UNIT 11 TIME SERIES ANALYSIS

STRUCTURE

11.0 OBJECTIVES

After studying this unit, you should be able to:

- define the concept of time series,
- appreciate the role of time series in short-term forecasting,
- explain the components of time series, and
- estimate the trend values by different methods.

11.1 INTRODUCTION

In the previous units, you have learnt statistical treatment of data collected for research work. The nature of data varied from case to case. You have come across quantitative data for a group of respondents collected with a view to understanding one or more parameters of that group, such as investment, profit, consumption, weight etc. But when a nation, state, an institution or a business unit etc., intend to study the behaviour of some element, such as price of a product, exports of a product, investment, sales, profit etc., as they have behaved over a period of time, the information shall have to be collected for a fairly long period, usually at equal time intervals. Thus, a set of any quantitative data collected and arranged on the basis of time is called 'Time Series'. Depending on the research objective, the unit of time may be a decade, a year, a month, or a week etc. Typical time series are the sales of a firm in successive years, monthly production figures of a cement mill, daily closing price of shares in Bombay stock market, hourly temperature of a patient.

Usually, the quantitative data of the variable under study are denoted by y_1, y_2, ...y_n and the corresponding time units are denotecd by t_1, t_2.....t_n. The variable 'y' shall have variations, as you will see ups and downs in the values. These changes account for the behaviour of that variable.

Instantly it comes to our mind that 'time' is responsible for these changes, but this is not true. Because, the time (t) is not the cause and the changes in the variable (y) are not the effect. The only fact, therefore, which we must understand is that there are a number of causes which affect the variable and have operated on it during a given time period. Hence, time becomes only the basis for data analysis.

Forecasting any event helps in the process of decision making. Forecasting is possible if we are able to understand the past behaviour of that particular activity. For understanding the past behaviour, a researcher needs not only the past data but also a detailed analysis of the same. Thus, in this unit we will discuss the need for analysis of time series, fluctuations of time series which account for changes in the series over a period of time, and measurement of trend for forecasting.

11.2 DEFINITION AND UTILITY OF TIME SERIES ANALYSIS

Based on the above discussion we can understand the definitions given by a few statisticians. They are:

"A time series consists of statistical data which are collected, recorded over successive increments".

"When quantitative data are arranged in the order of their occurrence, the resulting statistical series is called a time series".

The analysis of time series is of great utility not only to research workers but also to economists, businessmen and scientists etc., for the following reasons:

1) It helps in understanding past behaviour of the variables under study.
2) It facilitates in forcasting the future behaviour with the help of the changes that have taken place in the past.
3) It helps in planning future course of action.
4) It helps in knowing current accomplishment.
5) It is helpful to make comparisons between different time series and significant conclusions drawn therefrom.

Thus we can say that the need for time series analysis arises in research because:

- we want to understand the behaviour of the variables under study,
- we want to know the expected quantitative changes in the variable under study, and
- we want to estimate the effect of various causes in quantitative terms.

In a nutshell, the time series analysis is not only useful for researchers, business research institutions, but also for Governments for devising appropriate future growth strategies.

11.3 COMPONENTS OF TIME SERIES

If you are informed that the price of one kilogram sunflower oil was Rs. 0.50 in the year1940 and in the year 1980 it was Rs. 30 and in the year 2004 it is reported to be Rs. 70, and if you are asked this question: shall sunflower oil be sold again in the future for either Rs. 0.50 or Rs. 30 per kg? Surely, your answer would be 'No'.

Another question: Shall sunflower oil be sold again in future for Rs. 60 per kg? No doubt, your answer would be 'Yes'. Have you ever thought about how you answered the above two questions? Probably you have not! The analysis of these answers shall lead us to arrive at the following observations:

– There are several causes which affect the variable gradually and permanently. Therefore we are prompted to answer 'No' for the first question.

– There are several causes which affect the variable for the time being only. For this reason we are prompted to answer 'Yes' for the second question.

The causes which affect the variable gradually and permanently are termed as "Long-Term Causes". The examples of such causes are: increase in the rate of capital formation, technological innovations, the introduction of automation, changes in productivity, improved marketing etc. The effect of long term causes is reflected in the tendency of a behaviour, to move in an upward or downward direction, termed as 'Trend' or 'Secular Trend'. It reveals as to how the time series has behaved over the period under study.

The causes which affect the variables for the time being only are labelled as "Short-Term Causes". The short term causes are further divided into two parts, they are 'Regular' and 'Irregular'. Regular causes are further divided into two parts, namely 'cyclical causes' and 'seasonal causes'. The cyclical variations are also termed as business cycle fluctuations, as they influence the variable. A business cycle is composed of prosperity, recession, depression and recovery. The periodic movements from prosperity to recovery and back again to prosperity vary both in time and intensity. The seasonal causes, like weather conditions, business climate and even local customs and ceremonies together play an important role in giving rise to seasonal movements to almost all the business activities. For instance, the yearly weather conditions directly affect agricultural production and marketing.

It is worthwhile to say that the seasonal variations analysis will be possible only if the season-wise data are available. This fact must be checked first. For analysing the seasonal effects various methods are available. Among them seasonal index by 'Ratio to Moving Average Method' is the most widely used. However, if collected data provides only yearly values, there is no possibility of obtaining seasonal variations. Therefore, the residual amount after eliminating trend will be the effect of irregular or random causes.

Irregular causes are also termed as 'Erratic' or 'Random' causes. Random variations are caused by infrequent occurrences such as wars, strikes, earthquakes, floods etc. These reasons either go very deep downwards or very high upwards.

The foregoing paragraphs have, in a way, led us to enumerate the components of the time series. These components form the basis for 'Time Series Analysis'.

Long-term causes	:	Secular Trend or Trend (T)
Short-term causes	:	
Regular	:	Cyclical (C)
	:	Seasonal (S)
Irregular or Random	:	Erratic (I)

11.4 DECOMPOSITION OF TIME SERIES

Decomposition and analysis of a time series are one and the same thing. The original data or observed data 'O' is the result of the effects generated by the long-term and short-term causes, namely, (1) Trend = T, (2) cyclical = C, (3) seasonal = S, and (4) Irregular = I. Finding out the values for each of the components is called decomposition of a time series. Decomposition is done either by the Additive model or the Multiplicative model of analysis. Which of these two models is to be used in analysis of time series depends on the assumption that we might make about the nature and relationship among the four components.

Additive Model: It is based on the assumption that the four components are independent of one another. Under this assumption, the pattern of occurrence and the magnitude of movements in any particular component are not affected by the other components. In this model the values of the four components are expressed in the original units of measurement. Thus, the original data or observed data, 'Y' is the total of the four component values, that is,

$$Y = T + S + C + I$$

where, T, S, C and I represent the trend variations, seasonal variations cyclical variations, and erratic variations, respectively.

Multiplicative Model: It is based on the assumption that the causes giving rise to the four components are interdependent. Thus, the original data or observed data 'Y' is the product of four component values, that is :

$$Y = T \times S \times C \times I$$

In this model the values of all the components, except trend values, are expressed as percentages.

In business research, normally, the multiplicative model is more suited and used more frequently for the purpose of analysis of time series. Because, the data related to business and economic time series is the result of interaction of a number of factors which individually cannot be held responsible for generating any specific type of variations.

Let us consider an example for construction of time series according to the Multiplicative Model. Table 11.1 presents trend, seasonal, and cyclical-erratic components of a hypothetical series.

Table 11.1: Hypothetical time series and its components (quarterly)

Year	Quarter	Series (O)	Components		
			Trend (T) (100 C1)	Seasonal (100 S)	Cyclical-erratic
1	1	79	80	120	82
	2	58	85	80	85
	3	84	90	92	102
	4	107	95	108	105
2	1	130	100	120	108
	2	93	105	80	132
	3	121	110	92	120
	4	161	115	108	130
3	1	216	120	120	150
	2	132	125	80	132
	3	150	130	93	125
	4	163	135	108	112
4	1	176	140	120	105
	2	112	145	80	97
	3	128	150	93	93
	4	142	155	108	85

According to multiplicative model

$$Y = T \times S \times C \times I$$

Thus, 79 (1 year and 1 quarter) = $80 \times \frac{120}{100} \times \frac{82}{100}$

130 (2 year and 1 quarter) = $100 \times \frac{120}{100} \times \frac{108}{100}$

Thus each quarterly figure (Y) is the product of the T, S, and CI. Such a synthetic composition looks like an actual time series and has encouraged use of the model as the basis for the analysis of time series data.

11.5 PRELIMINARY ADJUSTMENTS

Before we proceed with the task of analysing a time series data, it is necessary to do relevant adjustments in the raw data. They are:

1) **Calender variations:** As we are aware, all the calender months do not have the same number of days. For instance, the production in the month of February may be less than other months because of fewer days and if we take the holidays into account the variation is greater. Therefore, adjustments for calender variations have to be made.

2) **Price changes:** As price level changes are inevitable, it is necessay to convert monetary values into real values after taking into consideration the price indices. In fact this is the process of deflating which will be discussed in Unit-12 (Index Numbers) of this course.

3) **Population changes:** Population grows constantly. This also calls for adjustment in the data for the population changes. In such cases, if necessary, per capita values may be computed (dividing original figures by the total population).

Self Assessment Exercise A

1) State whether the following statements are 'True' or 'False'.

   a) Time is the cause for the ups and downs in the values of the variable under study.

   b) The variable under study in time series analysis is denoted by 'y'.

   c) 'Trend' values are a major component of the time series.

   d) Analysis of time series helps in knowing current accomplishment.

   e) Weather conditions, customs, habits etc., are causes for cyclical variations.

   f) The analysis of time series is done to know the expected quantity change in the variable under study.

2) Why do we analyse a time series?

   .....

   .....

   .....

3) List out the components of a time series.

   .....

   .....

   .....

11.6 METHODS OF MEASUREMENT OF TREND

The effect of long-term causes is seen in the trend values we compute. A trend is also known as 'secular trend' or 'long-term trend' as well. There are several methods of isolating the trend of which we shall discuss only two methods which are most frequently used in the business and economic time series data analysis. They are: Free Hand Method, and Method of Least Square.

11.6.1 Free Hand Method

In this method, the first requirement is that a graph is drawn of the original data. After plotting the data on the graph paper, without the help of any numerical calculations, a free hand straight line is drawn through the graph ensuring that it passes (as closely as possible) through the middle of the entire graph of the time series. This is, thus, the easiest and quickest method of estimating secular trend. Even though the straight line is drawn on personal judgments, it requires a careful inspection of the overall behaviour of movements in that time series graph.

Though this method is very simple, it does not have a common acceptance because it gives varying trend values for the same data when efforts are made by different persons or even by the same persons at different times. It is to be noted that free-hand method is highly subjective and therefore, different researchers may draw different trend lines from the same data set. Hence, it is not advisable to use it as a basis for forecasting, particularly, when the time series is subject to very irregular movements. Let us consider an illustration to draw a trend line by free-hand method.

Illustration 1

From the following data, find the trend line by using Free hand (graphline) Method

Years:	1994, 1995, 1996, 1997, 1998, 1999, 2000, 2001, 2002, 2003
Foodgrain production *(lakh tonnes)*	35, 55, 40, 85, 135, 110, 130, 150, 130, 120

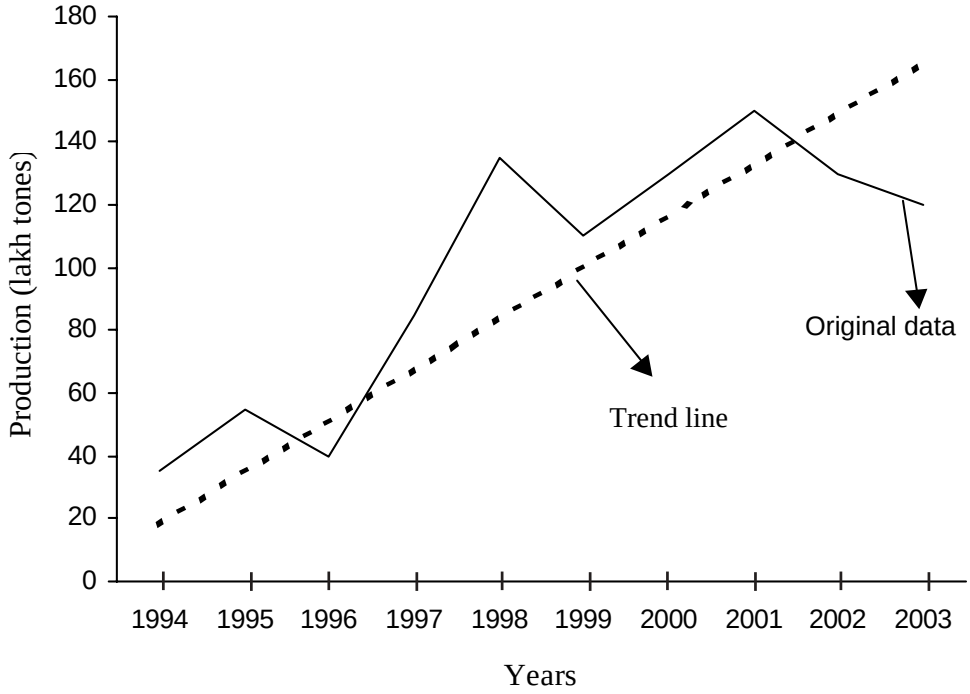

Fig. 1: Food Grain Production (in lakh tons)

11.6.2 Least Square Method

This is also known as straight line method. This method is most commonly used in research to estimate the trend of time series data, as it is mathematically designed to satisfy two conditions. They are:

1) Sum of $(Y-Y_c) = 0$, and
2) Sum of $(Y-Y_c)^2$ = least

The straight line method gives a line of best fit on the given data. The straight line which can satisfy the above conditions and make use of the regression equation, is given by :

$Y_c = a + bx$

where, 'Y_c' represents the trend value of the time series variable y, 'a' and 'b' are constant values of which 'a' is the trend value at the point of origin and 'b' is the amount by which the trend value changes per unit of time, and 'x' is the unit of time (value of the independent variable).

The values of constants, 'a' and 'b', are determined by the following two normal equations.

$\Sigma y = na + b \Sigma x$ .....(i)

$\Sigma xy = a\Sigma x + b \Sigma x^2$ .....(ii)

The process of finding values of constants a and b can be made simple by using a shortcut method, that is, by taking the origin year in such a way that it gives the total of 'x' (Σx) equal to 'zero'. This becomes possible if we take the median year as origin period. Thus, the negative values in the first half of the series balance out the positive values in the second half. Thus, the earlier normal equation shall be changed as follows, with reference to $\Sigma x = 0$.

$\Sigma y = a$ (as Σbx becomes zero)

$\Sigma xy = b\Sigma x^2$ (as $a\Sigma x$ becomes zero)

Therefore, the values of two constants are obtained by the following formulae:

$$a = \frac{\Sigma y}{N}, \quad \text{and} \quad b = \frac{\Sigma xy}{\Sigma x^2}$$

It is to be noted that when the number of time units involved is even, the point of origin will have to be chosen between the two middle time units.

Let us consider an illustration to understand the procedure for estimation of the trend by using the method of least squares.

Illustration 2

The decision making body of a fertilizer firm producing fertilizers wants to predict future sales trend for the years 2006 and 2008 based on the analysis of its past sales pattern. The sales of the firm for the last 7 years, for this purpose, are given below:

Years	Sales (in '000 tonnes)
1998	70
1999	75
2000	90
2001	98
2002	85
2003	91
2004	100

Solution: To find the straight line equation ($Y_c = a + bx$) for the given time series data, we have to substitute the values of already arrived expression, that is:

$$a = \frac{\Sigma y}{N}, \quad \text{and} \quad \frac{\Sigma xy}{\Sigma x^2}$$

In order to make the total of x = 'zero', we must take median year (i.e., 2001) as origin. Study the following table carefully to understand the procedure for fitting the straight line.

Table 11.1: Computation of Trend

Year	Sales ('000 tons)	x	x^2	xy	Trend (Y_c) a+bx
1998	70	–3	9	–210	74.5
1999	75	–2	4	–150	78.6
2000	90	–1	1	–90	82.8
2001	98	0	0	0	87.0
2002	85	1	1	85	91.2
2003	91	2	4	182	95.4
2004	100	3	9	300	99.5
N = 7	$\Sigma y = 609$	$\Sigma x = 0$	$\Sigma x^2 = 28$	$\Sigma xy = 117$	609.0

$$a = \frac{\Sigma y}{N} = \frac{609}{7} = 87; \quad b = \frac{\Sigma xy}{\Sigma x^2} = \frac{117}{28} = 4.18$$

Thus, the straight line trend equation is

$Y_c = 87 + 4.18x$

From the above equation, we can also find the monthly increase in sales as follows:

$$\frac{4.180}{12} = 348.33 \text{ tons}$$

The reason for this is that the trend values increased by a constant amount 'b' every year. Hence the annual increase in sales is 4.18 thousand tons.

Trend values are to be obtained as follows:

Y_{1998} = 87 + 4.18 (–3) = 74.5

Y_{1999} = 87 + 4.18 (–2) = 78.6 and so on .....

Predicting with decomposed components of the time series: The management wants to estimate fertiliser sales for the years 2006 and 2008.

Estimation of sales for 2006, 'x' would be 5 (because for 2004 'x' was 3).

Y_{2006} = 87 + 4.18 (5) = 107.9 thousand tonnes.

Estimation of sales for 2008, 'x' would be 7

Y_{2008} = 87 + 4.18 (7) = 116.3 thousand tonnes.

Self Assessment Exercise B

1) State whether the following statements are 'True' or 'False'.

   a) The free hand method gives different straight lines for the same data when efforts are made by different persons.

   b) The multiplicative model is based on the assumption that the causes giving rise to the four components are dependent.

   c) A free hand curve is drawn without any numerical calculations for trend estimation.

   d) The total of the difference between original data and trend values (obtained by straight line method) will never be zero.

   e) In the least square trend equation $Y_c = a + bx$, if b is positive it indicates a rising trend.

   f) the additive model of time series analysis is expressed as: Y = T + S + C + I.

2) Enumerate the methods of isolating trend.

   .....

   .....

   .....

3) Foodgrain production (in lakh tonnes) is given below (figures are imaginary). Find the Trend by using a) Graphic method (free hand) b) Straight Line Method. Tabulate the trend values. c) Predict the production for the year 2010.

Years	Production
1996	40
1997	60
1998	45
1999	83
2000	130
2001	135
2002	150
2003	120
2004	200

Years	y	x	x^2	xy	y_c

11.7 LET US SUM UP

This unit has introduced you to the concept of time series and its analysis with a view to making more accurate and reliable forcasts for the future.

A set of quantitative data arranged on the basis of TIME are referred to as 'Time Series'. The analysis of time series is done to understand the dynamic conditions for achieving the short-term and long-term goals of institution(s). With the help of the techniques of time series analysis the future pattern can be predicted on the basis of past trends.

The quantitative values of the variable under study are denoted by y_1, y_2, y_3..... and the corresponding time units are denoted as x_1, x_2, x_3..... . The variable 'y' shall have variations, you will see ups and downs in the values. There are a number of causes during a given time period which affect the variable. Therefore, time becomes the basis of analysis. Time is not the cause and the changes in the values of the variable are not the effect.

The causes which affect the variable gradually and permanently are termed as Long-term causes. The causes which affect the variable only for the time being are termed as Short-term causes. The time series are usually the result of the effects of one or more of the four components. These are trend variations (T), seasonal variations (S), Cyclical variations (C) and Irregular variations (I).

When we try to analyse the time series, we try to isolate and measure the effects of various kinds of these components on a series.

We have two models for analysing time series:

1) Additive model, which considers the sum of various components resulting in the given values of the overall time series data and symbolically it would be expressed as: $Y = T + C + S + I$.

2) The multiplicative model assumes that the various components interact in a multiplicative manner to produce the given values of the overall time series data and symbolically it would be expressed as : $y = T \times C \times S \times I$.

The trend analysis brings out the effect of long-term causes. There are different methods of isolating trends, among these we have discussed only two methods which are usually used in research work, i.e. free hand and least square methods.

Long-term predictions can be made on the basis of trends, and only the least square method of trend computation offers this possibility.

11.8 KEY WORDS

Time Series : is the data on any variable accumulated at regular time intervals.

Secular Trend : A type of variation in a time series, the long-term tendency of a time series to grow or decline over a period of time.

Seasonal Variation : Patterns of change in a time series within a year and the same changes tend to be repeated from year to year.

Cyclical Variations : A type of variation in a time series, in which the values of variables vary up and down around the secular trend line.

Irregular Variations : A type of element of a time series, refers to such variations in business activity which do not repeat according to a definite pattern and the values of variables are completely unpredictable.

11.9 ANSWERS TO SELF ASSESSMENT EXERCISES

A) 1) a) False b) True c) True d) True e) False f) True

3) Secular trend, Seasonal variation, Cyclical variation, and Irregular variation

B) 1) a) True b) False c) True d) True e) False f) True

3) $Y_1 = 107 + 18.03\ x$

Estimated production for 2010 is 287.3 lakh tonnes.

11.10 TERMINAL QUESTIONS/EXERCISES

1) What is time series? Why do we analyse a time series?

2) Explain briefly the components of time series.

3) Explain briefly the additive and multiplicative models of time series. Which of these models is more commonly used and why?

4) From the following data, obtain the trend line by Freehand Method for further analysis.

Years	1996	1997	1998	1999	2000	2001	2002	2003
'y'	24	28	38	33	49	50	66	68

5) The production (in thousand tons) in a sugar factory during 1994 to 2001 has been as follows:

Year	1994	1995	1996	1997	1998	1999	2000	2001
Production	35	38	49	41	56	58	76	75

(*Hint:* The point of origin must be taken between 1997 and 1998).

i) Find the trend values by applying the method of least square.

ii) What is the monthly increase in production?

iii) Estimate the production of sugar for the year 2008.

6) The following data relates to a survey of used car sales in a city for the period 1993-2001. Predict sales for 2006 by using the linear trend equation.

Years	1993	1994	1995	1996	1997	1998	1999	2000	2001
Sales	214	320	305	298	360	450	340	500	520

Note: These questions/exercises will help you to understand the unit better. Try to write answers for them. But do not submit your answers to the university for assessment. These are for your practice only.

11.11 FURTHER READING

A number of good text books are available for the topics dealt with in this unit. The following books may be used for more indepth study.

Mentgomery, D.C. and L.A. Johnson, 1996, '*Forecasting and Time Series Analysis*' McGraw Hill : New York.

Chandan, J.S., 2001, *Statistics for Business and Economics*, Vikas Publishing House Pvt. Ltd., New Delhi.

Gupta, S.P. and H.P. Gupta, 2001, *Business Statistics*, S. Chand, New Delhi.