16 ANALYSIS OF TIME SERIES

16.1 TIME SERIES

The sequence $y_1, y_2, \cdots, y_n$ of n observations of a variable Y, recorded in accordance with their time of occurrence $t_1, t_2, \cdots, t_n$, is called a *time series.* Symbolically, the variable Y can be expressed as a function of time t as

$$y = f(t) + e$$

where $f(t)$ is a completely determined or specified sequence that follows a systematic pattern of variation and e is a random error that follows an irregular pattern of variation.

Signal. The signal is a systematic component of variation in a time series.

Noise. The noise is an irregular component of variation in a time series.

Therefore, a time series is a sequence of observations, on a variable, that are arranged in chronological order. The observations in a time series are usually made at equidistant points of time. Examples of a time series are: the hourly temperature recorded at a weather bureau, the total annual yield of wheat over a number of years, the monthly sales of a fertilizer at a store, the enrolment of students in various years in a college, the daily sales at a departmental store, *etc.*

16.1.1 Historigram. A *historigram* is a graphic representation of a time series that reveals the changes occurred at different time periods. A first step in the prediction or forecast of a time series involves an examination of the set of past observations. The construction of a historigram involves the following steps:

(*i*) Using an appropriate scale, take time t along *x-axis* as an independent variable.

(*ii*) Using an appropriate scale, plot the observed values of variable Y as a dependent variable against the given points of time.

(*iii*) Join the plotted points by line segments to get the required historigram.

***Example* 16.1** *Draw a historigram to show the population of Pakistan in various census years*

Census year	Population (million)
1951	33.44
1961	42.88
1972	65.31
1981	83.78
1998	130.58

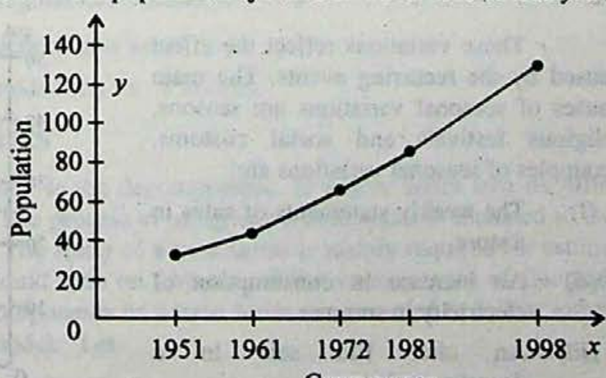

Fig. 16.1 Population of Pakistan

Solution. The population of Pakistan in different census years is represented by a historigram as shown in Fig. 16.1.

16.2 COMPONENTS OF A TIME SERIES

The examples of time series suggest that a typical time series may be composed of the following four components:

(*i*) Secular trend (T)

(*ii*) Seasonal variations (S)

(*iii*) Cyclical fluctuations (C)

(*iv*) Irregular movements (I)

These are the basic components of a time series, each of which is regarded as the result of a well defined distinct cause. A time series is not necessarily composed of all these four components.

16.2.1 Secular (Long-term) Trend. The *secular trend* is a line or curve that shows the general tendency of a time series. It represents a relatively smooth, steady, and gradual movement of a time series in the same direction (upward or downward). It shows the general increase or decrease in a sequence of observations, and reflects the effect of the forces operating over a fairly long period of time. Examples of secular trend are:

(*i*) The decline in death rate due to advancement in science.

(*ii*) A continually increasing demand for smaller automobiles.

(*iii*) A need for increased wheat production due to a constant increase in population.

16.2.2 Seasonal Variations. The *seasonal variations* are short term movements that represent the regularly recurring changes in a time series. These variations indicate a repeated pattern of identical changes in the data that tend to recur regularly during a period of one year or less. These changes are repeated with the same pattern within a specific time period, called the *periodicity*. Seasonal variations may have the fixed periodicity, such as daily, weekly, monthly, or yearly *etc.* These changes are periodic in nature and their influence; upon a specific time series is fairly regular, both in respect of length (*time*) and amplitude (*size*).

These variations reflect the effect caused by the recurring events. The main causes of seasonal variations are seasons, religious festivals and social customs. Examples of seasonal variations are:

(*i*) The weekly statements of sales in a store.

(*ii*) An increase in consumption of electricity in summer.

(*iii*) An after Eid sale in a departmental store.

(*iv*) An increase in sales of cold drinks during summer.

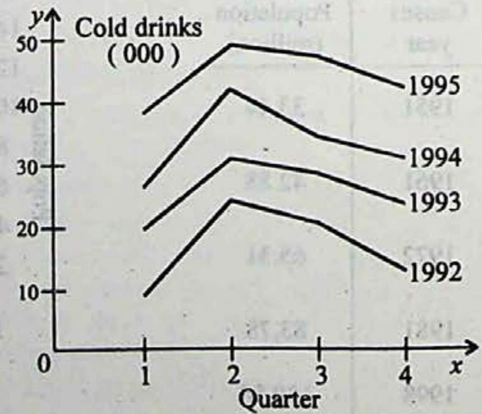

Fig. 16.2 The seasonal pattern of cold drinks sales

16.2.3 Cyclical Fluctuations. The *cyclical fluctuations* are the long term oscillations about the trend. These are the periodic up-and-down movements in a time series that tend to recur over a long period of time. The cyclic patterns tend to vary in length (time) and amplitude (size) and they are differentiated from the seasonal variations by the fact that they do not have a fixed periodicity. Although, these variations are recurring yet are less predictable than seasonal variations and secular trend, therefore, they have a more dangerous effect on a business and economic activity. These fluctuations reflect the effect caused by a so called *business cycle*. A business cycle has the following four phases:

(*i*) Trough (Depression)

(*ii*) Expansion (Recovery)

(*iii*) Peak (Boom or Prosperity)

(*iv*) Recession (Contraction)

A *trough* is the lowest point relative to the rest of the particular cycle. After the downswing has run its course, the *expansion* phase reverses direction and starts rising. The upswing eventually levels off and reaches its *peak*. This is the highest point relative to the particular cycle. Finally, the upswing starts to turn downward. We refer to this following phase as a *recession.*

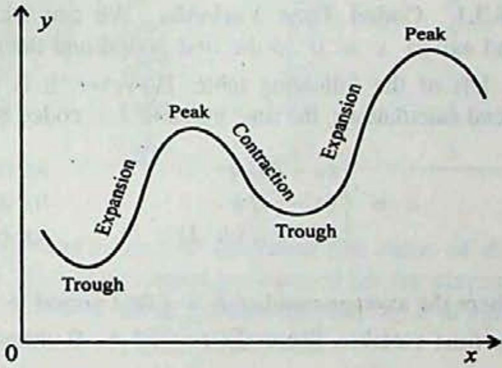

Fig. 16.3 The four phases of business cycle

16.2.4 Irregular Movements. The *irregular movements* are unpredictable changes that indicate the effect of random events. The examples of random events are wars, floods, earthquakes, strikes, fires, elections *etc*. The irregular movements are unsystematic, non-recurring, accidental and unusual in nature. These variations are also known as *erratic, accidental* or *random* variations. Examples of irregular movements are:

(*i*) A steel strike, delaying production for a week.

(*ii*) A fire in a factory delaying production for 3 weeks.

16.3 ANALYSIS OF TIME SERIES

The *analysis of a time series* is the decomposition of a time series into its different components for their separate study. The process of analysing a time series is intended to isolate and measure its various components. The study of a time series is mainly required for estimation and forecasting. An ideal forecast should base on forecasts of the various types of fluctuations. While performing the analysis, the components of a time series are assumed to follow either the multiplicative model or the additive model. Let

Y = Original observation,

T = Trend component, S = Seasonal component,

C = Cyclical component, I = Irregular component.

In the *multiplicative model*, it is assumed that the value Y of a composite series is the product of the four components T, S, C and I. Symbolically,

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

where the component T is given in original units of Y but the other components S, C and I are expressed as percentage unitless index numbers.

In the *additive model*, it is assumed that the value Y of the composite series is the sum of the four components T, S, C and I. Symbolically,

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

where the components T, S, C and I all are given in the original units of Y. Conventionally, the multiplicative model is considered as the standard model for analysis of a time series.

16.3.1 Coded Time Variable. We can take the origin at the beginning of a time series and assign $x = 0$ to the first period and then number other periods as $1, 2, 3, \cdots$ as shown at left of the following table. However, it is important to note that in order to simplify the trend calculations, the time variable t is coded by

$$x = \begin{cases} (t - \bar{t})/h & \text{for odd number of periods} \\ (t - \bar{t})/h & \text{for even number of periods in units of } h \text{ period} \\ (t - \bar{t})/(h/2) & \text{for even number of periods in units of } h/2 \text{ period} \end{cases}$$

where the average number $\bar{t}$ = (first period + last period)/2 and h is the constant interval in the time variable. Since $\Sigma(t - \bar{t}) = 0$, then we get $\sum x = 0 = \sum x^3 = \sum x^5 = \cdots$, and so on.

The odd number of years in period 1980 — 1984 at the middle of the following table has $\bar{t} = (1980 + 1984)/2 = 1982$ as the middle point. The code for the year t is $x = t - \bar{t}$. For $t = 1982$, we have $x = t - \bar{t} = 1982 - 1982 = 0$. Thus, the coded year is zero at $\bar{t}$. For $t = 1980$, we have $x = 1980 - 1982 = -2$. Actually, the only computation we need is that for $\bar{t}$. Thus after entering $x = 0$ at the middle of an odd number of years, we assign $-1, -2, \cdots$ and so on for the years before the middle year, and $1, 2, \cdots$ and so on for the years after the middle year as shown in the following table.

The even number of years in period 1980 — 1985 at the right of the following table has $\bar{t} = (1980 + 1985)/2 = 1982.5$ as the middle point. So $x = 0$ half way between the years 1982 and 1983. For $t = 1982$, we have $x = t - \bar{t} = 1982 - 1982.5 = -0.5$. Then after considering $x = 0$ at the middle of an even number of years, we assign $-0.5, -1.5, -2.5, \cdots$ and so on for the years before the middle year and $0.5, 1.5, 2.5, \cdots$ and so on for the years, after the middle year as shown in the following table.

However, to avoid decimals in the coded years we can take the unit of measurement as 1/2 year. Then after considering $x = 0$ at the middle of an even number of years, we assign $-1, -3, -5, \cdots$ and so on for the years before the middle year and $1, 3, 5, \cdots$ and so on for the years after the middle year as shown in the following table.

Table: Coded Year Number

Origin at beginning. The starting year is coded $x = 0$		Odd number of years. The middle year is coded $x = 0$		Even number of years. $x = 0$ is at the centre of two middle years		
Year t	Coded year in units as one year x	Year t	Coded year in units as one year $x = t - \bar{t}$	Year t	Coded year in units as one year $x = t - \bar{t}$	Coded year in units as 1/2 year $x = \frac{t - \bar{t}}{1/2}$
1980	0	1980	−2	1980	−2.5	−5
1981	1	1981	−1	1981	−1.5	−3
1982	2	1982	0	1982	−0.5	−1
1983	3	1983	1	1983	0.5	1
1984	4	1984	2	1984	1.5	3
				1985	2.5	5

16.4 ESTIMATION OF SECULAR TREND

It has been earlier stated that one component force that determine the value of the variable at any period of time is the secular trend. The secular trend is measured for the purpose of prediction or projection into the future. The secular trend can be represented either by a straight line or by some type of smooth curve. It is measured by the following methods:

(*i*) Method of free hand curve

(*ii*) Method of semi-averages

(*iii*) Method of moving averages

(*iv*) Method of least squares

16.4.1 Method of free hand curve. The secular trend is measured by the method of free hand curve in the following steps.

(*i*) Using an appropriate scale, take the time periods along *x-axis*, as an independent variable.

(*ii*) Using an appropriate scale, plot the points for observed values of the variable Y as a dependent variable against the given time periods.

(*iii*) Join these plotted points by line segments to get a historigram.

(*iv*) Keeping in view the up and down fluctuations of the graph, draw a free hand smooth curve or a straight line through the historigram in a way such that it indicates the general trend of the time series.

(*v*) Instead of locating the line simply by eye looking at the graph, the average $\bar{y}$ of original values may be used as the trend value $\bar{y}'$ at the middle of the time period. Plot this average in the middle of the time period and the required trend line or curve should be drawn through this point, as it is a reasonable condition that $\bar{y}$ should be equal to $\bar{y}'$.

(*vi*) Read off the trend values for different time periods from this trend line or curve.

If a straight line is used for locating the trend, then it becomes easy to estimate the rate of change (slope of the line b) by measuring the difference $y'_{x+1} - y'_x$ of the trend values for any two consecutive time periods x and $x + 1$. Symbolically it is expressed as $b = y'_{x+1} - y'_x$. Then the equation of the trend line is summarised in the slope intercept from as $y' = a + bx$ with origin at any time period, so that, a = trend value for the origin.

If the historigram indicates a non-linear trend, then in such situations it is generally preferred to use a curve instead of a straight line to show the secular trend.

Merits.

(*i*) The free hand curve method is a simple, easy and quick method for measuring secular trend.

(*ii*) The trend line or curve smoothes out seasonal variations.

(*iii*) A good fitted trend line or curve can give a close approximation to the trend based on a mathematical model.

Demerits.

(*i*) It is a rough and crude method. It is greatly affected by the personal bias, *i. e.*, different persons may fit different trends to the same data.

(*ii*) It requires too much practice to get a good fit.

(*iii*) The free hand curve method is subject to personal bias, so it is unable to give reliable estimates.

***Example* 16.2** *The following time series shows the number of road accidents in Punjab for the year* 1972 *to* 1978.

Year	1972	1973	1974	1975	1976	1977	1978
Number of accidents	2493	2638	2699	3038	3745	4079	4688

(*i*) *Obtain the historigram showing the number of road accidents and a free hand trend line by drawing a straight line.*

(*ii*) *Find the trend values for this time series.*

Solution. (*i*)

Year	Value y	Total	Mean	Trend value
1972	2493			2200
1973	2638			2550
1974	2699			2950
1975	3038	23380	3340	3340
1976	3745			3650
1977	4079			4050
1978	4688			4400

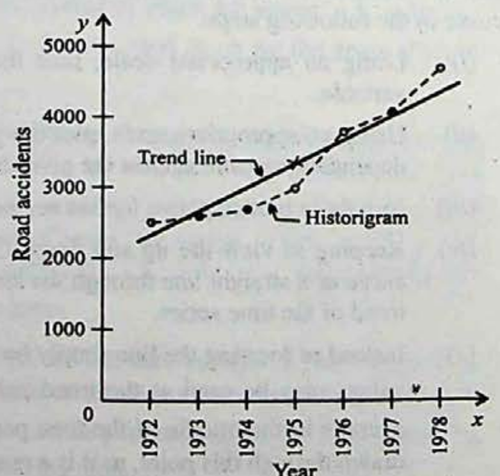

Fig. 16.4 Number of road accidents

(ii) Reading off the trend line, we get the trend values.

16.4.2 Method of Semi-averages. The secular trend is measured by the method of semi-averages in the following steps.

(i) Divide the observed values of the time series into two equal periods. If the number of observed values is odd then it is advisable either to omit the middle value altogether or to include the middle value in each half.

(ii) Take the average of each part and place these average values against the mid points of the two parts, and the average value of each part should be considered equal to the average value of its respective trend values.

(iii) Plot the semi-averages on the graph of the original values.

(iv) Draw the required trend line through these two plotted points, and extend it to cover the whole period.

(v) With two points located on the straight line, it is simple to compute the slope and y-intercept of the line. This slope gives the estimate of the rate of change of values. Now, the trend values are found either by reading off the semi-average trend line or by the estimated straight line as explained below

Semi-average Trend Line. Let y_1' and y_2' be the semi averages placed against the times x_1 and x_2, and the estimated straight line (in slope-intercept from) $y' = a + bx$ is to pass through the points (x_1, y_1') and (x_2, y_2'). The constants a (y-intercept) and b (slope of the line) can be easily determined. The equation of the line passing through the points (x_1, y_1') and (x_2, y_2') can be written as

$$y' - y_1' = \frac{y_2' - y_1'}{x_2 - x_1}(x - x_1)$$

$$y' - y_1' = b(x - x_1) \qquad \text{where} \quad b = \frac{y_2' - y_1'}{x_2 - x_1}$$

$$y' = (y_1' - bx_1) + bx$$

$$y' = a + bx \qquad \text{where} \quad a = y_1' - bx_1$$

If the number of time units in the observed time series is even, then the following formula may be used to find the slope of the trend line.

$$b = \frac{1}{n/2}\left(\frac{S_2}{n/2} - \frac{S_1}{n/2}\right)$$

$$= \frac{1}{n/2}\left(\frac{S_2 - S_1}{n/2}\right) = \frac{(S_2 - S_1)}{n^2/4}$$

$$= \frac{4(S_2 - S_1)}{n^2}$$

where S_1 = sum of y-values for the first half of the period.

S_2 = sum of y-values for the second half of the period.

n = number of time units covered by the time series.

Merits.

(*i*) The method of semi-averages is simple, easy and quick.

(*ii*) It gives an objective result.

(*iii*) It smoothes out seasonal variations.

(*iv*) It gives better approximation to the trend because it is based on a mathematical model as compared to free hand method.

Demerits.

(*i*) The arithmetic mean, which is used to average the two halves of the observed values, is highly affected by extreme values.

(*ii*) This method can only be applied if the trend is linear or approximately linear.

(*iii*) This method is not appropriate if the trend is not linear.

Example 16.3 *The following table shows the property damaged by road accidents in Punjab for the years* 1973 *to* 1979.

Year	1973	1974	1975	1976	1977	1978	1979
Property damaged	201	238	392	507	484	649	742

(*i*) *Obtain the semi-averages trend line.*

(*ii*) *Find out the trend values.*

Solution. (*i*) Let $x = t - 1973$.

Year t	Property damaged y	Semi-total	Semi-average	Coded year $x = t - 1973$	Trend value $y' = 190 + 87x$
1973	201			0	$190 + 87(0) = 190$
1974	238	831	277	1	$190 + 87(1) = 277$
1975	392			2	$190 + 87(2) = 364$
1976	507			3	$190 + 87(3) = 451$
1977	484			4	$190 + 87(4) = 538$
1978	649	1875	625	5	$190 + 87(5) = 625$
1979	742			6	$190 + 87(6) = 712$

The semi averages trend line is

$$y' = a + bx$$

Taking the origin at 1973, we have

$y_1' = 277, \quad x_1 = 1$

$y_2' = 625, \quad x_2 = 5$

$$b = \frac{y_2' - y_1'}{x_2 - x_1}$$

$$= \frac{625 - 277}{5 - 1} = 87$$

$$a = y_1' - b\,x_1$$

$$= 277 - 87(1) = 190$$

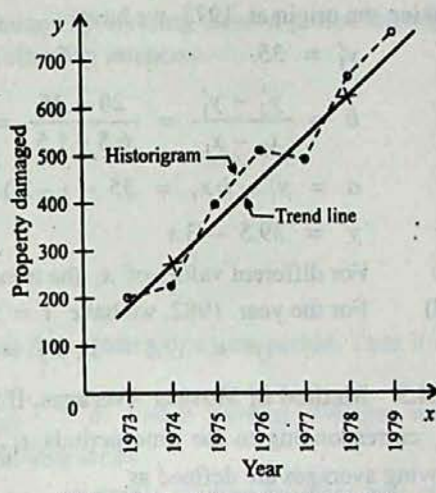

Fig. 16.5 Property damaged

The semi-averages trend line is

$y' = 190 + 87\,x$ with origin at 1973

(ii) For different values of x, the trend values are obtained as shown in the table.

***Example* 16.4** *The following table gives the number of books (in 000's) sold at a book stall for the year 1973 to 1981.*

Year	1973	1974	1975	1976	1977	1978	1979	1980	1981
Number of books (000's)	42	38	35	25	32	24	20	19	17

(*i*) *Find the equation of the semi-averages trend line.*

(*ii*) *Compute out the trend values.*

(*iii*) *Estimate the number of books sold for the year* 1982.

Solution. (*i*) Let $x = t - 1973$.

Year t	No. of books y	Semi-total	Semi-average	Coded year $x = t - 1973$	Trend value $y' = 39.5 - 3x$
1973	42			0	$39.5 - 3(0) = 39.5$
1974	38			1	$39.5 - 3(1) = 36.5$
1975	35	140	35	2	$39.5 - 3(2) = 33.5$
1976	25			3	$39.5 - 3(3) = 30.5$
1977	32			4	$39.5 - 3(4) = 27.5$
1978	24			5	$39.5 - 3(5) = 24.5$
1979	20			6	$39.5 - 3(6) = 21.5$
1980	19	80	20	7	$39.5 - 3(7) = 18.5$
1981	17			8	$39.5 - 3(8) = 15.5$

The semi-averages trend line is

$y' = a + b\,x$

Taking the origin at 1973, we have

$$y_1' = 35, \quad x_1 = 1.5 \quad \text{and} \quad y_2' = 20, \quad x_2 = 6.5,$$

$$b = \frac{y_2' - y_1'}{x_2 - x_1} = \frac{20 - 35}{6.5 - 1.5} = -3$$

$$a = y_1' - b x_1 = 35 - (-3)(1.5) = 39.5$$

$$y' = 39.5 - 3x \qquad \text{with origin at 1973}$$

(*ii*) For different values of x, the trend values are obtained as shown in the table.

(*iii*) For the year 1982, we have $x = 1982 - 1973 = 9$. Then

$$y' = 39.5 - 3(9) = 12.5$$

16.4.3 Method of Moving Averages. If the observed values of a variable Y are $y_1, y_2, \cdots, y_n$ corresponding to the time periods $t_1, t_2, \cdots, t_n$ respectively, then the k-period simple moving averages are defined as

$$a_1 = \frac{1}{k}\sum_{i=1}^{k} y_i, \qquad a_2 = \frac{1}{k}\sum_{i=2}^{k+1} y_i,$$

$$a_3 = \frac{1}{k}\sum_{i=3}^{k+2} y_i, \quad \cdots, \quad a_m = \frac{1}{k}\sum_{i=m}^{n} y_i$$

where $a_1, a_2, a_3, \cdots, a_m$ is the sequence of k-period simple moving averages. That is, the k-period simple moving averages are calculated by averaging first k observations and then repeating this process of averaging the k observations by dropping each time the first observation and including the nest one that has not been previously included. This process is continued till the last k observations have been averaged. For example, the 3-period simple moving averages are given as

$$a_1 = \frac{1}{3}(y_1 + y_2 + y_3) = \frac{1}{3}\sum_{i=1}^{3} y_i$$

$$a_2 = \frac{1}{3}(y_2 + y_3 + y_4) = \frac{1}{3}\sum_{i=2}^{4} y_i$$

$$a_3 = \frac{1}{3}(y_3 + y_4 + y_5) = \frac{1}{3}\sum_{i=3}^{5} y_i$$

and so on. Each of these simple moving average of the sequence $a_1, a_2, a_3, \cdots$ is placed against the middle of each successive group. For practical purposes the k-period moving successive totals $S_1, S_2, S_3 \cdots$ are obtained by the following relations

$$S_1 = \sum_{i=1}^{k} y_i$$

$$S_2 = S_1 + y_{k+1} - y_1$$

$$S_3 = S_2 + y_{k+2} - y_2$$

and so on. The k-period simple moving averages are obtained by dividing these k-period moving successive totals $S_1, S_2, S_3 \cdots$ by k as given in the following relations.

$$a_1 = \frac{S_1}{k}$$

$$a_2 = a_1 + \frac{y_{k+1} - y_1}{k}$$

$$a_3 = a_2 + \frac{y_{k+2} - y_2}{k}$$

and so on. Each moving average should be placed against the middle of its time period. Then it is obvious that

(*i*) When k is odd, the sequence $a_1, a_2, a_3, \cdots$ of simple moving averages will correspond directly to the observed values in the time series.

(*ii*) When k is even, the sequence $a_1, a_2, a_3, \cdots$ of simple moving averages will not correspond directly to the observed values in the time series and will be placed in the middle of two time periods. It is then sometimes necessary to centralize these averages so that they should correspond to the observed values in the time series. For centralization, further 2-period moving averages of the former k-period moving averages are computed which are called k-period centred moving averages.

Smoothing of a Time Series. The smoothing of a time series is a process of eliminating the unwanted fluctuations in a time series. The moving averages tend to reduce the variation present among the observed values of a time series, so they are used to eliminate the unwanted fluctuations. Thus the moving averages may be used in smoothing of a time series. They eliminate the effect of periodic fluctuations if an appropriate period moving averages are calculated. For this purpose the period of the moving average is chosen such that it should be equal to the period of at least one cycle. The secular trend is measured by taking the following steps.

(*i*) Find the moving averages of an appropriate period.

(*ii*) Plot the points representing these moving averages on the graph of the observed time series and join these points by the line segments.

(*iii*) The graph of the moving averages indicates the secular trend by eliminating the periodic fluctuations

The period of moving averages should be decided in the light of the periodicity of a time series. Because only the moving averages, calculated by using the time period which approximately coincides with the periodicity of the time series, would eliminate, nearly completely, all its regular fluctuations and show a trend.

Merits.

(*i*) The method of moving averages is easy and simple.

(*ii*) The moving averages of an appropriate period eliminate the periodic fluctuations, so it may be used to eliminate cyclical and seasonal fluctuations.

Demerits.

(*i*) The method of moving averages does not give the trend values at the beginning and at the end of the original time series.

(*ii*) The moving averages are highly affected by the extreme observations however the affect may be reduced by using the geometric mean as average.

(*iii*) The method of moving averages does not provide a mathematical equation for the trend, therefore, the forecasting is only subjective.

(*iv*) The selection of inappropriate period of moving averages may generate the cycles which are not present in the observed time series.

Example 16.5 *The following table shows the production of silver utensils (in thousands) at a certain factory in Gujranwala.*

Year	Utensils (000)
1970	170.0
1971	154.8
1972	156.5
1973	158.9
1974	140.3
1975	154.2
1976	160.7
1977	178.3

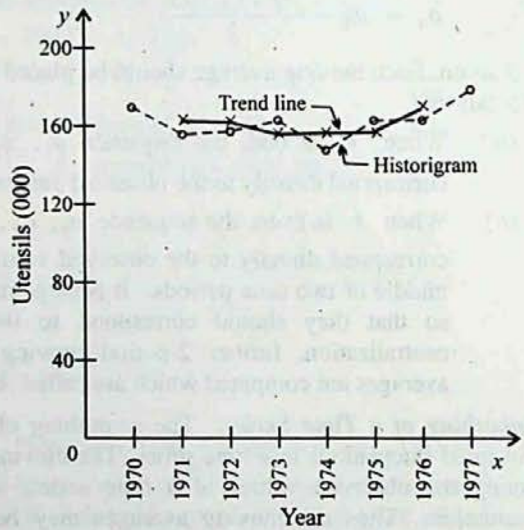

Fig. 16.6 Production of silver utensils

(*i*) *Calculate 3 year simple moving averages for the following time series.*

(*ii*) *Also plot actual data and moving averages on a graph.*

Solution.

Year	Production y	3-year moving total	3-year moving average
1970	170.0		
1971	154.8	481.3	160.43
1972	156.5	470.2	156.73
1973	158.9	455.7	151.90
1974	140.3	453.4	151.13
1975	154.2	455.2	151.73
1976	160.7	493.2	164.40
1977	178.3		

Example 16.6 *The following table shows the food grain price index number of quarters for the years* 1962 *and* 1963.

Year	Quarter I	Quarter II	Quarter III	Quarter IV
1962	93	97	96	93
1963	97	102	106	98

Calculate four quarter moving average centred.

Solution. The four quarter centred moving averages are obtained as under:

(1) Year	(2) Quarter	(3) Price index number y	(4) 4-quarter moving total	(5) 4-quarter moving average (4) ÷ 4	(6) 2-quarter moving total of (5)	(7) 4-quarter centred moving average (6) ÷ 2
1962	I	93				
	II	97				
			379	94.75		
	III	96			190.50	95.25
			383	95.75		
	IV	93			192.75	96.38
			388	97.00		
1963	I	97			196.50	98.25
			398	99.50		
	II	102			200.25	100.12
			403	100.75		
	III	106				
	IV	98				

Alternately, for the sake of convince, the four quarter centred moving averages may be calculated as shown in the table given below:

(1) Year	(2) Quarter	(3) Price index number y	(4) 4-quarter moving total	(5) 4-quarter centred moving total	(6) 4-quarter centred moving average (5) ÷ 8
1962	I	93			
	II	97			
			379		
	III	96		762	95.25
			383		
	IV	93		771	96.38
			388		
1963	I	97		786	98.25
			398		
	II	102		801	100.12
			403		
	III	106			
	IV	98			

16.4.4 Method of Least Squares. For situations in which it is desirable to have a mathematical equation to describe the secular trend of a time series, the most commonly used method is to fit a straight line $\hat{y} = a + bx$, a second degree parabola $\hat{y} = a + bx + cx^2$, *etc.*, where y is the value of a time series variable, x representing the time and all others are constants. For determining the values of the constants appearing in such an equation, the most widely used method is the method of least squares, because it is a practical method that provides best fit according to a reasonable criterion. The principle of least squares says that "*the sum of squares of the deviations of the observed values from the corresponding expected values should be least*".

Among all the trend lines approximating a given time series data, the trend line is called a least squares fit for which the sum of the squares of the deviations of the observed values from their corresponding expected values is the least. The method of least squares consists of minimizing the sum of the squares of these deviations. To avoid the personal bias in measuring the secular trend this method is used to find a trend line approximating a given time series.

Secular Trend* — *Linear. It is useful to describe the trend in a time series where the amount of change is constant per unit time.

Let $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$ be the n pairs of observed sample values of a time series variable y, with x representing the coded time value. We can plot these n points on a graph. Because of the fact that $y_1, y_2, \cdots, y_n$ are observed values of a time series variable, these points will not necessarily lie on a straight line. Let us suppose that we want to fit a straight line expressed in slope-intercept form as

$$\hat{y} = a + bx$$

This line will be called the least squares line if it makes $\sum (y - a - bx)^2$ minimum. The method of least squares yields the following normal equations.

$$\sum y = na + b\sum x, \qquad \sum xy = a\sum x + b\sum x^2$$

The normal equations give the values of a and b as

$$b = \frac{n\sum xy - (\sum x)(\sum y)}{n\sum x^2 - (\sum x)^2}, \qquad a = \frac{\sum y - b\sum x}{n} = \bar{y} - b\bar{x}$$

However, if $\sum x = 0$, then the usual normal equations reduce to

$$\sum y = na, \qquad \sum xy = b\sum x^2$$

Therefore, the values of a and b also reduce to

$$a = \frac{\sum y}{n} = \bar{y}, \qquad b = \frac{\sum xy}{\sum x^2}$$

The trend values $\hat{y}$ are computed from the least squares line $\hat{y} = a + bx$ by substituting the values of x corresponding to the different time periods. The secular trend can be indicated on a graph by plotting these estimated vales against their respective time periods.

***Properties*:**

(*i*) The least squares line always passes through the point $(\bar{x}, \bar{y})$ called the centre of gravity of the data.

(*ii*) The sum of the deviations $\sum(y - \hat{y})$ of the observed values y from their corresponding expected values $\hat{y}$ is zero, *i. e.*,

$$\sum(y - \hat{y}) = 0 \quad \Rightarrow \quad \sum y = \sum \hat{y}$$

(*iii*) The sum of squares of the deviations $\sum(y - \hat{y})^2$ measures how well the trend line fits the data. A smaller $\sum(y - \hat{y})^2$ means the better fit.

***Example* 16.7** *The following table shows the production of steel in a steel mill for the time period* 1977 *to* 1983.

Year	1977	1978	1979	1980	1981	1982	1983
Production (000 tons)	12.7	10.1	13.0	13.2	12.6	14.2	13.7

Find the linear trend by the method of least squares by taking the origin:

(*i*) *at the beginning period of the time period,*

(*ii*) *at the middle of the time period* 1977 — 83.

Calculate the trend values in both cases.

Solution. (*i*) Taking the origin at the beginning period, 1977 (*i. e.,* July 1, 1977), we have $x = t - 1977$.

Year t	Production y	Coded year $x = t - 1977$	xy	x^2	Trend value $\hat{y} = 11.628 + 0.386x$
1977	12.7	0	0	0	11.628 + 0.386 (0) = 11.628
1978	10.1	1	10.1	1	11.628 + 0.386 (1) = 12.014
1979	13.0	2	26.0	4	11.628 + 0.386 (2) = 12.400
1980	13.2	3	39.6	9	11.628 + 0.386 (3) = 12.786
1981	12.6	4	50.4	16	11.628 + 0.386 (4) = 13.172
1982	14.2	5	71.0	25	11.628 + 0.386 (5) = 13.558
1983	13.7	6	82.2	36	11.628 + 0.386 (6) = 13.944
Total	89.5	21	279.3	91	

The least squares trend line is

$$\hat{y} = a + bx$$

The least squares estimates a and b are

$$b = \frac{n\sum xy - (\sum x)(\sum y)}{n\sum x^2 - (\sum x)^2} = \frac{7(279.3) - (21)(89.5)}{7(91) - (21)^2} = 0.386$$

$$a = \frac{\sum y - b\sum x}{n} = \frac{89.5 - (0.386)(21)}{7} = 11.628$$

The best fitted line is

$$\hat{y} = 11.628 + 0.386x \qquad \text{with origin at } 1977$$

For different values of x, the trend values are obtained as shown in the table.

(*ii*) We have $\bar{t} = (1977 + 1983)/2 = 1980$. Taking the origin at the middle of the time period at 1980 (*i. e.,* July 1, 1980), we have $x = t - \bar{t} = t - 1980$.

Year t	Production y	Coded year $x = t - 1980$	xy	x^2	Trend values $\hat{y} = 12.786 + 0.386x$
1977	12.7	− 3	− 38.1	9	12.786 + 0.386 (− 3) = 11.628
1978	10.1	− 2	− 20.2	4	12.786 + 0.386 (− 2) = 12.014
1979	13.0	− 1	− 13.0	1	12.786 + 0.386 (− 1) = 12.400
1980	13.2	0	0	0	12.786 + 0.386 (0) = 12.786
1981	12.6	1	12.6	1	12.786 + 0.386 (1) = 13.172
1982	14.2	2	28.4	4	12.786 + 0.386 (2) = 13.558
1983	13.7	3	41.1	9	12.786 + 0.386 (3) = 13.944
Total	89.5	0	10.8	28	

The least squares trend line is $\hat{y} = a + bx$

Since $\sum x = 0$, the least squares estimates a and b are

$$a = \frac{\sum y}{n} = \frac{89.5}{7} = 12.786$$

$$b = \frac{\sum xy}{\sum x^2} = \frac{10.8}{28} = 0.386$$

The best fitted line is

$\hat{y} = 12.786 + 0.386\,x$ with origin at 1980

For different values of x, the trend values are obtained as shown in the table.

Example 16.8 *The consumer price index for medical care (medical cost) are given in the following table for the years* 1980 *to* 1987. *The base period* 1979 *is assigned the value* 100 *which actually means* 100 %.

Year	1980	1981	1982	1983	1984	1985	1986	1987
Production (000 tons)	106.0	111.1	117.2	121.3	125.2	128.0	132.6	138.0

Find a least squares linear trend,

(*i*) *by taking the origin at the middle of the time period with unit of measurement as* 1 *year*

(*ii*) *with unit of measurement as* 1/2 *year.*

Compute the trend values in both cases.

Solution. (*i*) We have, $\bar{t} = (1980 + 1987)/2 = 1983.5$. Taking the origin at the middle of the years 1983 and 1984 (*i. e.*, January 1, 1984), with unit of measurement as 1 year, we have $x = t - \bar{t} = t - 1983.5$.

Year t	Production y	Coded year $x = t - 1983.5$	xy	x^2	Trend value $\hat{y} = 122.42 + 4.38\,x$
1980	106.0	−3.5	−371.00	12.25	122.42 + 4.38 (− 3.5) = 107.09
1981	111.1	−2.5	−277.75	6.25	122.42 + 4.38 (− 2.5) = 111.47
1982	117.2	−1.5	−175.80	2.25	122.42 + 4.38 (− 1.5) = 115.85
1983	121.3	−0.5	−60.65	0.25	122.42 + 4.38 (− 0.5) = 120.23
1984	125.2	0.5	62.60	0.25	122.42 + 4.38 (0.5) = 124.61
1985	128.0	1.5	192.00	2.25	122.42 + 4.38 (1.5) = 128.99
1986	132.6	2.5	331.50	6.26	122.42 + 4.38 (2.5) = 133.37
1987	138.0	3.5	483.00	12.25	122.42 + 4.38 (3.5) = 137.75
Total	979.4	0	183.9	42	

The least squares trend line is

$\hat{y} = a + bx$

Since $\sum x = 0$, the least squares estimates a and b are

$$a = \frac{\sum y}{n} = \frac{979.4}{8} = 122.42$$

$$b = \frac{\sum xy}{\sum x^2} = \frac{183.9}{42} = 4.38$$

The best fitted line is

$\hat{y} = 122.42 + 4.38\,x$ with origin at middle of the years 1983 and 1984 and unit of measurement as 1 year

For different values of x, the trend values are obtained as shown in the table.

(ii) We have, $\bar{t} = (1980 + 1987)/2 = 1983.5$. Taking the origin at the middle of the years 1983 and 1984 (*i. e.*, January 1, 1984), with unit of measurement as 1/2 year, we have

$$x = \frac{t-\bar{t}}{1/2} = \frac{t-1983.5}{1/2}$$

Year t	Production y	Coded year $x = \frac{t-1983.5}{1/2}$	xy	x^2	Trend value $\hat{y} = 122.42 + 2.19\,x$
1980	106.0	−7	−742.0	49	122.42 + 2.19 (−7) = 107.09
1981	111.1	−5	−555.5	25	122.42 + 2.19 (−5) = 111.47
1982	117.2	−3	−351.6	9	122.42 + 2.19 (−3) = 115.85
1983	121.3	−1	−121.3	1	122.42 + 2.19 (−1) = 120.23
1984	125.2	1	125.2	1	122.42 + 2.19 (1) = 124.61
1985	128.0	3	384.0	9	122.42 + 2.19 (3) = 128.99
1986	132.6	5	663.0	25	122.42 + 2.19 (5) = 133.37
1987	138.0	7	966.0	49	122.42 + 2.19 (7) = 137.75
Total	979.4	0	367.8	168	

The least squares trend line is $\hat{y} = a + b\,x$

Since $\sum x = 0$, the least squares estimates a and b are

$$a = \frac{\sum y}{n} = \frac{979.4}{8} = 122.42$$

$$b = \frac{\sum xy}{\sum x^2} = \frac{367.8}{168} = 2.19$$

The best fitted line is

$\hat{y} = 122.42 + 2.19\,x$ with origin at middle of the years 1983 and 1984 and unit of measurement as 1/2 year

For different values of x, the trend values are obtained as shown in the table.

***Example* 16.9** *The consumer price index numbers y for medical care (medical cost) were given for the years* 1980 — 1987. *The base period* 1979 *was assigned the value* 100. *The least squares linear trend, with x measured from the middle of* 1983 *and* 1984 *(i. e., January* 1, 1984 *), and unit of measurement as* 1/2 *year is*

$$\hat{y} = 122.42 + 2.19\,x$$

(i) Compute the trend values.

(ii) *Predict the consumer price index number for the year* 1988.

(iii) *In which year can we expect the index of medical cost to be double than that of* 1979 *assuming the present trends.*

Solution. The least squares linear trend is

$\hat{y} = 122.42 + 2.19\,x$ with origin at middle of the years 1983 and 1984 and unit of measurement as 1/2 year

We have $\bar{t} = (1980 + 1987)/2 = 1983.5$. Then

$$x = \frac{t - \bar{t}}{1/2} = \frac{t - 1983.5}{1/2}$$

(i) For different values of x, the trend values are shown in the table.

Year t	Coded year $x = \frac{t - 1983.5}{1/2}$	Trend value $\hat{y} = 122.42 + 2.19\,x$
1980	−7	122.42 + 2.19 (−7) = 107.09
1981	−5	122.42 + 2.19 (−5) = 111.47
1982	−3	122.42 + 2.19 (−3) = 115.85
1983	−1	122.42 + 2.19 (−1) = 120.23
1984	1	122.42 + 2.19 (1) = 124.61
1985	3	122.42 + 2.19 (3) = 128.99
1986	5	122.42 + 2.19 (5) = 133.37
1987	7	122.42 + 2.19 (7) = 137.75

(ii) For $t = 1988$, we have $x = \frac{t - 1983.5}{1/2} = \frac{1988 - 1983.5}{1/2} = 9$

The estimated consumer price index for the year 1988 is

$$\hat{y} = 122.42 + 2.19\,(9) = 142.13$$

(iii) Price index for 1979 is 100. Expected price index for t is 200.

Now $200 = 122.42 + 2.19\,x \quad \Rightarrow \quad x = 35.4$

But $x = \frac{t - 1983.5}{1/2}$

$$35.4 = \frac{t - 1983.5}{1/2} \Rightarrow 17.7 = t - 1983.5 \Rightarrow t = 2001$$

Shifting of the Origin. While shifting the origin of a given trend line k units from the previous origin, we substitute $x + k$ or $x - k$ in the given trend line, for x depending upon whether the new origin is forward or backward of the previous origin and then find the trend line with new origin.

Thus in shifting the origin of a given linear trend the only change that take place is the change in the y-intercept. If we are to shift the origin k units forward, then to obtain the value of new y-intercept the previous y-intercept a is to be increased by k times the slope b and if we are to shift the origin k units backward, then to obtain the value of new y-intercept the previous y-intercept a is to be decreased by k times the slope b. That is, the value of the y-intercept of the new trend line would be the trend value at the new origin based on the previous trend line.

Thus, if with previous origin the trend line is $\hat{y} = a + bx$, then with new origin k units from the previous origin, the trend line is

$$\hat{y} = a + b(x \pm k) = (a \pm bk) + bx.$$

***Example* 16.10** *Suppose that the linear trend equation is* $\hat{y} = 110 + 1.5x$, *with origin at* 1980 *and unit of measurement for* x *is one year. Shift the origin at* 1985.

Solution. The linear trend equation is

$$\hat{y} = 110 + 1.5x \qquad \text{with origin at the year 1980}$$

For shifting the origin at 1985, replace x by $(x + 5)$

$$\begin{aligned}\hat{y} &= 110 + 1.5(x + 5)\\ &= 110 + 1.5x + 7.5\\ &= 117.5 + 1.5x \qquad \text{with origin at the year 1985}\end{aligned}$$

Secular Trend — Nonlinear : Many times a straight line will not describe accurately the long-term movement of a time series. In such situations by a careful look at the graph of a time series we might detect some curvature and decide to fit a curve instead of a straight line.

Second degree curve (Parabola). This curve is useful to describe the trend in a time series where change in the amount of change is constant per unit time. The equation of the quadratic (parabolic) trend is

$$\hat{y} = a + bx + cx^2$$

The method of least squares given the normal equations as

$$\begin{aligned}\sum y &= na + b\sum x + c\sum x^2\\ \sum xy &= a\sum x + b\sum x^2 + c\sum x^3\\ \sum x^2 y &= a\sum x^2 + b\sum x^3 + c\sum x^4\end{aligned}$$

However, if $\sum x = 0 = \sum x^3$, then the usual normal equations reduce to

$$\begin{aligned}\sum y &= na + c\sum x^2\\ \sum xy &= b\sum x^2\\ \sum x^2 y &= a\sum x^2 + c\sum x^4\end{aligned}$$

which give the values of a, b *and* c as

$$c = \frac{n\sum x^2 y - (\sum x^2)(\sum y)}{n\sum x^4 - (\sum x^2)^2}$$

$$a = \frac{\sum y - c\sum x^2}{n}$$

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

***Example* 16.11** *Given the following time series.*

Year	1931	1933	1935	1937	1939	1941	1943	1945
Price index	96	87	91	102	108	139	307	289

(*i*) *Fit a second-degree curve (parabola) taking the origin at* 1938.

(ii) *Find the trend values.*

(iii) *What would have been the equation of parabola if origin were at 1933.*

Solution. (i) We have $\bar{t} = (1931 + 1945)/2 = 1938$. Let $x = t - \bar{t} = t - 1938$

Year t	Price index y	Coded year $x = t - 1938$	x^2	x^4	xy	x^2y	Trend value $\hat{y}$
1931	96	−7	49	2401	−672	4704	100.3
1933	87	−5	25	625	−435	2175	83.0
1935	91	−3	9	81	−273	819	81.8
1937	102	−1	1	1	−102	102	96.7
1939	108	1	1	1	108	108	127.7
1941	139	3	9	81	417	1251	174.7
1943	307	5	25	625	1535	7675	237.8
1945	289	7	49	2401	2023	14161	317.0
Sum	1219	0	168	6216	2601	30995	1219.0

The quadratic trend is

$$\hat{y} = a + bx + cx^2$$

Since $\sum x = 0 = \sum x^3$, the least squares estimates a, b and c are

$$c = \frac{n\sum x^2 y - (\sum x^2)(\sum y)}{n\sum x^4 - (\sum x^2)^2} = \frac{8(30995) - (168)(1219)}{8(6216) - (168)^2} = 2.01$$

$$a = \frac{\sum y - c\sum x^2}{n} = \frac{1219 - (2.01)(168)}{8} = 110.2$$

$$b = \frac{\sum xy}{\sum x^2} = \frac{2601}{168} = 15.48$$

The best fitted curve is

$$\hat{y} = 110.2 + 15.48x + 2.01x^2 \qquad \text{with origin at the year 1938}$$

(ii) For different values of x, the trend values are obtained as shown in the table.

(iii) For shifting the origin at 1933, replace x by $(x - 5)$

$$\begin{aligned}\hat{y} &= 110.2 + 15.48(x - 5) + 2.01(x - 5)^2 \\ &= 110.2 + 15.48(x - 5) + 2.01(x^2 - 10x + 25) \\ &= 110.2 + 15.48x - 77.4 + 2.01x^2 - 20.1x + 50.25 \\ &= 83.05 - 4.62x + 2.01x^2 \qquad \text{with origin at the year 1933}\end{aligned}$$

***Example* 16.12** *Given the following time series.*

Year	1931	1933	1935	1937	1939	1941	1943	1945
Price index	96	87	91	102	108	139	307	289

(i) *Fit a straight line taking the origin at 1938.*

(ii) *Fit a second-degree curve (parabola) taking the origin at 1938.*

(iii) *Which is the better fitted trend.*

Solution. *(i)* We have $\bar{t} = (1931 + 1945)/2 = 1938$. Let $x = t - \bar{t} = t - 1938$

Year t	Price index y	Coded year $x = t - 1938$	x^2	x^4	xy	$x^2 y$	y^2
1931	96	−7	49	2401	−672	7404	9216
1933	87	−5	25	625	−435	2175	7569
1935	91	−3	9	81	−273	819	8281
1937	102	−1	1	1	−102	102	10404
1939	108	1	1	1	108	108	11664
1941	139	3	9	81	417	1251	19321
1943	307	5	25	625	1535	7675	94249
1945	289	7	49	2401	2023	14161	83521
Sum	1219	0	168	6216	2601	30995	244225

The linear trend is

$$\hat{y} = a + bx$$

Since $\sum x = 0$, the least squares estimates a and b are

$$a = \frac{\sum y}{n} = \frac{1219}{8} = 152.38$$

$$b = \frac{\sum xy}{\sum x^2} = \frac{2601}{168} = 15.48$$

The best fitted line is

$$\hat{y} = 152.38 + 15.48\,x \qquad \text{with origin at the year 1938}$$

The sum of squares of residuals is

$$\sum e^2 = \sum y^2 - a\sum y - b\sum xy$$

$$= 244225 - 152.38\,(1219) - 15.48\,(2601) = 18210.3$$

The quadratic trend is

$$\hat{y} = a + bx + cx^2$$

Since $\sum x = 0 = \sum x^3$, the least squares estimates a, b and c are

$$c = \frac{n\sum x^2 y - (\sum x^2)(\sum y)}{n\sum x^4 - (\sum x^2)^2} = \frac{8(30995) - (168)(1219)}{8(6216) - (168)^2} = 2.01$$

$$a = \frac{\sum y - c\sum x^2}{n} = \frac{1219 - (2.01)(168)}{8} = 110.2$$

$$b = \frac{\sum xy}{\sum x^2} = \frac{2601}{168} = 15.48$$

The best fitted curve is

$$\therefore \hat{y} = 110.2 + 15.48\,x + 2.01\,x^2 \qquad \text{with origin at the year 1938}$$

The sum of squares of residuals is

$$\sum e^2 = \sum y^2 - a\sum y - b\sum xy - c\sum x^2 y$$
$$= 244225 - 110.2(1219) - 15.48(2601) - 2.01(30995)$$
$$= 7327.77$$

(iii) Since the sum of squares of the residuals for quadratic trend is smaller than the sum of squares of the residuals for linear trend, therefore quadratic trend is better fitting trend.

Merits.

(i) The method of least squares gives the most satisfactory measurement of the secular trend in a time series, when the distribution of the deviations is approximately normal.

(ii) The least squares estimates are unbiased estimates of the parameters.

(iii) The superiority of this method lies in that the computations needed to determine the linear, exponential or quadratic trend have been reduced to formulae.

Demerits.

(i) The method of least squares gives too much weight to extremely large deviations form the trend.

(ii) The least squares line is the best only for the period to which it has reference.

(iii) The elimination or addition for a few more time periods may change its position.

(iv) The only real criterion for the selection of a method of measuring trend is the judgement as to how well the trend line follows the general movement of the time series.

Uses of Secular Trend.

(i) The secular trend may be used either in determining how a time series has grown in the past or in making a forecast.

(ii) The trend line is used to adjust a series to eliminate the effect of the secular trend in order to isolate non-trend fluctuations.

Exercise 16.1

1. *(a)* What is meant by a time series? What are different movements that may be present in a time series.? Describe each of them carefully.

(b) Explain the difference between histogram and historigram.

(c) Describe the following terms:

(i) Secular trend. *(ii)* Seasonal variations

(iii) Cyclical fluctuations. *(iv)* Irregular movements.

2. *(a)* Describe various methods of measuring secular trend in a time series. Discuss the merits and demerits of the methods of smoothing the data.

(b) Plot the original time series to obtain a historigram.

Year	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992
Value	50.0	36.5	43.0	44.5	38.9	38.1	32.6	38.7	41.7	41.1	33.8

Draw a free-hand trend of the following data on the same graph paper:

3. (*a*) What do you understand by the method of semi-averages utilized for smoothing of a time series. Give an example?

(*b*) The following table shows the property damaged by road accidents in Punjab for the years 1972 to 1982.

Year	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981	1982
Property damaged	213	203	238	392	507	441	649	473	342	365	330

Using the method of semi-averages, find the linear trend.
($y' = 270.2 + 20.2\,x$ with origin at 1972)

(*c*) The following table gives the number of books (in 000's) sold at a book stall for the year 1970 to 1981.

Year	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981
Number of books(000)	15	18	17	42	38	40	25	20	20	16	19	17

Using semi-average method, find the trend line. Compute the trend values.
($y' = 32.01 - 1.47\,x$ with origin at 1970)

4. (*a*) What are moving averages? How is a time series smoothed by moving average method? Give an example.

(*b*) Draw a historigram of the following time series. Determine a trend line by a simple moving averages of 5-year from the following data:

Year	1921	1922	1923	1924	1925	1926	1927	1928	1929	1930
Value	102	108	130	140	158	180	196	210	220	230

(127.6, 143.2, 160.8, 176.8, 192.8, 207.2)

5. (*a*) Calculate 7-day moving averages for the following record of attendances:

Week	Sun	Mon	Tues	Wed	Thurs	Fri	Sat
1	24	50	30	48	54	55	62
2	28	52	41	42	50	41	42

Plot the given data and moving averages on the same graph.
(46.14, 46.71, 47.00, 48.57, 47.71, 47.14, 45.14, 42.29)

(*b*) The following table shows the United States average monthly production of bituminous coal in millions of short tons for the years 1981–91. Construct (*i*) 4-year moving averages (*ii*) 4-year centred moving averages

Year	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991
Production	50.0	36.5	43.0	44.5	38.9	38.1	38.7	32.6	41.1	41.7	33.8

{ (*i*) 43.5, 40.7, 41.1, 40.0, 37.1, 37.6 38.6 37.3 (*ii*) 42.1, 40.9, 40.6 38.6, 37.4, 38.1 38.0}

6. (*a*) Compute 4-month centred moving averages from the following:

Month	Jan	Feb	Mar	April	May	June	July	Aug	Sept	Oct
Value	23	26	28	30	31	35	37	32	34	38

(27.75, 29.88, 32.12, 33.50, 34.12, 34.88)

(b) Find 4-quarter centred moving averages for the following data.

Year	Quarter I	Quarter II	Quarter III	Quarter IV
1948	71	72	78	84
1949	72	69	75	79
1950	73	80	85	86

Plot the original data and the trend values on a graph.
(76.38, 76.12, 75.38, 74.38, 73.88, 75.38, 78.0, 80.12)

7. (a) Explain the method of least squares utilized for finding a secular trend in a time series.

(b) Given the following time series.

Year	1968	1969	1970	1971	1972	1973	1974	1975	1976
Value	3	4	6	8	7	7	10	13	12

Determine the linear trend using least squares method by taking the origin at the beginning period of the time period. Estimate the value for the year 1978.
($\hat{y} = 3.11 + 1.17\,x$ with origin at 1968; 14.78)

8. (a) The following time series shows the number of road accidents in Punjab for the years 1977 to 1987.

Year	1977	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987
Number of accidents	2493	2639	2669	3038	3745	4079	4683	4845	4505	4793	4728

(*i*) Use the method of least squares to fit a straight line taking the origin at the middle of the time period.

(*ii*) Find the trend values for this time series.

(*iii*) Estimate the number of road accidents in 1989.

{ (*i*) $\hat{y} = 3837.91 + 271.37\,x$ with origin at 1982; (*ii*) 2481.05, 2752.42, 3023.79, 3295.16, 3566.54, 3837.91, 4109.28, 4380.65, 4652.02, 4923.39, 5194.76; (*iii*) 5737.50 }

(b) Fit a straight line $\hat{y} = a + b\,x$ from the following results, for the years 1985—95 (both inclusive). Find out the trend values of y as well.

$$\sum x = 0, \quad \sum y = 438.9, \quad \sum x^2 = 110, \quad \sum xy = -84.4$$

($\hat{y} = 39.9 - 0.77\,x$ with origin at 1990; 43.75, 42.98, 42.21, 41.44, 40.67, 39.90, 39.13, 38.36, 37.59, 36.82, 36.05)

9. (a) Fit a straight line to the following data taking the origin at the middle of the time period and unit of measurement as 1/ 2 year and find the trend values:

Year	1980	1981	1982	1983	1984	1985
Production (000)	10	12	8	10	14	16

($\hat{y} = 11.66 + 0.54\,x$ with origin at the middle of 1982 and 1983 and unit of measurement as 1/ 2 year; 8.96, 10.04, 11.12, 12.20, 13.28, 14.36) (*b*) Fit a straight line to following data. Plot on the same graph paper the actual and trend values.

Year	1970	1971	1972	1973	1974	1975	1976	1977
Value	12	15	18	25	20	22	26	30

($\hat{y} = 21 + 1.12\,x$ with origin at January 1, 1974; 13.16, 15.40, 17.64, 19.88, 22.12, 24.36, 26.60, 28.84)

(*c*) For the following time series, determine the trend by using the method of

(*i*) semi-average,

(*ii*) 3-year moving average,

(*iii*) least-squares for fitting a straight line.

Year	1968	1969	1970	1971	1972	1973	1974	1975	1976
Value	2	4	6	8	7	6	8	10	12

Which of the trend do you prefer, and why?
{ (*i*) $\hat{y} = 3.8 + 0.8\,x$ with origin at 1968, (*ii*) 4.0, 6.0, 7.0, 7.0, 7.0, 8.0, 10.0; (*iii*) $\hat{y} = 7 + x$ with origin at 1972; Least squares trend }

10. (*a*) Fit a second degree curve to the following time series and find the trend values.

Year	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990
Production	23.2	31.4	39.8	50.2	62.9	76.0	92.0	105.7	122.8	131.7	151.1

($\hat{y} = 76.64 + 13.0\,x + 0.3974\,x^2$ with origin at 1985 and unit of x as 1 year; 21.6, 31.0, 41.2 52.2, 64.0, 76.6, 90.0, 104.2, 119.2, 135.0, 151.6)

(*b*) Fit a quadratic curve to the following time series.

Year	1924	1927	1930	1933	1936	1939	1942
Index of coal price	187	142	133	129	136	169	279

Use your results to estimate the values of the index for 1935.
($\hat{y} = 119.9 + 11.89\,x + 11.99\,x^2$ with origin at 1933 and unit of x as 3 years; $\hat{y} = 133.16$)

11. (*a*) Fit a second degree curve to the following time series

Year	1980	1981	1982	1983	1984	1985	1986	1987
Quantum Index	100	87	96	102	139	210	289	307

($\hat{y} = 131.8 + 16.89\,x + 1.64\,x^2$ with origin at the January 1, 1984 and unit of x as $1/2$ year)

(*b*) Fit a quadratic curve (parabola) to the following data. Compute the trend values.

Year	1931	1933	1935	1937	1939	1941	1943	1945
Price index	96	87	91	102	108	139	307	289

($\hat{y} = 110.16 + 15.48\,x + 2.01\,x^2$ with origin at 1938 and unit of x as one year; 100.3, 83.0, 81.8, 96.7, 127.7, 174.7, 237.7, 317.0)

12. (a) The following are the annual profits in thousands of rupees in certain business:

Year	1977	1978	1979	1980	1981	1982	1983
Profit	88	101	105	91	113	120	132

(*i*) Fit a linear trend by the method of least-squares and make an estimate of the profits in 1985.

(*ii*) Fit a parabolic trend.

(*iii*) Determine which is the better fitting trend.

{ (*i*) $\hat{y} = 107.14 + 6.36\,x$ with origin at 1980 and unit of x as 1 year; $\hat{y} = 139.04$, (*ii*) $\hat{y} = 103.24 + 6.36\,x + 0.976\,x^2$ with origin at 1980 and unit of x as 1 year }

(b) Fit a quadratic trend from the following results, for the years 1985—95 (both inclusive).

$$\sum x = \sum x^3 = 0, \quad \sum x^2 = 110, \quad \sum x^4 = 1958,$$

$$\sum y = 410, \quad \sum xy = 601, \quad \sum x^2 y = 4587$$

Find out the trend values of y as well. Estimate the trend value for the year 1996. ($\hat{y} = 31.6 + 5.46\,x + 0.568\,x^2$ with origin at 1990 and unit of x as 1 year; 18.50, 18.85, 20.33, 22.95, 26.71, 31.60, 37.63, 44.79, 53.09, 62.53, 73.10; 84.81)

13. (a) Suppose that the linear trend equation is $\hat{y} = 50 + 2\,x$, with origin at 1983 and unit of measurement for x is one year. Shift the origin at 1980.

($\hat{y} = 44 + 2\,x$, with origin at the year 1980)

(b) If the linear trend in the data for the years 1960 to 1965 both inclusive with origin at the middle of 1962 and 1963 is $\hat{y} = 1306.667 + 73.428\,x$, the unit of x being one year, then determine the trend line with origin at 1960 and hence determine the trend values.

($\hat{y} = 1123.097 + 73.428\,x$; 1123.097, 1196.525, 1269.953, 1343.381, 1416.809, 1490.237)

(c) The parabolic trend equation for the projects of a company (in thousand rupees) is $\hat{y} = 10.4 + 0.6\,x + 0.7\,x^2$, with origin at 1980 and unit of measurement for x is one year. Shift the origin to 1975.

($\hat{y} = 24.9 - 6.4\,x + 0.7\,x^2$)

◆

Exercise 16.2

Objective Questions

1. With which particular characteristic movement of a time series would you mainly associate each of the following:

(*i*) Increased demand for foot-wears before Eid. (*S*)

(*ii*) The decline in death rate due to advancement in science. (*T*)

(*iii*) A steel strike, delaying production for a week. (*I*)

(*iv*)	Rise in the prices of certain consumer goods due to tax increase in the annual budget.	(*S*)
(*v*)	An era of prosperity in a business.	(*C*)
(*vi*)	The festival sale.	(*S*)
(*vii*)	The production of sugar recorded for 1986, 1987, · · ·, 1992.	(*T*)
(*viii*)	The weekly statement of the sale of pens.	(*S*)
(*ix*)	A fire in a factory delaying production for 3 weeks.	(*I*)
(*x*)	An after Eid sale in a departmental store	(*S*)
(*xi*)	A need for increased wheat production due to a constant increase in population.	(*T*)
(*xii*)	The monthly rainfall in inches in a city over a 5-year period.	(*S*)
(*xiii*)	A recession in a business.	(*C*)
(*xiv*)	An increase in employment during summer months.	(*S*)
(*xv*)	A continually increasing demand for smaller automobiles.	(*T*)

2. State whether the following statements are true or false.

(*i*)	The graph of a time series is called histogram.	(*false*)
(*ii*)	Secular trend is a short term variation.	(*false*)
(*iii*)	Seasonal variations are regular in nature.	(*true*)
(*iv*)	Secular trend has booms and depressions.	(*false*)
(*v*)	Irregular variations are not regular in nature.	(*true*)
(*iv*)	The increase in the school fee in private schools is an irregular variation.	(*false*)
(*vii*)	The increase in the number of patients in the hospitals is like secular trend in a time series.	(*true*)
(*viii*)	The increase in the number of patients of heat stroke in summer is like secular trend in the time series.	(*false*)
(*ix*)	The secular trend is measured by a straight line when a time series has an upward trend.	(*false*)
(*x*)	The secular trend is measured by semi-averages method when trend is linear.	(*true*)
(*xi*)	The straight line is fitted to a time series when the movements in the time series are linear.	(*true*)
(*xii*)	In the measurement of secular trend by the method of least squares, the number of years must be odd.	(*false*)
(*xiii*)	For a least squares linear trend $\hat{y} = a + b\,x$, the b is a variable and $\hat{y}$ is the slope of the line.	(*false*)
(*xiv*)	Seasonal variations can be measured only when the time series contains yearly values.	(*false*)

3. ***Multiple choice :*** Select a suitable answer:

(*i*) The graph of a time series is called

(*a*) histogram (*b*) polygon (*c*) straight line (*d*) historigram

(*ii*) The secular trend is measured by the method of semi-averages when:

(*a*) time series contains yearly values (*b*) trend is linear

(*c*) time series contains odd number of values (*d*) none of them

(*iii*) In the measurement of secular trend the moving averages

(*a*) give the trend in a straight line (*b*) measure the seasonal variations

(*c*) smooth out a time series (*d*) none of them

(*iv*) For a least squares linear trend $\hat{y} = a + bx$, the b is the:

(*a*) variable (*b*) intercept (*c*) trend (*d*) slope

(*v*) For a least squares linear trend $\hat{y} = a + bx$,

(*a*) $\sum y < \sum \hat{y}$ (*b*) $\sum \hat{y} = 0$ (*c*) $\sum y = \sum \hat{y}$ (*d*) none of them

(*vi*) For a least squares linear trend $\hat{y} = a + bx$, the $\sum(y - \hat{y})^2 = 0$ when

(*a*) all the *y*-values lie on the line. (*b*) all the *y*-values are positive.

(*c*) all the *y*-values lie above the line. (*d*) none of them.

{ (*i*) *d*, (*ii*) *b*, (*iii*) *c*, (*iv*) *d*, (*v*) *c*, (*vi*) *a* }