UNIT 20 TIME SERIES ANALYSIS

Objectives

After completion of this unit, you should be able to :

- appreciate the role of time series analysis in short term forecasting
- decompose a time series into its various components
- understand auto-correlations to help identify the underlying patterns of a time series
- become aware of stochastic models developed by Box and Jenkins for time series analysis
- make forecasts from historical data using a suitable choice from available methods.

Structure

20.1 INTRODUCTION

Time series analysis is one of the most powerful methods in use, especially for short term forecasting purposes. From the historical data one attempts to obtain the underlying pattern so that a suitable model of the process can be developed, which is then used for purposes of forecasting or studying the internal structure of the process as a whole. We have already seen in Unit 17 that a variety of methods such as subjective methods, moving averages and exponential smoothing, regression methods, causal models and time-series analysis are available for forecasting. Time series analysis looks for the dependence between values in a time series (a set of values recorded at equal time intervals) with a view to accurately identify the underlying pattern of the data.

In the case of quantitative methods of forecasting, each technique makes explicit assumptions about the underlying pattern. For instance, in using regression models we had first to make a guess on whether a linear or parabolic model should be chosen and only then could we proceed with the estimation of parameters and model-development. We could rely on mere visual inspection of the data or its graphical plot to make the best choice of the underlying model. However, such guess work, through not uncommon, is unlikely to yield very accurate or reliable results. In time series analysis, a systematic attempt is made to identify and isolate different kinds of patterns in the data. The four kinds of patterns that are most frequently encountered are horizontal, non-stationary (trend or growth), seasonal and cyclical. Generally, a random or noise component is also superimposed.

We shall first examine the method of decomposition wherein a model of the time-series in terms of these patterns can be developed. This can then be used for forecasting purposes as illustrated through an example.

A more accurate and statistically sound procedure to identify the patterns in a time-series is through the use of auto-correlations. Auto-correlation refers to the correlation between the same variable at different time lags and was discussed in Unit 18. Auto-correlations can be used to identify the patterns in a time series and suggest appropriate stochastic models for the underlying process. A brief outline of common processes and the Box-Jenkins methodology is then given.

Finally the question of the choice of a forecasting method is taken up. Characteristics of various methods are summarised along with likely situations where these may be applied. Of course, considerations of cost and accuracy desired in the forecast play a very important role in the choice.

20.2 DECOMPOSITION METHODS

Economic or business oriented time series are made up of four components -- trend. seasonality, cycle and randomness. Further, it is usually assumed that the relationship between these four components is multiplicative as shown in equation 20.1.

$X_t = T_t S_t C_t R_t$ ...(20.1)

where

X_t is the observed value of the time series

T_t denotes trend

S_t denotes seasonality

C_t denotes cycle

and

R_t denotes randomness.

Alternatively, one could assume an additive relationship of the form

$X_t = T_t + S_t + C_t + R_t$

But additive models are not commonly encountered in practice. We shall, therefore, be working with a model of the form (20.1) and shall systematically try to identify the individual components.

You are already familiar with the concept of moving averages, If the time series represents a seasonal pattern of L periods, then by taking a moving average of L periods, we would get the mean value for the year. Such a value will obviously be free of seasonal effects, since high months will be offset by low ones. If M_t denotes the moving average of equation (20.1), it will be free of seasonality and will contain little randomness (owing to the averaging effect). Thus we can write

$M_t = T_t C_t$ ....(20.2)

The trend and cycle components in equation (20.2) can be further decomposed by assuming some form of trend.

- One could assume different kinds of trends, such as
- linear trend, which implies a constant rate of change (Figure I)
- parabolic trend, which implies a varying rate of change (Figure II)
- exponential or logarithmic trend, which implies a constant percentage rate of change (Figure III).
- an S curve, which implies slow initial growth, with increasing rate of growth followed by a declining growth rate and eventual saturation (Figure IV).

Figure I: Linear Trend

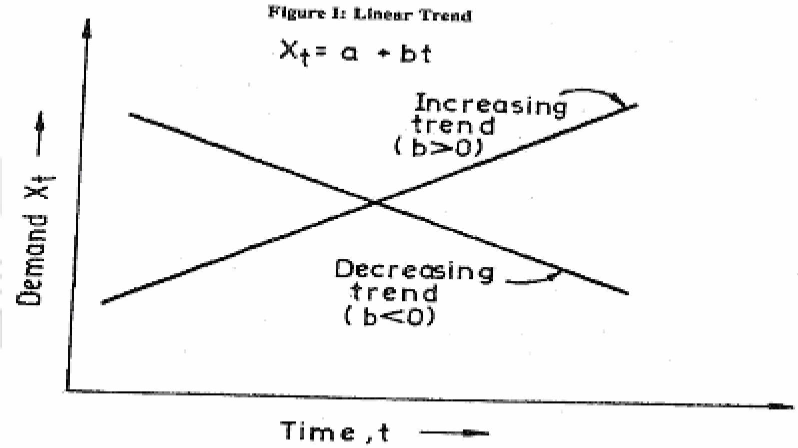


Figure II: Parabolic Trend

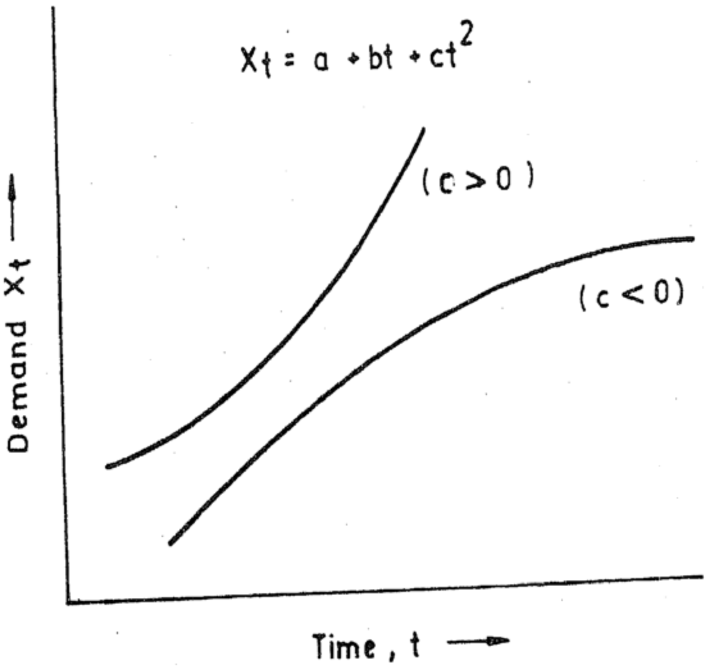


Figure III: Exponential Trend

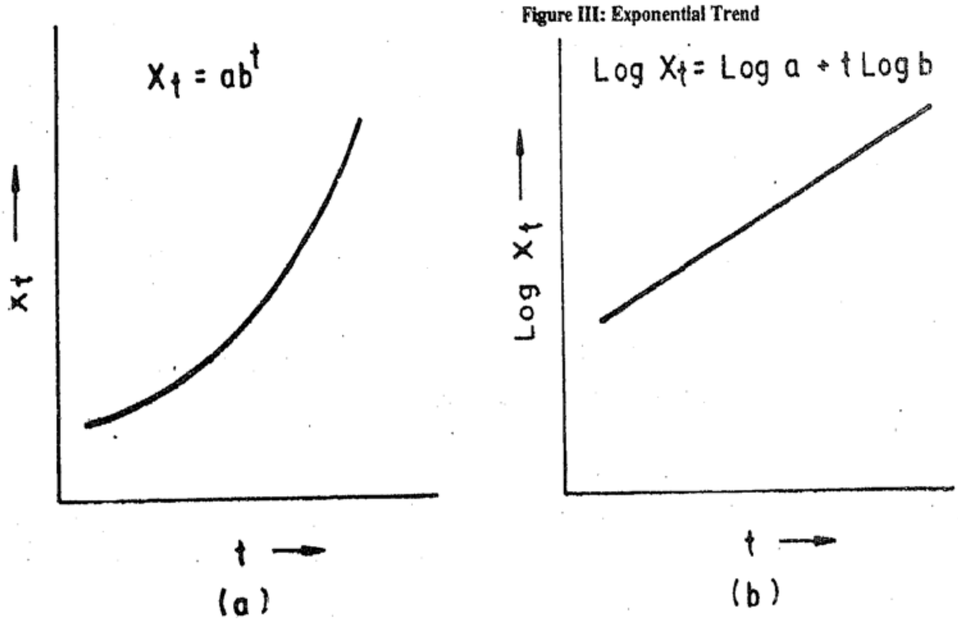
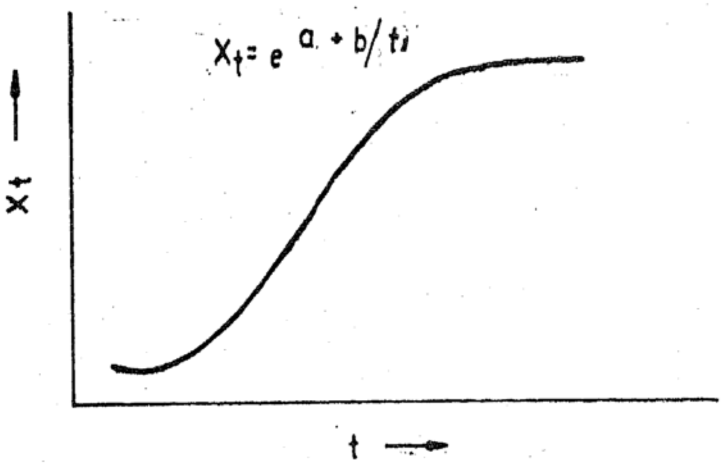


Figure IV: A Typical S Curve



Knowing the pattern of the trend, the appropriate mathematical function could be determined from the data by using the methods of regression, as outlined in Unit 19. This would establish the values of parameters of the chosen trend model. For example, assuming a linear trend gives

$$T_t = a + bt \qquad \ldots(20.3)$$

The cycle component C_t can now be isolated from the trend T_t in equation (20.2) by the use of equation (20.3) as follows

$$\frac{M_t}{T_t} = \frac{M_t}{a + bt} \qquad \ldots(20.4)$$

As already indicated, if a linear trend is not adequate, one may wish to specify a non-linear one. Any pattern for the trend can be used to separate it from the cycle. In practice, however, it is often difficult to separate the two, and one may prefer to work with the trend cycle figures of equation (20.2). The isolation of the trend will add little to the overall ability to forecast. This will become clear when we take up an example problem for solution.

To isolate seasonality one could simply divide the original series (equation 20.1) by the moving average (equation 20.2) to obtain

$$\frac{X_t}{M_t} = \frac{T_t S_t C_t R_t}{T_t C_t} = S_t R_t \qquad \ldots(20.5)$$

Finally, randomness can be eliminated by averaging the different values of equation (20.5). The averaging is done on the same months or seasons of different years (for example the average of all Januaries, all Februaries, ... all Decembers). The result is a set of seasonal values free of randomness, called seasonal indices, which are widely used in practice.

In order to forecast, one must reconstruct each of the components of equation (20.1). The seasonality is known through averaging the values in equation (20.5) and the trend through (20.3). The cycle of equation (20.4) must be estimated by the user and the randomness cannot be predicted.

For details of how to predict the cycle term, the interested reader may refer to Makridakis and Wheelwright (1978). To illustrate the application of this procedure to actual forecasting of a time series, an example will now be considered.

20.3 EXAMPLE OF FORECASTING USING DECOMPOSITION

(This example has been adapted from Srivastava et al.) (1978)

An Engineering firm producing farm equipment wants to predict future sales based on the analysis of its past sales pattern. The sales of the company for the last five years are given in Table 1.

Table 1 : Quarterly Sales of an Engineering Firm during 1983-87 (Rs. in lakhs)

Year	Quarters			
	I	II	III	IV
1983	5.5	5.4	7.2	6.0
1984	4.8	5.6	6.3	5.6
1985	4.0	6.3	7.0	6.5
1986	5.2	6.5	7.5	7.2
1987	6.0	7.0	8.4	7.7

The procedure involved in the study consists of

a) deseasonalising the time series which is done by constructing a moving average M and taking the ratio $\frac{X_t}{M_t}$ which we know from equation (20.5) represents the seasonality and randomness.

b) fitting a trend line of the type $T_t = a + bt$ to the deseasonalised time series

c) identifying the cyclical variation around the trend line

d) use the above information for forecasting sales for the next year

Deseasonalising the Time Series

The moving averages and the ratios of the original variable to the moving average have first to the computed.

This is done in Table 2

Table 2: Computation of moving averages M_t and the ratios X_t /M_t

Year	Quarter	Actual Sales	4 Quarter Moving Total	Centred Moving Total	Centred Moving Average (M_t)	$\frac{X_t}{M_t}$
1983	I	5.5				
	II	5.4				
	III	7.2		23.8	6.0	1.200
	IV	6.0	24.1	23.5	5.9	1.017
1984	I	4.8	23.4	23.2	5.8	0.828
	II	5.6	23.6	22.5	5.6	1.000
	III	6.3	22.7	21.9	5.5	1.145
	IV	5.6	22.3	21.9	5.5	1.018
1985	I	4.0	21.5	22.6	5.7	0.702
	II	6.3	22.2	23.4	5.9	1.068
	III	7.0	22.9	24.4	6.1	1.148
	IV	6.5	23.8	25.1	6.3	1.032
1986	I	5.2	25.0	25.5	6.4	0.813
	II	6.5	25.2	26.1	6.5	1.000
	III	7.5	25.7	26.8	6.7	1.119
	IV	7.2	26.4	27.5	6.9	1.043
1987	I	6.0	27.2	28.2	7.1	0.845
	II	7.0	27.7	28.9	7.2	0.972
	III	8.4	28.6			
	IV	7.7	29.1			

It should be noticed that the 4 Quarter moving totals pertain to the middle of two successive periods. Thus the value 24.1 computed at the end of Quarter IV, 1983 refers to middle of Quarters II, III, 1983 and the next moving total of 23.4 refers to the middle of Quarters III and IV, 1983. Thus, by taking their average we obtain the centred moving total of $\frac{(24.1+23.4)}{2} = 23.75 \cong 23.8$ to be placed for Quarter III, 1983. Similarly for the other values in case the number of periods in the moving total or average is odd, centering will not be required.

The seasonal indices for the quarterly sales data can now be computed by taking averages of the X_t/M_t ratios of the respective quarters for different years as shown in Table 3.

Table 3: Computation of Seasonal Indices

Year	Quarters			
	I	II	III	IV
1983	-	-	1.200	1.017
1984	0.828	1.000	1.145	1.018
1985	0.702	1.068	1.148	1.032
1986	0.813	1.000	1.119	1.043
1987	0.845-	0.972	-	-
Mean	0.797	1.010	1.153	1.028
Seasonal Index	0.799	1.013	1.156	1.032

The seasonal indices are computed from the quarter means by adjusting these values of means so that the average over the year is unity. Thus the sum of means in Table 3 is 3.988 and since there are four Quarters, each mean is adjusted by multiplying it with the constant figure of 4/3.988 to obtain the indicated seasonal indices. These seasonal indices can now be used to obtain the deseasonalised sales of the firm by dividing the actual sales by the corresponding index as shown in Table 4.

Table 4: Deseasonalised Sales

Year	Quarter	Actual Sales	Seasonal index	Deseasonalised Sales
1983	I	5.5	0.799	6.9
	II	5.4	1.013	5.3
	III	7.2	1.156	6.2
	IV	6.0	1.032	5.8
1964	I	4.8	0.799	6.0
	II	5.6	1.013	5.5
	111	6.3	1.156	5.4
	IV	5.6	1.032	5.4
1985	1	4.0	0.799	5.0
	11	6.3	1.013	6.2
	III	7.0	1.156	6.0
	IV	6.5	1.032	6.3
1986	I	5.2	0.799	6.5
	II	6.5	1.013	6.4
	111	7.5	1.156	6.5
	IV	7.2	1.032	7.0
1967	1	6.0	0.799	7.5
	II	7.0	1.013	6.9
	III	8.4	1.156	7.3
	IV	7.7	1.032	7.5

Fitting a Trend Line

The next step after deseasonalising the data is to develop the trend line. We shall here use the method of least squares that you have already studied in Unit 19 on regression. Choice of the origin in the middle of the data with a suitable scaling simplifies computations considerably. To fit a straight line of the form Y = a + bX to the deseasonalised sales, we proceed as shown in Table 5.

Table 5: Computation of Trend

Table 5: Computation of Trend

Year	Quarter	Deseasonalised Sales (Y)	X	X^2	XY
1983	I	6.9	−19	361	−131.1
	II	5.3	−17	289	−90.1
	III	6.2	−15	225	−93.0
	IV	5.8	−13	169	−75.4
1984	I	6.0	−11	121	−66.0
	II	5.5	−9	81	−49.5
	III	5.4	−7	49	−37.8
	IV	5.4	−5	25	−27.0
1985	I	5.0	−3	9	−15.0
	II	6.2	−1	1	−6.2
	III	6.0	1	1	6.0
	IV	6.3	3	9	18.9
1986	I	6.5	5	25	32.5
	II	6.4	7	49	44.8
	III	6.5	9	81	58.5
	IV	7.0	11	121	77.0
1987	I	7.5	13	169	97.5
	II	6.9	15	225	103.5
	III	7.3	17	289	124.1
	IV	7.5	19	361	142.5
Total		125.6	0	2660	114.2

$$a = \frac{\Sigma Y}{n} = \frac{125.6}{20} = 6.3$$

$$b = \frac{\Sigma XY}{\Sigma X^2} = \frac{114.2}{2660} = 0.04\ X$$

∴ the trend line is $Y = 6.3 + 0.04\ X$.

Identifying Cyclical Variation

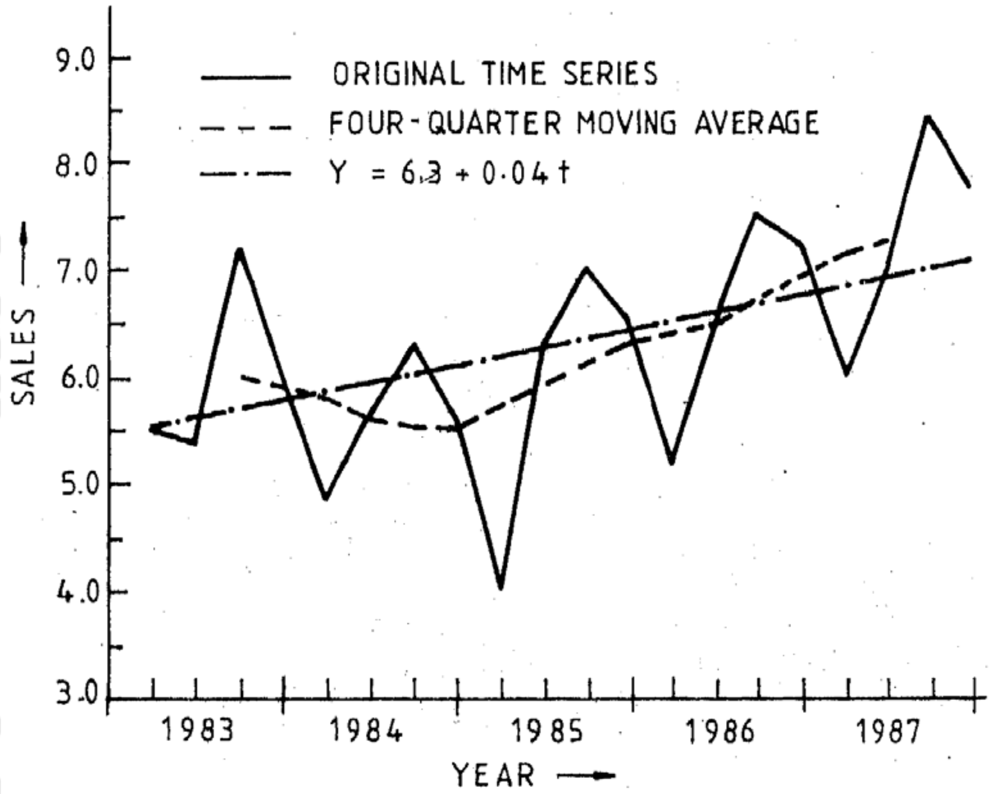
The cyclical component is identified by measuring deseasonalised variation around the trend line, as the ratio of the actual deseasonalised sales to the value predicted by the trend line. The computations are shown in Table 6.

Table 6: Computation of Cyclical Variation

Year	Quarter	Deseasonalised Sales (Y)	Trend a+bX	$\frac{Y}{a+bX}$
1983	I	6.9	5.54	1.245
	II	5.3	5.62	0.943
	III	6.2	5.70	1.088
	IV	5.8	5.78	1.003
1984	I	6.0	5.86	1.024
	II	5.5	5.94	0.926
	III	5.4	6.02	0.897
	IV	5.4	6.10	0.885
1985	I	5.0	6.18	0.809
	II	6.2	6.26	0.990
	III	6.0	6.34	0.946
	IV	6.3	6.42	0.981
1986	I	6.5	6.50	1.000
	II	6.4	6.58	0.973
	III	6.5	6.66	0.976
	IV	7.0	6.74	1.039
1987	I	7.5	6.82	1.110
	II	6.9	6.90	1.000
	III	7.3	6.98	1.046
	IV	7.5	7.06	1.062

The random or irregular variation is assumed to be relatively insignificant. We have thus described the time series in this problem using the trend, cyclical and seasonal components. Figure V represents the original time series, its four quarter moving average (containing the trend and cycle components) and the trend line.

Figure V: Time Series with Trend and Moving Averages



Forecasting with the Decomposed Components of the Time Series

Suppose that the management of the Engineering firm is interested in estimating the sales for the second and third quarters of 1988. The estimates of the deseasohalised sales can be obtained by using the trend line

Y = 6.3 + 0.04(23)

= 7.22 (2nd Quarter 1988)

and Y = 6.3 + 0.04 (25)

= 7.30 (3rd Quarter 1988)

These estimates will now have to be seasonalised for the second and third quarters respectively. This can be done as follows :

For 1988 2nd quarter

seasonalised sales estimate = 7.22 x 1.013 = 7.31

For 1988 3rd quarter

seasonalised sales estimate = 7.30 x 1.56

= 8.44

Thus, on the basis of the above analysis, the sales estimates of the Engineering firm for the second and third quarters of 1988 are Rs. 7.31 lakh and Rs. 8.44 lakh respectively.

These estimates have been obtained by taking the trend and seasonal variations into account. Cyclical and irregular components have not been taken into account. The procedure for cyclical variations only helps to study past behaviour and does not help in predicting the future behaviour.

Moreover, random or irregular variations are difficult to quantify.

20.4 USE OF AUTO-CORRELATIONS IN IDENTIFYING TIME SERIES

While studying correlation in Unit 18, auto-correlation was defined as the correlation of a variable with itself, but with a time lag. The study of auto-correlation provides very valuable clues to the underlying. pattern of a time series. It can also be used to estimate the length of the season for seasonality. (Recall that in the example problem considered in the previous. section, we assumed that a complete season consisted of four quarters.)

When the underlying time series represents completely random data, then the graph of auto-correlations for various time lags stays close to zero with values fluctuating both on the +ve and -ve side but staying within the control limits. This in fact represents a very convenient method of identifying randomness in the data.

If the auto-correlations drop slowly to zero, and more than two or three differ significantly from zero, it indicates the presence of a trend in the data. This trend can be-removed by differentiating (that is taking differences between consecutive values and constructing a new series).

A seasonal pattern in the data would result in the auto-correlations oscillating around zero with some values differing significantly from zero. The length of seasonality can be determined either from the number of periods it takes for the auto-correlations to make a complete cycle or by the tine lag giving the largest auto Correlation.

For any given data, the plot of auto-correlation for van us time lags is diagnosed to identify which of the above basic patterns (or a combination of these patterns) it follows. This is broadly how auto-correlations are used to identify the structure of the underlying model to be chosen. The underlying mathematics and computational burden tend to be heavy and involved. Computer routines for carrying out computations are available. The interested reader may refer to Makridakis and Wheelwright for further details.

20.5 AN OUTLINE OF BOX-JENKINS MODELS FOR TIME SERIES

Box and Jenkins (1976) have proposed a sophisticated methodology for stochastic model building and forecasting using time series. The purpose of this section is merely to acquaint you with some of the terms, models and methodology developed by Box and Jenkins.

A time series may be classified as stationary (in equilibrium about a constant mean value) or non-stationary (when the process has no natural or stable mean). In stochastic model building the non-stationary processes often converted to a stationary one by differencing. The two major classes of models used popularly in time series analysis are Auto-regressive and Moving Average models.

Auto-regressive Models

In such models, the current value of the process is expressed as a finite, linear aggregate of previous values of the process and a random shock or error a_t. Let us denote the value of a process at equally spaced times t, t-1, t - 2... by Z_t, Z_{t-1}, Z_{t-2} ……

also let Z_t, Z_{t-1} Z_{t-2} …be the deviations from the process mean, m. That is $\overline{Z}_t = Z_t - m$. Then

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \ldots + \phi_p Z_{t-p} + a_t \qquad \ldots (20.6)$$

is called an auto-regressive (AR) process of order p. The reason for this name is that equation (20.6) represents a regression of the variable Z_t on successive values of itself. The model contains p + 2 unknown parameters m, ϕ_1, ϕ_2,......ϕ_p, σ_2 a which in practice have to be estimated from the data.

The additional parameter σ^2_a is the variance of the random error component.

Moving Average models

Another kind of model of great importance is the moving average model where Z_t is made linearly dependent on a finite number q of previous a's (error terms)

Thus

$$Z_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \theta_q a_{t-q} \qquad \ldots (20.7)$$

is called a moving average (MA) process of order q. The name "moving average" is somewhat misleading because the weights 1, - θ_1, - θ_2, ..., _ θ_q which multiply the a's, need not total unity nor need they be positive. However, this nomenclature is in common use and therefore we employ it. The model (20.7) contains q + 2 unknown parameters m, θ_1, θ_2,..... θ_q, σ^2 a which in practice have to be estimated from the data.

Mixed Auto-regressive-moving average models :

It is sometimes advantageous to include both auto-regressive and moving average terms in the model. This leads to the mixed auto-regressive-moving average (ARMA) model.

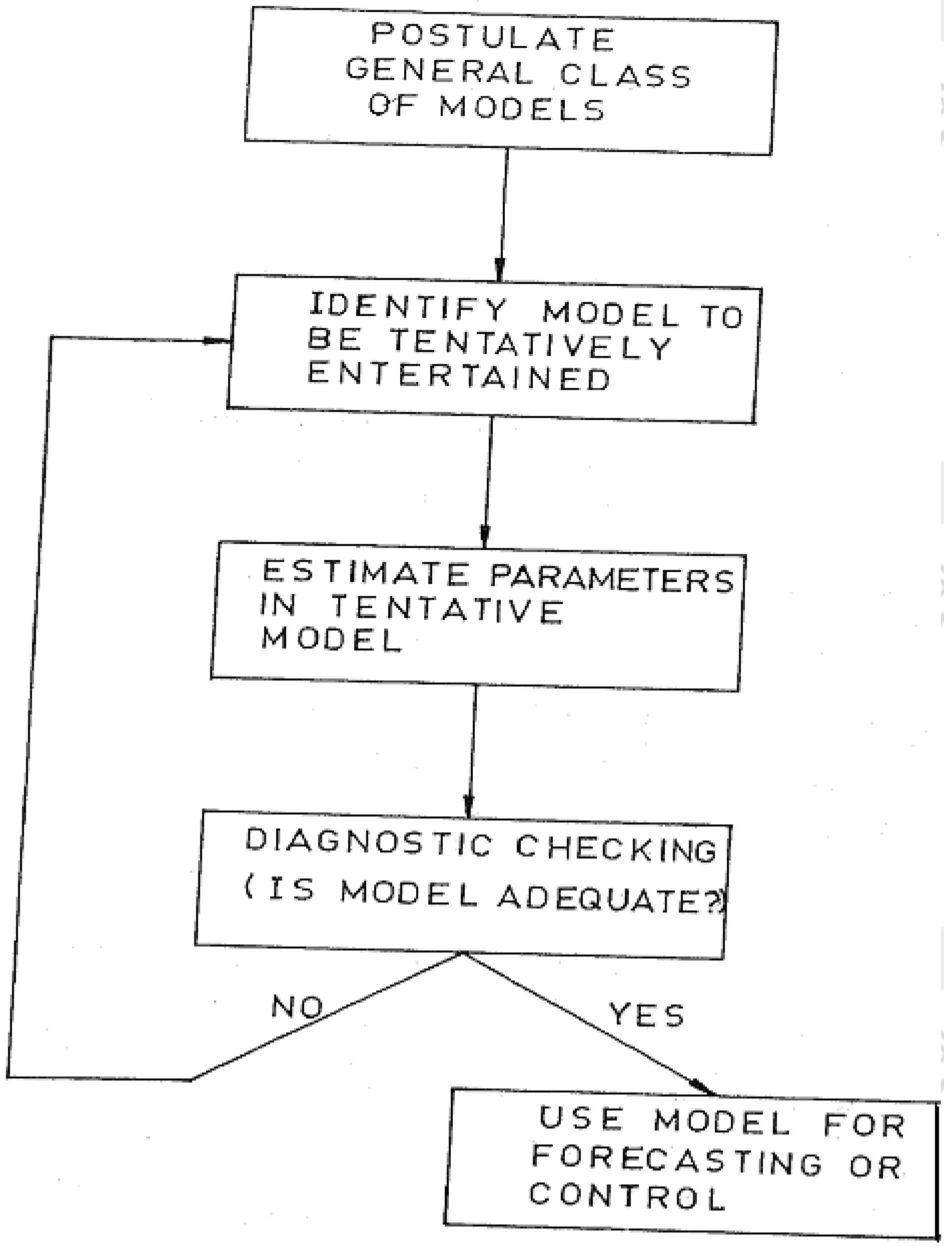
$$Z_t = \phi_1 Z_{t-1} + \ldots \phi_p Z_{t-p} + a_t - \theta_1 a_{t-1} - \ldots - \theta_q a_{t-q} \qquad \ldots (20.8)$$

In using such models in practice p and q are not greater than 2.

For non-stationary processes the most general model used is an auto-regressive integrated moving average (ARIMA) process of order (p, d, q) where d represents the degree of differencing to achieve stationarity in the process.

The main contribution of Box and Jenkins is the development of procedures for identifying the ARMA model that best fits a set of data and for testing the adequacy of that model. The various stages identified by Box and Jenkins in their interactive approach to model building are shown in Figure VI. For details on how such models are developed refer to Box and Jenkins.

Figure VI: The Box-Jenkins Methodology



20.6 SUMMARY

Some procedures for time series analysis have been described in this unit with a view to making more accurate and reliable forecasts of the future. Quite often the question that puzzles a person is how to select an appropriate forecasting method. Many times the problem context or time horizon involved would decide the method or limit the choice of methods. For instance, in new areas of technology forecasting where historical information is scanty, one would resort to some subjective method like opinion poll or a DELPHI study. In situations where one is trying to control or manipulate a factor a causal model might be appropriate in identifying the key variables and their effect on the dependent variable.

In this particular unit, however, time series models or those models where historical data on demand or the variable of interest is available are discussed. Thus we are dealing with projecting into the future from the past. Such models are short term forecasting models.

The decomposition method has been discussed. Here the time series is broken up into seasonal, trend, cycle and random components from the given data and reconstructed for forecasting purposes. A detailed example to illustrate the procedure is also given.

Finally the framework of stochastic models used by Box and Jenkins for time series analysis has been outlined. The AR, MA, ARMA and ARIMA processes in Box-Jenkins models are briefly described so that the interested reader can pursue a detailed study on his own.

20.7 SELF-ASSESSMENT EXERCISES

1 What do you understand by time series analysis? How would you go about conducting such an analysis for forecasting the sales of a product in your firm?

2 Compare time series analysis with other methods of forecasting, briefly summarising the strengths and weaknesses of various methods.

3 What would be the considerations in the choice of a forecasting method?

4 Find the 4-quarter moving average of the following time series representing the quarterly production of coffee in an Indian State.

	Production (in Tonnes)			
Year	Quarter I	Quarter II	Quarter III	Quarter IV
1983	5	1	10	17
1984	7	1	10	16
1985	9	3	8	18
1986	5	2	15	19
1987	8	4	14	21

5 Given below is the data of production of a certain company in lakhs of units

Year	1981	1982	1983	1984	1985	1986	1987
Production	15	14	18	20	17	24	27

a) Compute the linear trend by the method of least squares.

b) Compute the trend values of each of the years.

6 Given the following data on factory production of a certain brand of motor vehicles, determine the seasonal indices by the ratio to moving average method for August and September, 1985.

	Production (in thousand units)											
Year	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
1985	7.92	7.81	7.91	7.03	7.25	7.17	5.01	3.90	4.64	7.03	6.88	6.14
1986	4.86	4.48	5.26	5.48	6.42	6.82	4.98	2.45	4.51	6.38	6.38	7.59

7 A survey of used car sales in a city for the 10-year period 1976-85 has been made. A linear trend was fitted to the sales for month for each year and the equation was found to be

$Y = 400 + 18\,t$

where t = 0 on January 1, 1981 and t is measured in $\frac{1}{2}$ year (6 monthly) units

a) use this trend to predict sales for June, 1990

b) If the actual sales in June. 1987 are 600 and the relative seasonal index for June sales is 1.20, what would be the relative cyclical, irregular index for June, 1987?

9 The monthly sales for the last one year of a product in thousands of units are given below :

Month	1	2	3	4	5	6	7	8	9	10	11	12
Sales	0.5	1.5	2.2	3.0	3.2	3.5	3.5	3.5	3.8	4.0	4.7	5.5

Compute the auto-correlation coefficients up to lag 4. What conclusion can be derived from these values regarding the presence of a trend in the data?

20.8 KEY WORDS

Auto-correiation : Similar to correlation in that it Describes the association between values of the same variable but at different time periods. Auto-correlation coefficients provide important information about the underlying patterns in the data.

Auto-regressive/Moving Average (ARMA) Models : Auto-regressive(AR) models assume that future values are linear combinations of past values. Moving Average (MA) models, on the other hand, assume that future values are linear combinations of past errors. .A combination of the two is called an "Auto-regressive/Moving Average (ARMA) model".

Decomposition : Identifying the trend, seasonality, cycle and randomness in a time series.

Forecasting : Predicting the future values of a variable based on historical values of the same or other variable(s). If the forecast is based simply on past values of the variable itself, it is called time series forecasting, otherwise it is a causal type forecasting.

Seasonal Index : A number with a base of 1.00 that indicates the seasonality for a given period in relation to other periods.

Time Series Model : A model that predicts the future by expressing it as a function of the past.

Trend : A growth or decline in the mean value of a variable over the relevant time span.

20.9 FURTHER READINGS

Box, G.E.P. and G.M. *Jenkinsx 1976. Time Series Analysis, Forecasting and Control,* Holden-Day: San Francisco.

Chambers, J.C., S.K. Mullick and D.D. Smith, 1974. *An Executive's Guide to Forecasting,* John Wiley: New York.

Makridakis, S. and S. Wheelwright, 1978, *interactive Forecasting: Univariate and Multivariate Methods,* Holden-Day: San Francisco.

Makridakis, S. and S. Wheelwright, 1978. *Forecasting: Methods and Applications,* John Wiley, New York.

Montgomery, D.C. and L.A. Johnson, 1976. *Forecasting and Time Series Analysis,* McGraw Hill: New York.

Nelson, C.R., 1973. *Applied Time Series Analysis for Managerial Forecasting,* Holden-Day: