

INDUSTRIAL STATISTICS AND OPERATIONAL MANAGEMENT

6 : FORECASTING TECHNIQUES

Dr. Ravi Mahendra Gor

Associate Dean

ICFAI Business School

ICFAI HUse,

Nr. GNFC INFO Tower

S. G. Road

Bodakdev

Ahmedabad-380054

Ph.: 079-26858632 (O); 079-26464029 (R); 09825323243 (M)

E-mail: ravigor@hotmail.com

Contents

[Introduction](#)

[Some applications of forecasting](#)

[Defining forecasting](#)

[General steps in the forecasting process](#)

[Qualitative techniques in forecasting](#)

[Time series methods](#)

The Naive Methods

Simple Moving Average Method

Weighted Moving Average

Exponential Smoothing

[Evaluating the forecast accuracy](#)

[Trend Projections](#)

Linear Regression Analysis

Least Squares Method for Linear Regression

[Decomposition of the time series](#)

[Selecting A Suitable Forecasting Method](#)

[More on Forecast Errors](#)

[Review Exercise](#)

CHAPTER 6

FORECASTING TECHNIQUES

6.1 Introduction:

Every manager would like to know exact nature of future events to accordingly take action or plan his action when sufficient time is in hand to implement the plan. The effectiveness of his plan depends upon the level of accuracy with which future events are known to him. But every manager plans for future irrespective of the fact whether future events are exactly known or not. That implies, he does try to *forecast* future to the best of his Ability, Judgment and Experience.

Virtually all management decisions depend on forecasts. Managers study sales forecasts, for example, to take decisions on working capital needs, the size of the work force, inventory levels, the scheduling of production runs, the location of facilities, the amount of advertising and sales promotion, the need to change prices, and many other problems.

For our purpose forecasting can be defined as attempting to predict the future by using qualitative or quantitative methods. In an informal way, forecasting is an integral part of all human activity, but from the business point of view increasing attention is being given to formal forecasting systems which are continually being refined. Some forecasting systems involve very advanced statistical techniques beyond the scope of this book, so are not included.

All forecasting methodologies can be divided into three broad headings i.e. forecasts based on:

<i>What people have done Examples:</i>	<i>What people say examples:</i>	<i>What people do examples:</i>
<i>Time Series Analysis</i>	Surveys	Testing Marketing
<i>Regression Analysis</i>	Questionnaires	Reaction tests

The data from past activities are cheapest to collect but may be outdated and past behavior is not necessarily indicative of future behavior.

Data derived from surveys are more expensive to obtain and needs critical appraisal - intentions as expressed in surveys and questionnaires are not always translated into action.

Finally, the data derived from recording what people actually do are the most reliable but also the most expensive and occasionally it is not feasible for the data to be obtained.

Forecasting is a process of estimating a future event by casting forward past data. The past data are systematically combined in a predetermined way to obtain the estimate of the future. **Prediction** is a process of estimating a future event based on subjective considerations other than just past data; these subjective considerations need not be combined in a predetermined way.

Thus forecast is an estimate of future values of certain specified indicators relating to a decisional/planning situation, In some situations forecast regarding single indicator is sufficient, where as, in some other situations

forecast regarding several indicators is necessary. The number of indicators and the degree of detail required in the forecast depends on the intended use of the forecast.

There are two basic reasons for the need for forecast in any field.

1. **Purpose** – Any action devised in the PRESENT to take care of some contingency accruing out of a situation or set of conditions set in future. These future conditions offer a purpose / target to be achieved so as to take advantage of or to minimize the impact of (if the foreseen conditions are adverse in nature) these future conditions.
2. **Time** – To prepare plan, to organize resources for its implementation, to implement; and complete the plan; all these need time as a resource. Some situations need very little time, some other situations need several years of time. Therefore, if future forecast is available in advance, appropriate actions can be planned and implemented ‘intime’.

6.2 Some Applications of Forecasting:

Forecasts are vital to every business organization and for every significant management decision.

We now will discuss some areas in which forecasting is widely used.

Sales Forecasting

Any company in selling goods needs to forecast the demand for those goods. Manufacturers need to know how much to produce. Wholesalers and retailers need to know how much to stock. Substantially understanding demand is likely to lead to many lost sales, unhappy customers, and perhaps allowing the competition to gain the upper hand in the marketplace. On the other hand, significantly overestimating demand also is very costly due to (1) excessive inventory costs, (2) forced price reductions, (3) unneeded production or storage capacity, and (4) lost opportunities to market more profitable goods. Successful marketing and production managers understand very well the importance of obtaining good sales forecasts.

For the production managers these sales forecast are essential to help trigger the forecast for production which in turn triggers the forecasting of the raw materials needed for production.

Forecasting the need for raw materials and spare parts

Although effective sales forecasting is a key for virtually any company, some organizations must rely on other types of forecasts as well. A prime example involves forecasts of the need for raw materials and spare parts.

Many companies need to maintain an inventory of spare parts to enable them to quickly repair either own equipment or their products sold or leased to customers.

Forecasting Economic Trends

With the possible exception of sales forecasting, the most extensive forecasting effort is devoted to forecasting economic trends on a regional, national, or even international level.

Forecasting Staffing Needs

For economically developed countries there is a shifting emphasis from manufacturing to services. Goods are being produced outside the country (where labor is cheaper) and then imported. At the same time, an increasing number of business firms are specializing in providing a service of some kind (e.g., travel, tourism, entertainment, legal aid, health services, financial, educational, design, maintenance, etc.). For such a company forecasting “sales” becomes forecasting the demand for services, which then translates into forecasting staffing needs to provide those services.

Forecasting in education environment

A good education institute typically plans its activities and areas concentration for the coming years based on the forecasted demand for its different activities. The institute may come out with a forecast that the future requirements of its students who graduate may be more in particular sector. This may call for the reorientation of the syllabus and faculty, development of suitable teaching materials/cases, recruitment of new faculty with specific sector-oriented background, experience and teaching skills. Alternatively, the management may decide that the future is more secure with the conventional areas of operation and it may continue with the original syllabus, etc.

Forecasting in a rural setting

Cooperative milk producers', union operates in a certain district. The products it manufactures, the production capacities it creates, the manpower it recruits, and many more decisions are closely linked with the forecasts of the milk it may procure and the different milk products it may see. Milk being a product which has a ready market, is not difficult to sell. Thus demand forecasting for products may not be a very dominant issue for the organization. However, the forecast of milk procurement is a crucial issue as raw milk is a highly perishable commodity and building up of adequate processing capacity is important for the dairy. The milk procurement forecast also forms an important input to the production planning process which includes making decisions on what to produce, how much and when to produce.

Ministry of Petroleum

The officials of this crucial ministry have to make decisions on the quantum of purchase to be made for various types of crude oils and petroleum products from different sources across the oil-exporting nations for the next few years. They also have to decide as to how much money has to be spent on development of indigenous sources. These decisions involve/need information on the future demand of different types of petroleum products and the likely change in the prices and the availability of crude oil and petroleum products in the country and the oil-exporting nations. This takes us back to the field of forecasting.

Department of Technology

The top officials of this department want to make decisions on the type of information technology to recommend to the union government for the next decade. But they are not very clear on the directions which will be taken by this year rapidly changing field. They decided to entrust this task to the information system group of a national management institute. The team leader decided to forecast the changing technology in this area with the help of a team of information technology experts throughout the country. This is again a forecasting problem although of a much different type. This field of forecasting is known as technological forecasting.

Forecasting is the basis of corporate long-run planning. In the functional areas of finance and accounting, forecasts provide the basis for budgetary planning and cost control. Marketing relies on sales forecasting to plan new products, compensate sales personnel, and make other key decisions. Productions and operations personnel use forecasts to make periodic decisions involving process selection, capacity planning, and facility layout, as well as for continual decisions about production planning, scheduling, and inventory.

As we have observed in the aforementioned examples, forecasting forms an important input in many business and social science-related situations.

6.3 Defining Forecasting:

A forecast is an estimate of a future event achieved by systematically combining and casting forward in predetermined way data about the past. It is simply a statement about the future. It is clear that we must distinguish between forecast per se and good forecasts. Good forecast can be quite valuable and would be worth a great deal. Long-run planning decisions require consideration of many factors: general economic conditions, industry trends, probable competitor's actions, overall political climate, and so on.

Forecasts are possible only when a history of data exists. An established TV manufacturer can use past data to forecast the number of picture screens required for next week's TV assembly schedule. But suppose a manufacturer offers a new refrigerator or a new car, he cannot depend on past data. He cannot forecast, but has to predict. For prediction, good subjective estimates can be based on the manager's skill experience, and judgment. One has to remember that a forecasting technique requires statistical and management science techniques.

In general, when business people speak of forecasts, they usually mean some combination of both forecasting and prediction. Forecasts are often classified according to time period and use. In general, short-term (up to one year) forecasts guide current operations. Medium-term (one to three years) and long-term (over five years) forecasts support strategic and competitive decisions.

Bear in mind that a perfect forecast is usually impossible. Too many factors in the business environment cannot be predicted with certainty. Therefore, rather than search for the perfect forecast, it is far more important to establish the practice of continual review of forecasts and to learn to live with inaccurate forecasts. This is not to say that we should not try to improve the forecasting model or methodology, but that we should try to find and use the best forecasting method available, within reason. Because forecasts deal with past data, our forecasts will be less reliable the further into the future we predict. That means forecast accuracy decreases as time horizon increases. The accuracy of the forecast and its costs are interrelated. In general, the higher the need for accuracy translates to higher costs of developing forecasting models. So how much money and manpower is budgeted for forecasting? What possible benefits are accrued from accurate forecasting? What are possible cost of inaccurate forecasting? The best forecast are not necessarily the most accurate or the least costly. Factors as purpose and data availability play important role in determining the desired accuracy of forecast.

When forecasting, a good strategy is to use two or three methods and look at them for the commonsense view. Will expected changes in the general economy affect the forecast? Are there changes in industrial and private consumer behaviors? Will there be a shortage of essential complementary items? Continual review and updating in light of new data are basic to successful forecasting. In this chapter we look at qualitative and quantitative forecasting and concentrate primarily on several quantitative time series techniques.

The following figure illustrates various methods of forecasting.

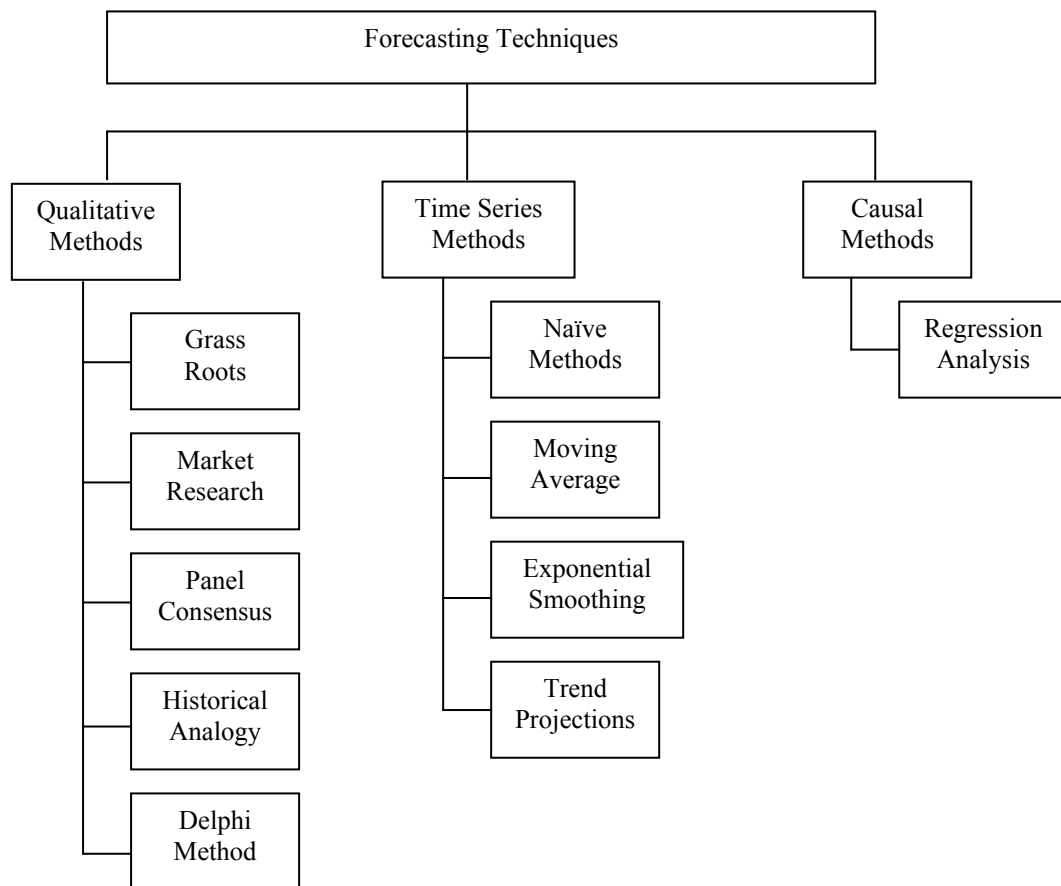


Fig. 6.1 Different Forecasting Methods

6.4 General Steps In The Forecasting Process

The general steps in the forecasting process are as follows:

- 1) *Identify the general need*
- 2) *Select the Period (Time Horizon) of Forecast*
- 3) *Select Forecast Model to be used:* For this, knowledge of various forecasting models, in which situations these are applicable, how reliable each one of them is; what type of data is required. On these considerations; one or more models can be selected.
- 4) *Data Collection:* With reference to various indicators identified-collect data from various appropriate sources-data which is compatible with the model(s) selected in step(3). Data should also go back that much in past, which meets the requirements of the model.
- 5) *Prepare forecast:* Apply the model using the data collected and calculate the value of the forecast.
- 6) *Evaluate:* The forecast obtained through any of the model should not be used, as it is, blindly. It should be evaluated in terms of 'confidence interval' – usually all good forecast models have methods of calculating upper value and the lower value within which the given forecast is expected to remain with

certain specified level of probability. It can also be evaluated from logical point of view whether the value obtained is logically feasible? It can also be evaluated against some related variable or phenomenon. Thus, it is possible, some times advisable to modify the statistically forecasted' value based on evaluation.

6.5 Qualitative Techniques In Forecasting

Grass Roots

Grass roots forecasting builds the forecast by adding successively from the bottom. The assumption here is that the person closest to the customer or end use of the product knows its future needs best. Though this is not always true, in many instances it is a valid assumption, and it is the basis for this method.

Forecasts at this bottom level are summed and given to the next higher level. This is usually a district warehouse, which then adds in safety stocks and any effects of ordering quantity sizes. This amount is then fed to the next level, which may be a regional warehouse. The procedure repeat until it becomes an input at the top level, which, in the case of a manufacturing firm, would be the input to the production system.

Market Research:

Firms often hire outside companies that specialize in market research to conduct this type of forecasting. You may have been involved in market surveys through a marketing class. Certainly you have not escaped telephone calls asking you about product preferences, your income, habits, and so on.

Market research is used mostly for product research in the sense of looking for new product ideas, likes and dislikes about existing products, which competitive products within a particular class are preferred, and so on. Again, the data collection methods are primarily surveys and interviews.

Panel Consensus:

In a panel consensus, the idea that two heads are better than one is extrapolated to the idea that a panel of people from a variety of positions can develop a more reliable forecast than a narrower group. Panel forecasts are developed through open meetings with free exchange of ideas form all levels of management and individuals. The difficulty with this open style is that lower employee levels are intimidated by higher levels of management. For example, a salesperson in a particular product line may have a good estimate of future product demand but may not speak up to refute a much different estimate given by the vice president of marketing. The Delphi technique (which we discuss shortly) was developed to try to correct this impairment to free exchange.

When decisions in forecasting are at a broader, higher level (as when introducing a new product line or concerning strategic product decisions such as new marketing areas) the term executive judgment is generally used. The term is self-explanatory: a higher level of management is involved.

Historical Analogy:

The historical analogy method is used for forecasting the demand for a product or service under the circumstances that no past demand data are available. This may specially be true if the product happens to be new for the organization. However, the organization may have marketed product(s) earlier which may be similar in some features to the new product. In such circumstances, the marketing personnel use the historical analogy between the two products and derive the demand for the new product using the historical data of the earlier

product. The limitations of this method are quite apparent. They include the questionable assumption of the similarity of demand behaviors, the changed marketing conditions, and the impact of the substitutability factor on the demand.

Delphi Method:

As we mentioned under panel consensus, a statement or opinion of a higher-level person will likely be weighted more than that of a lower-level person. The worst case is where lower level people feel threatened and do not contribute their true beliefs. To prevent this problem, the Delphi method conceals the identity of the individuals participating in the study. Everyone has the same weight. A moderator creates a questionnaire and distributes it to participants. Their responses are summed and given back to the entire group along with a new set of questions.

The Delphi method was developed by the Rand Corporation in the 1950s. The step-by-step procedure is

- 1) Choose the experts to participate. There should be a variety of knowledgeable people in different areas.
- 2) Through a questionnaire (or e-mail), obtain forecasts (and any premises or qualification captions for the forecasts) from all participants.
- 3) Summarize the results and redistribute them to the participants along with appropriate new questions.
- 4) Summarize again, refining forecasts and conditions, and again develop new questions.
- 5) Repeat Step 4 if necessary. Distribute the final results to all participants.

The Delphi technique can usually achieve satisfactory results in three rounds. The time required is a function of the number of participants, how much work is involved for them to develop their forecasts, and their speed in responding.

We now discuss the quantitative methods of forecasting

6.6 Time-Series Methods

In many forecasting situations enough historical consumption data are available. The data may relate to the past periodic sales of products, demands placed on services like transportation, electricity and telephones. There are available to the forecaster a large number of methods, popularly known as the time series methods, which carry out a statistical analysis of past data to develop forecasts for the future. The underlying assumption here is that past relationships will continue to hold in the future. The different methods differ primarily in the manner in which the past values are related to the forecasted ones.

A time series refers to the past recorded values of the variables under consideration. The values of the variables under consideration in a time-series are measured at specified intervals of time. These intervals may be minutes, hours, days, weeks, months, etc. In the analysis of a time series the following four time-related factors are important.

- 1) **Trends:** These relate to the long-term persistent movements/tendencies/changes in data like price increases, population growth, and decline in market shares. An example of a decreasing linear trend is shown in Fig. 6.2

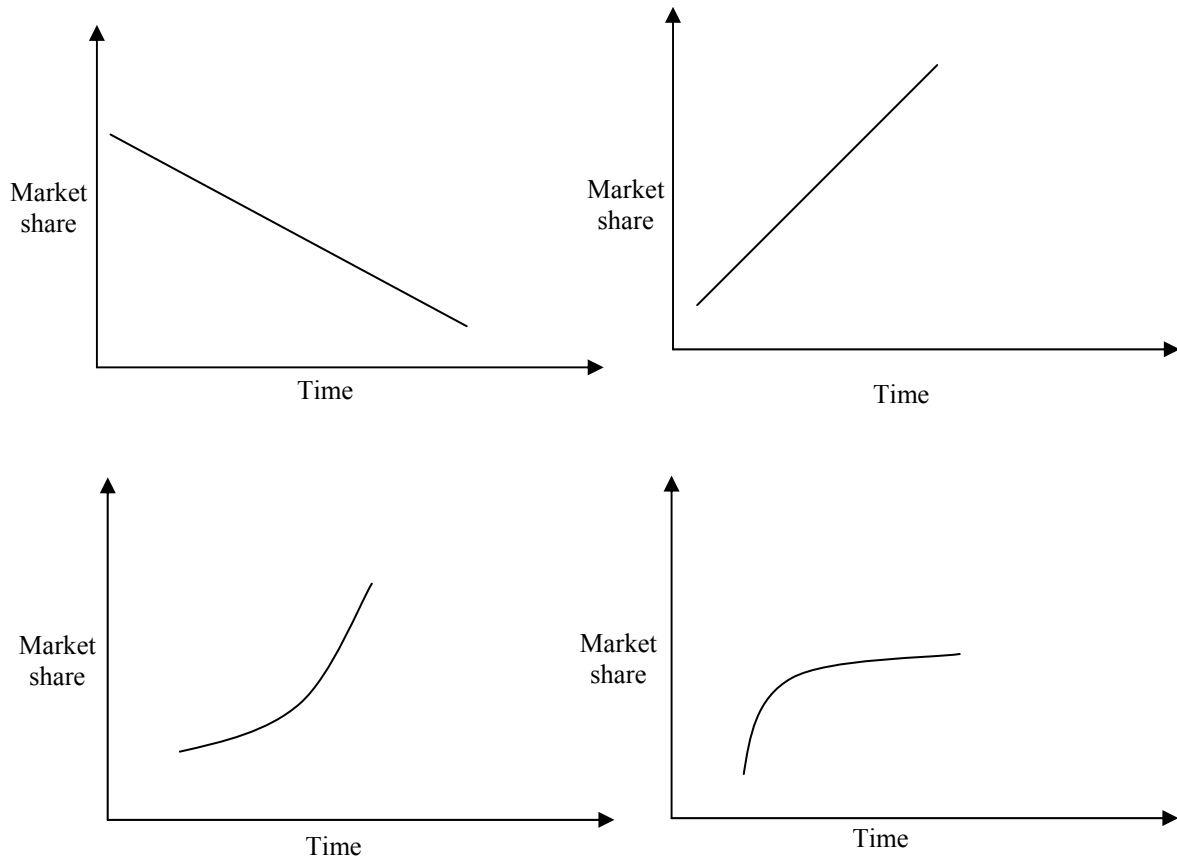


Fig. 6.2

- (2) **Seasonal variations:** There could be periodic, repetitive variations in time-series which occur because of buying or consuming patterns and social habits, during different times of a year. The demand for products like soft drinks, woolens and refrigerators, also exhibits seasonal variations. An illustration of seasonal variations is provided in Fig. 6.3

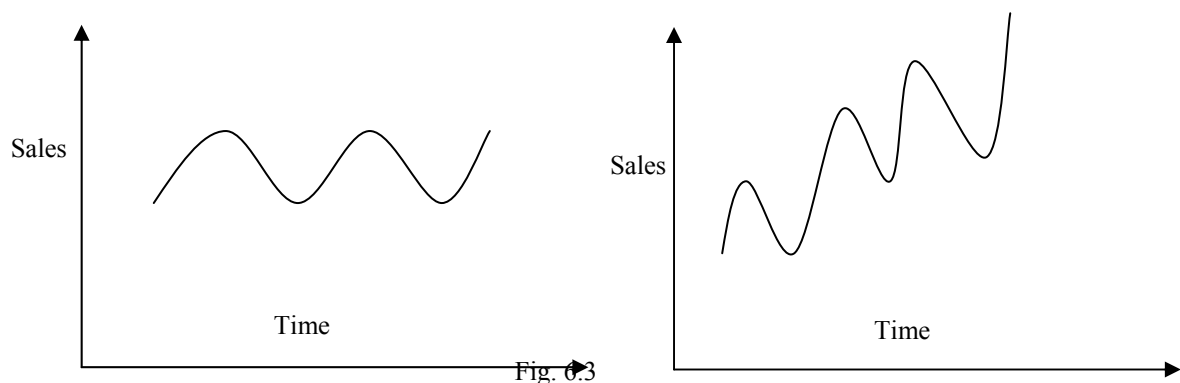


Fig. 6.3

- (3) **Cyclical variations:** These refer to the variations in time series which arise out of the phenomenon of business cycles. The business cycle refers to the periods of expansion followed by periods of contraction.

The period of a business cycle may vary from one year to thirty years. The duration and the level of resulting demand variation due to business cycles are quite difficult to predict.

- (4) **Random or irregular variations:** These refer to the erratic fluctuations in the data which cannot be attributed to the trend, seasonal or cyclical factors. In many cases, the root cause of these variations can be isolated only after a detailed analysis of the data and the accompanying explanations, if any. Such variations can be due to a wide variety of factors like sudden weather changes, strike or a communal clash. Since these are truly random in nature, their future occurrence and the resulting impact on demand are difficult to predict. The effect of these events can be eliminated by smoothing the time series data. A graphical example of the random variations is given in Fig. 6.4.

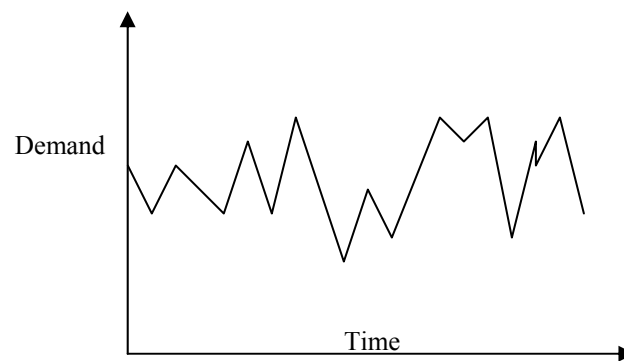


Fig. 6.4

The historical time series, as obtained from the past records, contains all the four factors described earlier. One of the major tasks is to isolate each of the components, as elegantly as possible. This process of desegregating the time series is called **decomposition**. The main objective here is to isolate the trend in time series by eliminating the other components. The trend line can then be used for projecting into the future. The effect of the other components on the forecast can be brought about by adding the corresponding cyclical, seasonal and irregular variations.

In most short-term forecasting situations the elimination of the cyclical component is not attempted. Also, it is assumed that the irregular variations are small and tend to cancel each other out over time. Thus, the major objective, in most cases, is to seek the removal of seasonal variations from the time series. This process is known as **deseasonalization** of the time series data.

There are a number of time-series-based methods. Not all of them involve explicit decomposition of the data. The methods extend from mathematically very simple to fairly complicated ones.

Let us also see some of the time series models are based on the trend lines of the data.

The *constant-level models* assume no trend at all in the data. The time series is assumed to have a relatively constant mean. The forecast for any period in the future is a horizontal line.

Linear trend models forecast a straight-line trend for any period in the future. Refer Fig. 6.2

Exponential trends forecast that the amount of growth will increase continuously. At long horizons, these trends become unrealistic. Thus models with a damped trend have been developed for longer-range forecasting. The amount of trend extrapolated declines each period in a damped trend model. Eventually, the trend dies out and the forecasts become a horizontal line. Refer Fig 6.2

The additive seasonal pattern assumes that the seasonal fluctuations are of constant size.

The *multiplicative pattern* assumes that the seasonal fluctuations are proportional to the data. As the trend increases, the seasonal fluctuations get larger. Refer Fig 6.3

6.6.1 The Naive Methods

The forecasting methods covered under this category are mathematically very simple. The simplest of them uses the most recently observed value in the time series as the forecast for the next period. Effectively, this implies that all prior observations are not considered. Another method of this type is the ‘free-hand projection method’. This includes the plotting of the data series on a graph paper and fitting a free-hand curve to it. This curve is extended into the future for deriving the forecasts. The ‘semi-average projection method’ is another naive method. Here, the time-series is divided into two equal halves, averages calculated for both, and a line drawn connecting the two semi averages. This line is projected into the future and the forecasts are developed.

Illustration 6.1: Consider the demand data for 8 years as given. Use these data for forecasting the demand for the year 1991 using the three naive methods described earlier.

<i>Year</i>	<i>Actual sales</i>
<i>1983</i>	100
<i>1984</i>	105
<i>1985</i>	103
<i>1986</i>	107
<i>1987</i>	109
<i>1988</i>	110
<i>1989</i>	115
<i>1990</i>	117

Solution: The forecasted demand for 1991, using the last period method = actual sales in 1990 = 117 units.

The forecasted demand for 1991, using the free-hand projection method = 119 units. (Please check the results using a graph papers!)

The semi-averages for this problem will be calculated for the periods 1983-86 and 1987-90. The resultant semi-averages are 103.75 and 112.75. A straight line joining these points would lead to a forecast for the year 1991. The value of this forecast will be = 120 units

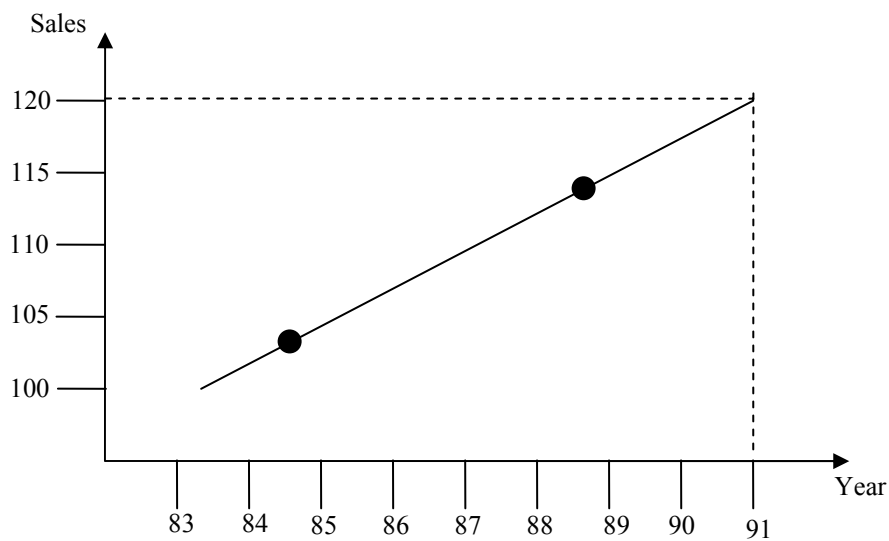


Fig. 6.5

6.6.2 Simple Moving Average Method

When demand for a product is neither growing nor declining rapidly, and if it does not have seasonal characteristics, a moving average can be useful in removing the random fluctuations for forecasting. Although moving averages are frequently centered, it is more convenient to use past data to predict the following period directly. To illustrate, a centered five-month average of January, February, March, April and May gives an average centered on March. However, all five months of data must already exist. If our objective is to forecast for June, we must project our moving average- by some means- from March to June. If the average is not centered but is at forward end, we can forecast more easily, though we may lose some accuracy. Thus, if we want to forecast June with a five-month moving average, we can take the average of January, February, March, April and May. When June passes, the forecast for July would be the average of February, March, April, May and June.

Although it is important to select the best period for the moving average, there are several conflicting effects of different period lengths. The longer the moving average period, the more the random elements are smoothed (which may be desirable in many cases). But if there is a trend in the data-either increasing or decreasing-the moving average has the adverse characteristic of lagging the trend. Therefore, while a shorter time span produces more oscillation, there is a closer following of the trend. Conversely, a longer time span gives a smoother response but lags the trend.

The formula for a *simple moving average* is

$$F_t = \frac{A_{t-1} + A_{t-2} + A_{t-3} + \dots + A_{t-n}}{n}$$

where, F_t = Forecast for the coming period, n = Number of period to be averaged and A_{t-1} , A_{t-2} , A_{t-3} and so on are the actual occurrences in the in the past period, two periods ago, three periods ago and so on respectively.

Illustration 6.2: The data in the first two columns of the following table depict the sales of a company. The first two columns show the month and the sales.

The forecasts based on 3, 6 and 12 month moving average and shown in the next three columns.

The 3 month moving average of a month is the average of sales of the preceding three months. The reader is asked to verify the calculations himself.

<i>Past sales of generators</i>		<i>Forecasts produced by</i>		
<i>Month</i>	Actual units sold	3 month moving average	6 month moving average	12 month moving average
<i>January</i>	450			
<i>February</i>	440			
<i>March</i>	460			
<i>April</i>	410	$(450+440+460)/3 = 450$		
<i>May</i>	380	$(440+460+410)/3 = 437$		
<i>June</i>	400	$(460+410+380)/3 = 417$		
<i>July</i>	370	397	423	
<i>August</i>	360	383	410	
<i>September</i>	410	377	397	
<i>October</i>	450	380	388	
<i>November</i>	470	407	395	
<i>December</i>	490	443	410	
<i>January</i>	460	470	425	424

The 6 month moving average is given by the average of the preceding 6 months actual sales.

For the month of July it is calculated as

$$\begin{aligned} \text{July's forecast} &= (\text{Sum of the actual sales from January to June}) / 6 \\ &= (450 + 440 + 460 + 410 + 380 + 400) / 6 \\ &= 423 \text{ (rounded)} \end{aligned}$$

For the forecast of January by the 12 month moving average we sum up the actual sales from January to December of the preceding year and divide it by 12.

Note:

1. A moving average can be used as a forecast as shown above but when graphing moving averages it is important to realize, that being averages, they must be plotted at the mid point of the period to which they relate.
2. Twelve-monthly moving averages or moving annual totals form part of a commonly used diagram, called the Z chart. It is called a Z chart because the completed diagram is shaped like a Z. The top part of the Z is formed by the moving annual total, the bottom part by the individual monthly figures and the sloping line by the cumulative monthly figures.

Illustration 6.3: Using the data given in the Illustration 1 forecast the demand for the period 1987 to 1991 using

- a. 3- year moving average and
- b. 5- year moving average

If we want to check the error in our forecast as $Error = Actual\ observed\ value - Forecasted\ value$

find which one gives a lower error in the forecast.

Year	Demand	Three year moving average		Five year moving average	
		forecast	error	forecast	error
1983	100	-	-	-	-
1984	105	-	-	-	-
1985	103	-	-	-	-
1986	107	102.6	4.4	-	-
1987	109	105.0	4.0	-	-
1988	110	106.3	3.7	104.8	5.2
1989	115	108.6	6.4	106.8	8.2
1990	117	111.3	5.7	108.8	8.2
1991	-	114.0	-	111.6	-

Here we observe that the forecast always lags behind the actual values. The lag is greater for the 5-year moving average based forecasts.

Characteristics of moving averages

- a. The different moving averages produce different forecasts.
- b. The greater the number of periods in the moving average, the greater the smoothing effect.
- c. If the underlying trend of the past data is thought to be fairly constant with substantial randomness, then a greater number of periods should be chosen.
- d. Alternatively, if there is thought to be some change in the underlying state of the data, more responsiveness is needed, therefore fewer periods should be included in the moving average.

Limitations of moving averages

- a. Equal weighting is given to each of the values used in the moving average calculation, whereas it is reasonable to suppose that the most recent data is more relevant to current conditions.
- b. An n period moving average requires the storage of $n - 1$ values to which is added the latest observation. This may not seem much of a limitation when only a few items are considered, but it becomes a significant factor when , for example, a company carries 25,000 stock items each of which requires a moving average calculation involving say 6 months usage data to be recorded.

- c. The moving average calculation takes no account of data outside the period of average, so full use is not made of all the data available.
- d. The use of the unadjusted moving average as a forecast can cause misleading results when there is an underlying seasonal variation.

6.6.3 Weighted Moving Average

Whereas the simple moving average gives equal weight to each component of the moving average database, a weighted moving average allows any weights to be placed on each element, providing, of course, that the sum of all weights equals 1. For example, a department store may find that in a four-month period, the best forecast is derived by using 40 percent of the actual sales for the most recent month, 30 percent of two months ago, 20 percent of three months ago, and 10 percent of four months ago. If actual sales experience was

Month 1	Month 2	Month 3	Month 4	Month 5
100	90	105	95	?

the forecast for month 5 would be

$$\begin{aligned}
 F_5 &= 0.40(95) + 0.30(105) + 0.20(90) + 0.10(100) \\
 &= 38 + 31.5 + 18 + 10 \\
 &= 97.5
 \end{aligned}$$

The formula for the *weighted moving average* is

$$F_t = w_1 A_{t-1} + w_2 A_{t-2} + w_3 A_{t-3} + \dots + w_n A_{t-n}$$

Where F_t = Forecast for the coming period, n = the total number of periods in the forecast.

w_i = the weight to be given to the actual occurrence for the period $t-i$

A_i = the actual occurrence for the period $t-i$

Although many periods may be ignored (that is, their weights are zero) and the weighting scheme may be in any order (for example, more distant data may have greater weights than more recent data), the sum of all the weights must equal 1.

$$\sum_{i=1}^n w_i = 1$$

Suppose sales for month 5 actually turned out to be 110. Then the forecast for month 6 would be

$$\begin{aligned}
 F_6 &= 0.40(110) + 0.30(95) + 0.20(105) + 0.10(90) \\
 &= 44 + 28.5 + 21 + 9 \\
 &= 102.5
 \end{aligned}$$

Choosing Weights : Experience and trial and error are the simplest ways to choose weights. As a general rule, the most recent past is the most important indicator of what to expect in the future, and, therefore, it should get higher weighting. The past month's revenue or plant capacity, for example, would be a better estimate for the coming month than the revenue or plant capacity of several months ago.

However, if the data are seasonal, for example, weights should be established accordingly. For example, sales of air conditioners in May of last year should be weighted more heavily than sales of air conditioners in December.

The weighted moving average has a definite advantage over the simple moving average in being able to vary the effects of past data. However, it is more inconvenient and costly to use than the exponential smoothing method, which we examine next.

6.6.4 Exponential Smoothing

In the previous methods of forecasting (simple and weighted moving average), the major drawback is the need to continually carry a large amount of historical data. (This is also true for regression analysis techniques, which we soon will cover) As each new piece of data is added in these methods, the oldest observation is dropped, and the new forecast is calculated. In many applications (perhaps in most), the most recent occurrences are more indicative of the future than those in the more distant past. If this premise is valid – “that the importance of data diminishes as the past becomes more distant” - then exponential smoothing may be the most logical and easiest method to use.

The reason this is called exponential smoothing is that each increment in the past is decreased by $(1-\alpha)$. If α is 0.05 for example, weights for various period would be as follows (α is defined below):

Weighting at $\alpha = 0.05$		
Most recent weighting	$= \alpha (1 - \alpha)^0$	0.0500
Data one time period older	$= \alpha (1 - \alpha)^1$	0.0475
Data two time periods older	$= \alpha (1 - \alpha)^2$	0.0451
Data three time periods older	$= \alpha (1 - \alpha)^3$	0.0429

Therefore, the exponents 0, 1, 2, 3 and so on give it its name.

The method involves the automatic weighting of past data with weights that decrease exponentially with time, i.e. the most current values receive a decreasing weighting.

The exponential smoothing technique is a weighted moving average system and the underlying principle is that the

New Forecast = Old Forecast + a proportion of the forecast error

The simplest formula is

$$New\ forecast = Old\ forecast + \alpha (Latest\ Observation - Old\ Forecast)$$

where α (alpha) is the smoothing constant.

Or more mathematically,

$$F_t = F_{t-1} + \alpha (A_{t-1} - F_{t-1})$$

i.e $F_t = \alpha A_{t-1} + (1 - \alpha) F_{t-1}$

Where

- F_t = The exponentially smoothed forecast for period t
- F_{t-1} = The exponentially smoothed forecast made for the prior period
- A_{t-1} = The actual demand in the prior period
- α = The desired response rate, or smoothing constant

The smoothing constant

The value of α can be between 0 and 1. The higher value of α (i.e. the nearer to 1), the more sensitive the forecast becomes to current conditions, whereas the lower the value, the more stable the forecast will be, i.e. it will react less sensitively to current conditions. An approximate equivalent of alpha values to the number of periods' moving average is given below:

α value	Approximate periods in equivalent Moving average
0.1	19
0.25	7
0.33	5
0.5	3

The total of the weights of observations contributing to the new forecast is 1 and the weight reduces exponentially progressively from the alpha value for the latest observation to smaller value for the older observations. For example, if the alpha value was 0.3 and June's sales were being forecast, then June's forecast is produced from averaging past sales weighted as follows.

$$\begin{aligned}
 &0.3 \text{ (May's Sales)} + 0.21 \text{ (April's Sales)} + 0.147 \text{ (March's Sales)} \\
 &+ 0.1029 \text{ (February Sales)} + 0.072 \text{ (January Sales)} \\
 &+ 0.050 \text{ (December Sales), etc.}
 \end{aligned}$$

In the above calculation, the reader will observe that $\alpha (1 - \alpha)^0 = 0.3$, $\alpha (1 - \alpha)^1 = 0.21$, $\alpha (1 - \alpha)^2 = 0.147$

$$\alpha (1 - \alpha)^3 = 0.1029 \text{ and so on.}$$

From this it will be noted that the weightings calculated approach a total of 1.

Exponential smoothing is the most used of all forecasting techniques. It is an integral part of virtually all computerized forecasting programs, and it is widely used in ordering inventory in retail firms, wholesale companies, and service agencies.

Exponential smoothing techniques have become well accepted for six major reasons:

1. Exponential models are surprisingly accurate
2. Formulating an exponential model is relatively easy
3. The user can understand how the model works
4. Little computation is required to use the model
5. Computer storage requirements are small because of the limited use of historical data
6. Tests for accuracy as to how well the model is performing are easy to compute

In the exponential smoothing method, only three pieces of data are needed to forecast the future: the most recent forecast, the actual demand that occurred for that forecast period and a **smoothing constant alpha** (α). This smoothing constant determines the level of smoothing and the speed of reaction to differences between forecasts and actual occurrences. The value for the constant is determined both by the nature of the product and by the manager's sense of what constitutes a good response rate. For example, if a firm produced a standard item with relatively stable demand, the reaction rate to difference between actual and forecast demand would tend to be

small, perhaps just 5 or 10 percentage points. However, if the firm were experiencing growth, it would be desirable to have a higher reaction rate, perhaps 15 to 30 percentage points, to give greater importance to recent growth experience. The more rapid the growth, the higher the reaction rate should be. Sometimes users of the simple moving average switch to exponential smoothing but like to keep the forecasts about the same as the simple moving average. In this case, α is approximated by $2 \div (n+1)$, where n is the number of time periods.

To demonstrate the method once again, assume that the long-run demand for the product under study is relatively stable and a smoothing constant (α) of 0.05 is considered appropriate. If the exponential method were used as a continuing policy, forecast would have been made for last month. Assume that last month's forecast (F_{t-1}) was 1,050 units. If 1,000 actually were demanded, rather than 1,050, the forecast for this month would be

$$\begin{aligned}
 F_t &= F_{t-1} + \alpha (A_{t-1} - F_{t-1}) \\
 &= 1,050 + 0.05 (1,000 - 1,050) \\
 &= 1,050 + 0.05 (-50) \\
 &= 1,047.5 \text{ units}
 \end{aligned}$$

Because the smoothing coefficient is small, the reaction of the new forecast to an error of 50 units is to decrease the next month's forecast by only 2.5 units.

Illustration 6.4: The data are given in the first two columns and the forecasts have been prepared using the values of α as 0.2 and 0.8.

Month	Actual units sold	Exponential Forecasts	
		$\alpha = 0.2$	$\alpha = 0.8$
January	450	-	-
February	440	450 **	450**
March	460	450 + 0.2 (440-450) = 448	450 + 0.8(440-450) =442
April	410	448 + 0.2 (460-448) = 450.4	442 + 0.8(460-442) =456.4
May	380	450.4 + 0.2 (410 - 450.4) = 442.32	456.4 + 0.8(410-456.4) =419.28
June	400	429.86	387.86
July	370	423.89	397.57
August	360	413.11	375.51
September	410	402.49	363.102
October	450	403.99	400.62
November	470	413.19	440.12
December	490	424.55	464.02
January	460	437.64	484.80

** In the above example, no previous forecast was available. So January sales were used as February's forecast. The reader is advised to check the calculations for himself

It is apparent that the higher α value, 0.8, produces a forecast which adjusts more readily to the most recent sales.

Extensions of exponential smoothing

The basic principle of exponential smoothing has been outlined above, but to cope with various problem such as seasonal factors strongly, rising or failing demand, etc many developments to the basic model have been made. These include double and triple exponential smoothing and correction for trend and delay factors, etc. These are outside the scope of the present book, so are not covered.

Characteristics of exponential smoothing

- a) Greater weight is given to more recent data
- b) All past data are incorporated there is no cut-off point as with moving averages
- c) Less data needs to be stored than with the longer period moving averages.
- d) Like moving averages it is an adaptive forecasting system. That is, it adapts continually as new data becomes available and so it is frequently incorporated as an integral part of stock control and production control systems.
- e) To cope with various problems (trend, seasonal factors, etc) the basic model needs to be modified
- f) Whatever form of exponential smoothing is adopted, changes to the model to suit changing conditions can simply be made by altering the α value.
- g) The selection of the smoothing constant α is done through trial-error by the researcher/analyst. It is done by testing several values of α (within the range 0 to 1) and selecting one which gives a forecast with the least error (one can take standard error). It has been found that values in the range 0.1 to 0.3 provide a good starting point.

Illustration 6.5: Data on production of an item for 30 periods are tabulated below. Determine which value of the smoothing constant (α), out of 0.1 and 0.3, will lead to the best simple exponential smoothing model. The first 15 values can be used for initialization of the model and then check the error in the forecast as asked after the table.

<i>Month</i>	<i>Production (in tones)</i>	<i>Month</i>	<i>Production (in tones)</i>	<i>Month</i>	<i>Production (in tones)</i>
1	30.50	11	25.70	21	27.60
2	28.80	12	30.90	22	29.90
3	31.50	13	31.50	23	30.20
4	29.90	14	28.10	24	35.50
5	31.40	15	30.80	25	30.80
6	33.50	16	29.50	26	28.80
7	25.70	17	29.80	27	30.80
8	32.10	18	30.00	28	32.20
9	29.10	19	29.90	29	31.20
10	30.80	20	31.50	30	29.80

Use the total
squared error or the mean squared error as the basis of comparison of the performances.

The total squared error is the sum of the squares of all the error figures for the period selected (here only the last 15 figures because the first 15 periods will initialize the forecast). Their mean is the mean squared error.

Solution: The following table give the forecasted values for the two different values of the smoothing constant for the first 15 periods.

<i>Month</i>	<i>Production (in tonnes)</i>	<i>$\alpha = 0.1$ Forecast</i>	<i>$\alpha = 0.3$ Forecast</i>
1	30.50	30.50	30.50
2	28.80	30.33	29.90
3	31.50	30.45	30.44
4	29.90	30.39	30.28
5	31.40	30.49	3.62
6	33.50	30.79	31.48
7	25.70	30.28	29.75
8	32.10	30.47	30.45
9	29.10	30.33	30.05
10	30.80	30.38	30.27
11	25.70	29.91	28.90
12	30.90	30.01	29.50
13	31.50	30.16	30.10
14	28.10	29.95	29.50
15	30.80	30.04	29.89

The following table now gives the forecasted values and also checks the errors in the forecast for the last 15 periods.

<i>Month</i>	<i>Production (in tonnes)</i>	<i>$\alpha = 0.1$</i>			<i>$\alpha = 0.3$</i>	
		Forecast	error	Squared error	forecast	Squared error
16	29.50	30.04	-0.54	0.29	29.89	0.15
17	29.80	29.98	-0.18	0.03	29.77	0.0
18	30.00	29.96	0.04	0.0	29.78	0.05
19	29.90	29.97	-0.07	0.0	29.85	0.0
20	31.50	29.96	1.54	2.37	29.86	2.68
21	27.60	30.12	-2.52	6.33	30.35	7.58
22	29.90	29.86	0.04	0.0	29.53	0.14
23	30.20	29.87	0.33	0.11	29.64	0.31
24	35.50	29.90	5.60	31.35	29.81	32.40
25	30.80	30.46	0.34	0.12	31.52	0.51
26	28.80	30.49	-1.69	2.87	31.30	6.25
27	30.80	30.32	0.48	0.23	30.55	0.06
28	32.20	30.37	1.83	3.34	30.63	2.48
29	31.20	30.56	0.64	0.42	31.10	0.01
30	29.80	30.62	-0.82	0.67	31.13	1.76
<i>TOTAL SQUARED ERROR</i>				48.13	54.41	
<i>MEAN SQUARED ERROR</i>				=48.13/15	=54.41/15	
				=3.20	=3.62	

The results here indicate that the forecast accuracy is better for $\alpha = 0.1$ as compared to 0.3. They also indicate that a search around 0.1 may lead to a more refined single exponential smoothing model.

6.7 Evaluating the forecast accuracy:

There are many ways to measure forecast accuracy. Some of these measures are the mean absolute forecast error, called the **MAD** (Mean Absolute Deviation), the mean absolute percentage error (**MAPE**) and the mean square error (**MSE**)

Error = Actual Observed value – Forecasted value

Absolute Percentage Error = (Error / Actual Observed Value) \times 100

MAD = the average of the absolute errors

MAPE = the average of the Absolute Percentage Errors

MSE = the average of the squared errors

It is common for two forecasting models to be ranked differently depending on the accuracy measure used. For example, model A may give a smaller MAD but a larger MSE than model B. Why? Because the MAD gives equal weight to each error. The MSE gives more weight to large errors because they are squared.

It is up to the manager, not the management scientist to decide which accuracy measure is most appropriate for his or her application. The MSE is most often used in practice.

Given a preferred accuracy measure, how do we know when our forecasts are good, bad, or indifferent? One way to answer this question is to compare the accuracy of a given model with that of a benchmark model. A handy benchmark is the naïve model, which assumes that the value of the series next period will be the same as it is this period; i.e., say $F_{t+1} = X_t$

where F is the forecast and X is the observed value. The subscript t is an index for the time period. The current period is $t + 1$.

The first step in any forecasting problem should be to use the naive model to compute the benchmark accuracy. A model which cannot beat the naive model should be discarded. Checking model accuracy against that of the naive model may seem to be a waste of time, but unless we do so, it is easy to choose an inappropriate forecasting model.

The mean error measures are computed only for the last half of the data. The forecasting models are evaluated by dividing the data in to two parts. The first part is used to fit the forecasting model. Fitting consists of running the model through the first part of the data to get “warmed up.” We call the fitting data the warm-up sample. The second part of the data is used to test the model and is called the forecasting sample. Accuracy in the warm-up sample is really irrelevant. Accuracy in the forecasting sample is more important because the pattern of the data often changes over time. The forecasting sample is used to evaluate how well the model tracks such changes. This point will be explained in detail in the next few sections.

There are no statistical rules on where to divide the data into warm-up samples and forecasting samples. There may not be enough data to have two samples. A good rule of thumb is to put at least six non seasonal data points or two complete seasons of seasonal data in the warm-up sample. If there are fewer data than this, there is no need to bother with two samples. In a long time series, it is common practice simply to divide the data in half.

Illustration 6.6: Data regarding the sales of a particular item in the 12 time periods is given below. The manager wants to forecast 1 time period ahead in order to plan properly. Determine the forecasts using

- (a) Naïve method
- (b) 3 period moving average
- (c) simple exponential smoothing taking $\alpha = 0.1$.

Also compute the errors MAD, MAPE, and MSE to check the forecasting accuracy for the last six periods.

Solution:

- (a) The following table shows the naïve forecasting model. In this model the next period's forecast is the present period's actual observed value

<i>Time period (T)</i>	<i>Demand (D)</i>	<i>Forecast (F)</i>	<i>Absolute Error (E = D - F)</i>	<i>Absolute Percentage error (E/D) × 100</i>	<i>Squared Error (E²)</i>
1	28	-	-	-	-
2	27	28	-	-	-
3	33	27	-	-	-
4	25	33	-	-	-
5	34	25	-	-	-
6	33	34	-	-	-
7	35	33	2	5.7%	4
8	30	35	5	16.7%	25
9	33	30	3	9.1%	9
10	35	33	2	5.7%	4
11	27	35	8	29.6%	64
12	29	27	2	6.9%	4
13		29			
<i>Total of time periods 7 to 12 to check forecasting accuracy</i>			22	73.7%	110
<i>MAD = 22 / 6 = 3.7</i>					
<i>MAPE = 73.7% / 6 = 12.3%</i>					
<i>MSE = 110 / 6 = 18.3</i>					

(b) The following table shows the 3 period moving average model.

<i>Time period (T)</i>	<i>Demand (D)</i>	<i>Forecast by 3 period moving average (F)</i>	<i>Absolute Error (E)</i>
1	28	-	-
2	27	-	-
3	33	-	-
4	25	$(28+27+33)/3 = 29.3$	4.3
5	34	$(27+33+25)/3 = 28.3$	5.7
6	33	30.7	2.3
7	35	30.7	4.3
8	30	34.0	4.0
9	33	32.7	0.3
10	35	32.7	2.3
11	27	32.7	5.7
12	29	31.7	2.7
13	-	30.3	-
$MSE (periods 7 to 12) = (4.3^2 + 4^2 + 0.3^2 + 2.3^2 + 5.7^2 + 2.7^2) / 6 = 13.3$			

(c) The simple exponential smoothing model with smoothing constant $\alpha = 0.1$ is presented below.

<i>Time period (T)</i>	<i>Demand (D)</i>	<i>Forecast by exponential smoothing with $\alpha = 0.1$ (F)</i>	<i>Absolute Error (E)</i>
1	28	30	2
2	27	$30 + 0.1(28-30) = 29.8$	2.8
3	33	$29.8 + 0.1(27-29.8) = 29.5$	3.5
4	25	$29.5 + 0.1(33-29.5) = 29.9$	4.9
5	34	29.4	4.6
6	33	29.9	3.1
7	35	30.2	4.8
8	30	30.7	0.7
9	33	30.6	2.4
10	35	30.8	4.2
11	27	31.2	4.2
12	29	30.8	1.8
13	-	30.6	
$MSE (periods 7 to 12) = (4.8^2 + 0.7^2 + 2.4^2 + 4.2^2 + 4.2^2 + 1.8^2) / 6 = 11.3$			

6.8 Trend Projections:

This time-series forecasting method fits a trend line to a series of historical data points and then projects the line into the future for medium- to long range forecasts. There are several mathematical trend equations that can be developed viz. linear, exponential, quadratic etc. Here we will concentrate only on the linear trends. Of the components of a time series, secular trend represents the long-term direction of the series. One way to describe the trend component is to fit a line visually to a set of points on a graph. Any given graph, however, is subject to slightly different interpretations by different individuals. We can also fit a trend line by the method of least squares. In our discussion, we will concentrate on the method of least squares because visually fitting a line to a time to series is not a completely dependable process.

Reasons for Studying Trends

There are three reasons why it is useful to study secular trends:

1. The study of secular trends allows us to describe a historical pattern.
2. Studying secular trends permits us to project past patterns, or trends, into the future.
3. In many situations, studying the secular trend of a time series allows us to eliminate the trend component from the series.

6.8.1 Linear Regression Analysis

Regression can be defined as a functional relationship between two or more correlated variables. It is used to predict one variable given the other. The relationship is usually developed from observed data. The data should be plotted first to see if they appear linear or if at least parts of the data are linear. Linear regression refers to the special class of regression where the relationship between variables forms a straight line.

The linear regression line is of the form $Y = a + bX$, where Y is the value of the dependent variable that we are solving for, a is the Y intercept, b is the slope, and X is the independent variable. (In time series analysis, X is units of time)

Linear regression is useful for long-term forecasting of major occurrences and aggregate planning. For example, linear regression would be very useful to forecast demands for product families. Even though demand for individual products within a family may vary widely during a time period, demand for the total product family is surprisingly smooth.

The major restriction in using linear regression forecasting is, as the name implies, that past data and future projections are assumed to fall about a straight line. Although this does limit its application, sometimes, if we use a shorter period of time, linear regression analysis can still be used. For example, there may be short segments of the longer period that are approximately linear.

Linear regression is used both for time series forecasting and for casual relationship forecasting. When the dependent variable (usually the vertical axis on the graph) changes as a result of time (plotted on the horizontal axis), it is time series analysis. When the dependent variable changes because of the change in another variable, this is a casual relationship (such as the demand of cold drinks increasing with the temperature).

We illustrate the linear trend projection with a hand fit regression line.

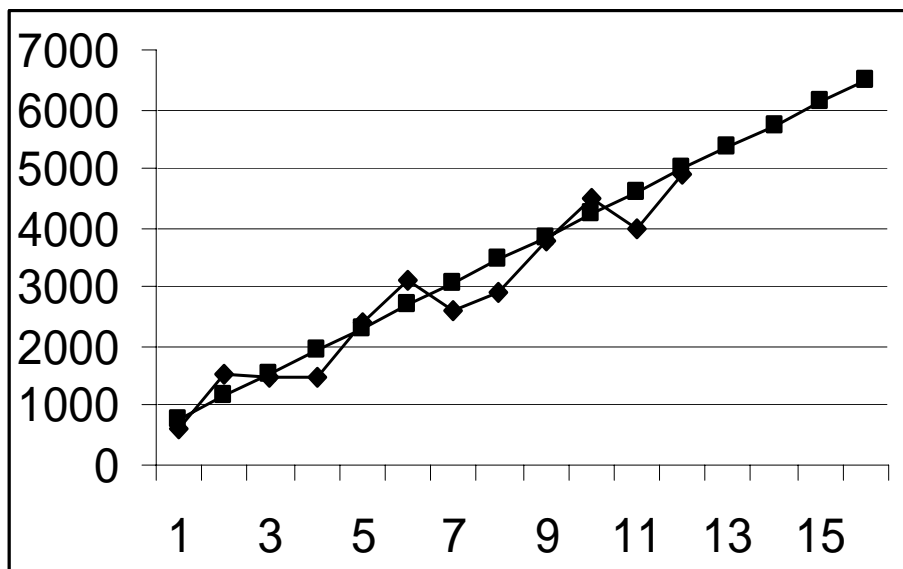
Illustration 6.7 : A firms sales for a product line during the 12 quarters of the past three years were as follows.

Quarter	Sales	Quarter	Sales
1	600	7	2600
2	1550	8	2900
3	1500	9	3800
4	1500	10	4500
5	2400	11	4000
6	3100	12	4900

Forecast the sales for the 13, 14, 15 and 16th quarters using a hand-fit regression equation.

Solution: The procedure is quite simple: Lay a straightedge across the data points until the line seems to fit well, and draw the line. This is the regression line. The next step is to determine the intercept a and slope b .

The following fig shows a plot of the data and the straight line we drew through the points.



The intercept a , where the line cuts the vertical axis, appears to be about 400. The slope b is the "rise" divided by the "run" (the change in the height of some portion of the line divided by the number of units in the horizontal axis). Any two points can be used, but two points some distance apart give the best accuracy because of the errors in reading values from the graph. We use values for the 1st and 12th quarters.

By reading from the points on the line, the Y values for quarter 1 and quarter 12 are about 750 and 4,950.

Therefore.

$$b = (4950 - 750) / (12 - 1) = 382$$

The hand-fit regression equation is therefore

$$Y = 400 + 382X$$

The forecasts for quarters 13 to 16 are

<i>Quarter</i>	<i>Forecast</i>
<i>13</i>	$400 + 382(13) = 5366$
<i>14</i>	$400 + 382(14) = 5748$
<i>15</i>	$400 + 382(15) = 6130$
<i>16</i>	$400 + 382(16) = 6512$

These forecasts are based on the line only and do not identify or adjust for elements such as seasonal or cyclical elements.

6.8.2 Least Squares Method for Linear Regression:

The least squares equation for linear regression is $Y = a + bX$

Where,

- Y = Dependent variable computed by the equation
- y = The actual dependent variable data point (used below)
- a = y intercept , b = Slope of the line , X = Time period

The least squares method tries to fit the line to the data that minimize the sum of the squares of the vertical distance between each data point and its corresponding point on the line.

If a straight line is drawn through general area of the points, the difference between the point and the line is $y - Y$. The sum of the squares of the differences between the plotted data points and the line points is

$$(y_1 - Y_1)^2 + (y_2 - Y_2)^2 + \dots + (y_{12} - Y_{12})^2$$

The best line to use is the one that minimizes this total.

As before, the straight line equation is

$$Y = a + bX$$

Previously we determined a and b from the graph. In the least squares method, the equations for a and b are

$$a = \bar{y} - b\bar{X} \quad \text{and} \quad b = \frac{\sum Xy - n\bar{X}\bar{y}}{\sum X^2 - n\bar{X}^2}$$

Where

a = Y intercept , b = slope of the line , n = number of data points.

We discuss the procedure to fit a straight line by the least squares method with the help of the following illustration and then we will compare the results obtained by hand fitting and the fitting by the method of least squares.

Illustration 6.8: Forecast the sales for the 13, 14, 15 and 16th quarters for the data given in illustration 7 using the least squares method. Also calculate the standard error of the estimate.

Solution: The following table shows the computations carried out for the 12 data points.

X	y	Xy	X ²	y ²	Y
1	600	600	1	360000	801.3
2	1550	3100	4	2402500	1160.9
3	1500	4500	9	2250000	1520.5
4	1500	6000	16	2250000	1880.1
5	2400	12000	25	5760000	2239.7
6	3100	18600	36	9610000	2599.4
7	2600	18200	49	6760000	2959.0
8	2900	23200	64	8410000	3318.6
9	3800	34200	81	14440000	3678.2
10	4500	45000	100	20250000	4037.8
11	4000	44000	121	16000000	4397.4
12	4900	58800	144	24010000	4757.1
Total :78	33350	268200	650	112502500	
$\bar{X} = 6.5$ $\bar{y} = 2779.17$ $b = 359.61$ $a = 441.66$					
$Y = 441.66 + 359.6 X$					
$S_{yx} = 363.9$					

The reader is advised to verify the calculations of a and b on his own.

Note that the final equation for Y shows an intercept of 441.6 and a slope of 359.6. The slope shows that for every unit change in X, Y changes by 359.6.

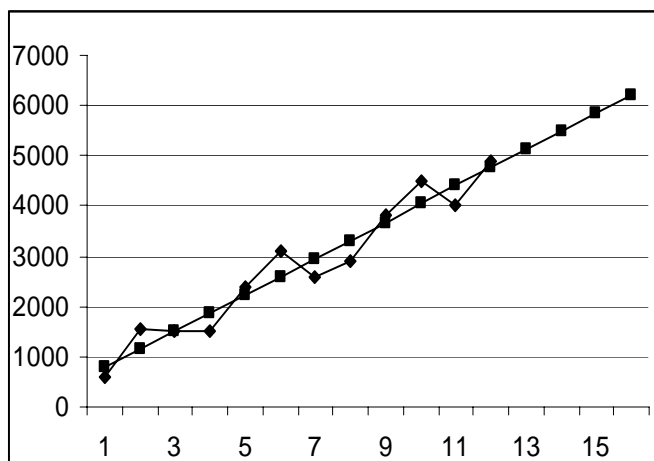
Strictly based on the equation, forecasts for periods 13 through 16 would be

$$Y_{13} = 441.6 + 359.6(13) = 5116.4$$

$$Y_{14} = 441.6 + 359.6(14) = 5476.0$$

$$Y_{15} = 441.6 + 359.6(15) = 5835.6$$

$$Y_{16} = 441.6 + 359.6(16) = 6195.2$$



The reader is also advised to verify the results for the forecasts for the above two illustrations 7 and 8.

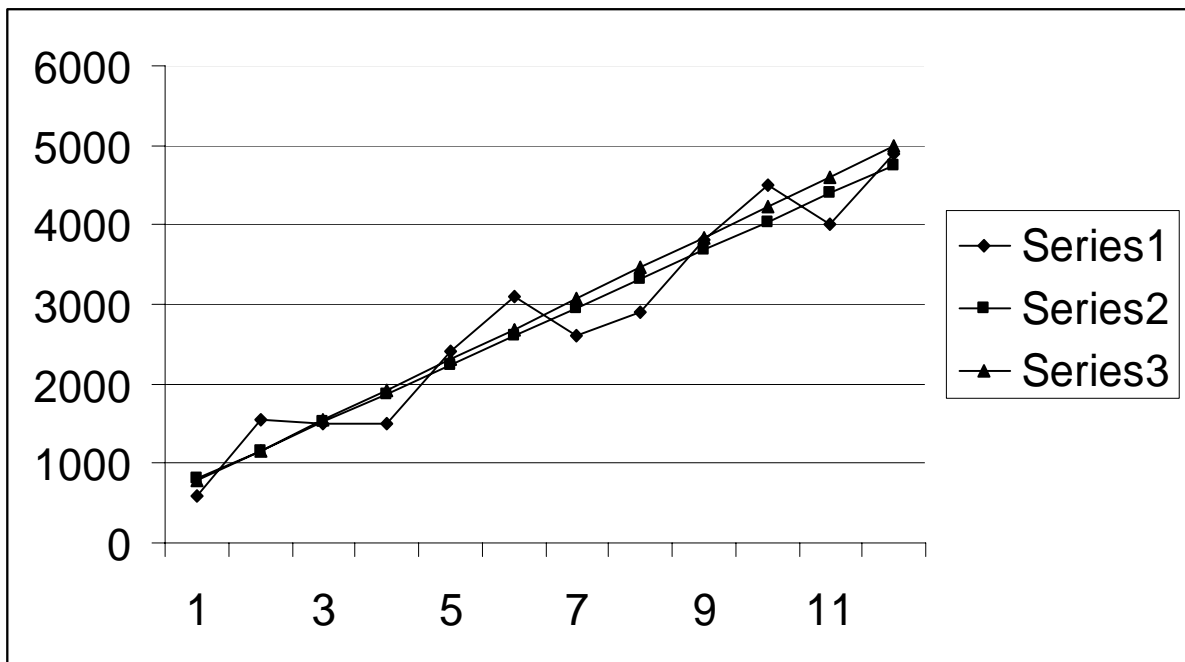
The standard error of the estimate is computed as S_{yX} which is given as follows

$$S_{yX} = \sqrt{\frac{(600 - 801.3)^2 + (1550 - 1160.9)^2 + \dots + (4900 - 4757.1)^2}{12}}$$

$$= 363.9$$

We have plotted the graph of the values of the actual demand, forecasted demand values by the two methods of hand fitting and the method of least squares.

Following is the graph.



Here, Series 1 is the actual observed demand values, Series 2 are the values based on calculations by the method of least squares and Series 3 are the values by hand fitting the trend line.

The following table compares the forecasted values for the 13th, 14th, 15th and the 16th quarters based on illustrations 7 and 8.

<i>quarter</i>	<i>Forecasts by the hand fit line (line going above in the figure) Series 3</i>	<i>Forecasts by the least squares method (line beneath the hand fit line) series 2</i>
13	5366	5116.4
14	5748	5476.0
15	6130	5835.6
16	6512	6195.2

Converting the time for ease in calculations:

Normally, we measure the independent variable time in terms such as weeks, months, and years. Fortunately, we can convert these traditional measures of time to a form that simplifies the computation, we call this process coding. To use coding here, we find the mean time and then subtract that value from each of the sample times. Suppose our time series consists of only three points, 1992, 1993, and 1994. If we had to place these numbers in the above least squares equations we would find the resultant calculations tedious. Instead, we can transform the values 1992, 1993, and 1994 into corresponding values of -1,0, and 1, where 0 represents the mean (1993), -1 represents the first year (1992 -1993= -1), and 1 represents the last year (1994 - 1993 = 1) or alternatively to 1, 2 and 3.

We need to consider two cases when we are coding time values. The first is a time series with an odd number of elements, as in the previous example. The second is a series with an even number of elements. Consider the following table.

In part (A), on the left, we have an odd number of years. Thus, the process is the same as the one we just described, using the years 1992, 1993, and 1994.

In part (B), on the right, we have an even number of elements. In cases like this, when we find the mean and subtract it from each element, the fraction 1/2 becomes part of the answer. To simplify the coding process and to remove the 1/2, we multiply each time element by 2. We will denote the "coded," or translated, time with a lowercase *x*.

<i>Part (A)</i> <i>Odd number of entries</i>			<i>Part (B)</i> <i>Even number of entries</i>			
<i>X</i>	$X - \bar{X}$	Coded time(<i>x</i>)	<i>X</i>	$X - \bar{X}$	$X - \bar{X} \times 2$	Coded time(<i>x</i>)
1989	1989 - 1992	-3	1990	1990 - 1992.5	-2.5 × 2 =	-5
	=			=		
1990	1990 - 1992	-2	1991	1991 - 1992.5	-1.5 × 2 =	-3
	=			=		
1991	1991 - 1992	-1	1992	1992 - 1992.5	-0.5 × 2 =	-1
	=			=		
1992	1992 - 1992	0	1993	1993 - 1992.5	0.5 × 2 =	1
	=			=		
1993	1993 - 1992	1	1994	1994 - 1992.5	1.5 × 2 =	3
	=			=		
1994	1994 - 1992	2	1995	1995 - 1992.5	2.5 × 2 =	5
	=			=		
1995	1995 - 1992	3				
	=					
$\sum X = 13944$			$\sum X = 11955$			
$\bar{X} = 1992$			$\bar{X} = 1992.5$			

We have two reasons for this translation of time. First, it eliminates the need to square numbers as large as 1992, 1993, 1994, and so on. This method also sets the mean year, \bar{x} , equal to zero and allows us to simplify the least squares equations of the Y intercept, a and the slope b as follows.

$$b = \frac{\sum xy}{\sum x^2}$$

$$a = \bar{y}$$

where x represents the coded values.

Illustration 6.9: The following table gives the number of items produced in a factory between 1988 and 1995.

Year	1988	1989	1990	1991	1992	1993	1994	1995
Production	98	105	116	119	135	156	177	208

Determine the equation that will describe the secular trend of production. Also project the production for 1996.

Solution: The following table calculates the necessary values.

		=	$\times 2$		
1988	98	-3.5	-7	-686	49
1989	105	-2.5	-5	-525	25
1990	116	-1.5	-3	-348	9
1991	119	-0.5	-1	-119	1
1992	135	0.5	1	135	1
1993	156	1.5	3	468	9
1994	177	2.5	5	885	25
1995	208	3.5	7	1456	49
$\sum X = 15932$	$\sum Y = 1114$			$\sum xY = 1266$	$\sum x^2 = 168$
$\bar{X} = 1991.5$	$\bar{Y} = 139.25$				

With these values, we can now compute b and a to find the slope and the Y intercept for the line describing the trend in production.

$$b = \frac{\sum xy}{\sum x^2}$$

$$a = \bar{y}$$

$$b = 1266 / 168 = 7.536$$

and $a = 139.25$

Thus, the general linear equation describing the secular trend in production is

$$\begin{aligned} Y &= a + bx \\ &= 139.25 + 7.536 x \end{aligned}$$

where Y = dependent variable computed by the equation OR the estimated production calculated

x = coded time value representing the number of half year intervals (a – sign indicates half year intervals before 1991.5, a + sign indicates half year intervals after 1991.5)

Now suppose we want to estimate the production for the year 1996. First, we must convert 1996 to the value of the coded time (in half year intervals)

$$\begin{aligned} x &= 1996 - 1991.5 \\ &= 4.5 \text{ years} \\ &= 9 \text{ half year intervals} \end{aligned}$$

Substituting this value in the equation for the secular trend, we get

$$\begin{aligned} Y &= 139.25 + 7.536 (9) \\ &= 139.25 + 67.82 \\ &= 207 \end{aligned}$$

Therefore we estimate the production for the year 1996 as 207.

Note: If the number of elements in our time series would have been odd, our procedure would have been the same except that we would have dealt with 1 year intervals.

6.9 Decomposition of the time series

We have seen earlier that observations taken over time (i.e. time series) often contain the four following characteristics :

- (a) A long-term trend (denoted by T)
- (b) Seasonal variations (denoted by S)
- (c) Cyclical variations (denoted by C)
- (d) Random or residual variations (denoted by R)

The methods covered so far do not make any attempt to isolate the individual factors, namely, seasonally, trend, cyclical and random variations, in the time series. But there are many situations where such a breaking down of the time series is possible and necessary. The decomposition methods basically operate on the principle that a time series is composed of the four factors stated earlier.

The decomposition methods assume the time series value at time t to be a function of the different components,

i.e.,
$$D_t = f(T_t, S_t, C_t, R_t)$$

where

T_t = trend value at period t

S_t = seasonal component at period t

C_t = cyclical component at period t , and

R_t = random variation at period t .

To make reasonably accurate forecasts, it is often necessary to separate out the above characteristics (i.e. T, S, C and R) from the raw data. This is known as time series decomposition or often just time series analysis. The separated elements are then combined to produce a forecast.

The functional form for the series used may either be additive or multiplicative. The multiplicative form (most commonly used) is written as follows:

$$D_t = T_t \times S_t \times C_t \times R_t$$

Here the components are expressed as percentages or proportions

The additive model takes the form

$$D_t = T_t + S_t + C_t + R_t$$

Here the components are expressed as absolute values

The multiplicative model is commonly used in practice and is more appropriate if the characteristics interact, e.g. where a higher trend value increases the seasonal variation. The additive model is more suitable if the component factors are independent, e.g. where the amount of seasonal variation is not affected by the value of the trend.

Of the four elements the most important are the first two; the trend and seasonal variation, so this book concentrates on these two.

Seasonal Factor (Or Index)

A seasonal factor is the amount of correction needed in a time series to adjust for the season of the year. Basically, the seasonal factor (or index) is the ratio of the amount sold during each season divided by the average for all seasons.

The following examples show how seasonal indices are determined and used to forecast.

Illustration 6.10: Assume that in the past years, a firm sold an average of 1000 units of air conditioners each year. On the average, 200 units were sold from the period January to March, 350 in April to June, 300 in July to September and 150 in October to December. Calculate the seasonal indices for each season (quarter). If the expected demand for the next year is believed to be 1100, forecast the demand in each quarter.

Solution: First we find the average sales for each season and then divide the sales of each season by that average.

	Past sales	Average sales for each season = Total / 4 = 1000 / 4	Seasonal factor
Jan-Mar	200	250	200 / 250 = 0.8
Apr-June	350	250	350 / 250 = 1.4
July-Sept	300	250	300 / 250 = 1.2
Oct-Dec	150	250	150 / 250 = 0.6
Total	1000		

Using these factors, if we expected demand for next year to be 1100 units, we would forecast the demand to occur as shown in the following table. First we distribute the demand equally among all the quarters and then divide those amounts by the seasonal index obtained in the previous step.

	<i>Expected demand For next year</i>	<i>Average sales for each season = 1100 / 4</i>	<i>Seasonal Factor</i>	<i>Next years seasonal forecast</i>
<i>Jan-Mar</i>		275	× 0.8	= 220
<i>Apr-June</i>		275	× 1.4	= 385
<i>July-Sept</i>		275	× 1.2	= 330
<i>Oct-Dec</i>		275	× 0.6	= 165
<i>Total</i>	1100			

The seasonal factors are updated periodically as the new data are available.

This is how we can forecast based on the past seasonal data and knowing the future expected demand.

One issue in the above illustration was that the expected demand for the next year was known to be 1100. When not known, we can compute the seasonal indices by even using a hand-fir straight line.

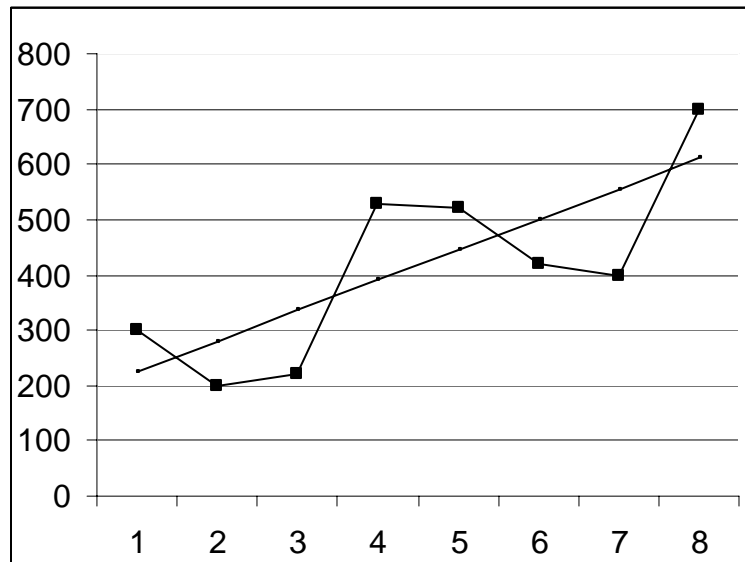
This procedure is given in the following illustration.

Illustration 6.11: Simply hand fit a straight line through the data points and measure the trend and the intercept from the graph. The history of the data is

<i>Year</i>	<i>Quarter</i>	<i>Sales</i>	<i>Year</i>	<i>Quarter</i>	<i>Sales</i>
2004	1	300	2005	1	520
2004	2	200	2005	2	420
2004	3	220	2005	3	400
2004	4	530	2005	4	700

Find the seasonal factors for the quarters and using them forecast the sales in the quarters of 2006.

Solution: First we plot the data points on the graph and then fit a straight line. (The reader can fit a straight line in a manner different from this. Naturally then his estimates are likely to vary a bit. You can also fit a straight line by the method of least squares).



The equation for the trend line is $T_t = 170 + 55t$, which is derived from the intercept 170 and a slope of $(610 - 170) / 8 = 55$.

Next we derive a seasonal index by comparing the actual data with the trend line as shown below.

<i>Quarter</i>	<i>Actual</i>	<i>Calculated from the trend</i> $170 + 55t$	<i>Ration of</i> <i>Actual / Trend</i>	<i>Seasonal factor</i> <i>(Average of the same quarters</i> <i>in both years)</i>
<i>2004</i>				
<i>1</i>	300	$= 170 + 55(1) = 225$	1.33	Q-1 : $(1.33 + 1.17)/2 = 1.25$
<i>2</i>	200	$= 170 + 55(2) = 280$	0.71	
<i>3</i>	220	$= 170 + 55(3) = 335$	0.66	
<i>4</i>	530	$= 170 + 55(4) = 390$	1.36	
<i>2005</i>				Q-3 : $(0.66 + 0.72)/2 = 0.69$
<i>1</i>	520	$= 170 + 55(5) = 445$	1.17	Q-4 : $(1.36 + 1.15)/2 = 1.25$
<i>2</i>	420	$= 170 + 55(6) = 500$	0.84	
<i>3</i>	400	$= 170 + 55(7) = 555$	0.72	
<i>4</i>	700	$= 170 + 55(8) = 610$	1.15	

We can now compute the forecast for the quarters in 2006 including trend and the seasonal factors using the following equation:

$$\text{Forecast} = \text{Trend value} \times \text{seasonal factor}$$

<i>Year 2006</i>	<i>trend value using the trend equation</i> $170 + 55t$ T_t	<i>Seasonal index for the quarter</i> S_t	<i>Forecast including the trend and seasonal factors</i> $T_t \times S_t$
<i>Q1</i>	$= 170 + 55(9) = 665$	1.25	831
<i>Q2</i>	$= 170 + 55(10) = 720$	0.78	562
<i>Q3</i>	$= 170 + 55(11) = 775$	0.69	535
<i>Q4</i>	$= 170 + 55(12) = 830$	1.25	1038

With this knowledge of seasonal factors we now go on and study the decomposition of time series and forecast into the future.

The following illustration shows how the trend (T) and seasonal variation (S) are separated out from a time series and how the seasonal indices are calculated.

We can calculate the seasonal indices for the time series data by two methods

Method 1. By using a trend line from the data by the least squares method and

Method 2. By using the method of ratio to moving average.

We illustrate the first method in the following illustration.

Illustration 6.12: Estimate the sales of air conditioners for the quarters 17, 18, 19, 20 i.e the four quarters of the year 2005 using the following data by the method of decomposition of time series.

<i>Sale of air conditioners in '000s</i>				
<i>Year</i>	Quarter 1	Quarter 2	Quarter 3	Quarter 4
<i>2001</i>	20	32	62	29
<i>2002</i>	21	42	75	31
<i>2003</i>	23	39	77	48
<i>2004</i>	27	39	92	53

It is apparent that there is a strong seasonal element in the above data (low in Quarter 1 and high in Quarter 3) and that there is a generally upward trend.

The steps in analyzing the data and preparing a forecast are:

- Step 1:** Calculate the trend in the data using the least squares method.
- Step 2:** Estimate the sales for each quarter using the regression formula established in Step 1.
- Step 3:** Calculate the percentage variation of each quarter's actual sales from the estimates, obtained in Step 2.
- Step 4:** Average the percentage variations from Step 3 for each quarter. This establishes the average seasonal variations.
- Step 5:** Prepare forecast based on trend \times percentage seasonal variations.

Step 1. We use the following least squares linear regression equations to compute the regression line $Y = a + bX$

$$\sum y = an + b\sum x$$

$$\sum xy = a\sum x + b\sum x^2$$

	<i>X (quarters)</i>	<i>Y (sales)</i>	<i>XY</i>	<i>X²</i>
<i>Year 2001</i>	1	20	20	1
	2	32	64	4
	3	62	186	9
	4	29	116	16
<i>Year 2002</i>	5	21	105	25
	6	42	252	36
	7	75	525	49
	8	31	248	64
<i>Year 2003</i>	9	23	207	81
	10	39	390	100
	11	77	847	121
	12	48	576	144
<i>Year 2004</i>	13	27	351	169
	14	39	546	196
	15	92	1380	225
	16	53	848	256
	$\sum X = 136$	$\sum Y = 710$	$\sum XY = 6661$	$\sum X^2 = 1496$

We get the following two equations

$$710 = 16a + 136b$$

$$6661 = 136a + 1496b$$

Solving them we get $a = 28.74$ and $b = 1.84$.

Thus, the trend line equation is $Y = 28.74 + 1.84X$

Step 2 and Step 3: Use the trend line to calculate the estimated sales for each quarter.

Express the actual value of sales as a percentage of this estimated sales. (Remember that this is similar to finding the ratio of actual to trend in the above two illustrations)

	<i>X (quarters)</i>	<i>Y (sales)</i>	<i>Estimated sales Using trend line</i>	<i>(actual/Estimated)% = (A/B)×100</i>
		<i>(A)</i>	<i>(B)</i>	
<i>Year 2001</i>	1	20	30.58	65
	2	32	32.42	99
	3	62	34.26	181
	4	29	36.10	80
<i>Year 2002</i>	5	21	37.94	55
	6	42	39.78	106
	7	75	41.62	180
	8	31	43.46	71
<i>Year 2003</i>	9	23	45.30	51
	10	39	47.14	83
	11	77	48.98	157
	12	48	50.82	94
<i>Year 2004</i>	13	27	52.66	51
	14	39	54.50	72
	15	92	56.34	163
	16	53	58.18	91

Step 4: Average the percentage variations to find the average seasonal variation.

<i>In %</i>	<i>Q1</i>	<i>Q2</i>	<i>Q3</i>	<i>Q4</i>
<i>2001</i>	65	99	181	80
<i>2002</i>	55	106	180	71
<i>2003</i>	51	83	157	94
<i>2004</i>	51	72	163	91
<i>Total</i>	222	360	681	336
<i>÷ 4 =</i>	56%	90%	170%	84%

These then are the average variations expected from the trend for each of the quarters; for example, on average the first quarter of each year will be 56% of the value of the trend. Because the variations have been averaged, the amounts *over* 100% (Q3 in this example) should equal the amounts below 100%. (Q1, Q2 and Q4 in this example); This can be checked by adding the average variations and verifying that they total 400% thus:

$$56\% + 90\% + 170\% + 84\% = 400\%.$$

On occasions, roundings in the calculations will make slight adjustments necessary to the average variations. We will discuss this in the next illustration.

Step 5. Prepare final forecasts based on the trend line estimates and the averaged seasonal variations.

The seasonally adjusted forecast is calculated thus:

$$\text{Seasonally adjusted forecast} = \text{Trend estimate} \times \text{Seasonal factor}$$

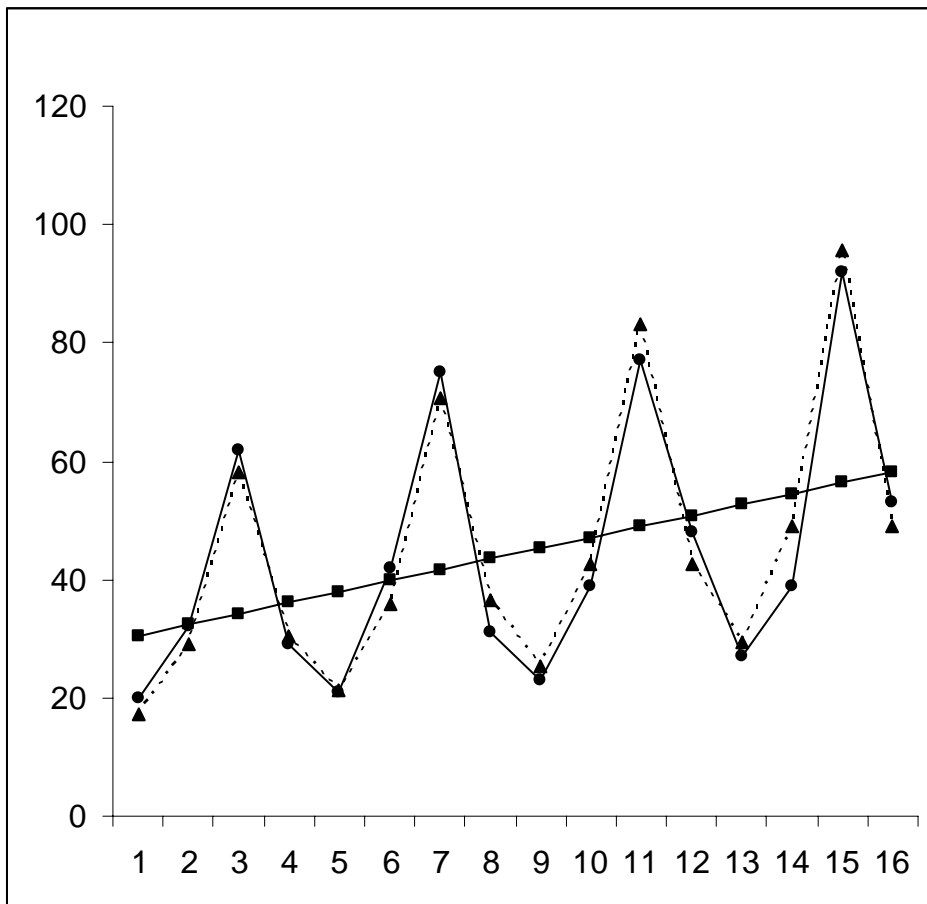
	<i>X (quarters)</i>	<i>Y (sales)</i>	<i>Estimated sales Using trend line</i>	<i>Seasonal factor</i>	<i>Seasonally adjusted forecast</i>
<i>Year 2001</i>	1	20	30.58	0.56	17.12
	2	32	32.42	0.90	29.18
	3	62	34.26	1.7	58.24
	4	29	36.10	0.84	30.32
<i>Year 2002</i>	5	21	37.94	0.56	21.24
	6	42	39.78	0.90	35.80
	7	75	41.62	1.7	70.75
	8	31	43.46	0.84	36.51
<i>Year 2003</i>	9	23	45.30	0.56	25.37
	10	39	47.14	0.90	42.43
	11	77	48.98	1.7	83.27
	12	48	50.82	0.84	42.69
<i>Year 2004</i>	13	27	52.66	0.56	29.49
	14	39	54.50	0.90	49.05
	15	92	56.34	1.7	95.75
	16	53	58.18	0.84	48.87

Extrapolation using the trend and seasonal factors

Once the formulae above have been calculated, they can be used to forecast (extrapolate) future sales.

Here it is required to estimate the sales for the year 2005 (i.e. Quarters 17, 18, 19 and 20 in our series) this is done as follows:

<i>Year 2005</i>	<i>Number of the quarter in the time series (X)</i>	<i>Estimated trend Using the trend equation $Y=28.74 + 1.84X$ (T)</i>	<i>Seasonal factor (S)</i>	<i>Forecasted sales $T \times S$</i>
<i>Q1</i>	17	$= 28.74 + 1.84(17) = 60.02$	0.56	33.61
<i>Q2</i>	18	$= 28.74 + 1.84(18) = 61.86$	0.90	55.67
<i>Q3</i>	19	$= 28.74 + 1.84(19) = 63.7$	1.7	108.29
<i>Q4</i>	20	$= 28.74 + 1.84(20) = 65.54$	0.84	55.05



The above figure shows the actual sales (●), the trend line (■) and the seasonally adjusted forecast (▲).

Notes :

- (a) Time series decomposition is not an adaptive forecasting system like moving averages and exponential smoothing.
- (b) Forecasts produced by such an analysis should always be treated with caution. Changing conditions and changing seasonal factors make long term forecasting a difficult task.
- (c) The above illustration has been an example of a multiplicative model. This is because the seasonal variations were expressed in percentage or proportionate terms. Similar steps would have been necessary if the additive model had been used except that the variations from the trend (i.e when we compute Actual / estimated in case of the above illustration) would have been the absolute values (i.e we do not compute the ratio now but we compute the absolute variation). For example, the first two variations would have been

$$Q1 : 20 - 30.58 = \text{absolute variation} = - 10.58.$$

$$Q2 : 32 - 32.42 = \text{absolute variation} = - 0.42 \text{ and so on.}$$

The absolute variations would have been averaged in the normal way to find the average absolute variation, (i.e. finding the seasonal factor) whether + or -, and these values would have been used to make the final seasonally adjusted forecasts.

Now we see another method to find the seasonal indices and forecast the future demand. This method is called the *ratio-to-moving average method*. This method provides an index that describes the degree of seasonal variation. The index is based on a mean of 100 with the degree of seasonality measured by variations away from the base. This process is the decomposition of a time series into a deseasonalized series.

Consider the following illustration.

Illustration 6.13 : A manager wanted to establish the seasonal pattern of the units of a particular product X demanded by his client. The following table contains the quarterly sales, that is, the average number of units sold during each quarter of the last 5 years.

<i>Sale of product X</i>				
<i>Year</i>	Quarter 1	Quarter 2	Quarter 3	Quarter 4
<i>2001</i>	1861	2203	2415	1908
<i>2002</i>	1921	2343	2514	1986
<i>2003</i>	1834	2154	2098	1799
<i>2004</i>	1837	2025	2304	1965
<i>2005</i>	2073	2414	2339	1967

Calculate the seasonal indices and deseasonalize the time series.

Solution: Refer to the following table to demonstrate the five steps required to compute a seasonal index.

Step 1. The first step in computing a seasonal index is to calculate the 4-quarter moving total for the time series. To do this, we total the values for the quarters during the first year, 2001 as $1,861 + 2,203 + 2,415 + 1,908 = 8,387$. A moving total is associated with the middle data point in the set of values from which it was calculated. So this moving total is of the middle part i.e quarter 2 and quarter 3 of year 2001.

We then take the mean of this total i.e we compute $8387 / 4 = 2096.75$. This value is the 4 quarter moving average and we place it in the next column.

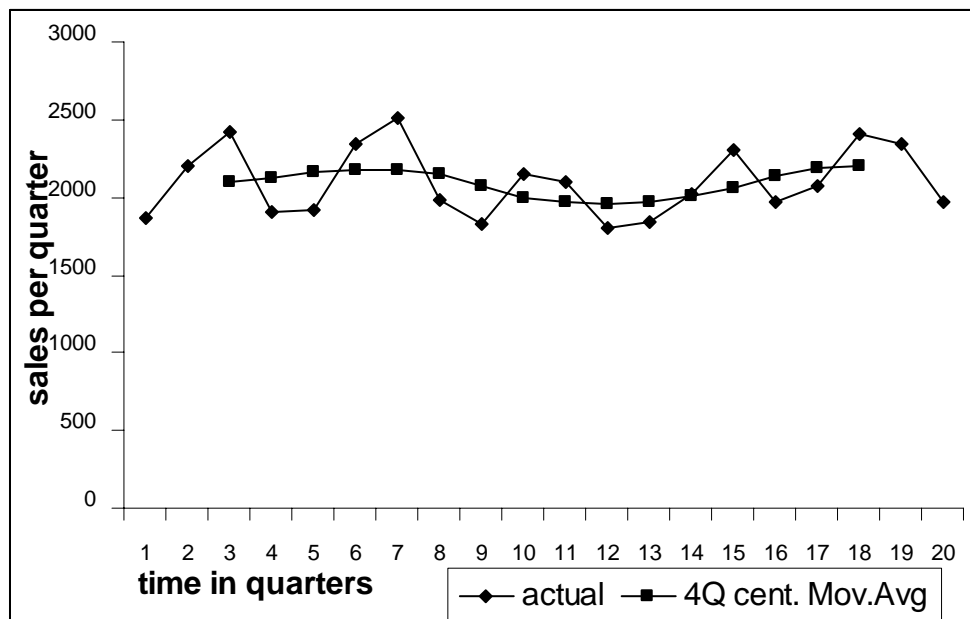
<i>Year</i>	<i>Quarter</i>	<i>Sales</i>	<i>4 quarter moving average</i>	<i>4 quarter centered moving average</i>	<i>Percentage of actual to moving average values $((3)/(5)) \times 100$</i>
<i>(1)</i>	<i>(2)</i>	<i>(3)</i>	<i>(4)</i>	<i>(5)</i>	<i>(6)</i>
2001	1	1861			
	2	2203			
	3	2415	2096.75	2104.250	114.8
	4	1908	2111.75	2129.250	89.6
2002	1	1921	2146.75	2159.125	89
	2	2343	2171.50	2181.250	107.4
	3	2514	2191.0	2180.125	115.3
	4	1986	2169.25	2145.625	92.6
2003	1	1834	2122.0	2070.00	88.6
	2	2154	2018.0	1994.625	108
	3	2098	1971.25	1971.625	106.4
	4	1799	1972.0	1955.875	92
2004	1	1837	1939.75	1965.500	93.5
	2	2025	1991.25	2012.00	100.6
	3	2304	2032.75	2062.250	111.7
	4	1965	2091.75	2140.375	91.8
2005	1	2073	2189.0	2193.375	94.5
	2	2414	2197.75	2198.0	109.8
	3	2339	2198.25		
	4	1967			

We now find the next moving total by dropping the 2001 quarter 1 value, 1,861, and adding the new 2002 quarter 1 value, 1,921. By dropping the first value and adding the fifth, we keep four quarters in the total. The four values added now are 2,203 + 2,415 + 1,908 + 1,921 = 8,447. We again find the average of this total giving us the 4 quarter moving average. We again place it in below the 4 quarter moving average we found out earlier. This process is continued over the time series until we have included the last value in the series.

Step 2.

In the second step, we center the 4-quarter moving average. The moving averages in column 4 all fall halfway between the quarters. We would like to have moving averages associated with each quarter. In order to center our moving averages, we associate with each quarter the average of the two 4-quarter moving averages falling just above and just below it. For the 2001 quarter 3, the resulting 4-quarter centered moving average is $(2,096.75 + 2,111.75)/2 = 2,104.25$. The other entries in column 5 are calculated the same way.

The above process when plotted on the graph illustrates how the moving average has smoothed the peaks and troughs of the original time series. The seasonal and irregular components have been smoothed. The smoothed line represents the cyclical and trend components of the series.



(Note: Whenever the number of periods for which we want indices is odd (7 days in a week or 3 four monthly data, etc.), would compute 7-day moving totals and moving averages, and the moving averages would already be centered (because the middle of a 7-day period is the fourth of those 7 days). However, when the number of periods is even (4 quarters, 12 months, 24 hours), then we must centre the moving averages)

Step 3.

Next, we calculate the percentage of the actual value to the moving-average value for each quarter in the time series having a 4-quarter moving-average entry. This step allows us to recover the seasonal component for the quarters. We determine this percentage by dividing each of the actual quarter values in column 3 of the Table by the corresponding 4-quarter centered moving-average values in column 5 and then multiplying the result by 100.

Step 4.

Collect all the percentage of actual to moving-average values in column 6 and arrange them by quarter. Then calculate the modified mean for each quarter. The modified mean is calculated by discarding the highest and lowest values for each quarter and averaging the remaining values. (You may not ignore some of the values but just take the average of all the means for a particular quarter) In the following table the process of finding the modified mean is shown.

<i>Year</i>	<i>Quarter 1</i>	<i>Quarter 2</i>	<i>Quarter 3</i>	<i>Quarter 4</i>
2001	-	-	114.8	89.6 ignore
2002	89.0	107.4	115.3 ignore	92.6 ignore
2003	88.6 ignore	108.0	106.4 ignore	92.0
2004	93.5	100.6 ignore	111.7	91.8
2005	94.5 ignore	109.8 ignore	-	-
Total of non ignored values	182.5	215.4	226.5	183.8
Modified Mean				
	Quarter 1	$182.5/2 = 91.25$		
	Quarter 2	$215.4/2 = 107.70$		
	Quarter 3	$226.5/2 = 113.25$		
	Quarter 4	$183.8/2 = 91.90$		
	Total	404.1		

The seasonal values we recovered for the quarters in column 6 of the table still contain the cyclical and irregular components of variation in the time series. By eliminating the highest and lowest values from each quarter, we reduce the extreme cyclical and irregular variations. When we average the remaining values, we further smooth the cyclical and irregular components. Cyclical and irregular variations tend to be removed by this process, so the modified mean is an index of the seasonality component.

Step 5.

The final step, demonstrated in the following table adjusts the modified mean slightly. Notice that the four indices in Table total 404.1. However, the base for an index is 100. Thus, the four quarterly indices should total 400, and their mean should be 100. To correct for this error, we multiply each of the quarterly indices in previous Table by an adjusting constant. This number is found by dividing the desired sum of the indices (400) by the actual sum (404.1). In this case, the result is 0.9899. The following table shows that multiplying the indices by the adjusting constant brings the quarterly indices to a total of 400. (Sometimes even after this adjustment, the mean of the seasonal indices is not exactly 100 because of accumulated rounding errors. In this case, however, it is exactly 100.)

Quarter	Unadjusted indices × Adjusting constant	= seasonal index
1	91.25 × 0.9899	90.3
2	107.70 × 0.9899	106.6
3	113.25 × 0.9899	112.1
4	91.90 × 0.9899	91.0
Total of the indices		400.0
Mean of the seasonal indices = 400/4 = 100		

(NOTE: In the previous illustration the sum of all the seasonal indices added to 400. Here it is not the case.)

The ratio-to-moving-average method just explained allows us to identify seasonal variation in a time series. The seasonal indices are used to remove the effects of seasonality from a time series. This is called deseasonalizing a time series. Before we can identify either the trend or cyclical components of a time series, we must eliminate seasonal variation.

(NOTE: The reader is reminded that in the earlier method we initially found a trend line and based on those trend values calculated the ratio of actual sales to that of estimated sales and hence found the seasonal index by taking the average of those values. After that we calculated the estimated future sales using the trend equation and adjusted the values obtained with the seasonal index. In this method, we now prepare the deseasonalized trend line and then forecast using the seasonal indices)

To deseasonalize a time series, we divide each of the actual values in the series by the appropriate seasonal index (expressed as a fraction of 100). To demonstrate, we shall deseasonalize the value of the first four quarters of the year 2001. In the following Table, we see the deseasonalizing process using the values for the seasonal indices from the previous table. Once the seasonal effect has been eliminated, the deseasonalized values that remain reflect only the trend, cyclical, and irregular components of the time series.

Year (1)	Quarter (2)	Actual Sales (3)	Seasonal index / 100 (4)	Deseasonalized sales = (3) / (4)
2001	1	1861	0.903	2061
	2	2203	1.066	2067
	3	2415	1.121	2154
	4	1908	0.910	2097

Once we have removed the seasonal variation, we can compute a deseasonalized trend line, which we can then project into the future. Suppose the manager in our example estimates from a deseasonalized trend line that the deseasonalized average sales for the fourth quarter of the next year 2006 will be 2,121. When this prediction has been obtained, management must then take the seasonality into account. To do this, it multiplies the

deseasonalized predicted average sales of 2,121 by the fourth-quarter seasonal index (expressed as a fraction of 100) to obtain a seasonalized estimate of $2121 \times 0.910 = 1,930$ units for the fourth-quarter average sales.

The reader is given as an exercise to deseasonalize the whole series and then calculate the deseasonalized trend line. Based on that, forecast the sales for the four quarters of the year 2006 and then use the seasonal indices for the quarters we have obtained here to forecast the seasonalized sales.

To sum up the decomposition of the time series:

The steps followed in applying the decomposition method are as follows:

- (1) First a moving average of length n is computed for the time series. The value of n is taken as equal to the length of seasonality. For instance, if we have quarterly data, $n = 4$, and for monthly data $n = 12$. This step reduces random variations and eliminates seasonality. Then we centre this moving average and obtain the centered moving average.
- (2) Next, the actual value of the time series for each period is compared with its centered moving average (obtained in step 1). The resulting ratio is the seasonality factor estimate:

$$S_t = \frac{D_t}{\text{moving average at } t}$$

- (3) For more than one year's data, the seasonality factor is averaged for each of the periods of the seasonal cycle. The factors should total up to the number of periods n per cycle. If they do not, then they should be adjusted so that they add up to n .
- (4) Next, the time series should be deseasonalized. The formula used here is:

$$T_t = \frac{D_t}{S_t}$$

- (5) Next, the trend is estimated for the deseasonalized data using regression analysis or by moving averages with trend adjustments. In case a linear trend results, then the resulting values are a (intercept) and b (slope).
- (6) The forecast for period $t + m$ can now be prepared using the relation

$$F_{t+m} = [a + b(t+m)]S_{t+m}$$

6.10 Selecting A Suitable Forecasting Method

In this section we discuss the factors that are important when they select a forecasting method.

User and System Sophistication

How sophisticated are the managers who are expected to use the forecasting results? It has been found that the forecasting method must be matched to the knowledge and sophistication of the user. Generally speaking, managers are reluctant to use results from techniques they do not understand. Another factor is the status of forecasting systems currently in use. The forecasting systems tend to evolve toward more mathematically sophisticated methods, they do not change in one grand step. So the method chosen must not be too advanced or sophisticated for its users or too far advanced beyond the current forecasting system.

Time and Resource Available

The selection of a forecasting method will depend on the time available in which to collect the data and prepare the forecast. This may involve the time of users, forecasters and data collectors. The time required is also closely related to the necessary resources and the costs of the forecasting method. The preparation of a complicated forecast for which most of the data must be collected may take several months and cost thousands of money value. For routine forecasts made by computerized systems, both the cost and the amount of time required may be very modest.

Use or Decision characteristics

The forecasting method must be related to the use or decisions required. The use is closely related to such characteristics as accuracy required, time horizon of the forecast, and number of items to be forecast. For example, inventory and scheduling decisions require highly accurate short range forecasts for a large number of items. Time-series methods are ideally suited to these requirements. On the other hand, decisions involving process and facility planning are long range in nature, they require less accuracy for, perhaps, a single estimate of total demand. Qualitative or casual methods tend to be more appropriate for those decisions. In this middle time range are aggregate planning and budgeting decisions which often utilize time-series or casual methods.

Data Availability

The choice of forecasting method is often constrained by available data. An econometric model might require data which are simply not available in the short run; therefore another method must be selected. The Box-Jenkins time-series method requires about 60 data points (5 years of monthly data).

Data Pattern

The pattern in the data will affect the type of forecasting method selected. If the time-series is flat, a first order method can be used. However, if the data show trends or seasonal patterns, more advanced methods will be needed. The pattern in the data will also determine whether a time-series method will suffice or whether casuals model are needed. If the data pattern is unstable over time, a qualitative method may be selected. Thus the data pattern is one of the most important factors affecting the selection of a forecasting method.

Another issue concerning the selection of forecasting methods is the difference between fit and prediction. When different models are tested, it is often thought that the model with the best fit to historical data (least error) is also the best predictive model. This is not true. For example, suppose demand observations are obtained over the last eight time periods and we want to fit the best time-series model to these data. A polynomial model of degree seven can be made to fit exactly through each of the past eight data points. But this model is not necessarily the best predictor of the future.

6.11 More on Forecast Errors

In using the word error, we are referring to the difference between the forecast value and what actually occurred. In statistics, these errors are called residuals. As long as the forecast value is within the confidence limits, as we discuss later in "Measurement of Error", this is not really an error. But common usage refers to the difference as an error.

Demand for a product is generated through the interaction of a number of factors too complex to describe accurately in a model. Therefore, all forecasts certainly contain some error. In discussing forecast errors, it is convenient to distinguish between sources of error and the measurement of error.

Sources Of Error

Errors can come from a variety of sources. One common source that many forecasters are unaware of is projecting past trends into the future. For example, when we talk about statistical errors in regression analysis, we are referring to the deviations of observations from our regression line. It is common to attach a confidence band (that is, statistical control limits) to the regression line to reduce the unexplained error. But when we then use this regression line as a forecasting device by projecting it into the future, the error may not be correctly defined by the projected confidence band. This is because the confidence interval is based on past data; it may not hold for projected data points and therefore cannot be used with the same confidence. In fact, experience has shown that the actual errors tend to be greater than those predicted from forecast models.

Errors can be classified as bias or random. *Bias errors* occur when a consistent mistake is made. Sources of bias include failing to include the right variables; using the wrong relationships among variables; employing the wrong trend line, mistakenly shifting the seasonal demand from where it normally occurs; and the existence of some undetected secular trend. *Random errors* can be defined as those that cannot be explained by the forecast model being used.

Measurement Of Error

Several common terms used to describe the degree of error are *standard error*, *mean squared error* (or *variance*), and *mean absolute deviation*. This error estimate helps in monitoring erratic demand observations. In addition, they also help to determine when the forecasting method is no longer tracking actual demand and it need to be reset. For this tracking signals are used to indicate any positive or negative bias in the forecast.

Standard error is discussed earlier. It being a square root of a function, it is often more convenient to use the function itself. This is called the mean square error (MSE) or variance.

The mean absolute deviation (MAD) is also important because of its simplicity and usefulness in obtaining tracking signals. MAD is the average error in the forecasts, using absolute values. It is valuable because MAD, like the standard deviation, measures the dispersion of some observed value from some expected value. The only difference is that like standard deviation, the errors are not squared.

When the errors that occur in the forecast are normally distributed (the usual case), the mean absolute deviation relates to the standard deviation as

$$1 \text{ standard deviation} = \sqrt{\frac{\pi}{2}} \times \text{MAD, or approximately } 1.25 \text{ MAD}$$

Conversely, $1 \text{ MAD} = 0.8 \text{ standard deviation}$

The standard deviation is the larger measure. If the MAD of a set of points was found to be 60 units, then the standard deviation would be 75 units. In the usual statistical manner, if control limits were set at plus or minus 3 standard deviations (or ± 3.75 MADs), then 99.7 percent of the points would fall within these limits.

A **tracking signal** is a measurement that indicates whether the forecast average is keeping pace with any genuine upward or downward changes in demand. As used in forecasting, the tracking signal is the number of mean absolute deviations that the forecast value is above or below the actual occurrence. The following figure shows a normal distribution with a mean of 0 and a MAD equal to 1. Thus, if we compute the tracking signal and find it equal to minus 2, we can see that the forecast model is providing forecasts that are quite a bit above the mean of the actual occurrences.

A tracking signal (TS) can be calculated using the arithmetic sum of forecast deviations divided by the mean absolute deviation:

$$\text{Tracking Signal } TS = RSFE / MAD$$

where

RFSE is the running sum of forecast errors, considering the nature of the error. (For example, negative errors cancel positive error and vice versa.)

MAD is the average of all the forecast errors (disregarding whether the deviations are positive or negative). It is the average of the absolute deviations.

For forecast to be “in control”, 89% of the errors are expected to fall within ± 2 MADs, 98% within 3 MADs, or 99% within 4 MADs.

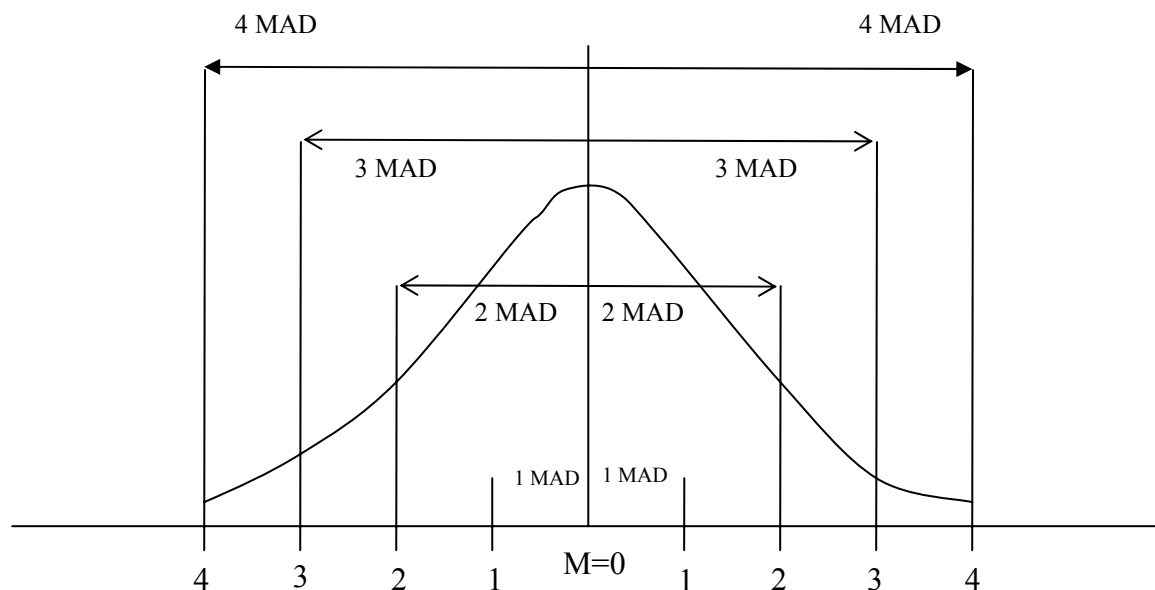


Fig. 6.

The following illustration shows the procedure for computing MAD and the tracking signal.

Illustration 6.14: Compute the Mean Absolute deviation (MAD), the running sum of forecast errors (RSFE) and hence obtain the Tracking Signal for a six month period for the following set of data where the period number, demand forecast (set at constant 1000) and the actual demand occurrences are given below.

Month	Demand forecast	Actual demand
1	1000	950
2	1000	1070
3	1000	1100
4	1000	960
5	1000	1090
6	1000	1050

Solution:

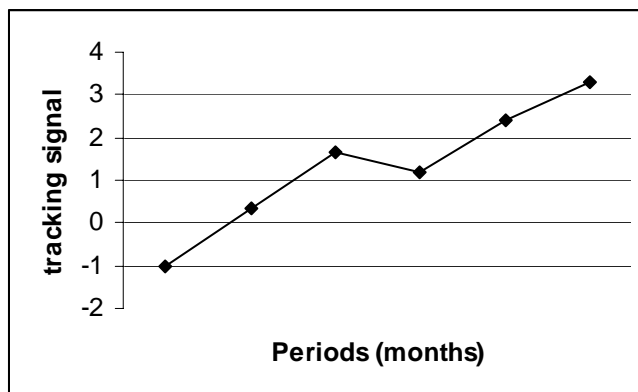
Month	Demand Forecast	Actual	Deviation	RSFE	Abs. Dev.	Sum of Abs. Dev.	MAD*	TS = RSFE / MAD
1	1,000	950	-50	-50	50	50	50	-1
2	1,000	1,070	+70	+20	70	120	60	0.33
3	1,000	1,100	+100	+120	100	220	73.3	1.64
4	1,000	960	-40	+80	40	260	65	1.2
5	1,000	1,090	+90	+170	90	350	70	2.4
6	1,000	1,050	+50	+220	50	400	66.7	3.3

For month 6, $MAD = 400 / 6 = 66.7$

For month 6, $TS = RSFE/MAD = 220 / 66.7 = 3.3$ MADs

Here, we say that the forecast, on the average, is off by 66.7 units and the tracking signal is equal to 3.3 mean absolute deviations.

We can get a better feel for what the MAD and tracking signal mean by plotting the points on a graph.



Note that it drifted from minus 1 MAD to plus 3.3 MADs. This happened because actual demand was greater than the forecast in four of the six periods. If the actual demand does not fall below the forecast to offset the

continual positive RSFE, the tracking signal would continue to rise and we would conclude that assuming a demand of 1,000 is a bad forecast.

Acceptable limits for the tracking signal depend on the size of the demand being forecast (high-volume or high-revenue items should be monitored frequently) and the amount of personnel time available (narrower acceptable limits cause more forecasts to be out of limits and therefore require more time to investigate).

The Percentages of Points included within the Control Limits for a Range of 1 to 4
MADs

<i>Number of MADs</i>	<i>Related Number of Standard Deviations</i>	<i>Percentage of Points lying within Control Limits</i>
± 1	0.798	57.048
± 2	1.596	88.946
± 3	2.394	98.334
± 4	3.192	99.856

In a perfect forecasting model, the sum of the actual forecast errors would be 0; the errors that result in overestimates should be offset by errors that are underestimates. The tracking signal would then also be 0, indicating an unbiased model, neither leading nor lagging the actual demands.

Exponentially smoothed MAD

Often MAD is used to forecast errors. It might then be desirable to make the MAD more sensitive to recent data. A useful technique to do this is to compute an exponentially smoothed MAD as a forecast for the next period's error range. The procedure is similar to single exponential smoothing, covered earlier in this chapter. The value of the MAD forecast is to provide a range of error. In the case of inventory control, this is useful in setting safety stock levels.

where

$$MAD_t = \alpha |A_{t-1} - F_{t-1}| + (1-\alpha) MAD_{t-1}$$

MAD_t = Forecast MAD for the t th period

α = Smoothing constant (normally in the range of 0.05 to 0.20)

A_{t-1} = Actual demand in the period $t - 1$

F_{t-1} = Forecast demand for period $t - 1$

REVIEW EXERCISE

Q. Demand for patient surgery at a hospital has increased steadily in the past few years, as seen in the following table.

<i>year</i>	<i>Outpatient surgeries performed</i>
<i>1</i>	45
<i>2</i>	50
<i>3</i>	52
<i>4</i>	56
<i>5</i>	58
<i>6</i>	

The director of medical services predicted six years ago that demand in year 1 would be 42 surgeries. Using exponential smoothing with a weight $\alpha = 0.20$, develop forecasts for years 2 through 6. What is the MAD?

Ans: 42.6, 44.1, 45.7, 47.7, 49.8, MAD = 7.78

Q. Room registrations in the Park Hotel have been recorded for the past nine years. Management would like to determine the mathematical trend of guest registration in order to project future occupancy. This estimate would help the hotel determine whether a future expansion will be needed. Given the following time series data, develop a regression equation relating registrations to time. Then forecast year 11's registration. Room registrations are in thousands.

Year 1 : 17 , Year 2 : 16 , Year 3 : 16 , Year 4 : 21, Year 5 : 20 ,

Year 6 : 20 , year 7 : 23 , Year 8 : 25 , Year 9 : 24

Ans: $b = 1.135$, $a = 14.545$, $Y = 14.545 + 1.135 X$, Reg. for the 11th year = 27030 guests

Q. Annual sales of Brand X over the last eleven years have been as follows:

<i>1995</i>	<i>1996</i>	<i>1997</i>	<i>1998</i>	<i>1999</i>	<i>2000</i>	<i>2001</i>	<i>2002</i>	<i>2003</i>	<i>2004</i>	<i>2005</i>
<i>50</i>	59	46	54	65	51	60	70	56	66	76

You are required to (a) calculate a three year moving average, (b) plot a the series and the trend on the same graph (c) produce a sales forecast for 2006 stating assumptions, if any.

Q. The following tabulations are actual sales of units for six months and a starting forecast in January.

(a) Calculate forecasts for the remaining five months using simple exponential smoothing with $\alpha = 0.2$.

(b) Calculate MAD for the forecasts

<i>Month</i>	<i>Actual</i>	<i>Forecast</i>
<i>January</i>	100	80
<i>February</i>	94	
<i>March</i>	106	
<i>April</i>	80	
<i>May</i>	68	
<i>June</i>	94	

Ans: 84, 86, 90, 88, 84

Q. Here are the actual tabulated demands for an item for a nine month period (January through September). The supervisor wants to test forecasting methods to see which method was better over the period.

<i>Month</i>	<i>Actual demand</i>	<i>Month</i>	<i>Actual demand</i>
<i>January</i>	110	June	180
<i>February</i>	130	July	140
<i>March</i>	150	August	130
<i>April</i>	170	September	140
<i>May</i>	160		

- Forecast April through September using a three month moving average.
- Use simple exponential smoothing with $\alpha = 0.3$ to estimate April through September.
- Use MAD to decide which method produced the better forecast over the six month period.

Ans: (a) 130, 150, 160, 170, 160, 150.

(b) 136, 146, 150, 159, 153, 146

(c) Exponential smoothing performed better

Q. A bakery markets cakes through a chain of food stores. It has been experiencing over and under production because of forecasting errors. The following data are its demand for cakes for the past four weeks. Cakes are made for the following day; for example, Sunday's cakes production is for Monday's sales, Monday's production is for Tuesday's sales and so on. The bakery is closed on Sunday. So Friday's production must satisfy demand for both Saturday and Sunday.

	<i>4 weeks ago</i>	<i>3 weeks ago</i>	<i>2 weeks ago</i>	<i>Last week</i>
<i>Monday</i>	2200	2400	2300	2400
<i>Tuesday</i>	2000	2100	2200	2200
<i>Wednesday</i>	2300	2400	2300	2500
<i>Thursday</i>	1800	1900	1800	2000
<i>Friday</i>	1900	1800	2100	2000
<i>Saturday</i>				
<i>Sunday</i>	2800	2700	3000	2900

Make a forecast for this week on the following basis:

- (a) Daily, using a simple four week moving average.
- (b) Daily, using a weighted average of 0.40, 0.30, 0.20, and 0.10 for the past four weeks.
- (c) The bakery is also planning its purchases for bread production. If bread demand had been forecast for last week at 22,000 loaves and only 21,000 loaves were actually demanded, what would the bakery's forecast be for this week using exponential smoothing with $\alpha = 0.10$?
- (d) Suppose, with the forecast made in (c) above, this week's demand actually turns out to be 22,500. What would the new forecast be for the next week?

Ans: (a) Monday: 2325, Tuesday: 2125, Wednesday: 2375, Thursday: 1875, Friday: 1950, Sat. and Sun: 2850.

(b) Monday: 2350, Tuesday: 2160, Wednesday: 2400, Thursday: 1900, Friday: 1980, Sat. and Sun: 2880

(c) 21900

(d) 21960

Q. Data collected on the yearly demand for 50 pound bags of fertilizer at ABC Fertilizer Company are shown in the following table. Develop a three year moving average to forecast sales. Then estimate demand again with a weighted moving average in which sales in the most recent year are given a weight of 2 and sales in the other two years are each given a weight of 1. Which method do you think is best?

<i>Year</i>	<i>Demand for fertilizer bags (in thousands)</i>	<i>Year</i>	<i>Demand for fertilizer bags (in thousands)</i>
1	4	7	7
2	6	8	9
3	4	9	12
4	5	10	14
5	10	11	15
6	8		

Ans: weighted average is slightly more accurate

Q. Sales of COLDWAVE air conditioners have grown steadily during the past five years.

<i>Year</i>	<i>Sales</i>
1	450
2	495
3	518
4	563
5	584
6	?

- (a) The sales manager has predicted that year 1's sales would be 410 air conditioners. Using exponential smoothing with a weight $\alpha = 0.3$, develop forecasts for years 2 through 6.
- (b) Using smoothing constants 0.6 and 0.9, develop a forecast for the sales.
- (c) which of the above three smoothing constants gives the most accurate forecast?
- (d) Use a three year moving average forecasting model to forecast the sales.
- (e) Using the trend projection method, develop a forecasting model for the sales.

Ans: (a) Year 1: 410, Year 2: 422, Year 3: 443.9, Year 4: 466.1, Year 5: 495, Year 6: 521.8

(c) MAD for $\alpha = 0.3$ is 74.56, MAD for $\alpha = 0.6$ is 51.8, MAD for $\alpha = 0.9$ is 38.1

(e) If the years are coded -2,-1,0,1 and 2 , $Y = 522 + 33.6X$; next years sales = 622.8

If the years are coded 1,2,3,4 and 5, $Y = 421.2 + 33.6X$

Q. In this problem, you are to test the validity of your forecasting model. Here are the forecasts for a model you have been using and the actual demands that occurred.

<i>Week</i>	<i>Forecast</i>	<i>Actual</i>
1	800	900
2	850	1000
3	950	1050
4	950	900
5	1000	900
6	975	1100

Compute the MAD and tracking signal. Then decide whether the forecasting model you have been using is giving reasonable results.

Ans: MAD : 104, TS = 3.1. The high TS value indicates the model is unacceptable.

Q. The following table shows predicted product demand using your particular forecasting method along with the actual demand that occurred.

<i>Forecast</i>	<i>Actual</i>
1500	1550
1400	1500
1700	1600
1750	1650
1800	1700

- (a) Compute the tracking signal using the mean absolute deviation and running sum of forecast errors
 (b) Discuss whether your forecasting method is giving good predictions.

Ans: (a) MAD = 90, TS = -1.67 (b) Model O.K since TS is -1.67

Q. The following table shows the past two years of quarterly sales information. Assume that there are both trend and seasonal factors and that the season cycle is one year. Use time series decomposition to forecast quarterly sales for the next year.

<i>Quarter</i>	<i>Sales</i>	<i>Quarter</i>	<i>Sales</i>
<i>1</i>	160	<i>5</i>	215
<i>2</i>	195	<i>6</i>	240
<i>3</i>	150	<i>7</i>	205
<i>4</i>	140	<i>8</i>	190

Ans: Q 9: 232 , Q 10 : 281 , Q 11 : 239 , Q 12 : 231.

Q. The managers of a company are preparing revenue plans for the last quarter of 2004/05 and for the first three quarters of 2005/06. The data below refer to one of the main products.

<i>Revenue</i>	<i>April-June</i>	<i>July-Sept</i>	<i>Oct-Dec</i>	<i>Jan-Mar</i>
	<i>Q1</i>	<i>Q2</i>	<i>Q3</i>	<i>Q4</i>
<i>2001/02</i>	49	37	58	67
<i>2002/03</i>	50	38	59	68
<i>2003/04</i>	51	40	60	70
<i>2004/05</i>	50	42	61	--

Calculate the four quarterly seasonal indices and decompose the time series.

Ans: Note: For the computation of the modified mean, the highest and the lowest values of the means in a quarter are not ignored. Adjustment factor: $4/3.994 = 1.0015$ and the Seasonal indices are
 Q1: 0.924, Q2: 0.719, Q3: 1.905, Q4: 1.262.