Business statist

Unit - V

Time Series Analysis

INTRODUCTION

" A time series is a set of statistical observations arranged in chronological order."

A series of observations, on a variable, recorded after successive intervals of time is called a time series. The successive intervals are usually equal time intervals, e.g., it can be 10 years, a year, a quarter, a month, a week, a day, and an hour, etc. The data on the population of India is a time series data where time interval between two successive figures is 10 years. Similarly figures of national income, agricultural and industrial production, etc., are available on yearly basis.

A time series is a sequence of measurements over time, usually obtained at equally spaced intervals

- Daily
- Monthly
- Quarterly
- Yearly.

USES OF TIME SERIES ANALYSIS

1. It helps in understanding past behaviour
2. It helps in planning future operations
3. It helps in evaluating current accomplishments
4. It facilitates comparison

OBJECTIVES OF TIME SERIES ANALYSIS

The analysis of time series implies its decomposition into various factors that affect the value of its variable in a given period. It is a quantitative and objective evaluation of the effects of various factors on the activity under consideration.

There are two main objectives of the analysis of any time series data:

(i) To study the past behaviour of data.

(ii) To make forecasts for future.

The study of past behaviour is essential because it provides us the knowledge of the effects of various forces. This can facilitate the process of anticipation of future course of events, and, thus, forecasting the value of the variable as well as planning for future.

COMPONENTS OF A TIME SERIES

Components of a time series Any time series can contain some or all of the following components:

1. Trend (T)
2. Cyclical (C)
3. Seasonal (S)
4. Irregular (I)

These components may be combined in different ways. It is usually assumed that they are multiplied or added, i.e.,

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

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

To correct for the trend in the first case one divides the first expression by the trend (T). In the second case it is subtracted.

Trend component

The trend is the long term pattern of a time series. A trend can be positive or negative depending on whether the time series exhibits an increasing long term pattern or a decreasing long term pattern. If a time series does not show an increasing or decreasing pattern then the series is stationary in the mean.

Cyclical component

Any pattern showing an up and down movement around a given trend is identified as a cyclical pattern. The duration of a cycle depends on the type of business or industry being analyzed.

Seasonal component

Seasonality occurs when the time series exhibits regular fluctuations during the same month (or months) every year, or during the same quarter every year. For instance, retail sales peak during the month of December.

Irregular component

This component is unpredictable. Every time series has some unpredictable component that makes it a random variable. In prediction, the objective is to "model" all the components to the point that the only component that remains unexplained is the random component.

TIME SERIES MODELS

The analysis of time series consists of two major steps:

1. Identifying the various forces (influences) or factors which produce the variations in the time series, and

2. Isolating, analysing and measuring the effect of these factors separately and independently, by holding other things constant.

The purpose of models is to break a time series into its components: Trend (T), Cyclical (C), Seasonality (S), and Irregularity (I).

Time series provides a basis for forecasting. There are many models by which a time series can be analysed; two models commonly used for a time series are discussed below.

1. Multiplicative Model

This is a most widely used model which assumes that forecast (Y) is the product of the four components at a particular time period. That is, the effect of four components on the time series is interdependent.

$Y = T \times C \times S \times I$ (Multiplicative model)

The multiplicative model is appropriate in situations where the effect of S, C, and I is measured in relative sense and is not in absolute sense. The geometric mean of S, C, and I is assumed to be less than one.

For example, let the actual sales for period 20 be Y20 = 423.36. Further let, this value be broken down into its components as: let trend component (mean sales) be 400; effect of current cycle (0.90) is to depress sales by 10 per cent; seasonality of the series (1.20) boosts sales by 20 per cent. Thus besides the random fluctuation, the expected value of sales for the

period is $400 \times 0.90 \times 1.20 = 432$. If the random factor depresses sales by 2 per cent in this period, then the actual sales value will be $432 \times 0.98 = 423.36$.

2. Additive Model

In this model, it is assumed that the effect of various components can be estimated by adding the various components of a time-series.

It is stated as:

Y=T + C + S + I (Additive model)

Here S, C, and I are absolute quantities and can have positive or negative values.

It is assumed that these four components are independent of each other. However, in real-life time series data this assumption does not hold good.

3. Mixedd Model

It is combination above two models.

MEASUREMENT OF TREND

The principal methods of measuring trend fall into following categories:

1. Free Hand Curve or Graphic method
2. Method of Averages
3. Method of least squares

The time series methods are concerned with taking some observed historical pattern for some variable and projecting this pattern into the future using a mathematical formula. These methods do not attempt to suggest why the variable under study will take some future value. This limitation of the time series approach is taken care by the application of a causal method. The causal method tries to identify factors which influence the variable is some way or cause it to vary in some predictable manner. The two causal methods, regression analysis and correlation analysis, have already been discussed previously. A few time series methods such as freehand curves and moving averages simply describe the given data values, while other methods such as semi-average and least squares help to identify a trend equation to describe the given data values.

1. Free Hand Curve or Graphic method

A freehand curve drawn smoothly through the data values is often an easy and, perhaps, adequate representation of the data. The forecast can be obtained simply by extending the trend line. A trend line fitted by the freehand method should conform to the following conditions: (i) The trend line should be smooth- a straight line or mix of long gradual curves. (ii) The sum of the vertical deviations of the observations above the trend line should equal the sum of the vertical deviations of the observations below the trend line. (iii) The sum of squares of the vertical deviations of the observations from the trend line should be as small as possible. (iv) The trend line should bisect the cycles so that area above the trend line should be equal to the area below the trend line, not only for the entire series but as much as possible for each full cycle.

Example 7.1: Fit a trend line to the following data by using the freehand method.

Year	1991	1992	1993	1994	1995	1996	1997	1998
Sales turnover : (Rs. in lakh)	80	90	92	83	94	99	92	104

Solution:

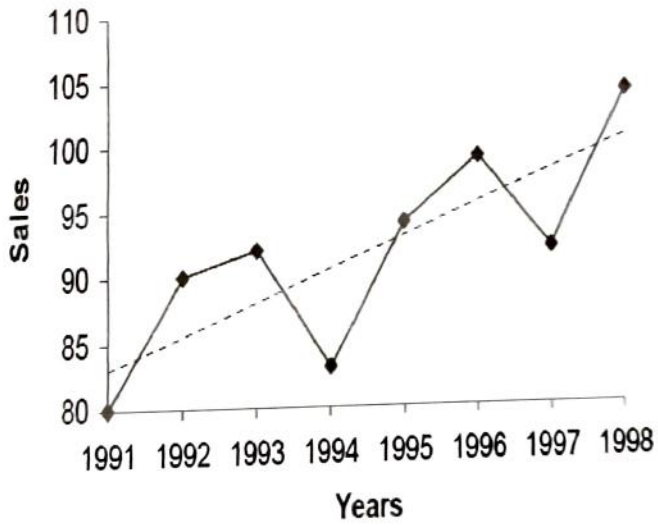


Figure 7.2 presents the freehand graph of sales turnover (Rs. in lakh) from 1991 to 1998. Forecast can be obtained simply by extending the trend line.

Merits

1. It is Simplest method of measuring of trend.
2. It is very flexible
3. Hands of experience

Limitations of freehand method

(i) This method is highly subjective because the trend line depends on personal judgement and therefore what happens to be a good-fit for one individual may not be so for another.

(ii) The trend line drawn cannot have much value if it is used as a basis for predictions.

(iii) It is very time-consuming to construct a freehand trend if a careful and conscientious job is to be done.

Method of Averages

The objective of smoothing methods into smoothen out the random variations due to irregular components of the time series and thereby provide us with an overall impression of the pattern of movement in the data over time. In this section, we shall discuss three smoothing methods.

(i) Moving averages

(ii) Weighted moving averages

(iii) Semi-averages

The data requirements for the techniques to be discussed in this section are minimal and these techniques are easy to use and understand.

Semi-averages:

When this method is used, the given data is divided into two parts, preferably with the same number of years. For example, if we are given data from 1994 to 2011, *i.e.*, over a period of 18 years, the two equal parts will be each nine years, *i.e.*, from 1994 to 2002 and from 2003 to 2011. In case of odd number of years like 9, 13, 17, etc., two equal parts can be made simply by omitting the middle year. For example, if data are given for 19 years from 1993 to 2011, the two equal parts would be from 1993 to 2001 and from 2003 to 2011—the middle year 2002 will be omitted.

After the data have been divided into two parts, an average (arithmetic mean) of each part is obtained. We thus get two points. Each point is plotted at the mid-point of the class interval covered by the respective part and then the two points are joined by a straight line which gives us the required trend line. The line can be extended downwards or upwards to get intermediate values or to predict future values.

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

Year	*Sales of Firm A (thousand units)*	*Year*	*Sales of Firm A (thousand units)*
2005	102	2009	108
2006	105	2010	116
2007	114	2011	112
2008	110		

Solution. Since seven years are given, the middle year shall be left out and an average of the first three years and the last three years shall be obtained. The average of the first three years is

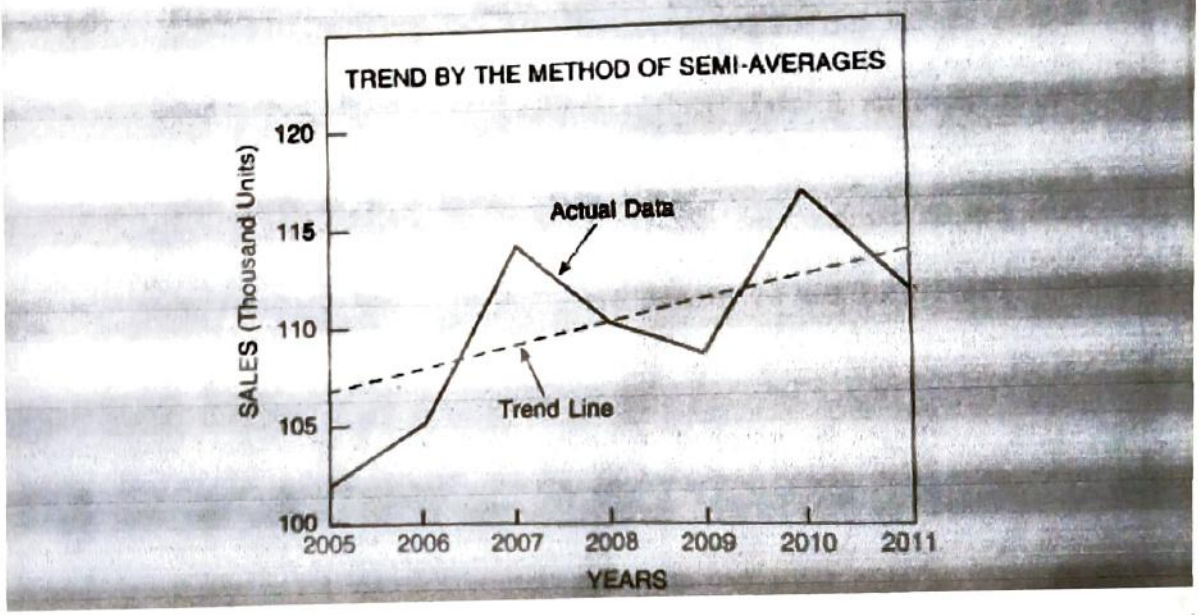
$$\frac{102 + 105 + 114}{3} = \frac{321}{3} = 107$$

and the average of the last three years is

$$\frac{108 + 116 + 112}{3} = \frac{336}{3} = 112.$$

Thus we get two points 107 and 112 which shall be plotted corresponding to their respective middle years, *i.e.*, 2006 and 2010. By joining these two points we shall obtain the required trend line. The line can be extended and can be used either for prediction or for determining intermediate values.

The actual data and the trend line are shown in the following graph :



Even Number of Years. When there are even number of years like 6, 8, 10 etc. equal parts can easily be formed and an average of each part obtained. However, when the average is to be centered there would be some problem in case the number of years is 8, 12, etc. For example, if the data relate to 2008, 2009, 2010 and 2011, which would be the middle year ? In such a case the average will be centered corresponding to 1st July 2009, *i.e.*, middle of 2009 and 2010.

The following example shall illustrate the point :

Illustration 3. Fit a trend line by the method of semi-averages to the data given below. Estimate the sales for the year 2012. If the actual sale for that year is Rs. 520 lakh, account for the difference between the two figures.

Year	*Sales (Rs. Lakh)*		*Year*	*Sales (Rs. Lakh)*	
2004	412		2008	470	
2005	438	$\frac{1748}{4} = 437$	2009	482	$\frac{1942}{4} = 485.5$
2006	444		2010	490	
2007	454		2011	500	

Solution. The average of the first four years is 437 and that of the last four years is 485.5. These two points shall be taken corresponding to the middle periods, *i.e.*, Ist July 2005 and Ist July 2009.

The estimated sales for 2012 by projecting the semi-average trend line is Rs. 518 lakh. The actual figure given to us is Rs. 520 lakh. The difference is due to the fact that time series analysis helps us to get the best possible estimates on certain assumptions which may come out to be true or not depending upon how far these assumptions have been realised in practice.

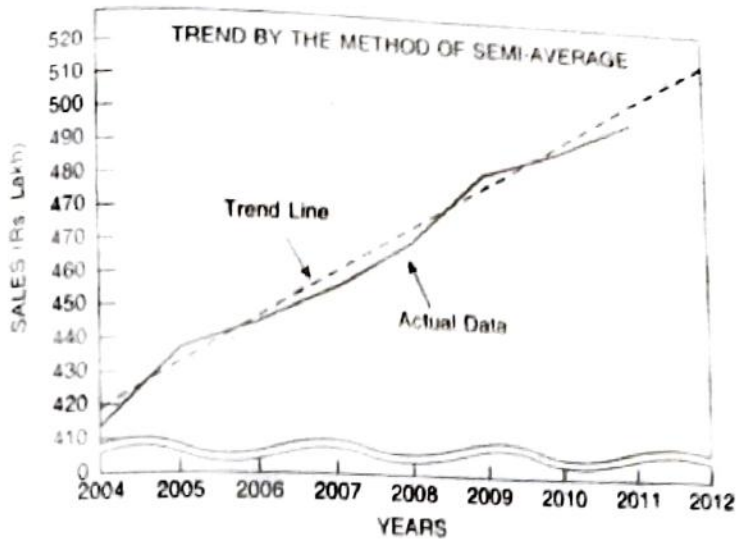


Illustration 4. The sale of a commodity in million tonnes varied from January 2011 to December 2011 in the following manner :

280	300	280	280	270	240
230	230	220	200	210	200

Fit a trend line by the method of semi-averages.

Solution.

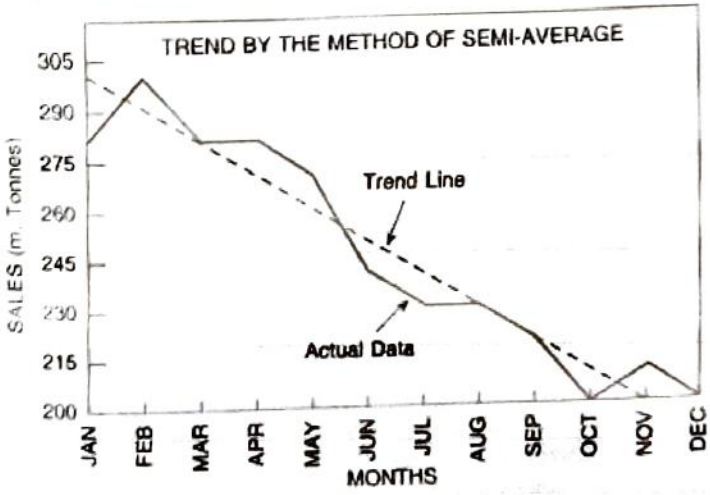
CALCULATION OF TREND VALUES BY THE METHOD OF SEMI-AVERAGES

Month	Sales in m. tonnes		Month	Sales in m. tonnes	
January	280		July	230	
February	300		August	230	
March	280	1,650 (Total) of first	September	220	1,290 (Total) of last
April	280	six months	October	200	six months
May	270		November	210	
June	240		December	200	

$$\text{Average of the first half} = \frac{1650}{6} = 275 \text{ tonnes.}$$

$$\text{Average of the second half} = \frac{1290}{6} = 215 \text{ tonnes.}$$

These two figures, namely, 275 and 215, shall be plotted at the middle of their respective periods, i.e., at the middle of March-April and that of September-October, 2011. By joining these two points, we get a trend line which describes the given data.



Merits of Semi-Averages

1. The method of semi-averages is simple, easy, and quick.
2. It smooths out seasonal variations
3. It gives a better approximation to the trend because it is based on a mathematical model.

Demeris of Semi-Averages

1. It is a rough and objective method.

2. The arithmetic mean used in Semi Average is greatly affected by very large or by very small values.
3. The method of semi-averages is applicable when the trend is linear. This method is not appropriate if the trend is not linear.

Method of Moving Average

Method of Moving Averages*

When a trend is to be determined by the method of moving averages, the average value for a number of years (or month or weeks) is secured, and this average is taken as the normal or trend value for the unit of time falling at the middle of the period covered in the calculation of the averages. The effect of averaging is to give a smoother curve, lessening the influence of the fluctuations that pull the annual figures away from the general trend.

While applying this method, it is necessary to select a period for moving average such as 3-yearly moving average, 5-yearly moving average, 8-yearly moving average, etc. The period of moving average is to be decided in the light of the length of the cycle. Since the moving average method is most commonly applied to data which are characterised by cyclical movements, it is necessary to select a period for moving average which coincides with the length of the cycle, otherwise the cycle will not be entirely removed. The danger is more severe, the shorter the time period represented by the average. When the period of moving average and the period of the cycle do not coincide, the moving average will display a cycle which has the same period as the cycle in the data, but which has less amplitude than the cycle in the data. Often we find that the cycles in the data are not of uniform length. In such a case we should take a moving average period equal to or somewhat greater than the average period of the cycle in the data. Ordinarily, the necessary period will

...ween three and ten years for general business series but even longer ...ds are required for certain types of data.

The 3 yearly moving average shall be computed as follows

$$\frac{a+b+c}{3}, \quad \frac{b+c+d}{3}, \quad \frac{c+d+e}{3}, \quad \frac{d+e+f}{3}$$

and for 5 yearly moving average

$$\frac{a+b+c+d+e}{5}, \quad \frac{b+c+d+e+f}{5}, \quad \frac{c+d+e+f+g}{5}$$

Illustration 5. (a) Calculate the 3-yearly moving averages of the production figures given ... and draw the trend.

Year	Production (in tonnes)	Year	Production (in tonnes)
1997	15	2005	63
1998	21	2006	70
1999	30	2007	74
2000	36	2008	82
2001	42	2009	90
2002	46	2010	95
2003	50	2011	102
2004	56		

Solution. CALCULATION OF 3-YEARLY MOVING AVERAGES

Year	Production Y (in tonnes)	3-yearly totals (in tonnes)	3-yearly moving average
1997	15	—	—
1998	21	66	22.00
1999	30	87	29.00
2000	36	108	36.00
2001	42	124	41.33
2002	46	138	46.00
2003	50	152	50.67
2004	56	169	56.33
2005	63	189	63.00
2006	70	207	69.00
2007	74	226	75.33
2008	82	246	82.00
2009	90	267	89.00
2010	95	287	95.67
2011	102	—	—

... Construct 5-yearly moving averages of the number of students studying in a college ... given below

Year	No. of students	Year	No. of students
2002	332	2007	405
2003	317	2008	410
2004	357	2009	427
2005	392	2010	405
2006	402	2011	438

Solution. CALCULATION OF 5-YEARLY MOVING AVERAGES

Year	No. of students	5-yearly total	5-yearly moving average
2002	332	—	—
2003	317	—	—
2004	357	1809	361.8
2005	392	1873	374.6
2006	402	1966	393.2
2007	405	2036	407.2
2008	410	2049	409.8
2009	427	2085	417.0
2010	405	—	—
2011	438	—	—

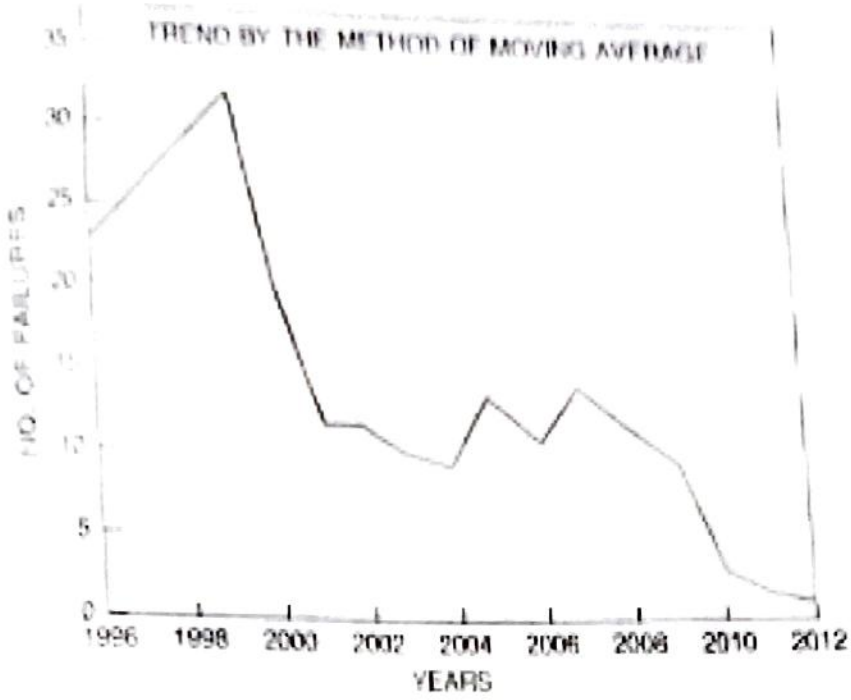
Illustration 6. Calculate 5-yearly and 7-yearly moving averages for the following data of the numbers of commercial and industrial failures in a country during 1996 to 2011 :

Year	*No. of failures*	*Year*	*No. of failures*
1996	23	2004	9
1997	26	2005	13
1998	28	2006	11
1999	32	2007	14
2000	20	2008	12
2001	12	2009	9
2002	12	2010	3
2003	10	2011	1

Also plot the actual and trend values on a graph.

Solution. CALCULATION OF 5-YEARLY AND 7-YEARLY MOVING AVERAGES

Year	*No. of failures*	*5-yearly moving total*	*5-yearly moving average*	*7-yearly moving total*	*7-yearly moving average*
1996	23			—	—
1997	26	—	—	—	—
1998	28	129	25.8	—	—
1999	32	118	23.6	153	21.9
2000	20	104	20.8	140	20.0
2001	12	86	17.2	123	17.6
2002	12	63	12.6	108	15.4
2003	10	56	11.2	87	12.4
2004	9	65	11.0	81	11.6
2005	13	57	11.4	81	11.6
2006	11	59	11.8	78	11.1
2007	14	59	11.8	71	10.1
2008	12	69	9.8	63	9.0
2009	9	39	7.8	—	—
2010	3	—	—	—	—
2011	1	—	—	—	—



Even Period of Moving Average. If the moving average is an even period moving average, say, four-yearly or six-yearly, the moving total and moving average which are placed at the centre of the time span from which they are computed fall between two time periods. This placement is incorrect since the moving average so placed would not coincide with the original time period. We, therefore, synchronise moving averages and original data. This process is called *centering*" and always consists of taking a two-period moving average of the moving averages.

Illustration 7. Estimate the trend values using the data given by taking a four-yearly moving average

Year	Value	Year	Value
1996	12	2003	100
1997	25	2004	82
1998	39	2005	65
1999	54	2006	49
2000	70	2007	34
2001	87	2008	20
2002	105	2009	7

[BA (H) Econ., Madras Univ., 2009]

Solution.

COTIMATING THE TREND VALUES

Year	Value	4-yearly moving totals	4-yearly moving average	4-yearly moving average centered
1996	12	—	—	—
1997	25	—	—	—
		← 130	32.5	
1998	39			← 39.75
		← 168	47.0	
1999	54			← 54.75

Contd

					4-yearly moving average centered
					70.75
2001	87				
					92.00
2002	105				
					90.75
2003	100				
					81.00
2004	82				
					65.75
2005	65				
					49.75
2006	49				
					34.75
2007	34				
2008	20				
2009					

Illustration 6 Assume a four-yearly cycle and calculate the trend by the method of moving averages from the following data relating to the production of tea in India :

Year	Production (in lbs.)	Year	Production (in lbs.)
2002	464	2007	540
2003	515	2008	557
2004	516	2009	571
2005	487	2010	586
2006	602	2011	612

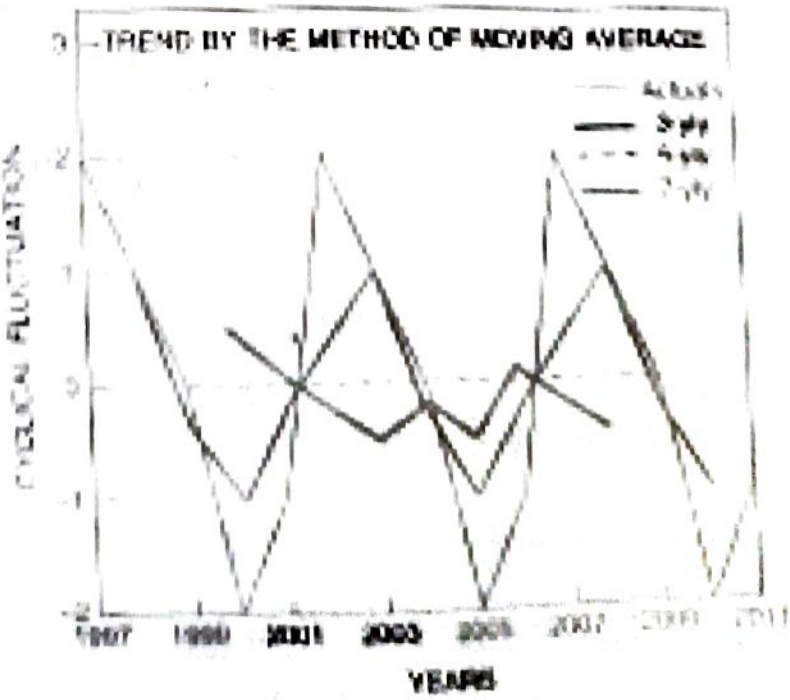
Solution. CALCULATION OF TREND BY THE MOVING AVERAGE METHOD

Year	Production (in lbs.)	4-yearly moving totals	4-yearly moving average	4-yearly moving average centred
2002	464	—	—	—
2003	515	—	—	—
		1984	491.00	
2004	516			495.75
		2002	500.50	
2005	487			503.63
		2027	506.75	
2006	602			511.63
		2066	516.50	
2007	540			522.00
		2170	542.50	
2008	557			552.00
		2254	563.50	
2009	571			572.50
		2326	581.50	
2010	586			
2011	612	—	—	—

Solution

CALCULATION OF THREE-YEARLY, FIVE-YEARLY AND SEVEN-YEARLY MOVING AVERAGES

Year	Cyclical fluctuations	3-yearly moving average	5-yearly moving average	7-yearly moving average
1997	+ 2			
1998	+ 1	+ 1.00	—	—
1999	0	0.33	0	
2000	+ 2	+ 1.00	0	+ 0.43
2001	- 1	- 0.33	0	+ 0.14
2002	- 2	+ 0.67	0	- 0.29
2003	+ 1	+ 1.00	0	- 0.43
2004	0	- 0.33	0	+ 0.14
2005	- 2	- 1.00	0	- 0.43
2006	- 1	- 0.33	0	+ 0.14
2007	+ 2	+ 0.67	0	- 0.29
2008	+ 1	+ 1.00	0	- 0.43
2009	0	- 0.33	0	
2010	- 2	- 1.00	—	—
2011	- 1	—	—	—



erits:

1. This method is easy to understand and easy to use because there are no mathematical complexities involved.

2. It is an objective method in the sense that anybody working on a problem with the method will get the same trend values. It is in this respect better than the free hand curve method.

3. It is a flexible method in the sense that if a few more observations are added, the entire calculations are not changed. This not with the case of semi-average method.

4. When the period of oscillatory movements is equal to the period of moving average, these movements are completely eliminated.

5. By the indirect use of this method, it is also possible to isolate seasonal, cyclical and random components.

Demerits:

1. It is not possible to calculate trend values for all the items of the series. Some information is always lost at its ends.

2. This method can determine accurate values of trend only if the oscillatory and the random fluctuations are uniform in terms of period and amplitude and the trend is, at least, approximately linear. However, these conditions are rarely met in practice. When the trend is not linear, the moving averages will not give correct values of the trend.

3. The trend values obtained by moving averages may not follow any mathematical pattern i.e. fails in setting up a functional relationship between the values of X(time) and Y(values) and thus, cannot be used for forecasting which perhaps is the main task of any time series analysis.

4. The selection of period of moving average is a difficult task and a great deal of care is needed to determine it.

5. Like arithmetic mean, the moving averages are too much affected by extreme values.