2·3·1. Mathematical Models for Time Series. The following are the two models commonly used for the decomposition of a time series into its components.

***(i) Decomposition by Additive Hypothesis* (or Additive Model).** According to the *additive model*, a time series can be expressed as

$$y_t = T_t + S_t + C_t + R_t \quad \cdots (2·1)$$

where y_t is the time-series value at time t, T_t represents the trend value, S_t, C_t and R_t represent the seasonal, cyclic and random fluctuations at time t. Obviously, the term S_t will not appear in a series of annual data. The additive model implicitly implies that seasonal forces (in different years), cyclical forces (in different cycles) and irregular forces (in different long term period) operate with equal absolute effect irrespective of the trend value. As such C_t (and S_t) will have positive or negative values, according as whether we are in an above normal or below normal phase of the cycle (and year) and the total of positive and negative values for any cycle (and any year) will be zero. R_t will also have positive or negative value and in the long-term (ΣR_t) will be zero. Occasionally, there may be a few isolated occurrences of extreme R_t of episodic nature.

The additive model assumes that all the four components of the time series operate independently of each other so that none of these components has any effect on the remaining three.

(ii) Decomposition by Multiplicative Hypothesis (or Multiplicative Model). On the other hand, if we have reasons to assume that the various components in a time series *operate proportionately* to the general level of the series, the traditional or classical multiplicative model is appropriate. According to the multiplicative model,

$$y_t = T_t \times S_t \times C_t \times R_t \quad \ldots (2·2)$$

where S_t, C_t and R_t, instead of assuming positive and negative value, are indices fluctuating above or below unity and the geometric means of S_t in a year, C_t in a cycle and R_t in a long-term period are unity. In a time series with both positive end negative values, the multiplicative model can not be applied unless the time series is translated by adding a suitable positive value. It may be pointed out that the multiplicative decomposition of a time series is same as the additive decomposition of logarithmic values of the original time series, *i.e.*,

$$\log y_t = \log T_t + \log S_t + \log C_t + \log R_t$$

In practice, most of the series relating to economic data conform to multiplicative model

Remarks **1.** *Limitations of the Hypothesis of Decomposition of a Time Series.* Hypothesis of decomposition presupposes that the trend and periodic components are determined by separate forces acting independently so that simple aggregation of the components could constitute the series. But in reality, it is possible that this year's value of the series will depend to some extent on last year's value so that trend and periodic movement will get inextricably mixed up and no meaningful separation of them will be possible. In such a case any variations of this year may affect the whole future course of the series and no meaningful separation of trend and periodic components will be possible.

2. Mixed Models. In addition to the additive and multiplicative models discussed above, the components in a time series may be combined in a large number of other ways. The different models, defined under different assumptions will yield different results. Some of the mixed models resulting from different combinations of additive and multiplicative models are given below :

$$\left.\begin{aligned} y_t &= T_t C_t + S_t R_t \\ y_t &= T_t + S_t C_t R_t \\ y_t &= T_t + S_t + C_t R_t \end{aligned}\right\} \quad \ldots (2·2a)$$

3. The model (2·1) or (2·2) can be used to obtain a measure of one or more of the components by elimination, *viz.*, subtraction or division. For example, if trend component (T_t) is known, then using multiplication model, it can be isolated from the given time series to give :

$$S_t \times C_t \times R_t = \frac{y_t}{T_t} = \frac{\text{Original values}}{\text{Trend values}} \qquad \text{... (2·2b)}$$

Thus, for the *annual data*, for which the seasonal component S_t is not there, we have

$$y_t = T_t \times C_t \times R_t \quad \Rightarrow \quad C_t \times R_t = \frac{y_t}{T_t} \qquad \text{... (2·2c)}$$

2·3·2. Uses of Time Series. The time series analysis is of greater importance not only to businessman or an economist but also to people working in various disciplines in natural, social and physical sciences. Some of its uses are enumerated below :

1. It enables us to *study the past behaviour of the phenomenon* under consideration, *i.e.*, to determine the type and nature of the variations in the data.

2. The segregation and study of the various components is of paramount importance to a businessman in the planning of future operations and in the formulation of executive and policy decisions.

3. It helps to compare the actual current performance of accomplishments with the expected ones (on the basis of the past performances) and analyse the causes of such variations, if any.

4. It enables us to predict or estimate or forcast the behaviour of the phenomenon in future which is very essential for business planning.

5. It helps us to compare the changes in the values of different phenomenon at different times or places, etc.

In the following sections we shall discuss various techniques for the measurement of different components

2·4. MEASUREMENT OF TREND

Trend can be studied and/or measured by the following methods :

(*i*) Graphic (or *Free-hand* Curve Fitting) Method,

(*ii*) Method of *Semi-Averages*,

(*iii*) Method of Curve Fitting by *Principle of Least Squares*, and

(*iv*) Method of *Moving Averages*.

We shall now discuss each of these methods in detail.

2·4·1. Graphic Method. A free-hand smooth curve obtained on plotting the values y_t against 't' enables us to form an idea about the general 'trend' of the series. Smoothing of the curve eliminates other components, *viz.* regular and irregular fluctuations.

This method does not involve any complex mathematical techniques and can be used to describe all types of trend, linear and non-linear. Thus, simplicity and flexibility are strong points of this method. Its main drawbacks are :

(*i*) The method is very subjective, *i.e.*, the bias of the person handling the data plays a very important role and as such different trend curves will be obtained by different persons for the same set of data. As such 'trend by inspection' should be attempted only by skilled and experienced statisticians and this limits the utility and popularity of the method.

(*ii*) It does not enable us to measure trend.

2·4·2. Method of Semi-averages. In this method, the whole data is divided into two parts with respect to time, *e.g.*, if we are given y_t for t from 1991-2002, *i.e.*, over a period of 12 years, the two equal parts will be the data from 1991 to 1996 and 1997 to 2002. In case of odd number of years the two parts are obtained by omitting the value corresponding to the middle year, *e.g.*, for the data from 1991-2001, the value corresponding to middle year, *viz.* 1996 being omitted. Next we compute the arithmetic mean for each part and plot these two averages (means) against the mid-values of the respective time-periods covered by each part. The line obtained on joining these two points is the required trend line and may be extended both ways to estimate intermediate or future values.

Remark. For even number of years like 8, 12, 16, etc. the centering of average of each part would create problems, *e.g.*, from the data 1997-2002 (n = 12), let the two averages be $\overline{X}_1$, (say) for period 1991-1996 and $\overline{X}_2$ (say), for the period 1991-2002. Here $\overline{X}_1$, will be plotted against the mean of two mid-values, *viz*. 1993 and 1994 for the period 1991-1996, *i.e.*, against 1st July 1993. Similarly, for the period 1997-2002.

Merits **1.** As compared with graphic method, the obvious advantage of this method is its objectivity in the sense that everyone who applies it would get the same results. Moreover, we can also estimate the trend values.

2. It is readily comprehensible as compared to the 'method of least squares' or the 'moving average method'.

Limitations. This method assumes linear relationship between the plotted points — which may not exist. Moreover, the limitations of arithmetic mean as an average also stand in its way.

Example 2·1. Fit a trend line to the following data by the method of semi-averages :

Year	*Bank Clearances (Rs. Crores)*	*Year*	*Bank Clearances (Rs. Crores)*
1992	53	*1999*	87
1993	79	*2000*	79
1994	76	*2001*	104
1995	66	*2002*	97
1996	69	*2003*	92
1997	94	*2004*	101
1998	105		

Solution. Here since n = 13 (odd), the two parts would consist of 1992 to 1997 and 1999 to 2004, the year 1998 being omitted.

$\overline{X}_1$ = Average sales for first part

$$= \frac{437}{6} = 72{\cdot}83 \text{ (Rs. crores)}$$

$\overline{X}_2$ = Average sales for second part

$$= \frac{560}{6} = 93{\cdot}33 \text{ (Rs. crores)}$$

As explained in Remark to § 2·4·2, $\overline{X}_1$ and $\overline{X}_2$ will be plotted against 1st July 1994 and 1st July 2001 respectively, as given in *Fig. 2·1*

Joining the points A [1994, $\overline{X}_1$] and *B* [2001, $\overline{X}_2$], we get the trend line [*Fig. 2·1*].

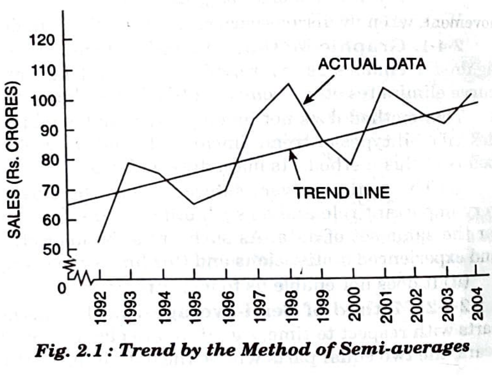

Fig. 2.1 : Trend by the Method of Semi-averages

2·4·3. Method of Curve Fitting by Principle of Least Squares. The principle of least squares is the most popular and widely used method of fitting mathematical functions to a given set of data. The method yields very correct results if sufficiently good appraisal of the form of the function to the fitted is obtained either by a scrutiny of the graphical plot of the values over time or by a theoretical understanding of the mechanism of the variable change. An examination of the plotted data often provides an adequate basis for deciding upon the type of trend to use. Apart from the usual arithmetic scales, semi-logarithmic or doubly-logarithmic scales may be used for the graphical representation of the data. The various types of curves that may be used to describe the given data in practice are :

(If y_t is the value of the variable corresponding to time t)

(*i*) *A straight line :* $y_t = a + bt$

(*ii*) *Second degree parabola :* $y_t = a + bt + ct^2$

(*iii*) *kth-degree polynomial :* $y_t = a_0 + a_1 t + a_2 t^2 + \dots + a_k t^k$

(*iv*) *Exponential curves :* $y_t = a\, b^t$

$$\Rightarrow \quad \log y_t = \log a + t \log b = A + Bt, \text{ (say).}$$

(*v*) *Second degree curve fitted to logarithms :*

$$y_t = a\, b^t\, c^{t^2}$$

$$\Rightarrow \quad \log y_t = \log a + t \log b + t^2 \log c = A + B\,t + C\,t^2, \text{ (say).}$$

(*vi*) *Growth curves :*

(*a*) $y_t = a + b\,c^t$ (Modified Exponential Curve)

(*b*) $y_t = ab^{c^t}$ (Gompertz curve)

$$\Rightarrow \quad \log y_t = \log a + c^t \log b = A + Bc^t \text{ (say)}$$

(c) $y_t = \dfrac{k}{1 + \exp(a + bt)}$ (*Logistic curve*)

Remark. For deciding about the type of trend to be fitted to a given set of data, the following points may be helpful :

(*i*) When the time series is found to be increasing or decreasing by equal absolute amounts, the straight line trend is used. In this case, the plotting of the data will give a straight line graph.

(*ii*) The logarithmic straight line (exponential curve $y_t = ab^t$) is used as an expression of the secular movement, when the series is increasing or decreasing by a constant percentage rather than a constant absolute amount. In this case, the data plotted on a semi-logarithmeic scale will give a straight line graph.

(*iii*) Second degree curve fitted to logarithms may be tried for trend fitting if the data plotted on a semi-logarithmic scale is not a straight lone graph but shows curvature, being concave either upward or downward.

Alternatively, approximations about the type of the curve to be fitted can be made by use of the following theorem based on finite differences :

"The nth differences $\Delta^n y_t$, $\Delta^n (\log y_t)$, $\Delta^n (1/y_t)$ of any general polynomial y_t of nth degree in t is constant and (n + 1)th differences are equal to zero."

For further guidelines, the following statistical tests based on the calculus of finite differences may be applied.

We know that for a polynomial y_t of nth degree in t,

$$\left.\begin{aligned} \Delta^r y_t &= \text{constant}, & r &= n \\ &= 0\,, & r &> n \end{aligned}\right\}$$

where Δ is the difference operator given by $\Delta y_t = y_{t+h} - y_t$, h being the interval of differencing and $\Delta^r y_t$ is the rth differences of y_t.

1. If Δy_t = constant, use straight line trend.
2. If $\Delta^2 y_t$ = constant, use a second degree (parabolic) trend.
3. If $\Delta(\log y_t)$ = constant, use exponential trend curve.
4. If $\Delta^2(\log y_t)$ = constant, use second degree curve fitted to logarithms.
5. The growth curves, *viz.*, modified exponential, Gompertz and Logistic curves can be approximated by the constancy of the ratios :

$$\frac{\Delta y_t}{\Delta y_{t-1}}, \quad \left\{\frac{\Delta(\log y_t)}{\Delta(\log y_{t-1})}\right\}, \quad \left\{\frac{\Delta(1/y_t)}{\Delta(1/y_{t-1})}\right\}$$

respectively, for all possible values of t.

The following tests may also be used :

6. If Δy_t tends to decrease by a constant percentage, use modified exponential curve.
7. if Δy_t resembles a skewed frequency curve, use a Gompertz curve or Logistic curve.

Fitting of Straight Line by Least Squares Method. Let the straight line trend between the given time-series values (y_t) and time t be given by the equation :

$$y_t = a + bt \qquad \dots (2\cdot3)$$

Principle of least squares consists in minimizing the sum of squares of the deviations between the given values of y_t and their estimates given by (2·3). In other words, we have to find a and b such that for given values of y_t corresponding to n different values of t,

$$E = \sum_t (y_t - a - bt)^2$$

is minimum. For a maxima or minima of E, for variations in a and b, we should have

$$\left.\begin{aligned} \frac{\partial E}{\partial a} &= 0 = -2\Sigma(y_t - a - bt) \\ \frac{\partial E}{\partial b} &= 0 = -2\Sigma t(y_t - a - bt) \end{aligned}\right\} \Rightarrow \left.\begin{aligned} \Sigma y_t &= na + b\Sigma t \\ \Sigma t_t &= a\Sigma t + b\Sigma t^2 \end{aligned}\right\}, \qquad \dots (2\cdot4)$$

which are the normal equations for estimating a and b.

The values of $\Sigma y_t, \Sigma t, \Sigma t^2$ are obtained from the given data and the equations (2·4) can now be solved for a and b. With these values of a and b, the line (2·3) gives the desired trend line.

Remark. The solution of normal equations (2·4) provides a minima of E. The proof is given below : The necessary and sufficient condition fo a minima of E for variations in a and b are :

$$(i)\ \frac{\partial E}{\partial a} = 0, \frac{\partial E}{\partial b} = 0 \ \dots (*) \quad \text{and} \quad (ii)\ \Delta = \begin{vmatrix} \frac{\partial^2 E}{\partial a^2} & \frac{\partial^2 E}{\partial a\, \partial b} \\ \frac{\partial^2 E}{\partial b\, \partial a} & \frac{\partial^2 E}{\partial b^2} \end{vmatrix} > 0 \text{ and } \frac{\partial^2 S}{\partial a^2} > 0 \qquad \dots (**)$$

From (2·4), we get

$$\frac{\partial^2 S}{\partial a^2} = 2n > 0 \ ; \quad \frac{\partial^2 S}{\partial b^2} = 2\Sigma t^2 > 0 \ ; \quad \frac{\partial^2 S}{\partial a\, \partial b} = \frac{\partial^2 E}{\partial b\, \partial a} = 2\,\Sigma t$$

$$\therefore \quad \Delta = \begin{vmatrix} 2n & 2\Sigma t \\ 2\Sigma t & 2\Sigma t^2 \end{vmatrix} = 4[n\Sigma t^2 - (\Sigma t)^2]$$

$$= 4n^2\left[\frac{\Sigma t^2}{n} - \left(\frac{\Sigma t}{n}\right)^2\right] = 4n^2 \operatorname{Var}(t) > 0$$

Hence, the solution of the least square equations (2·4), satisfies (*) and (**) and, therefore, provides a minima of E.

Fitting of Second Degree (Parabolic) Trend. Let the second degree parabolic trend curve be :

$$y_t = a + bt + ct^2 \qquad \dots (2\cdot5)$$

Proceeding similarly as in the case of a straight line, the normal equations for estimating a, b and c are given by :

$$\left.\begin{aligned} \sum y_t &= na + b\sum t + c\sum t^2 \\ \sum t\, y_t &= a\sum t + b\sum t^2 + c\sum t^3 \\ \sum t^2 y_t &= a\sum t^2 + b\sum t^2 + c\sum t^4 \end{aligned}\right\}, \qquad \dots (2\cdot6)$$

the summation being taken over the values of the time series.

Fitting of Exponential Curve :

$$y_t = a\, b^t \qquad \dots (2\cdot7)$$

$$\log y_t = \log a + t \log b$$

$$\Rightarrow \quad Y = A + Bt \text{ (say)}, \qquad \dots (2\cdot7a)$$

$$\Rightarrow \quad \text{where } Y = \log y_t \;,\; A = \log a \;,\; B = \log b. \qquad \dots (2\cdot7b)$$

(2·7*a*) is a straight line in t and Y and thus the normal equations for estimating A and B are

$$\left.\begin{aligned} \Sigma Y &= nA + B\Sigma t, \\ \Sigma tY &= A\Sigma t + B\Sigma t^2 \end{aligned}\right\} \qquad \dots (2\cdot7c)$$

These equations can be solved for A and B and finally on using (2·7*b*), we get

$$a = \text{antilog } (A) \;;\; b = \text{antilog } (B).$$

Second Degree Curve Fitted to Logarithms. Suppose the trend curve is :

$$Y_t = a\, b^t c^{t^2} \qquad \dots (2\cdot8)$$

Taking logarithms of both sides, we get

$$\log y_t = \log a + t \log b + t^2 \log c$$

$$\Rightarrow \quad Y_t = A + B\,t + C\,t^2 \qquad \dots (2\cdot8a)$$

$$\text{where} \quad Y_t = \log y_t \;;\; A = \log a \;;\; B = \log b \text{ and } C = \log c \qquad \dots (2\cdot8b)$$

Now, (2·8*a*) is a second degree parabolic curve in Y_t and t and can be fitted by the technique already explained. We can finally obtain

$$a = \text{Antilog } (A) \;;\; b = \text{Antilog } (B) \text{ and } c = \text{Antilog } (C).$$

With these values of a, b and c, the curve (2·8) becomes the best second degree curve fitted to logarithms.

Remark. The method of curve fitting by the principle of least squares is used quite often in trend analysis particularly when one is interested in making projections for future times. Obviously, the reliability of the estimated (projected) values primarily depends upon the appropriatness of the form of the mathematical function fitted to the given data. If the function is determined on the ad-hoc basis by the scrutiny of the plotted values, the projections based on it may be valid for the near future while, if the study of physical mechanism of the variable change forms the basis of the selection of function, then there is very little likelihood that the function will change for sufficiently long period and hence in this case reliable long term projections can be made.

Merits and Drawbacks of Trend Fitting by the Principle of Least Squares.

Merits. The method of least squares is the most popular and widely used method of fitting mathematical functions to a given set of observations. It has the following advantages :

1. Because of its mathematical or analytical character, this method completely eliminat the element of subjective judgement or personal bias on the part of the investigator.

2. Unlike the method of moving averages [discussed in § 2·4·5], this method enables us t compute the trend values for all the given time periods in the series.

3. The trend equation can be used to estimate or predict the values of the variable for an period t in future or even in the intermediate periods of the given series and the forecast values are also quite reliable.

4. The curve fitting by the principle of least squares is the only technique which enables us to obtain the rate of growth per annum, for yearly data, if linear trend is fitted.

Drawbacks **1.** The method is quite tedious and time-consuming as compared with other methods. It is rather difficult for a non-mathematical person (layman) to understand and use.

2. The addition of even a single new observation necessitates all calculations to be done afresh.

3. Future predictions or forecasts based on this method are based only on the long term variation, *i.e.*, trend and completely ignore the cyclical, seasonal and irregular fluctuations.

4. The most serious limitation of this method is the determination of the type of the trend curve to be fitted, *viz.*, whether we should fit a linear or a parabolic trend or some other more complicated trend curve.

5. It cannot be used to fit growth curves like Modified Exponential curve, Gompertz curve and Logistic curve, to which most of the economic and business time series data conform.

Example 2·2. *In a certain industry, the production of a certain commodity (in '000 units) during the years 1994—2004 is given in the adjoining table :*

Year	*Production ('000 units)*	*Year*	*Production ('000 units)*
1994	*66·6*	*2000*	*93·2*
1995	*84·9*	*2001*	*111·6*
1996	*88·6*	*2002*	*88·3*
1997	*78·0*	*2003*	*117·0*
1998	*96·8*	*2004*	*115·2*
1999	*105·2*		

(i) *Graph the data.*

(ii) *Obtain the least square line fitting the data and construct the graph of the trend line.*

(iii) *Compute the trend values for the year 1994-2004 and estimate the production of commodity during the years 2005 and 2006, if the present trend continues.*

(iv) *Eliminate the trend.*

Solution. Here $n = 11$, *i.e.*, odd and, therefore, we shift the origin to the middle time period, *viz.*, the year 1999. Let $x = t - 1999$ ...(1)

TABLE 2·1 : COMPUTATION OF TREND LINE

Year (*t*)	*Production ('000 units)* (y_t)	x	$x\,y_t$	x^2	*Trend values ('000 units)* (t_e) $y_e = 95{\cdot}49 + 3{\cdot}95x$
1994	66·6	– 5	– 333·0	25	75·74
1995	84·9	– 4	– 339·6	16	79·69
1996	88·6	– 3	– 265·8	9	83·64
1997	78·0	– 2	– 156·0	4	87·59

1998	96·8	– 1	– 96·8	1	91·54
1999	110·2	0	0	0	95·49
2000	93·2	1	93·2	1	99·44
2001	111·6	2	223·2	4	103·39
2002	88·3	3	264·9	9	107·34
2003	117·0	4	468·0	16	111·29
2004	115·2	5	576·0	25	115·24
Total	1,050·4	0	434·1	110	

Let the least square line of y_t on x be : $y_t = a + bx$ (origin : July 1999) ... (2)

The normal equations for estimating a and b are

$$\Sigma y_t = na + b\,\Sigma x \quad \text{and} \quad \Sigma x.y_t = a\Sigma x + b\Sigma x^2$$

$$\Rightarrow \quad 1050 = 11a \qquad \Big| \qquad \Rightarrow \quad 434{\cdot}1 = 110b$$

$$\Rightarrow \quad a = \frac{1050{\cdot}4}{11} = 95{\cdot}49 \qquad \Big| \qquad \Rightarrow \quad b = \frac{434{\cdot}1}{110} = 3{\cdot}95$$

Hence, the least square line fitting the data is : $y_t = 95{\cdot}49 + 3{\cdot}95x,$ ... (3)

where origin is July 1999 and x unit = 1 year.

Trend values for the years 1994 to 2004 are obtained on putting $x = -5, -4, -3, \ldots, 4, 5$ respectively in (3) and have been tabulated in the last column of the *Table 2·1*.

Estimate for 2005. Taking $x = 2005$ in (1), we get $x = 2005 - 1999 = 6$

Hence the estimate production of the commodity for 2005 is obtained on putting $x = 6$ in (***) and is given by :

$$(\hat{y}_e)_{2005} = 95.49 + 3.95 \times 6 = 119.19 \text{ ('000 units)}.$$

Similarly, $(\hat{y}_e)_{2006} = 95{\cdot}49 + 3{\cdot}95 \times 7 = 123{\cdot}14$ ('000 units)

The graph of the original data and the trend line is given in *Fig. 2·2* :

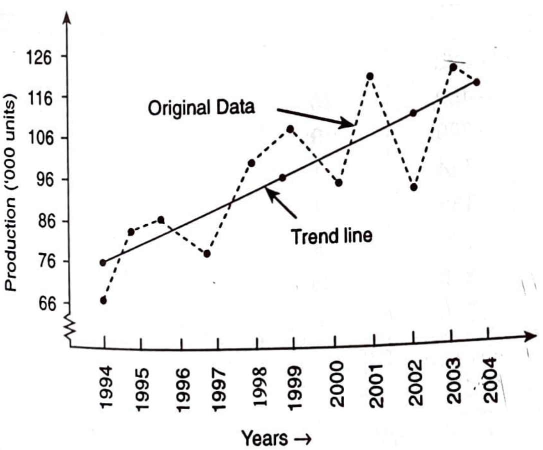

Fig. 2·2.

Assuming multiplicative model, the trend values are eliminated on dividing the given values (y_t) by the corresponding trend values (y_e). However, if we assume the additive model, the trend eliminated values are given by ($y_t - y_e$). The resulting values contain short-term (seasonal and cyclic) variations and irregular variations. Trend eliminated values are given in *Table 2·2*.

TABLE 2·2 : ELIMINATION OF TREND

Year	Trend Eliminated Values Based on *Additive Model* $(y_t - y_e)$	Trend Eliminated Values Based on *Multiplicative Model* $(y_t \div y_e)$
1994	66·6 – 75·74 = – 9·14	66·6/75·74 = 0·880
1995	84·9 – 79·69 = 5·21	84·9/79·69 = 1·065
1996	88·6 – 83·64 = 4·96	1·059
1997	78·0 – 87·59 = – 9·59	0·891
1998	96·8 – 91·54 = 5·26	1·057
1999	110·2 – 95·49 = 14·71	1·154
2000	93·2 – 99·44 = – 6·24	0·937
2001	111·6 – 103·39 = 8·21	1·079
2002	88·3 – 107·34 = – 19·04	0·823
2003	117·0 – 111·29 = 5·71	1·051
2004	115·2 – 115·24 = – 0·04	0·999

Example 2·3. *Fit a straight line trend by the method of least squares to the following data relating to the sales of a leading departmental store. Assuming that the same rate of change continues, what would be predicted earnings for the year 2006 ?*

Year	:	*1997*	*1998*	*1999*	*2000*	*2001*	*2002*	*2003*	*2004*
Sales (Crores Rs.)	:	*76*	*80*	*130*	*144*	*138*	*120*	*174*	*190*

Solution. Here $n = 8$, *i.e.*, even. Hence we shift the origin to the arithmetic mean of the two middle years, *viz.*, 2000 and 2001. We define

$$x = \frac{t - \frac{1}{2}(2000 + 2001)}{\frac{1}{2}\text{(Interval)}} = \frac{t - 2000{\cdot}5}{\frac{1}{2} \times 1} = 2t - 4001 \quad \ldots(1)$$

where x values are in units of six months (half year).

TABLE 2·3 : COMPUTATION OF LINEAR TREND

Year (t)	*Sales (Crores Rs.)* y_t	x	$x\,y_t$	x^2	*Trend values (Crores Rs.)* $y_e = 131{\cdot}5 + 7{\cdot}33x$
1997	76	– 7	– 532	49	80·19
1998	80	– 5	– 400	25	94·85
1998	130	– 3	– 390	9	109·51
2000	144	– 1	– 144	1	124·17
2001	138	1	138	1	138·83
2002	120	3	360	9	153·49
2003	174	5	870	25	168·15
2004	190	7	1330	49	182·81
Total	$\Sigma y_t = 1052$	$\Sigma x = 0$	$\Sigma x\,y_t =$ 1,232	$\Sigma x^2 = 168$	

Let the linear trend equation between y_t and x be :

$$y_t = a + bx, \; x = 2(t - 2000{\cdot}5)$$

Since $\sum x = 0$, the normal equations for estimating a and b are :

$$a = \frac{\sum y_t}{n} = \frac{1052}{8} = 131{\cdot}5\ , \quad b = \frac{\sum x y_t}{\sum x^2} = \frac{1232}{168} = 7{\cdot}33$$

Hence the least square trend line becomes : $y_t = 131{\cdot}5 + 7{\cdot}33x$ ... (3)

where $b = 7{\cdot}33$ units represent half yearly increase in the earnings.

The trend values for the year 1997 to 2004 can now be obtained from (3) on putting it $x = -7, -5, ..., 5, 7$ respectively, as shown in the last column of the above *Table 2·3.*

Estimate for 2006 : When $t = 2006$, we get from (1), $x = 2(2006 - 2000{\cdot}5) = 11$

Hence the predicted sales for 2006 are : $y_e = 131{\cdot}5 + 7{\cdot}33 \times 11 = 212{\cdot}13$ (Crores Rs.)

Example 2·4. *Below are given the figures of production (in thousand tonnes) of a fertiliser factory :*

Year :	*1995*	*1997*	*1998*	*1999*	*2000*	*2001*	*2004*
Production ('000 tonnes) :	*77*	*88*	*94*	*85*	*91*	*98*	*90*

(i) Fit a straight line by the 'Least Squares Method' and tabulate the trend values.

(ii) Eliminate the trend, assuming additive model. What components of the time series are thus left over ?

(iii) What is the monthly increase in the production ?

Solution. (i)

TABLE 2·4 : COMPUTATION OF TREND VALUES

Year (t)	*Production (y_t)*	$x = t - 1999$	xy_t	x^2	*Trend values ('000 tonnes)* $y_e = 88{\cdot}8 + 1{\cdot}37x$	*Elimination of Trend*
1995	77	– 4	– 308	16	83·32	– 6·32
1997	88	– 2	– 176	4	86·06	+ 1·94
1998	94	– 1	– 94	1	87·43	+ 6·57
1999	85	0	0	0	88·80	– 3·80
2000	91	1	91	1	90·17	+ 0·83
2001	98	2	196	4	91·54	+ 6·46
2004	90	5	450	25	96·65	– 5·65
Total	623	1	159	51	622·97	

Let the trend equation be $y_t = a + bx$, [origin : July 1999]

Normal equations for estimating a and b are

$$\left.\begin{aligned} \sum y_t &= na + b\sum x \\ \sum x\, y_t &= a\sum x + b\sum x^2 \end{aligned}\right\} \Rightarrow \left\{\begin{aligned} 623 &= 7a + b \\ 159 &= a + 51b \end{aligned}\right.$$

Solving for a and b, we get : $a = 88{\cdot}80$ and $b = 1.37$

$\therefore$ Trend equation is : $y_t = 88{\cdot}8 + 1{\cdot}37x\ ;\ x = t - 1999$ ... (*)

Substituting the values of x, *viz.*, – 4, – 2, etc. successively, we get the required trend values as shown in the last but one column of *Table 2·4*.

(ii) Assuming additive model for the time series, the trend values are eliminated by subtracting them from the given values, as shown in the last column of *Table 2·4*. The resulting values give the short-term fluctuations which change with a period of more than one year.

(*iii*) Yearly increase in the production of fertiliser, as provided by linear trend $y_t = a + bx$ is 'b' = 1·37 thousand tonnes.

$$\therefore \quad \text{Monthly increase in production} = \frac{1{\cdot}37}{12} = 0{\cdot}114 \text{ thousand tonnes.}$$

Example 2·5. **Fit a straight line trend to the following data by the method of least squares and obtain two monthly trend values for Nov. 2000 and Sept. 2001.**

Year	:	1996	1997	1998	1999	2000	2001	2002	2003	2004
Average Monthly Profit (crores Rs.)	:	12·6	14·8	18·6	14·8	16·6	21·2	18·0	17·4	15·8

Solution. Let the straight line trend of y_t on x be given by :

$y_t = a + bx$, where the origin is July 2000 and x unit = 1 year. ... (1)

The normal equations for estimating a and b in (1) are :

$$\Sigma y_t = na + b\,\Sigma x \quad \text{and} \quad \Sigma x y_t = a\Sigma x + b\Sigma x^2 \qquad \text{... (2)}$$

TABLE 2·5 : FITTING STRAIGHT LINE TREND

Year (t)	$x = t - 2000$	*Average monthly profit (in crores Rs.)* (y_t)	x^2	$x y_t$	*Trend values (crores Rs.)* $y_e = 16{\cdot}64 + 0{\cdot}43x$
1996	– 4	12·6	16	– 50·4	14·92
1997	– 3	14·8	9	– 44·4	15·35
1998	–2	18·6	4	– 37·2	15·78
1999	– 1	14·8	1	– 14·8	16·21
2000	0	16·6	0	0	16·64
2001	1	21·2	1	21·2	17·07
2002	2	18·0	4	36·0	17·50
2003	3	17·4	9	52·2	17·93
2004	4	15·8	16	63·2	18·36
Total	0	149·8	60	25·8	

Substituting the values in (2), we get

$$149{\cdot}8 = a(9) + b(0) \quad \Rightarrow \quad a = (149{\cdot}8/9) = 16{\cdot}64$$

and

$$25{\cdot}8 = a(0) + b(60) \quad \Rightarrow \quad b = (25{\cdot}8/60) = 0{\cdot}43$$

$\therefore$ The trend equation is : [From (1)]

$$y_t = 16{\cdot}64 + 0{\cdot}43x \text{ ; (Origin : July 2000, } x \text{ unit} = 1 \text{ year)}$$

Since y represent the *monthly average* for each year and the unit of x is 12 months, the trend of monthly average increases by 0·43 in 12 months, *i.e.*, (0·43/12) per month. So the trend equation for *monthly values* is :

$$y_t = 16{\cdot}64 + \frac{0{\cdot}43}{12}x \quad \Rightarrow \quad y_t = 16{\cdot}64 + 0{\cdot}036x \qquad \text{... (3)}$$

(origin : 1st July 2000, x unit = 1 month).

In order to make this equation useful for estimating monthly trend values, the origin is to be shifted to the middle of a month. Since July 2000 is selected as origin, and we have to shift the origin half a month later, x should be replaced by $x + (1/2)$. The transformed monthly trend equation is :

$$y_t = 16{\cdot}64 + 0{\cdot}036\left(x + \tfrac{1}{2}\right) \Rightarrow y_t = 16{\cdot}658 + 0{\cdot}036\,x$$

[(origin : 15 July 2000 ; unit of x = 1 month ; unit of y_t = Monthly (crores Rs.)]

Now we find the trend values, for November 2000, and September 1999. Since, Nov. 2000 is 4 months, (*i.e.*, 4 units) ahead origin, putting $x = 4$ in the trend equation (3) :

$$(y_e)_{Nov.\ 2000} = 16{\cdot}658 + 0{\cdot}036 \times 4 = 16{\cdot}802) \text{ (crores Rs.)}$$

Similarly, Sept. 1999 is 10 months behind the origin, putting $x = -10$, we have (in 3) :

$$(y_e)_{Sept.\ 1999} = 16{\cdot}658 + 0{\cdot}036\,(-10) = 16{\cdot}298 \text{ (crores Rs.)}$$

Example 2·6. *The following figures are the production data of a certain factory manufacturing air-conditioners :*

Year :	*1990*	*1991*	*1992*	*1993*	*1994*	*1995*	*1996*	*1997*	*1998*	*1999*	*2000*
Production ('000 units)	*17*	*20*	*19*	*26*	*24*	*40*	*35*	*55*	*51*	*74*	*79*

Fit the second degree parabolic trend curve to the above data and obtain the trend values.

Solution. Let the second degree parabolic tend curve be :

$y_t = a + bx + x^2$, where $x = t - 1995$ ... (*)

TABLE 2·6 : COMPUTATION OF PARABOLIC TREND VALUES

Year (t)	*Production ('000 units) (y_t)*	$x = t - 1995$	x^2	x^3	x^4	xy	x^2y	*Trend Values* $y_e = 34 + 6.28x + 0{\cdot}6x^2$
1990	17	– 5	25	– 125	625	– 85	425	17·60
1991	20	– 4	16	– 64	256	– 80	320	18·48
1992	19	– 3	9	– 27	81	– 57	171	20·56
1993	26	– 2	4	– 8	16	– 52	104	23·90
1994	24	– 1	1	– 1	1	– 24	24	28·32
1995	40	0	0	0	0	0	0	34·00
1996	35	1	1	1	1	35	35	40·88
1997	55	2	4	8	16	110	220	48·96
1998	51	3	9	27	81	153	459	58·24
1999	74	4	16	64	256	296	1184	68·72
2000	79	5	25	125	625	395	1975	80·40
Total	440	$\Sigma x = 0$	$\Sigma x^2 = 110$	$\Sigma x^3 = 0$	$\Sigma x^4 = 1{,}958$	$\Sigma xy = 691$	$\Sigma x^2y = 4{,}917$	

The normal equations for estimating a, b and c in (*) are :

$$\left.\begin{aligned} \textstyle\sum y_t &= na + b\textstyle\sum x + c\textstyle\sum x^2 \\ \textstyle\sum xy_t &= a\textstyle\sum x + b\textstyle\sum x^2 + c\textstyle\sum x^3 \\ \textstyle\sum x^2y_t &= a\textstyle\sum x^2 + b\textstyle\sum x^3 + c\textstyle\sum x^4 \end{aligned}\right\} \Rightarrow \left\{\begin{aligned} 440 &= 11a + 110\,c & \dots (1) \\ 691 &= 110b & \dots (2) \\ 4{,}917 &= 110\,a + 1{,}958\,c & \dots (3) \end{aligned}\right.$$

From (2), we get $b = (691/110) = 6{\cdot}28.$

Multiplying (1) by 10 and then subtracting from (3), we get

$$4{,}917 - 440 \times 10 = (110a + 1{,}958c) - (110a + 1{,}100c)$$

$$\Rightarrow \quad 517 = 858c \quad \Rightarrow \quad c = 0\cdot60.$$

Substituting in (1), we get $a = \dfrac{440 - 110c}{11} = \dfrac{440 - 110 \times 0\cdot60}{11} = \dfrac{374}{11} = 34$

Substituting the values of a, b and c in (**), we get the required trend equation as :

$$y_t = 34 + 6\cdot28\,x + 0\cdot06\,x^2 \;;\quad x = t - 1995 \qquad \ldots(**)$$

The trend values y_e can be computed on putting $x = -5, -4, -1, 0, 1, \ldots, 4, 5$ in (**) and are given in the last column of the *Table*. 2·6

Example* 2·7. *You are given the population figures of India as follows :

Census year (x)	:	1911	1921	1931	1941	1951	1961	1971
Population (in crores)	:	25·0	25·1	27·9	31·9	36·1	43·9	54·7

Fit an exponential trend $y = ab^x$ to the above data by the method of least squares and find the trend values. Estimate the population in 1981, 2001 and 2011

Solution. Taking logarithm of both sides of the equation $y = a\,b^{\,x}$, we get

$$\log y = \log a + x \log b \quad \Rightarrow \quad v = A + Bx \qquad \ldots(1)$$

where $v = \log y$, $A = \log a$ and $B = \log b$. Now (1) represents a linear trend between v and x.

The arithmetic for fitting the linear trend (1) to the given data can be reduced to a great extent if we shift the origin in x to 1941 and change the scale by defining a new variable u as follows :

$$u = [(x - 1941) / 10], \quad \text{so that } \Sigma u = 0$$

Thus the linear trend $v = A + Bu$ between v and u is equivalent to the exponential trend

$$y = ab^{\,u}, \quad [(u = (x - 1941) / 10]$$

where $A = \log a$ and $B = \log b$. $\qquad \ldots[(2)]$

By the principle of least squares, the normal equations for estimating A and B in (2) are given by :

$$\Sigma v = nA + B\Sigma u \quad \text{and} \quad \Sigma uv = A\Sigma u + B\Sigma u^2$$

Since $\Sigma u = 0$, these equations give

$$A = \frac{\Sigma v}{n} = \frac{\Sigma v}{7}, \quad B = \frac{\Sigma uv}{\Sigma u^2} \qquad \ldots(3)$$

TABLE 2·7 : FITTING OF EXPONENTIAL TREND

Year (x)	Population (in crores) (y_t)	$u = \dfrac{x - 1941}{10}$	$v = \log y$	u^2	uv
1911	25·0	−3	1·3979	9	−4·1937
1921	25·1	−2	1·3997	4	−2·7994
1931	27·9	−1	1·4456	1	−1·4456
1941	31·9	0	1·5038	0	0
1951	36·1	1	1·5575	1	1·5575
1961	43·9	2	1·6425	4	3·2850
1971	54·7	3	1·7380	9	5·2140
Total		0	10·6850	28	1·6178

On using (3), we get

$$A = \frac{10\cdot6850}{7} = 1\cdot5264$$

$$\Rightarrow \quad a = \text{Antilog } A = 33\cdot60$$

$$B = \frac{1\cdot6178}{18} = 0\cdot0577$$

$$\Rightarrow \quad b = \text{Antilog } B = 1\cdot142$$

Substituting the values of a and b in (2), the exponential trend fitted to the given data is :

$$y = 33\cdot60\,(1\cdot142)^{(x - 1941/10)}$$

To obtain the trend values y for different x, we use the linear trend (**),

$v = A + Bu \Rightarrow v = 1{\cdot}5264 + 0{\cdot}0577u$

Substituting the appropriate values of u from -3 to 3 in the above equations, we get the corresponding values of v *and* finally the trend values y are obtained from the fact that

$v = \log y \quad \Rightarrow \quad y = \text{Antilog}\,(v),$

as shown in the *Table 2·8* :

Hence, on assuming the exponential trend $y = ab^x$, the estimated population for 1981, 2001 and 2011 is 57·18 crores, 74·57 crores and 85·17 crores respectively.

TABLE 2·8 : COMPUTATION OF EXPONENTIAL TREND

Year	u	$0{\cdot}0577\,u$	$u = 1{\cdot}5264 + 0{\cdot}0577u$	*Trend Values* y_e = *Antilog* (v)
1911	−3	−0·1731	1·3533	22·56
1921	−2	−0·1154	1·4160	25·76
1931	−1	−0·0577	1·4687	29·43
1941	0	0	1·5264	33·50
1951	1	0·0577	1·5841	38·38
1961	2	0·1154	1·6418	43·83
1971	3	0·1731	1·6995	50·06
1981	4	0·2308	1·7572	57·18
2001	6	0·3462	1·8726	74·57
2011	7	0·4039	1·9303	85·17

2·4·4. Growth Curves and Their Fitting. The various growth curves, *viz.* the modified exponential, Gompertz and Logistic curves as given in (*vi*) § 2·4·3 cannot be determined by the principle of least squares. Special techniques have been devised for fitting these curves to the given set of data. In the following sections we shall discuss these curves and their fitting in detail.

Modified Exponential Curve and its Fitting. As already pointed out modified exponential curve is given by $y_t = a + bc^t, a > 0\,;$ ... (2·9)

where y_t represents the time series value at the time t and a, b, c are constants, called its parameters.

Taking first difference of (2·9), we get

$$\Delta y_t = y_{t+h} - y_t = bc^t\,(c^h - 1)$$

where 'h' is the interval of differencing.

Similarly

$$\Delta y_{t-h} = y_t - y_{t-h} = bc^{t-h}\,(c^h - 1)$$

$$\therefore \qquad \frac{\Delta y_t}{\Delta y_{t-h}} = c^h, \text{ a constant.}$$

Thus, the most striking feature of the modified exponential curve is that the first differences of the consecutive value of y_t corresponding to equivalent values of t change by a constant ratio. This implies that the first differences of y_t when plotted on a semi-logarthmic graph paper, lie on a straight line. It may be pointed out that in (2·9), the constant 'a' is always positive and $y_t = a$ is the only asymptote of the curve.

We discuss below two methods of fitting modified exponential curve.

1. *Method of Three Selected Points.* We take three ordinates y_1, y_2, y_3, (say), corresponding to three equidistant values of t, (say) t_1, t_2 and t_3 respectively such that

$$t_2 - t_1 = t_3 - t_2$$

Substituting the values of $t = t_1$, t_2 and t_3 in (2·9), we get respectively

$$y_1 = a + b\,c^{t_1}\ ,\quad y_2 = a + b\,c^{t_2}\ ,\quad y_3 = a + b\,c^{t_3} \qquad \text{... (2·10)}$$

$$\Rightarrow \quad y_2 - y_1 = b(c^{t_2} - c^{t_1}) = bc^{t_1}\,(c^{t_2 - t_1} - 1) \qquad \text{... (2·10a)}$$

and $$y_3 - y_2 = b(c^{t_3} - c^{t_2}) = bc^{t_2}(c^{t_3 - t_2} - 1)$$ ... (2·10b)

Dividing, we get

$$\frac{y_3 - y_2}{y_2 - y_1} = c^{\,t_2 - t_1} \quad \Rightarrow \quad c = \left[\frac{y_3 - y_2}{y_2 - y_1}\right]^{1/(t_2 - t_1)} \qquad [\because t_2 - t_1 = t_3 - t_2]$$ ... (2·11)

Substituting the value of c in (2·10a), we get

$$y_2 - y_1 = b\left[\frac{y_3 - y_2}{y_2 - y_1}\right]^{t_1/(t_2 - t_1)}\left[\frac{y_3 - y_2}{y_2 - y_1} - 1\right] \Rightarrow b = \frac{(y_2 - y_1)^2}{y_3 - 2y_2 + y_1}\left[\frac{y_2 - y_1}{y_3 - y_2}\right]^{t_1/(t_2 - t_1)}$$ ... (2·11a)

Substituting for b and c in (2·10), we get

$$a = y_1 - bc^{t_1} = y_1 - \frac{(y_2 - y_1)^2}{y_3 - 2y_2 + y_1} \quad \Rightarrow \quad = \frac{y_1 y_3 - y_2^2}{y_3 - 2y_2 + y_1}$$ ... (2·11b)

Substituting for a, b and c from (2·11), (2·11a) and (2·11b) in (2·9), we get the equation of the modified exponential curve fitted to the given time-series data; y_1, y_2, y_3 being ordinates of the free hand curve corresponding to three selected points $t = t_1$, t_2 and t_3.

2. *Method of Partial Sums.* The given time-series data are split up into three equal parts each containing, (say) n consecutive values of y_t correspoding to $t = 1, 2, \ldots, n$; $t = n + 1, n + 2, \ldots, 2n$; and $t = 2n + 1, 2n + 2, \ldots, 3n$. Let S_1, S_2 and S_3 represent the partial sums of the three parts respectively so that

$$S_1 = \sum_{t=1}^{n} y_t, \qquad S_2 = \sum_{t=n+1}^{2n} y_t, \qquad S_3 = \sum_{t=2n+1}^{3n} y_t$$ ... (2·12)

Substituting for y_t from (2·9), we get

$$S_1 = \sum_{t=1}^{n} (a + bc^t) = na + b(c + c^2 + \ldots + c^n) = na + bc\left(\frac{c^n - 1}{c - 1}\right)$$ ... (2·13)

Similarly, we shall get $$S_2 = na + bc^{n+1}\left(\frac{c^n - 1}{c - 1}\right)$$ ... (2·13a)

and $$S_3 = na + bc^{2c+1}\left(\frac{c^n - 1}{c - 1}\right)$$ ... (2·13b)

Subtracting (2·13) from (2·13a) and (2·13a) from (2·13b), we get respectively

$$S_2 - S_1 = bc\frac{(c^n - 1)^2}{(c - 1)}$$ ... (2·14)

and $$S_3 - S_2 = bc^{n+1}\frac{(c^n - 1)^2}{(c - 1)}$$ ... (2·14a)

Dividing (2·14a) by (2·14), we have

$$\frac{S_3 - S_2}{S_2 - S_1} = c^n \quad \Rightarrow \quad c = \left(\frac{S_3 - S_2}{S_2 - S_1}\right)^{1/n}$$ ... (2·15)

Substituting for c^n in (2·14), we get

$$S_2 - S_1 = \frac{bc}{c - 1}\left[\frac{S_3 - S_2}{S_2 - S_1} - 1\right]^2 \quad \Rightarrow \quad b = \frac{(c - 1)(S_2 - S_1)^3}{c(S_3 - 2S_2 + S_1)}$$ ... (2·15a)

Finally, substituting the values of b and c in (2·13), we get

$$a = \frac{1}{n}\left[S_1 - \frac{bc}{c - 1}(c^n - 1)\right] = \frac{1}{n}\left[S_1 - \frac{(S_2 - S_1)^3}{(S_3 - 2S_2 + S_1)^2}(c^n - 1)\right]$$ [From (2·15a)]

$$= \frac{1}{n}\left[S_1 - \frac{(S_2-S_1)^3}{(S_3-2S_2+S_1)^2}\left\{\frac{S_3-S_2}{S_2-S_1}-1\right\}\right] = \frac{1}{n}\left[S_1 - \frac{(S_2-S_1)^2}{S_3-2S_2+S_1}\right] \quad \text{[From (2·15)]}$$

$$= \frac{1}{n}\left[\frac{S_1S_3-S_2^2}{S_3-2S_2+S_1}\right] \quad \text{... (2·15}b\text{)}$$

Fitting of Grompertz Curve. Grompertz curve is given by the equation

$$y_t = ab^{ct} \quad \text{... (2·16)}$$

here y_t is the time series value at time t and a, b, c are its parameters.

$$\log y_t = \log a + \log b \,.\, c^t$$

i.e.,

$$Y_t = A + Bc^t, \quad \text{... (2·16}a\text{)}$$

where $Y_t = \log y_t$, $A = \log a$ and $B = \log b$.

(2·16a) is the equation of a modified exponential curve and the constants A, B and c can estimated by the method of three selected points or by the method of Partial Sums as plained above. Finally, the constants of the Grompertz curve are given by

$$a = \text{antilog } A \quad \text{and} \quad b = \text{antilog } B.$$

Logistic Curve. This is a particular form of complex types of growth curve. A symmetric gistic curve, also known as Pearl-Reed curve is given by :

$$y = y_t = \frac{k}{1+\exp(a+bt)}, \quad b < 0 \quad \text{... (2·17)}$$

ere a, b and k are constants and y_t is the value of the time series at the time t.

17) can also be written as : $y_t = \dfrac{k}{1+e^a.\,e^{bt}} = \dfrac{k}{1+c\,e^{bt}}$, $b < 0$ ... (2·17a)

Also from (2·17), we have

$$\frac{1}{y_t} = \frac{1}{k}\left[1+e^{a+bt}\right] = \frac{1}{k} + \frac{1}{k}\,.\,e^{a}.\,e^{bt} = A + Bc^t, \quad \text{... (2·17}b\text{)}$$

ere $A = \dfrac{1}{k}$, $B = \dfrac{1}{k}\,e^a$, $c = e^b$, are constants.

Thus, the reciprocal of y_t follows modified exponential law. Hence, the given time series ervations y_t will follow Logistic law if their reciprocal $1/y_t$ follows modified exponential

Accordingly the fist differences $\Delta(1/y_t)$ change by a constant ratio. In other words, the differences of the receiprocals of the given observations when plotted on a semi-rithmic graph paper, will exhibit a straight line.

Derivation of (2·17). The exponential straight line

$$\log y_t = A_1 + B_1.\,t \quad \Rightarrow \quad y = y_t = ab^t$$

simple form of the growth curve called the *simple exponential.* This form gives :

$$\frac{dy}{dt} = ab^t \log b = y \log b = \alpha y \text{ (say)},$$

ate of growth of y per unit of time is directly proportional to y. But in practice this rate owth cannot be in the same proportion always. It will continue upto certain level, called *vel of saturation,* after which it starts declining.

Thus, in general, we may take

$$\frac{dy}{dt} = \alpha y (\beta - y) ; \alpha > 0, \beta > 0 \quad \dots (2\cdot18)$$

The factor y is called the *momentum factor* which increases with time t and the factor $(\beta - y)$ is known as the *retarding factor* which decreases with time. When the process of growth approaches the *saturation level* β, the rate of growth tends to zero. The principle depicted by (2·18) is called *Robertson's Law*. We shall now solve (2·18) as a differential equation in y and t.

We have

$$\frac{dy}{y(\beta - y)} = \alpha\, dt \quad \Rightarrow \quad \frac{1}{\beta}\left[\frac{1}{y} + \frac{1}{\beta - y}\right] dy = \alpha\, dt \quad \Rightarrow \quad \left[\frac{1}{y} + \frac{1}{\beta - y}\right] dy = \alpha \beta\, dt$$

Integrating, we get

$$\log\left(\frac{y}{\beta - y}\right) = \alpha\beta t + \gamma, \text{ where } \gamma \text{ is the constant of integration.}$$

$$\therefore \quad \frac{y}{\beta - y} = \exp(\alpha\beta t + \gamma) = e^{\alpha\beta t} \cdot e^{\gamma} \quad \Rightarrow \quad \beta - y = \delta y e^{-\varepsilon t}$$

or

$$y = \frac{\beta}{1 + \delta e^{-\varepsilon t}}, \varepsilon > 0 \quad \dots (2\cdot19)$$

where $\delta = e^{-\gamma}$ and $\varepsilon = \alpha\beta > 0$. The equation (2·19) is of the same form as (2·17*a*).

Also from (2·19), we have

$$\frac{1}{y} = \frac{1}{\beta} + \frac{\delta}{\beta} e^{-\varepsilon t} = A_1 + B_1 c_1^{\,t} \quad \dots (2\cdot19a)$$

where $A_1 = \frac{1}{\beta}$, $B_1 = \frac{\delta}{\beta}$ and $c_1 = e^{-\varepsilon}$, which is of the form (2·17*b*).

Properties of Logistic Curve. Logistic curve satisfies Robertson's law (2·18). Differentiating (2·18) w.r.t. t, we get

$$\frac{d^2y}{dt^2} = \alpha\left[(\beta - y)\frac{dy}{dt} - y\frac{dy}{dt}\right] = \alpha(\beta - 2y)\frac{dy}{dt}$$

$$= \alpha^2 y (\beta - y)(\beta - 2y) \quad \dots (2\cdot20)$$

$$\frac{d^2y}{dt^2} > 0 \text{ if and only if } \beta - 2y > 0 \quad \Rightarrow \quad y < \frac{\beta}{2}$$

and

$$\frac{d^2y}{dt^2} < 0 \text{ if and only if } \beta - 2y < 0 \quad \Rightarrow \quad y > \frac{\beta}{2}$$

Thus, the logistic curve (2·19) has an increasing rate for $y < \beta / 2$ and it has a declining rate for $y > \beta / 2$. Moreover

$$\frac{d^2y}{dt^2} = 0, \text{ at } y = \beta / 2$$

This implies that the logistic curve (2·19) has a point of inflexion at $y = \beta / 2$. This point of inflexion is the *critical point* wherefrom the increasing rate of the curve starts to decline. We may also observe that the line $y = \beta$ is an asymptote to the curve since $\lim_{t \to \infty} y = \beta$.

The shape of the curve is thus an elongated *S* as shown in *Fig. 2·3* :

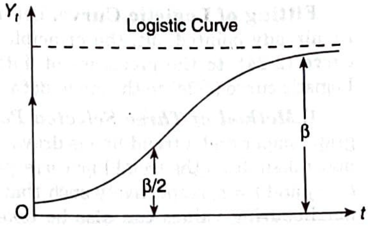

Fig. 2·3

The output of many industries display the same kind of trend. When the industry *comes into being*, the technical methods of production are not yet sufficiently formed. The production costs are high and the market demand is still small and so *production develops slowly*. Next, production grows at an increasing rate as a result of perfection of *manufacturing method, the transition to mass production and the increasing market*. In turn, there is a period of saturation of the market (*i.e.*, nearly everyone who could afford the given commodity already owns it) and accordingly the increase in output becomes slower and slower and in the end almost stops altogether. Output is stabilised at a constant level sufficient just to replace the goods used up.

***Remarks* 1.** The Logistic curve (2·17) can also be written as follows :

$$y_t = \frac{L}{1+\exp\,[\alpha(\beta - t)]}, \qquad \text{... (2·21)}$$

where L, α and β are constants. This curve is concave upward for $t < \beta$ and convex upward for $t > \beta$. The point of inflexion is at $t = \beta$ where the ordinate y_t is $L/2$. The curve thus looks like an elongated letter *S*.

Unlike the modified exponential curve which has only one asymptote, the Logistic curve has two asymptotes at the two ends. $y = L$ and $y = 0$ are the upper and lower asymptotes to the Logistic curve (2·21).

2. For the Logistic curve (2·16), *i.e.*,

$$y = y_t = \frac{k}{1+e^{a+bt}}, \quad b < 0, \qquad \text{... (*)}$$

the rate of growth is given by

$$\frac{dy}{dt} = \frac{-k}{(1+e^{a+bt})^2}\cdot b\,.\,e^{a+bt} = -b\left(\frac{k}{1+e^{a+bt}}\right)\left(\frac{1}{1+e^{a+bt}}\right).\,e^{a+bt}$$

$$= -by\,.\,\frac{y}{k}\left(\frac{k}{y}-1\right) = -by\left(1-\frac{y}{k}\right) \qquad \text{[From (*)]}$$

Thus, y has an extremum at $\quad \dfrac{dy}{dt} = 0 \;\Rightarrow\; y = 0 \;\text{ or }\; y = k$

In other words, $\quad k = \max\,(y_t) \qquad \text{... (2·22a)}$

$$\frac{d^2y}{dt^2} = -b\left[\frac{dy}{dt}\left(1-\frac{y}{k}\right)-\frac{y}{k}\cdot\frac{dy}{dt}\right] = -b\,\frac{dy}{dt}\left(1-\frac{2y}{k}\right) \qquad \text{... (2·22b)}$$

For point of inflexion, we have $\quad \dfrac{d^2y}{dt^2} = 0 \;\Rightarrow\; y = \dfrac{k}{2}$

Thus, (2·16) has point of inflexion at the time t such that

$$\frac{k}{2} = \frac{k}{1+e^{a+bt}} \;\Rightarrow\; e^{a+bt} = 1 \;\Rightarrow\; a+bt = 0 \quad \therefore \quad t = -a/b \qquad \text{... (2·23)}$$

In view of (2·22*a*); Logistic curve (2·16) may be written as :

$$y_t = \frac{k}{1+e^{a+bt}}\;;\quad b < 0,\ k = \max\,(y_t) \qquad \text{... (2·24)}$$

Fitting of Logistic Curve. Let us now consider the fitting of the Logistic curve (2·2 As already pointed out, the principle of least squares cannot be applied to fit the Logis curve (2·24) to the given set of data. We discuss below various methods of fitting t Logistic curve (2·24) to the given data.

1. *Method of Three Selected Points.* The given time-series data is first plotted on graph paper and a trend line is drawn by the freehand method. Three ordinates y_1, y_2, y_3, a now taken from the trend line corresponding to selected equidistant points of time, say $t =$ $t = t_2$ and $t = t_3$ respectively such that $t_2 - t_1 = t_3 - t_2$. The sum or average of more than o neighbouring values can also be taken with advantage. Values must be equidistant. F population data, geometric mean may be used.

Substituting the values of $t = t_1, t_2$ and t_3 in (2·24), we get respectively

$$y_1 = \frac{k}{1 + e^{a+bt_1}}, \quad y_2 = \frac{k}{1 + e^{a+bt_2}}, \quad y_3 = \frac{k}{1 + e^{a+bt_3}}$$

$$\Rightarrow \quad a + bt_1 = \log_e\left(\frac{k}{y_1} - 1\right); \; a + bt_2 = \log_e\left(\frac{k}{y_2} - 1\right); \; a + bt_3 = \log_e\left(\frac{k}{y_3} - 1\right)\Big\} \qquad ...(*$$

$$\Rightarrow \quad b(t_2 - t_1) = \log\left[\frac{(k/y_2) - 1}{(k/y_1) - 1}\right] \text{ and } b(t_3 - t_2) = \log\left[\frac{(k/y_3) - 1}{(k/y_2) - 1}\right]\Big\} \qquad ...(*$$

Since the points are equidistant, *i.e.*, $t_2 - t_1 = t_3 - t_2$, we get

$$\log\left[\frac{(k/y_2) - 1}{(k/y_1) - 1}\right] = \log\left[\frac{(k/y_3) - 1}{(k/y_2) - 1}\right] \Rightarrow \left(\frac{k}{y_3} - 1\right)\left(\frac{k}{y_1} - 1\right) = \left(\frac{k}{y_2} - 1\right)^2$$

$$\Rightarrow \quad y_2^2 (k - y_3)(k - y_1) = y_1 y_3 (k - y_2)^2 \Rightarrow y_2^2 [k^2 - k(y_1 + y_3) + y_1 y_3] = y_1 y_3 (k^2 + y_2^2 - 2y_2$$

$$\Rightarrow \quad k^2 (y_2^2 - y_1 y_3) = k\,[y_2^2 (y_1 + y_3) - 2y_1 y_2 y_3]$$

Since $k \neq 0$,
$$k = \frac{y_2^2 (y_1 + y_3) - 2y_1 y_2 y_3}{y_2^2 - y_1 y_3} \qquad ...(2·26$$

From (**) and (*), we get respectively

$$b = \frac{1}{t_2 - t_1} \log_e\left[\frac{(k - y_2) y_1}{(k - y_1) y_2}\right] \quad ...(2·26a) \quad \text{and} \quad a = \log_e\left(\frac{k - y_1}{y_1}\right) - bt_1 \qquad ...(2·26b$$

Example 2·8. *Given the three selected points y_1, y_2 and y_3 corresponding to $t_1 = 2$, $t_2 = 30$ and $t_3 = 58$ as follows :*

$$t_1 = 2, \quad y_1 = 55·8 \; ; \quad t_2 = 30, \quad y_2 = 138·6 \; ; \quad t_3 = 58 \; ; \quad y_3 = 251·8$$

Fit the Logistic curve by the method of selected points. Also obtain the trend values for $t = 5, 18, 25, 35, 46, 50, 54, 60, 66, 70$.

Solution. Let the equation of the logistic curve be :
$$y_t = \frac{k}{1 + e^{a+bt}} \qquad ...(*)$$

Then using (2·26), (2·26*a*) and (2·26*b*), we get

$$k = \frac{y_2^2 (y_1 + y_3) - 2y_1 y_2 y_3}{y_2^2 - y_1 y_3} = \frac{3987987·70 - 2348005·97}{19209·96 - 8470·44} = \frac{1639981·72}{10739·52} = 152·7$$

$$b = \left[\log_e\left\{\frac{y_1 (k - y_2)}{y_2 (k - y_1)}\right\}\right]\frac{1}{t_2 - t_1} = \left[\log_{10}\left\{\frac{786·78}{13430·34}\right\} \times \log_e 10\right]\frac{1}{28}$$

$$= (2{\cdot}8958 - 4{\cdot}1280)\,\frac{2{\cdot}3026}{28} = -\,0{\cdot}1013$$

And $\quad a = \log_e\left(\dfrac{k}{y_1} - 1\right) - bt_1 = (\log_{10} 1{\cdot}7365)\, 2{\cdot}3026 + 0{\cdot}2026$

$$= 0{\cdot}2396 \times 2{\cdot}3026 + 0{\cdot}2026 = 0{\cdot}7543$$

Hence the required (fitted) equation of the Logistic curve is : $\quad y_t = \dfrac{152{\cdot}7}{1 + e^{\cdot 7543 - \cdot 1013t}}$

Trend Values. In (*), let us take

$$e^{a+bt} = \mu \quad \Rightarrow \quad \log_e \mu = a + bt \quad \Rightarrow \quad \log_e \mu = 0{\cdot}7543 - 0{\cdot}1013t$$

Now $\quad \log_{10}\mu = \dfrac{\log_e \mu}{\log_e 10} = \dfrac{\log_e \mu}{2{\cdot}3026} = \dfrac{0{\cdot}7543 - 0{\cdot}1013\,t}{2{\cdot}3026}$

Finally, the trend values y_t are given by : $y_t = \dfrac{k}{1 + e^{a+bt}} = \dfrac{k}{1+\mu}$, and are obtained in the last column of *Table 2·9*.

TABLE 2·9 : COMPUTATION OF TREND VALUES BY LOGISTIC CURVE

Period t	$\log_e \mu = 0{\cdot}7543 - 0{\cdot}1013t$	$\log_{10}\mu = \dfrac{(2)}{2{\cdot}3026}$	μ = *Antilog* {(3)}	$y_t = \dfrac{k}{1+\mu}$
(1)	(2)	(3)	(4)	(5)
5	0·2478	0·1076	1·2810	66·944
18	− 1·0691	$-0{\cdot}4640 = \bar{1}{\cdot}5367$	0·3434	113·667
25	− 1·7782	$-0{\cdot}7722 = \bar{1}{\cdot}2278$	0·1690	130·624
35	− 2·7912	$-1{\cdot}2122 = \bar{2}{\cdot}7878$	0·0613	143·880
46	− 3·9055	$-1{\cdot}6961 = \bar{2}{\cdot}3039$	0·0201	149·691
50	− 4·3107	$-1{\cdot}8721 = \bar{2}{\cdot}1279$	0·0134	150·681
54	− 4·7195	$-2{\cdot}0481 = \bar{3}{\cdot}9519$	0·0089	151·353
60	− 5·3237	$-2{\cdot}3120 = \bar{3}{\cdot}6880$	0·0049	151·955
66	− 5·9315	$-2{\cdot}5760 = \bar{3}{\cdot}4240$	0·0027	152·289
70	− 6·3367	$-2{\cdot}7520 = \bar{3}{\cdot}2480$	0·0018	152·426

2. *Yule's Method.* Let us suppose that the value of k is approximately known or obtained other methods. Then the logistic curve (2·24) contains two parameters a and b, and two riables t and y_t. Hence the principle of least squares can be used to estimate a and b. We ve from (2·24),

$$a + bt = \log\left(\frac{k}{y} - 1\right) \quad \text{or} \quad v = a + bt \qquad \dots (2{\cdot}27)$$

ere $v = \log(k/y - 1)$. (2·27) represents a linear trend between v and t, and according to the nciple of least squares, the normal equations for estimating a and b are :

$$\Sigma\, v = na + b\Sigma\, t \;\text{ and }\; \Sigma t\, v = a\Sigma\, t + b\Sigma\, t^2$$

3. *Hotelling's Method.* A very elegant and ingeneous method for fitting a Logistic curve iven by Hotelling. We have [*c.f.* (2·22)]