_t\right)^2}{T}$$

Alternatively, Theil's U statistic can be used as a measure of forecasting accuracy. This method that measures the accuracy of forecast is by using U-Statistic called Theil's U-Test. The Theil's U-Statistic is defined as:

The formula for calculating Theil's U statistic:

$$U = \frac{\sqrt{\frac{1}{T} \sum_{t=1}^{T} \left(Y_t \quad \hat{Y}_t\right)^2}}{\sqrt{\frac{1}{T} \sum_{t=1}^{n-1} \hat{Y}_t^{\,2}} \quad \sqrt{\frac{1}{t} \sum_{t=1}^{T} Y_t^{\,2}}}$$

The scaling values of U always lie between 0 and 1. If $U = 0$, then the Theil Inequality Coefficient $U = 0$, the actual values of the series are equal to the estimated values $\hat{X}_t = X_t$ for all t. This case presents a perfect fit between the actual and predicted values.

Theil's U statistic is a relative accuracy measure that compares the forecasted results with the results of forecasting with minimal historical data. It also squares the deviations to give more weight to large errors and to exaggerate errors, which can help eliminate methods with large errors.

Interpreting Theil's U

If $\boldsymbol{U} < 0$, then the forecasting technique is better than guessing

If $U > 0$, then the forecasting technique is worse than guessing.

If $U = 0$, then the forecasting technique is about as good as guessing.

Exercise: The following table shows the actual values and the corresponding forecasts using some forecasting method. Then compute the U-statistic and comment on the forecast accuracy.

T	1	2	3	4	5	6	7	8	9	10
Y_t	22	23	39	37	38	47	43	49	61	63
$\hat{Y}_t$	24	28	32	36	40	44	48	52	56	60

Exercise

1. Assume that you have fit a model for a time series as:

$$X_t = 0.9X_{t-1} + 0.7\varepsilon_{t-1} \quad 0.2\varepsilon_{t-2} + \varepsilon_t$$

and suppose that you are at the end of time period T = 10.

   A. What is the equation for forecasting the time series in period 11, 12 and 13?

   B. Suppose that you know the observed value of X_{10} was 25 and forecast error in period 10 and 9 were 1.2 and -0.9, respectively. Determine forecasts for periods 11, 12, 13 and 14 from this model at origin 10.

2. The following ARIMA model has been fit to a time series:

$$X_t = 25 + 0.8X_{t-1} \quad 0.2\varepsilon_{t-2} + \varepsilon_t$$

A. Suppose that we are at the end of time period $T = 100$ and we know that the forecast for period 100 was 130 and the actual observed value was $X_t = 140$. Determine forecasts for periods 101, 102, 103, . . . from this model at origin 100.

B. What is the shape of the forecast function from this model?

C. Suppose that the observation for time period 101 turns out to be $X_{101} = 132$. Revise your forecasts for periods 102, 103, . . . using period 101 as the new origin of time.

3. **Consider the time series model** $X_t = 200 + 0.7X_{t-1} + \varepsilon_t$

   A. Is this a stationary time series process?

   B. What is the mean of the time series?

   C. If the current observation is $X_{120} = 750$, would you expect the next observation to be above or below the mean?

4. **Consider the time series model** $X_t = 50 + 0.8X_{t-1} + \varepsilon_t \quad 0.2\varepsilon_{t-1}$

   A. Is this a stationary time series process?

   B. What is the mean of the time series?

   C. If the current observation is $X_{100} = 270$, would you expect the next observation to be above or below the mean?

Time Series Analysis*

Time series analysis is a statistical technique that deals with time series data, or trend analysis. Time series data means that data is in a series of particular time periods or intervals. The data is considered in three types:

- Time series data: A set of observations on the values that a variable takes at different times.
- Cross-sectional data: Data of one or more variables, collected at the same point in time.
- Pooled data: A combination of time series data and cross-sectional data.

Terms and concepts:

- Dependence: Dependence refers to the association of two observations with the same variable, at prior time points.
- Stationarity: Shows the mean value of the series that remains constant over a time period; if past effects accumulate and the values increase toward infinity, then stationarity is not met.
- Differencing: Used to make the series stationary, to De-trend, and to control the auto-correlations; however, some time series analyses do not require differencing and over-differenced series can produce inaccurate estimates.
- Specification: May involves the testing of the linear or non-linear relationships of dependent variables by using models such as ARIMA, ARCH, GARCH, VAR, Co-integration, etc.
- Exponential smoothing in time series analysis: This method predicts the one next period value based on the past and current value. It involves averaging of data such that the nonsystematic components of each individual case or observation cancel out each other. The exponential smoothing method is used to predict the short term predication. Alpha, Gamma, Phi, and Delta are the parameters that estimate the effect of the time series data. Alpha is used when seasonality is not present in data. Gamma is used when a series has a trend in data. Delta is used when seasonality cycles are present in data. A model is applied according to the pattern of the data. Curve fitting in time series analysis: Curve fitting regression is used when data is in a non-linear relationship. The following equation shows the non-linear behavior:

The dependent variable, where the case is the sequential case number.

Curve fitting can be performed by selecting "regression" from the analysis menu and then selecting "curve estimation" from the regression option. Then select "wanted curve linear," "power," "quadratic," "cubic," "inverse," "logistic," "exponential," or "other."

ARIMA:

ARIMA stands for autoregressive integrated moving average. This method is also known as the Box-Jenkins method.

Identification of ARIMA parameters:

Autoregressive component: AR stands for autoregressive. Autoregressive parameters is denoted by p. When p =0, it means that there is no auto-correlation in the series. When p=1, it means that the series auto-correlation is till one lag.

Integrated: In ARIMA time series analysis, integrated is denoted by d. Integration is the inverse of differencing. When d=0, it means the series is stationary and we do not need to take the difference of it. When d=1, it means that the series is not stationary and to make it stationary, we need to take the first difference. When d=2, it means that the series has been differenced twice. Usually, more than two-time difference is not reliable.

Moving average component: MA stands for moving the average, which is denoted by q. In ARIMA, moving average q=1 means that it is an error term and it is auto-correlation with one lag.

In order to test whether or not the series and their error term are auto correlated, we usually use W-D test, ACF, and PACF.

Decomposition: Refers to separating a time series into trend, seasonal effects, and remaining variabilityAssumptions:

Stationarity: The first assumption is that the series are stationary. Essentially, this means that the series are normally distributed and the mean and variance are constant over a long time period.

Uncorrelated random error: We assume that the error term is randomly distributed and the mean and variance are constant over a time period. The Durbin-Watson test is the standard test for correlated errors.

No outliers: We assume that there is no outlier in the series. Outliers may affect conclusions strongly and can be misleading.

Random shocks (a random error component): If shocks are present, they are assumed to be randomly distributed with a mean of 0 and a constant variance.