MAS 309: TIME SERIES ANALYSIS AND FORECASTING
COURSE OUTLINE

- Definition of time series
- Importance of time series
- Measurement of seasonal variation
- Measurement of Regular, cyclical and irregular fluctuations
- Second degree parabola
- Measuring trends by Logarithms
- Fitting a cubic polynomial by method of least squares
- Useful time series models
- Forecasting

INTRODUCTION

Definition 1

A time series is a collection of observation made sequentially over a period of time.

Definition 2

It's an arrangement of statistical data in accordance to with the time of occurrence.

According to Yule and Kendall

When we observe numerical features of an individual or a population at different point of time the drift set constitute of a time series.

The theory and practice is different from other branches of statistics in that;

i) In time series the observation are dependent or correlated with time and the order of observation is important hence classical procedures and techniques that rely on independent assumption are no longer applicable.
ii) Asymptotic theory tends to be meaningful in practice. This is not the case with time series. In fact the temporal or spatial dependence in time series is what is of interest and importance.

Importance of Time Series Analysis

It plays a significant role in Business decision making. The following facts prove its importance;

i) Helps in the understanding of past behavior. The past trend helps in predicting the future behavior.
ii) It enables us to predict or forecast the behavior of the phenomenon in future, which is very essential for better value.
iii) Helps in making comparative studies in the values of different phenomenon at different times or places
iv) Helps in the evaluation of current achievements

v) The segregation and study of the various components of time series is of paramount importance to better man in planning of future operations and the formulation of executive and policy decisions.
vi) According to Hirch the main objective of analyzing time series is to understand, interpret and evaluate changes in economic phenomenon in the hope of more correctly anticipating the course of future events.

COMPONENTS OF TIME SERIES.

The main components are:

a) Secular Trends
b) Seasonal variation
c) Cyclical variation
d) Irregular variation

These are explained as under:

a) Secular Trends

The general tendency of time series data to increase or stagnate during a long period of time is called secular trend or simple trend. This phenomenon is observed in most of the series relating to economics e.g. an upward tendency is usually observed in time series relating to population, production and sales of products, prices, income, currency in circulation, Bank deposit etc.

While a downward tendency is noticed in the time series relating to deaths, epidemics, TB etc due to advancement in medical technology, improved medical facilities, better sanitation, diet etc.

Secular trend is regular, smooth and long time movement of a statistical series; it reveals the general tendency of the data.

b) Seasonal Variation

It represents a periodic movement where the period is not longer than 1 year. The factors that mainly cause this type of variation in time series are the climatic changes of the different seasons and the customs and habits which people follow at different times.

The short range stock and brisk period of business activity at different seasons of the year, production and consumption of commodities, sales and profits of a company, Bank clearing and Bank deposits etc are attributed to seasonal variations.

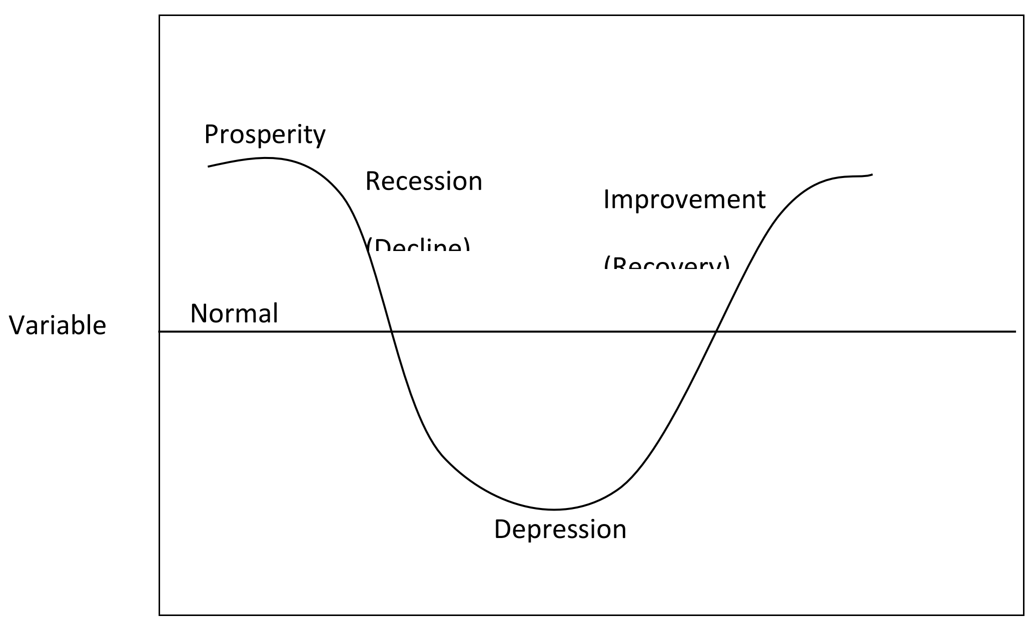
The main objective of the measurement of seasonal variations is to isolate them from the trend and study the effects. A study of the season patterns is extremely useful to business men, producers, sales managers etc. In planning future operations and in formulation of policy decisions regarding purchases, production, inventory control, personnel requirements, selling and advertising programs.

c) **Cyclical Variation.**

Cyclical variation or fluctuations are another type of periodic movements with a period more than 1 year. Such movements are barely regular in nature, one complete period is called a cycle. Cyclical variation is not as regular as seasonal fluctuations but the sequence of change marked by prosperity, decline, depression and recovery remains more or less regular. Cyclical variations are found to exist in most of the business and economic time series, the four phases of business cycle are usually completed over a period of (8-10) years. They are:

i) Prosperity
ii) Decline/Recession
iii) Depression
iv) Recovery / Improvement

Phases of Business cycle



d) **Irregular Variations**

Irregular variation or movements are such variations which are completely unpredictable in character. These powerful unforeseen variations are usually caused by factors which are either partly or wholly uncountable and are caused by such unforeseen events like war, floods, strikes and lockouts etc.

This may sometimes be as a result of many small forces each of which may have a negligible effect but their combined effect is not negligible.

Random movement doesn't reveal any pattern of the repetitive tendency and can be considered as residual variations.

MODEL FOR ANALYSIS OF TIME SERIES

The analysis of time series consist of isolating or breaking down the observation series into various component parts these components are secular, seasonal, cyclical and irregular variations. To achieve this breakdowns we must make assumptions about the relationship existing among the assumed to have either the multiplicative relationship (multiplicative model) or the additive relationship of model.
In the multiplicative model each observed value y_t at any time t is the product of all the four components T, C, S, & I Symbolically
$y_t = T \times S \times C \times I$
$y_t = TSCI$
Where the observed values and T values are stated in original units but the other values components i.e. S C & I are expressed in percentage
On the other hand the additive model each value that is observed is thought to be the sum of all the values of the components.
$y_t = T + S + C + I$
Where the components T, S, C & I are assumed to be mutually independent. This analysis of time series is a decomposition of observed series into, up to four distinct components according to some models.
Before analyzing the time series is often desirable to adjust the data for calendar variations, for holidays, for price changes and so on. Such adjustments help in removing the effects of certain false differences

Measurement of secular Trend

There are four methods commonly used to measure secular trend.

i) Method of free hand drawing
ii) Method of semi-average
iii) Method of moving average
iv) Method of least squares

i) Method of Freehand drawing

According to this method the time series is plotted on the graph paper taking time on the horizontal scale and observation on the vertical scale. Then a freehand smooth curve is drawing through the plotted points. This curve smooth out the irregularities and describes long term changes. The method eliminates the short-term change and gives the long-term changes.

Sometimes instead of drawing a free hand curve a straight line is drawn in such a way that it may not pass through all the plotted points. This is called trend line. When a trend line is fitted with the free hand method an attempt should be made to make it conform as much as possible with the following conditions;

i) It should be smooth either a straight line or a combination of long gradual curves.

ii) The sum of the vertical deviations from the trend of the annual observation above the trend line should equal the sum of the vertical deviations from the trend of observation below the trend line.

iii) The sum of the sq of the vertical deviation of the observation from the trend should be as small as possible

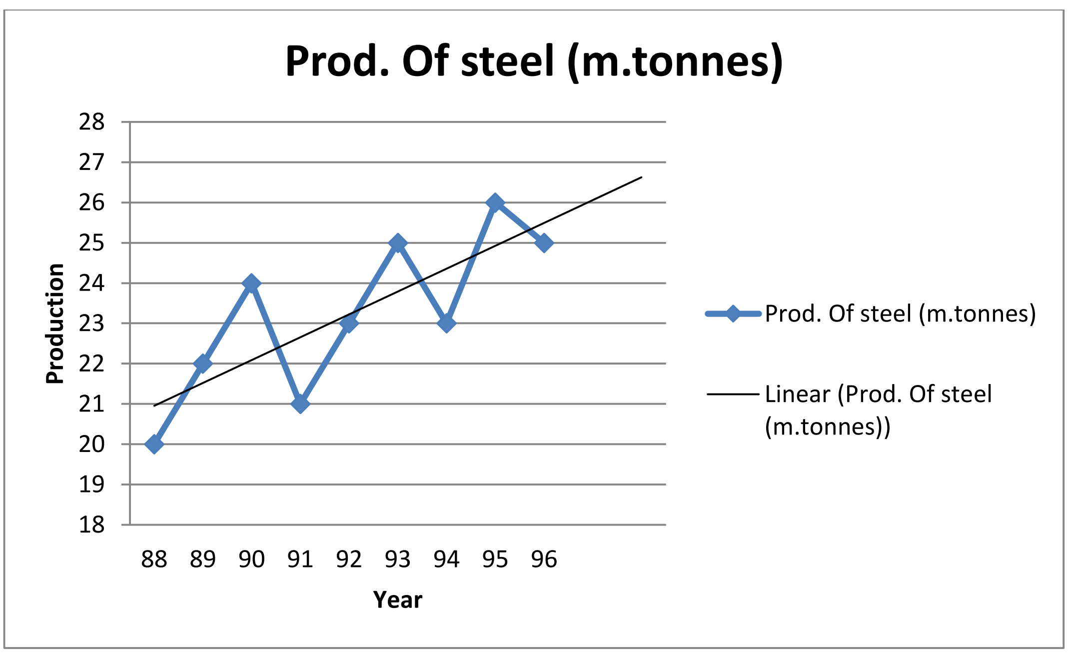
iv) The trend should bisect the circles. The area above the trend equals the area below the trend not only for the entire series but as much as possible for each full cycle. This last condition can't be always be made fully but a careful attempt should be made to observe it as closely as possible.

Example 1.

Fit a trend line to the following data by the freehand method

Year	88	89	90	91	92	93	94	95	96
Prod. of steel (m. tones)	20	22	24	21	23	25	23	26	25

The trend line drawn by the freehand method can be extended to predict future values. However since the freehand method of curve fitting may be subjective this method shouldn't be used as a basis for prediction



Merits and Limitations

Merits

1. It's the simplest method of measuring trend

2. It's very flexible in that it can be used regardless of whether the trend is a straight line or curve
3. The trend line drawn by a statistician experienced in computing trends and having knowledge of the economic history of the concerned or industry under analysis may be a better expression of the secular movement that a trend fitted by the use of a rigid mathematical formula which while providing a good fit to the points may have no other logical justification. Although this method is not recommended for beginners it has considerable merit in the hand of experienced statisticians and is widely used in mathematical situations.

Limitations

1. It's highly subjective, because the trend line depends on the personal judgment of the investigator and different people may draw different trend line from the same set of data. Moreover the work can't be left to clerks and must be handled by experienced people who are well conversant with the history of the particular concerned.
2. Since freehand curve is subjective it doesn't have much value if it's used as a basis for prediction.
3. Although this method appears simple and direct, it takes a lot of time to construct a free hand trend if a careful and conscientious job is done. It's only after a long experience in trend fitting that a statistician should attempt to fit a trend line by inspection.

ii) Method of moving Averages

In this method instead of taking the actual values relating to the particular years, we take into account the moving average of the values of three, Four, Five or more years. In order to compute three-years moving average, the values against the first, second and third years are averaged. This average value is written against middle year, that is, against the second year. Similarly the values against the second, third and fourth year are averaged and average value is written against the third year and so on. If it is required to compute 4-year moving average, then central moving average is computed. This moving average values are plotted in the same graph paper on which the time series have been drawn. This would smooth out irregularities eliminating short-time changes and will show long time tendency.

Example

Compute the moving average for five years and plot the data on a graph.

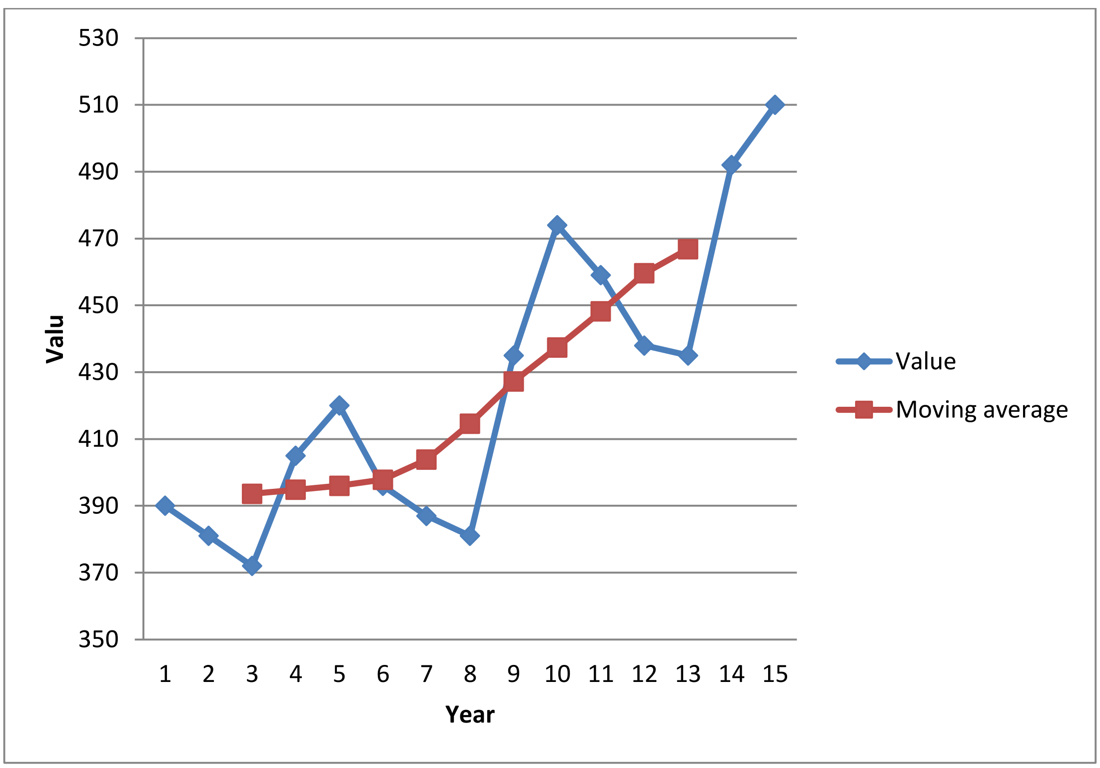
Year	value	year	value	Year	value
1	390	6	396	11	459
2	381	7	387	12	438
3	372	8	381	13	435
4	405	9	435	14	492
5	420	10	474	15	510

Solution

Computation of five year moving average.

Year (1)	Value (2)	5 yearly moving totals (3)	5 yearly moving averages(trend values)

			$4 = (3) \div 5$
1	390		
2	381		
3	372	1968	393.6
4	405	1974	394.8
5	420	1980	396.0
6	396	1989	397.8
7	387	2019	403.8
8	381	2073	414.6
9	435	2136	427.2
10	474	2187	437.4
11	459	2241	448.2
12	438	2298	459.6
13	435	2334	466.8
14	492		
15	510		



If the period of moving average is even, 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 face between two time periods. This placement is inconvenient since the moving averages so placed would not coincide with an original time period. We therefore, synchronize moving averages and original data. This process is called centering and consists of taking a two-period moving average of the moving averages.

Example.

The following table relate to the prices of a commodity for three years.

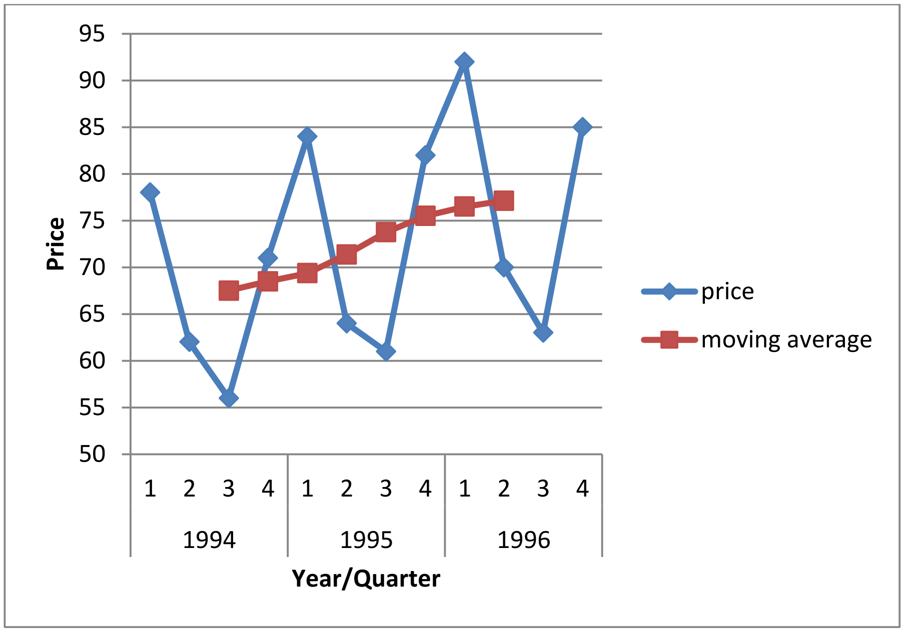
	Price (Ksh)			
Year/ Quarter	**First**	**second**	**Third**	**fourth**
1994	78	62	56	71
1995	84	64	61	82
1996	92	70	63	85

Find the centered 4-quarterly moving average.

Solution

Computation of moving averages

Year and quarter (1)		Price (Sh) (2)	4-quarter moving total (3)	2-Items moving total (4)	4-quarter moving average centred. $(5) = (4) \div 8$
1994	1	78	-	-	-
	2	62	-	-	-
			267		
	3	56		540	67.50
			273		
	4	71		548	68.50
			275		
1995	1	84		555	69.375
			280		
	2	64		571	71.375
			291		
	3	61		590	73.75
			299		
	4	82		604	75.50
			305		
1996	1	92		612	76.50
			307		
	2	70		617	77.125
			310		
	3	63	-	-	-
	4	85	-	-	-



Merits and limitations.

Merits

1. This period is simple as compared to the method of least squares.
2. It is flexible method of measuring trend. If a few more figures are added to the data, the entire calculations are not changed. We only get some more trend values.
3. If the period of moving average happens to coincide with the period of cyclical fluctuations in the data, such fluctuations are automatically eliminated.
4. The moving average has the advantage that it follows the general movements of the data, and that its shape is determined by the data rather than the statisticians choice of a mathematical function.

Limitations.

1. Trend values cannot be computed for all the years. The longer the period f moving average, the greater the number of years for which trend values cannot be obtained. For example, in a three yearly moving average, trend values cannot be obtained for the first year and last year, in five yearly averages for the first two years and the last two years, and so on. It is often these extreme years in which we are most interested.
2. Great care has to be exercised in selecting the period of moving average. No hard fast rules are available for the choice of the period and the one has to use his own judgment.
3. Since the moving average is not represented by a mathematical function, this method cannot be used in forecasting which one of the main objectives of trend analysis is.
4. Although theoretically we say that if the period of moving average happens to coincide with the period of cycles, the cyclical fluctuations are completely eliminated, but in practice since the cycles are by no means perfectly periodic, the lengths of the various cycles in any given series will usually vary considerably and therefore no moving average can completely remove the cycles.

iii) <u>Method of semi average</u>

This method consists of separating the data into two parts (Preferably equal) and averaging them for each part. These averages are chosen as points on a graph paper against the mid-points of the time interval in each part.

The straight line joining these two points gives the trend line. The distance of the trend line from the horizontal axis gives the trend values.

The logic behind this method is that the actual trend be a straight line. This method will give quite satisfactory results. The trend line can be extended downwards or upwards to get intermediate values or to predict future values.

Example

Fit a trend line to the following data by the method of semi-averages

Year	Sales of firm A(thousands)	Year	Sales of firm A(thousands)
1997	102	**2001**	108
1998	105	**2002**	116
1999	114	**2003**	112
2000	110		

Solution

Since seven years are given, the middle year shall be omitted and an average of the first three years and the last three years shall be obtained.

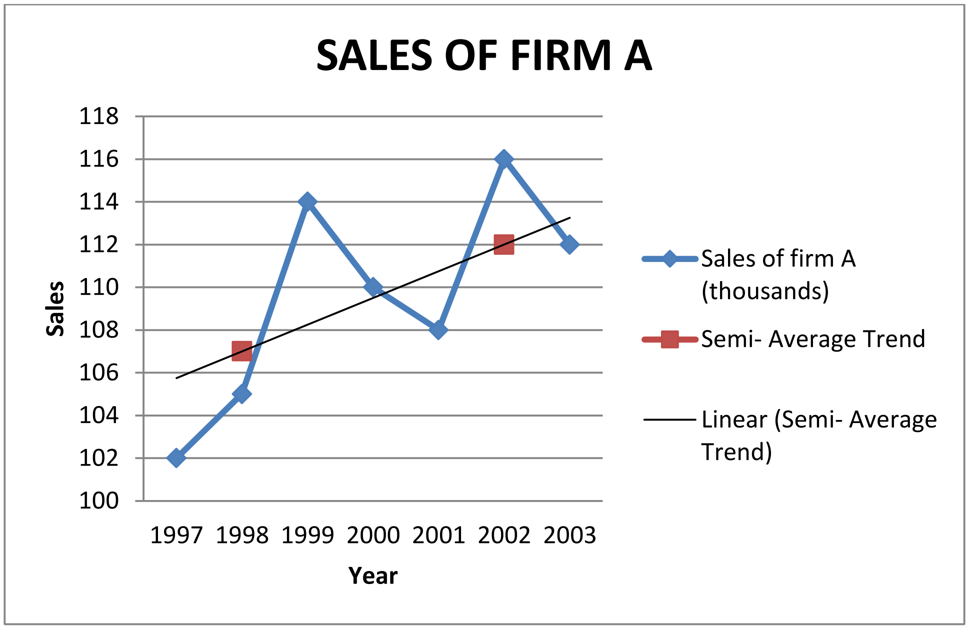
The average of the 1st three years

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

The average of the 2nd three years.

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

Thus we get two points 107 and 112 which shall be plotted corresponding to their respective middle years i.e. 1998 and 2002. By joining these two points we shall obtain the required trends 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 graph



Where there are even number of years like 6,8,10 etc two 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 problems in case the number of years is 8, 12 etc.

Example

If the data relate to 1998-2003 which would be the middle year? In such a case the average will be centered corresponding to 1st July, 2000, i.e. Middle of 2000 and 2001

Merits and Limitations

Merits

1. This method is simple to understand compared to the moving average method and methods of least square
2. This is an objective method of measuring trend as everyone who applies the method bound to get the same results. (Of course, leaving aside the arithmetical mistakes)

Limitations.

1. This method assumes straight line relationship between the plotted points regardless of the feet whether that relationship exists or NOT.
2. The limitations of arithmetic average shall automatically apply. If there are extremes in either half of both halves of the series, then the trend line is NOT true picture of the growth factors. This danger is greatest when the time period represented by the average is small. Consequently trend value obtain are not precise enough for the purpose either of forecasting the future trend or of eliminating trend from original data.

   For the above reasoning if the arithmetic average of the data is to be used in estimating the secular movement it is sometimes better to use moving average than the semi averages.

Exercise

Suppose the following sales (in £ '00' to nearest £ 10) were recorded for a firm.

Week 1

Day	Mon	Tue	Wed	Thus	Fri.
Sales Y	250	320	340	520	410

Week 2

Day	Mon	Tue	Wed	Thus	Fri.
Sales Y	260	380	410	670	420

Obtain a semi-Average trend.

Question 2

Given the data below
UK outward movement of passengers by sea.

	Year1				Year2				Year 3			
Quarter	1	2	3	4	1	2	3	4	1	2	3	4
No. of pass (millions)	2.2	5.0	7.9	3.2	2.9	5.2	8.2	3.8	3.2	5.8	9.1	4.1

a) Use the method of semi-averages to obtain and plot a trend line.
b) Draw up a table showing the original data (y) values against the trend (t) values (obtained from the graph)

iv) Method of least squares

When this method is applied a trend line is fitted to the data in such a manner that the following two conditions are satisfied.

i) $\Sigma\,(Y - Y_c) = 0$

i.e. the sum of the squares of the derivations of the actual values of Y and the computed values of Y is zero

ii) $\Sigma\,(Y - Y_c)^2 = 0$ is least.

i.e. the sum of the squares of the derivations of the actual and computed values is least from this line. That is why this method is called the method of least squares. The line obtained by this method is known as the line of best fit.

The method of least squares can be used either to fit a straight line trend or a parabolic trend. The straight line trend is represented by the equation

$$y_c = a + bX$$

Where y_c denotes the trend (computed) values to distinguish them from the actual y values a is the y-intercept or the value of y variable when $X = 0$, b represents slope of the line or the amount change in y variable that is associated with a change of one unit in X variable. The X variable in time series analysis represents time.

In order to determine the value of the constants a & b. The following two normal equations are to be solved.

$$\Sigma Y = Na + b\Sigma X$$

$$\Sigma XY = a\Sigma X + b\Sigma X^2$$

Where N represents numbers of years (Months or any other time period) for which data is given.

Example.

Below are given the figures of production (in tones) in a sugar factory.

Year	1998	1999	2000	2001	2002	2003	2004
Prod. in m. tones	80	90	92	83	94	99	92

i) Fit a straight line trend to these figures
ii) Plot these figures on the graph and show the trend line
iii) Estimate the likely sales of the company during 2006.

solution

i) Fitting the straight line trend. Recall $Y_c = a + bX$

Year	**Prod. in m. tones(Y)**	**Deviation from middle year (X)**	XY	X^2	**Trend value Y_c**
1998	80	-3	-240	9	84
1999	90	-2	-180	4	86
2000	92	-1	-92	1	88
2001	83	0	0	0	90
2002	94	1	94	1	92
2003	99	2	198	4	94
2004	92	3	276	9	96
$N = 7$	$\Sigma Y = 630$	$\Sigma X = 0$	$\Sigma XY = 56$	$\Sigma X^2 = 28$	$\Sigma Y_c = 630$

The equation of the straight line is $Y_c = a + bX$

Since $\Sigma X = 0$

$a = \frac{\Sigma Y}{N} = \frac{630}{7}$

$= 90$

$a = \frac{\Sigma XY}{\Sigma X^2} = \frac{56}{28} = 2$

Hence the equation of the straight line trend is

$y_c = 90 + 2X$ i.e $y_c = 2X + 90$

For $X = -3$ $y_c = 2(-3) + 90 = 84$

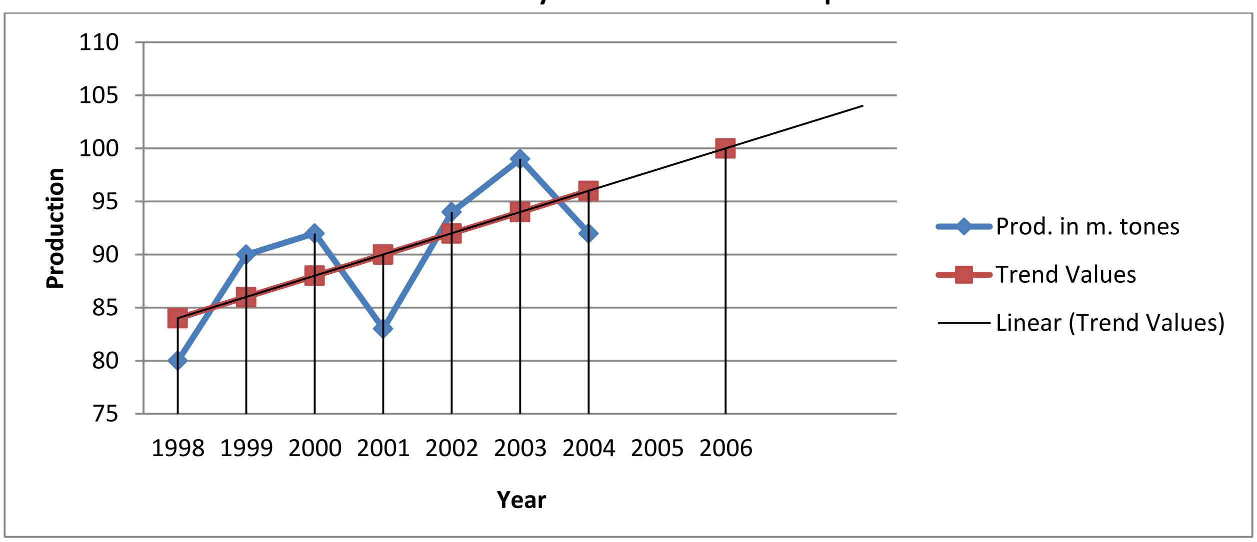
$X = -2$ $y_c = 2(-2) + 90 = 86$

$X = -1$ $y_c = 2(-1) + 90 = 88$

Similarly by putting $X = 0, 1, 2, 3$ we can obtain other trend value.

ii) The graph of above data is given below

Linear trend by the method of least squares.



iii) For 2005 X would be +4. Putting $X = 4$ in the equation

$y_c = 2(4) + 90 = 98$ m. tones.

Exercises

Question 1

Apply the method of least squares to obtain the trend values from the data below.

Year	Sales (Kg)	Year	Sales(Kg)
2000	100	2003	140
2001	120	2004	80
2002	110		

Predict the value of the sales for year 2006.

Question 2

Calculate the trend values by the method of least squares from the data below and estimate the values for sales of the year 2007.

Year	2000	2001	2002	2003	2004
Sales of T.V sets in('000') Ksh	12	18	20	23	27

Question 3

Fit a straight line trend by the method of least squares of the following data, and find the trend values.

Year	1998	1999	2000	2001	2002	2003
Sales of air-conditioned('ooo')Ksh.	10	13	16	21	24	30

Merits and limitations

Merits.

1. This is a mathematical method of measuring trend and as such there is no possibility of subjectiveness.
2. The line obtained by this method is called the line of best fit because it is this line from where the sum of the positive and negative deviations is zero and the sum of the squares of the deviations is least i.e. $\Sigma(Y - Y_c) = 0$ and $\Sigma(Y - Y_c)^2$ least.

Limitations

Mathematical curves are useful to describe the general movement of a time series, but it is doubtful whether any analytical significance should be attached to them, except in special cases. It is seldom possible to justify on theoretical grounds any real dependence of a variable with the passage of time.

MEASUREMENTS OF SEASONAL VARIATIONS

1. Simple average method.
2. Link relative method.
3. Ratio to moving average method.
4. Ratio to trend method.

These are explained as under;

1. Simple average method.

This method involves the following steps;

1. Add up all the values for each month for all the years separately.
2. Divide these totals by the number of years in order to get monthly average for various months.
3. Add up all the monthly averages and divide this sum by 12 in order to get the average of monthly averages.
4. Calculate the percentage of various monthly averages to the average of monthly averages i.e.

$$\frac{montly\ average\ for\ january}{Average\ of\ montly\ averages} \times 100$$

These percentages give the measure of seasonal variations.

Example.

Consumption of monthly electric power in millions kw hours for street lighting in a country during 1992-1996 is given below.

Year	Jan	Feb	Mar	April	May	Jun	Jul	Aug	Sept	Oct	Nov	Dec
1992	318	288	278	250	231	216	223	245	269	302	325	347
1993	342	309	299	268	249	236	242	262	288	321	342	361
1994	367	328	320	287	269	251	259	284	309	245	367	394
1995	392	349	342	311	290	273	282	305	328	364	389	417
1996	420	378	370	334	314	296	305	330	356	396	422	452

Find out seasonal variation by the method of monthly averages.

Solution

Month	Consumption of monthly electric Power 1992	1993	1994	1995	1996	Monthly total for 5 years	Five Yearly Average	Percentage
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Jan	318	342	367	392	420	1839	367.8	116.1
Feb	281	309	328	349	378	1645	329.0	103.9
Mar	278	299	320	342	370	1609	321.8	101.6
Apr	250	268	287	311	334	1450	290.0	91.6
May	231	249	269	290	314	1353	270.6	85.4
Jun	316	236	251	273	296	1272	245.4	80.3
Jul	223	242	259	282	305	1311	262.2	82.8
Aug	245	262	284	305	330	1426	285.2	90.1
Sept	269	288	309	328	356	1550	310.0	97.9
Oct	302	321	245	364	396	1728	345.6	109.1
Nov	325	342	367	389	422	1845	369.0	116.5
Dec	347	364	394	417	452	1974	394.6	124.7
					Total	19002	3800.4	1200
					Average	1583.5	316.7	100

The above calculations are explained below

1. Column No. 7 gives the total for each month for five.
2. In column No.8 each total of column No. 7 has been divided by five to obtain an average for each month.
3. The average of monthly averages is obtained by dividing the total of monthly average by 12.
4. In column No. 9 each monthly averages has been exposed as a percentage of the average of monthly averages. Thus the percentage for January

   $$= \frac{367.8}{316.7} \times 100 \qquad = 116.1$$

   Percentage for February $= \frac{329.0}{316.7} \times 100 \qquad = 103.9$

   If instead of monthly data, we are given weekly or quarterly data, we shall compute weekly or quarterly averages by following the same procedure as explained above.

2. Link relative method

In order to calculate seasonal variations indices by this method, following steps are taken

1. Express the value for each month as percentage of the previous month in order to get the link relatives i.e.

   $$\frac{Current\ months\ figures}{Previous\ months\ figures} \times 100$$

2. Calculate the average of the link relative for each month by using preferably the median, because it is less distorted by the extreme values. Convert these averages into chain relatives on the base of the first month.
3. Calculate the chain relative of the first month on the base of the last month. It will be seen that a discrepancy will exist between this chain relative.
4. Computed by the previous method. This difference between the two values represents the trend increment. It is, therefore, necessary to adjust these chain relative.
5. Divide this difference by the number of months for the purpose of adjustment.
6. Multiply the resultant figure by 1,2,3,4,……..etc , and subtract these respectively from the 2^{nd} ,3^{rd}, 4^{th} etc, month chain relative. These are adjusted chain relatives.
7. Express these adjusted chain relatives as the percentages of their averages in order to get the seasonal indices.

Example

Apply method of link relatives to the following data and calculate seasonal indices

Quarter	Quarterly Figures				
	1992	1993	1994	1995	1996
I	6.0	5.4	6.8	7.2	6.6
II	6.5	7.9	6.6	5.8	7.4
III	7.8	8.4	9.3	7.5	8.0
IV	8.7	7.3	6.4	8.5	7.1

Solution

Year	Quarter			
	I	II	III	IV
1992	-	108.3	120.0	111.5
1993	62.1	146.3	106.3	86.9
1994	93.2	95.6	143.1	68.8
1995	112.5	80.6	129.3	113.3
1996	77.6	110.6	109.6	88.8
Arithmetic average	$\frac{345.4}{4} = 86.35$	$\frac{541.4}{5} = 108.28$	$\frac{608.3}{5} = 121.66$	$\frac{469.3}{5} = 93.86$
Chain relative	100	$\frac{100 \times 108.28}{100} = 108.28$	$\frac{121.66 \times 108.28}{100} = 131.73$	$\frac{93.86 \times 131.73}{100} = 123.64$
Corrected chain relative	100	$108.28 - 1.675 = 106.605$	$131.73 - 3.35 = 128.35$	$123.64 - 5.025 = 118.615$
Seasonal indices	100	$\frac{106.605 \times 100}{113.4} = 94.0$	$\frac{128.38 \times 100}{113.4} = 113.21$	$\frac{118.615 \times 100}{113.4} = 104.60$

In the above table the correction factor has been calculated as follows;
Chain relative of the first quarter (on the basis of first quarter)= 100
Chain relative of the first quarter (on the basis of last quarter)

$$= \frac{86.35 \times 123.6}{100} = 106.7$$

The difference between these chain relatives = 106.7 – 100 = 6.7

Difference per quarter $= \frac{6.7}{4} = 1.675$

Adjusted chain relatives are obtained by subtracting $1 \times 1.675; 2 \times 1.675; 3 \times 1.675$ from the chain relatives of the 2nd, 3rd, and 4th quarters, respectively.

Seasonal variation indices have been calculated as follows.

$$\frac{100+106.605+128.38+118.615}{4} = \frac{453.6}{4} = 113.4$$

Seasonal variation index = $\frac{correct\ chain\ relative \times 100}{113.4}$

3. <u>Ratio to moving Average Method</u>

In order to calculate seasonal variations, the following steps are taken

1. Tabulate the original data year between year for all the months
2. Compute the moving average of the data
3. Calculate the deviations month wise by subtracting the moving average figures from the original figures in order to get total fluctuations.
4. Calculate the average fluctuations for each month in order to get seasonal variations. On the other hand, if it is desired to have seasonal variation indices, the following steps are taken:

1. Calculate the moving average of the data
2. Calculate the ratio of the actual values of the data to the corresponding moving average value for each month.
3. Calculate the average of these ratios for each month of all years by using either the mean of the median.
4. Express each of these monthly averages as a percentage of the average of all the 12 months average, in order to get the seasonal variation indices.

<u>Example</u>

Calculate the seasonal indices for the data gives below by the method of ratio to moving averages.

Output of salt in million tones				
Year	Quarters			
	I	II	III	IV
1991	68	62	61	63
1992	65	58	56	61
1993	68	63	63	67
1994	70	59	56	62
1995	60	55	51	58

Solution

Calculation of seasonal indices by the method of ratio to moving average.

Year	Quartile	Output of salt in millions of tons	4-quarterly moving total	2-quaterly moving total of cel. 4	4-quarterly moving average cel.5 $\div$ 8	Quarterly figures expressed as percentage of moving average.
1991	I	68	-	-	-	
	II	62	-	-	-	
	III	61	254	505	63.1	96.6
	IV	63	251	498	62.2	101.3
1992	I	65	247	489	61.1	106.4
	II	58	242	482	60.2	96.3
	III	56	240	483	60.4	92.7
	IV	61	243	491	60.1	101.5
1993	I	68	248	503	62.9	108.2
	II	63	255	516	64.5	97.7
	III	63	261	524	65.5	96.2
	IV	67	263	522	65.2	102.8
1994	I	70	259	511	63.9	109.5
	II	59	252	499	62.4	94.6
	III	56	247	484	60.5	92.6
	IV	62	237	470	58.8	105.4
1995	I	60	233	461	57.6	104.2
	II	55	228	452	56.5	97.3
	III	51	224	-	-	-
	IV	55	-	-	-	-

Calculation of seasonal index

Year	Percentage to moving average				Total
	Quarters				
	I	II	III	IV	
1991	-	-	96.6	101.3	
1992	106.4	96.3	92.7	101.5	
1993	108.2	97.7	96.2	102.8	
1994	109.5	94.6	92.6	105.4	
1995	104.2	97.3	-	-	
Total	428.3	385.9	378.1	411.0	**1603.3**
Average	**107.1**	**96.5**	**94.5**	**102.7**	**400.8**
Seasonal Index	**106.9**	**96.3**	**94.3**	**102.5**	**400.0**

Arithmetic mean of averages $= \frac{400.8}{4} = 100.2$

Seasonal indices are calculated by expressing each quarterly average as percentage of 100.2

4. Ratio to trend method

The ratio to trend method not only measures the seasonal variations but also the combined cyclical and accidental variations. This method involves the following steps:

1. Fit a curve to the given set of data by the method of last squares.
2. Calculate the trend values for the entire data.
3. Calculate the ratio of each original value to its representative trend. It means that if the original value is denoted by (v) and its representative trend by (T), then it is required to calculate the ratio $\frac{v}{T}$
4. Express these ratios $\frac{v}{T}$ as the percentages.
5. Calculate the average of the ratios $\frac{v}{T}$ for each month for the whole period (years) using some convenient average.
6. Convert the resultant figures into seasonal indices by the method explained earlier.
   If it is required to measure the cyclical and accidental variations, the indices of seasonal variations are subtracted from their respective ratios $\frac{v}{T}$ for each month. The resultant series thus obtained would represent the cyclical and accidental variations of the data.

Example

Calculate the seasonal variation of the following data by using ratio to trend method.

Year	Quarters			
	I	II	III	IV
1992	30	40	36	34
1993	34	52	50	44
1994	40	58	54	48
1995	54	76	68	62
1996	80	92	86	82

Solution

in order to calculate the seasonal variation by the ratio to trend method, first calculate the trend values for the yearly data and then convert them into quarterly trend values because, if trend values are calculated directly from these they will be influenced by the seasonal variations.

Calculation of yearly trend values

Year	X	Yearly total	Yearly average Y	XY	X^2	Trend values
1992	-2	140	35	-70	4	32
1993	-1	180	45	-45	1	44
1994	0	200	50	0	0	56
1995	1	260	65	65	1	68
1996	2	340	85	170	4	80
Total			280	120	10	

Equation of straight line is of the form

$$Y = a + bx$$

Normal equations are:

$$\Sigma Y = na$$

$$\Sigma XY = b\Sigma X^2$$

Substituting the values

$280 = 5a \qquad a = 56$

$120 = 10b \qquad b = 12$

Equation of the straight line best fitted to the data is of the form $Y = 56 + 12X$

Now trend values can be calculated by putting $X = -2, -1, 1$ ... etc in the equation quarterly increment

$$= \frac{12}{4} = 3$$

Calculation of quarterly Trend Values

In order to calculate the quarterly trend values, let us consider 1992. The trend value for the middle quarter i.e. half of 2nd and half of 3rd is 32. Since the quarterly increment is 3, therefore the trend value of 2nd quarter is $32-\frac{3}{2}=30.51$, third quarter is $32+\frac{3}{2}=33.5$, for 4th quarter is $33.5+3=36.5$ and for the first quarter is therefore $30.5-3=27.5$. In this way quarterly trend values for the rest of the year are calculated. After calculating the quarterly trend values, they are expressed as the percentage of corresponding trend values as follows:

Year	Trend values T				Given values as percentage of trend values i.e. $\frac{V}{T}\times 100$			
	Quarters				Quarters			
	I	II	III	IV	I	II	III	IV
1992	27.5	30.5	33.5	36.5	110.0	131.1	107.5	93.1
1993	39.5	42.5	45.5	48.5	86.1	122.4	109.9	90.7
1994	51.5	54.5	57.5	60.5	77.7	106.4	93.9	79.3
1995	63.5	66.5	69.5	72.5	85.0	114.3	97.8	85.5
1996	75.5	78.5	81.5	84.5	106.0	117.2	105.5	97.0
Total					464.8	591.4	514.6	445.6
Average					92.96	118.28	102.92	89.12
Seasonal index					92.2	117.3	102.1	88.4

Total of averages of percentage = 92.96+118.28+102.92+89.12

= 403.28

Average of average percentages $=\frac{403.28}{4}=100.82$

In order to calculate the seasonal indices for the various quarters, the average percentage of quarters are exposed as the percentage of their own average i.e. 100.82.

MEASUREMENT OF THE REGULAR AND IRREGULAR FLUCTUATIONS.

The short time changes consist of both regular and irregular fluctuations. If by the use of some technique we are in a position to find out the regular fluctuations in a time series, we can find the irregular fluctuations by subtracting the regular fluctuation from a time series. Regular fluctuations are calculated by averaging the seasonal or monthly figures. When these regular fluctuations are subtracted from the short time changes, we are left with irregular fluctuations. The following example will clarify the method of isolating the regular fluctuations from irregular fluctuations in the short time change.

Example

Using the data given below explain clearly how you would determine

i) The trend
ii) The short-time oscillations, and in the latter, isolate the regular from the irregular fluctuation.

Year	summer	monsoon	autumn	winter
1	30	81	62	119
2	33	104	86	171
3	42	153	99	221
4	56	172	129	235
5	67	201	136	302

Solution

Year (1)	Seasons (2)	Values (3)	4–seasons totals (4)	2 – seasons totals of Col. (4) (5)	4 – seasons moving average (6)	Short – time oscillation (7)	Regular seasonal fluctuations (8)	Irregular fluctuations (9)
1	Summer	30	-	-	-	-	-	-
	Monsoon	81	-	-	-	-	-	-
	Autumn	62	292	587	73.4	-11.4	-19.2	7.8
	Winter	119	295	613	76.6	42.4	68.6	-26.2
2	Summer	33	318	660	82.5	-49.5	-74.6	25.1
	Monsoon	104	342	734	92.0	12.0	25.5	-13.5
	Autumn	86	394	797	99.6	-13.6	-19.2	5.6
	Winter	171	403	855	106.9	64.6	68.6	-4.6
3	Summer	42	452	917	114.6	-72.6	-74.6	2.0
	Monsoon	153	465	980	122.5	30.5	25.5	5.0
	Autumn	99	515	1044	130.5	-31.5	-19.2	-12.3
	Winter	221	529	1077	134.6	86.4	68.6	17.8
4	Summer	56	548	1126	140.5	-84.5	-74.6	-9.9
	Monsoon	72	578	1170	146.2	26.8	25.5	1.3
	Autumn	129	592	1195	149.4	-20.4	-19.2	-1.2
	Winter	235	603	1235	154.4	81.6	68.6	13.0
5	Summer	67	632	1271	158.9	-91.6	-74.6	-17.3
	Monsoon	201	639	1345	168.1	32.9	25.5	7.4
	Autumn	136	706	-	-	-	-	-
	Winter	302	-	-	-	-	-	-

Calculation of regular seasonal fluctuations:

Year	Summer	Monsoon	Autumn	Winter
1	-	-	-11.4	42.4
2	-49.5	12.0	-13.6	64.1
3	-72.6	30.5	-31.5	86.4
4	-84.5	26.8	-20.4	81.6
5	-91.9	32.9	-	-
Total	**-298.5**	**102.2**	**-76.9**	**274.5**
Average	**-74.6**	**25.5**	**-19.2**	**68.6**

Deseasonalizations of data

If the original quarterly or monthly data are divided by corresponding seasonal index numbers, the resulting data are said to be deseasonalized or adjusted for seasonal variations.
Symbolically:

Deseasonalized data $= \frac{Y}{S} = \frac{TSCI}{S} = TCI$ so that the deseasonalized data contains trend, cyclical and irregular movement. We know that $Y = T + S + C + I$. Thus the data can be deseasonalized when the seasonal effects are subtracted from the corresponding original values.

MEASUREMENT OF CYCLICAL AND IRREGULAR VARIATIONS

We know that economic time series are typically the product of various factors. These factors are i) secular trend (T); ii) cyclical movements (C); iii) seasonal variations (S); iv) Irregular fluctuations (I)
Therefore original data $Y = T \times C \times S \times I$

Methods of measuring cyclical variations

There are four different methods which are commonly used for measuring cyclical variations.
They are:

i) Residual method
ii) Direct method
iii) Harmonic analysis
iv) Cyclical average method

But for the time being we shall deal with the first method only.

Residual method

This method is most commonly used in the study of cyclical variation. According to this method, the original data are first successively deseasonalized trend eliminated and cyclical relatives obtained by smoothing it further.

Procedure

1) Let us regard that the original data (Y) are the product of secular trend (T), cyclical movement (C), seasonal variation (S) and irregular fluctuations (I).
   Therefore $Y = T \times C \times S \times I$.
2) Deseasonalized the data by dividing the original data by the seasonal index
   i.e. Deseasonalised data $= \frac{T \times S \times C \times I}{S} = T \times C \times I$

3) Cyclical irregular movements are obtained by dividing the seasonal adjusted data by the trend values and expressed as percentage.

   i.e. cyclical irregular movement $= \frac{T \times C \times I}{T} = C \times I$ It does not make any difference whether the seasonal variations is first eliminated and then the trend or vice versa.

4) If the irregular movements are of minor nature to remove them smooth the series $C \times I$ by taking moving averages with a suitable chosen period and then plot the smoothed series.

Example

From the following series of annual data, for which the trend values have already been calculated, isolate cyclical fluctuations.

Year	Actual value	Trend value	Year	Actual value	Trend value
1984	170	228	1991	298	311
1985	231	239	1992	340	323
1986	261	251	1993	273	334
1987	267	263	1994	210	346
1988	278	275	1995	158	358
1989	302	287	1996	173	370
1990	299	299			

Solution

We isolate cyclical fluctuations by the residual method as shown below.

Year	Actual value A	Trend value T	Deviation of actual from trend $A - T$	Cycles $\frac{A-T}{T} \times 100$
1984	170	228	-58	-25.0
1985	231	239	-8	-3.4
1986	261	251	10	4.0
1987	267	263	4	1.5
1988	278	275	3	1.1
1989	302	287	15	5.2
1990	299	299	0	0
1991	298	311	-13	-4.2
1992	340	323	17	5.3
1993	273	334	-61	-18.3
1994	210	346	-136	-39.3
1995	158	358	-200	-56.0
1996	173	370	-197	-53.2

Note: Three yearly moving averages of the last column could also be found out to get better results.

Exercises

1. Draw a free hand trend for the following time series.

Year	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996
Value	815	885	775	781	841	765	810	820	845	780

2. Draw the trend by semi-average method for the following data

Year	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996
Value	14.0	16.2	18.1	24.6	30.8	42.4	50.8	40.8	50.2	50.4

3. Find out the trend values for the following time series of steel production by the method of moving averages using 5-year period for your purpose:

Year	production (tones)	Year	production (tones)
1979	351	1988	500
1980	366	1989	518
1981	361	1990	455
1982	362	1991	502
1983	400	1992	540
1984	419	1993	557
1985	410	1994	571
1986	420	1995	586
1987	450	1996	612

4. Below are given the figures of production of a sugar factory

Year	1991	1992	1993	1994	1995	1996
Production 'ooo' tones	92	83	94	92	92	110

5. Fit a straight line trend by the method of least squares.

Year	1988	1989	1990	1991	1992	1993	1994	1995	1996
Milk consumption (million liters)	102.3	101.9	105.8	112.0	114.8	118.7	124.5	129.9	134.8

6. Calculate the trend values by the method of moving averages, assuming a four yearly cycle, from the following data relating to sugar production in Kenya.

Year	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996
Sugar production (million kg)	37.4	31.1	38.7	39.5	47.9	42.6	48.4	64.6	58.4	38.6	51.4	84.4

7. The following table shows the number of letters posted in a particular area during a typical period of four weeks. Assuming that the trend value during the period remains the same, calculate seasonal indices (hereby daily indices) as percentage of the grand average.

Week	Sun	Mon	Tue	Wed	Thus	Fri	Sat	Total
1	18	161	170	164	153	181	76	923
2	18	165	169	147	148	190	80	927
3	21	162	169	153	155	190	82	922
4	20	165	170	155	150	180	85	925
Total	77	653	678	619	606	741	323	3,697

8. Using the data given below, explain how you would determine seasonal fluctuations in a time series.

Year	Summer	Monsoon	Autumn	Winter
1	30	81	62	119
2	33	104	86	171
3	42	153	99	221
4	56	172	129	235
5	67	201	136	302

9. The following information has been supplied by the sales department (sales are in units).

Year	Quarter			
	1	2	3	4
1993	100	125	127	102
1994	104	128	130	107
1995	110	131	133	107
1996	109	132		

You are required to

a) Calculate a four quarterly moving average of the above series.
b) Calculate the sales corrected for seasonal movements;
c) Plot the actual sales and the sales corrected for seasonal movements on a single graph and
d) Comment on your findings

10. Calculate the seasonal averages and seasonal indices for the following data of production of a commodity (in thousands of tones) for the years 1994,1995,1996

Month	1994	1995	1996	Month	1994	1995	1996
Jan	15	23	25	July	20	22	30
Feb	16	22	25	Aug	28	28	34
March	18	28	35	Sept	29	32	38
April	18	27	36	Oct	33	37	47
May	23	31	36	Nov	33	34	41
June	23	28	30	Dec	38	44	53

11. From the following data, obtain the seasonal indices by any method that you consider suitable:

Year/Quarter	I	II	III	IV
1993	97	88	76	94
1994	101	93	79	98
1995	106	96	83	103
1996	110	101	98	106

1. <u>Second Degree parabola</u>

The simplest example of the non-linear trend is the second degree parabola, the equation of which is written in the form $Y_c = a + bX + cX^2$

When numerical values for constants a, b and c has been derived, the trend value for any year may be computed substituting in the equation the value of X for that year. The value of a, b and c can be determined by solving the following the following three normal equations simultaneously:

(i) $\Sigma Y = Na + b\Sigma X + c\Sigma X^2$

(ii) $\Sigma XY = a\Sigma X + b\Sigma X^2 + c\Sigma X^3$

(iii) $\Sigma X^2 Y = a\Sigma X^2 + b\Sigma X^3 + c\Sigma X^4$

Note that the first equation is merely the summation of the given function; the second is the summation of X multiplied into the given function, and the third is the summation of X^2multiplied into given function.

When time origin is taken between two middle years ΣX and ΣX^3 would be zero. In that case, the above equation is reduced to:

i) $\Sigma Y = Na + c\Sigma X^2$

ii) $\Sigma XY = b\Sigma X^2$

iii) $\Sigma X^2 Y = a\Sigma X^2 + c\Sigma X^4$

The value of b now directly from equations (ii) and that of a and c by solving (i) and (iii) simultaneously. Thus,

$$a = \frac{\Sigma Y - c\Sigma X^2}{N}; \qquad b = \frac{\Sigma XY}{\Sigma X^2} \text{ and } \qquad c = \frac{N\Sigma X^2 Y - \Sigma X^2 \Sigma Y}{N\Sigma X^4 - (\Sigma X^2)^2}$$

Example 9

The price (in Rs) of a commodity during 1999 – 2004 is given below. Fit a parabola $Y = a + bX + cX^2$ to this data. Estimate the price of the year 2007.

Year	Price	Year	Price
1999	100	2002	140
2000	107	2003	181
2001	128	2004	192

Also plot the actual and trend value on a graph

Solution

To determine the value of a, b and c, we solve the following normal equations;

$\Sigma Y = Na + b\Sigma X + c\Sigma X^2$

$\Sigma XY = a\Sigma X + b\Sigma X^2 + c\Sigma X^3$

$\Sigma X^2 Y = a\Sigma X^2 + b\Sigma X^3 + c\Sigma X^4$

Year	Price (Y)	X	X^2	X^3	X^4	XY	X^2Y	Trend values (Y_c)
1999	100	-2	4	-8	16	-260	400	97.744
2000	107	-1	1	-1	1	-107	107	110.426
2001	128	0	0	0	0	0	0	126.680
2002	140	+1	1	+1	1	+140	140	146.506
2003	181	+2	4	+8	16	+362	724	169.904
2004	192	+3	9	+27	81	+576	1728	196.874
N=6	ΣY=848	ΣX=3	ΣX^2=19	ΣX^3=27	ΣX^4=115	ΣXY =771	ΣX^2Y=3099	ΣY_c=848.874

$848 = 6a + 3b + 19c$ (i)

$771 = 3a + 19b + 27c$ (ii)

$3{,}099 = 19a + 27b + 115c$ (iii)

Solving equations (i) and (ii), we get

$35b + 35c = 694$ (iv)

Multiplying equation (ii) by 19 and equation (iii) by 3 and subtracting, we get

$53.52 = 280b + 168c$ (v)

Solving equation (iv) and (v), we get

$c = 1.786$

Subtracting the value of c in equation (iv), we get

$b = 18.04$

Putting the value of b and c in equation (i), we get

$a = 126.68$

Thus, $a = 126.68$, $b = 18.04$ and $c = 1.786$.

Subtracting the values in the equations.

$$Y = 126.68 + 18.04X + 1.786X^2$$

When $X = -2,\ Y = 126.68 + 18.04(-2) + 1.786(-2)^2$

$$= 126.68 - 36.08 + 7.144 \quad = 97.744$$

When $X = -1, \quad Y = 126.68 + 18.04(-1) + 1.786(-1)^2$

$$= 126.68 - 18.04 + 1.786 \quad = 110.426.$$

When $X = 0,\ Y = 126.68$

When $X = 1, \quad Y = 126.68 + 18.04 + 1.786 = 146.506$

When $X = 2, \quad Y = 126.68 + 18.04(2) + 1.786(2)^2$

$$= 126.68 + 36.08 + 7.144 = 169.904$$

When $X = 3, \quad Y = 126.68 + 18.04(3) + 1.786(3)^2$

$= 126.68 + 54.12 + 16.074 = 196.874$

Price for 2007

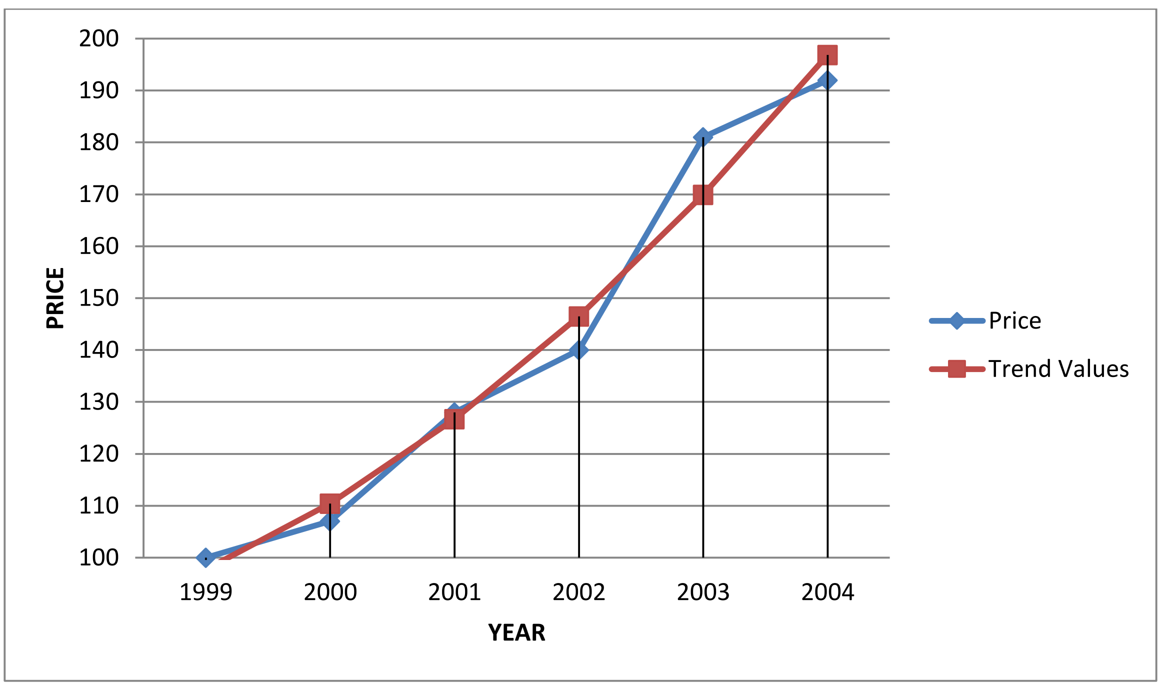
For 2007 X will be 6.

When $X = 6, \quad Y = 126.68 + 18.04(6) + 1.786(6)^2$

$$= 126.68 + 108.24 + 64.296 = 299.216$$

Thus the likely price of the commodity for the year 2007 is Rs.299.216

The graph of the actual trend values is given below.



2. Measuring Trends by Logarithms

The trends discussed so far were plotted on arithmetic scales. Trends may also be plotted on as semi-logarithmic (or semi-log) chart in the form of a straight line or a non-linear curve. A straight line on the semi-log chart shows the increase of Y values of a time series at a constant rate. A straight line on all arithmetic charts indicates the increase at a constant (amount). When it is a non-linear curve on the semi-log chart on upward curve shows the increase at varying rates, depending on the shapes of the slope. The steeper the slope, the higher is the rate of increase

The types of trend usually computed by logarithms are:

1. Exponential Trends, and
2. Growth curves.

3. Exponential Trends

The equations of the exponential curves is of the form

$$Y = ab^X$$

Putting the equations in logarithmic form, we get

$$\log Y = \log a + X \log b$$

When plotted on a semi-logarithmic graph, the curve gives a straight line; however, on an arithmetic chart the value gives a non-linear trend. In order to find out the values of a, b, the normal equations to be solved are:

$\Sigma \log \mathrm{Y} = \mathrm{N} \log \mathrm{a} + \log \mathrm{b} \Sigma \mathrm{X}$

$\Sigma(X.\log Y) = \log \mathrm{a}\, \Sigma \mathrm{X} + \log \mathrm{b}\, \Sigma \mathrm{X}^2$

When derivations are taken from middle year, i.e., $\Sigma \mathrm{X} = 0$, the above equation take the following form:

$\Sigma \log \mathrm{Y} = \mathrm{N} \log \mathrm{a}$

$\Sigma(X.\log Y) = \log \mathrm{b} \Sigma \mathrm{X}^2$ or $\log a = \frac{\Sigma \log \mathrm{Y}}{N}$; and $\log b = \frac{\Sigma(X.\log Y)}{\Sigma \mathrm{X}^2}$

Take the antilog of these expressions to arrive at the actual trend values.

Example 10

The sales of a company in lakhs of rupees for the year 1999 to 2005 are given below.

Years	1999	2000	2001	2002	2003	2004	2005
Sales(m. tones)	32	47	65	92	132	190	275

Estimate sales figures for the year 2008 using an equation of the form, $Y = ab^x$, where $x =$ years and $y =$ sales.

Solution

Fitting equation of the form $Y = ab^x$

Year	sales Y	Deviations X	X^2	$\log \mathrm{Y}$	$\mathrm{X}.\log \mathrm{Y}$
1999	32	-3	9	1.5051	-4.5153
2000	47	-2	4	1.6721	-3.3442
2001	65	-1	1	1.8129	-1.8129
2002	92	0	0	1.9638	0
2003	132	1	1	2.1206	+2.1206
2004	190	2	4	2.2788	+4.5576
2005	275	3	9	2.4393	+7.3179
$N = 7$	$\Sigma Y = 833$	$\Sigma \mathrm{X} = 0$	$\Sigma X^2 = 28$	$\Sigma \log Y = 13.7926$	$\Sigma X \, logY = 4.3237$

We have to fit the equation $Y = ab^x$. It can be written as $\log y = \log a + \log b$

Since deviations are taken from middle year, $\Sigma X = 0$

$$\log a = \frac{\Sigma \log Y}{N} = \frac{13.7926}{7} = 1.97.$$

$$\log b = \frac{\Sigma(X.\log Y)}{\Sigma X^2} = \frac{4.3237}{28} = 0.1544$$

Hence $\log Y = 1.97 + 0.1544X$

For, 2008 X would be +6.

$$\text{Log}\, Y = 1.97 + 0.1544(6) = 2.8964$$

$Y =$ Antilog 2.8964 =787.77

Thus the estimated figure sale for the year 2008 is 787.77 m. tones.

Example 11

Fit a logarithmic straight line to the following data.

Years	2000	2001	2002	2003	2004	2005
Productions (m. tones)	64	70	75	82	88	95

Solution:

Year	production Y	X	X^2	$\log Y$	$X \log Y$
2000	64	-3	9	1.8062	-5.4186
2001	70	-2	4	1.8451	-3.6902
2002	75	-1	1	1.8751	-1.8751
2003	82	0	0	1.9138	0
2004	88	+1	1	1.9445	+1.9445
2005	95	+2	4	1.9777	+3.9554
$N = 6$	$\Sigma Y = 474$	$\Sigma X = -3$	$\Sigma X^2 = 19$	$\Sigma \log Y = 11.3624$	$\Sigma X\ log Y = 5.084$

Fitting of Logarithmic straight line.

The logarithmic straight line trend is given by

$\log y = \log a + \log b$

The normal equations are:

$$\Sigma \log Y = N \log a$$

$$\Sigma X \log Y = \log a\, \Sigma + \log b\, \Sigma X^2$$

Substituting the values

$11.3624 = 6\log a - 3 \log b$

$-5.084 = -3\log a + 19 \log b$

Multiplying equation (ii) by 2 and adding to (i)

$$11.3624 = 6 \log a - 3 \log b$$

$$-10.168 = -6 \log a + 38 \log b$$

$$38 \log b = 1.1944$$

$$\log b = \frac{1.1944}{35} = 0.034$$

Putting the value of log b in equation (i)

$$11.3624 = 6\log a - 0.102$$

$6\log a = 11.4644$
$\log a = 1.911$
$\log y = 1.911 - 0.034X$

FITTING A CUBIC POLYNOMIAL BY METHOD OF LEAST SQUARES

If the time series contains certain fluctuations or cycles that tends to recur, the effects of this cycles can be eliminated by taking M.A where the no. of years in the average equals the period of the cycles

This property of eliminating unwanted fluctuations or cycles is called the smoothing of the time series. Moving average can be only smooth as time series but not eliminate the short term fluctuations entirely.

Suppose the original time series is U, U_2, U_3 ,…………. or U_t, U_{t+1}, U_{t+2}, …………….

If you want a 3 point M.A then the new series will be U_1, U_2, U_3 U_4 and U_5

i.e. $U_1 + U_2 + U_3/3$
$U_2 + U + U_4/3$
$U_3 + U_4 + U_5/3$

The M.A method for determining the trend consist of fitting a polynomial of power p to a smaller convenient set of points equivalent to the order of the moving average p<m order. The operations iss repeated with m terms starting from the 2nd , 3rd observations etc.

Example

Fit a cubic polynomial say $U_t = a_0 + a_1 t + a_2 t^2 + a_3 t^3$to a set of seven points by the method of least squares.

solution

The problem reduces to determining the coefficient of weights.

Let the actual observations of time series be U_1, U_2, U_3, U_4 , U_5 U_6 and U_7 can be coded as

U_1	-	X_{-3}
U_2	-	X_{-2}
U_3	-	X_{-1}
U_4	-	X_0
U_5	-	X_1
U_6	-	X_2
U_7	-	X_3

The sum of squares of deviation S will be

$S = \sum_{t=-3}^{3} (X_t - a_0 + a_1 t + a_2 t^2 + a_3 t^3)^2$ The conditions for the minimum are

$\frac{2s}{290}, \frac{2s}{290}, \frac{2s}{290}, \frac{2s}{290} = 0$

The normal equations are

$$\sum_{t=-3}^{3} X_t = \sum_{t=-3}^{3} [a_0 + a_1 t + a_2 t^2 + a_3 t^3]$$
$$\sum_{t=-3}^{3} t\, X_t = \sum_{t=-3}^{3} t\ (a_0 + a_1 t + a_2 t^2 + a_3 t^3)$$
$$\sum_{t=-3}^{3} t^2\, X_t = \sum_{t=-3}^{3} t^2\ (a_0 + a_1 t + a_2 t^2 + a_3 t^3)$$

$$\sum_{t=-3}^{3} t^3\, X_t = \sum_{t=-3}^{3} t^3\ (a_0 + a_1 t + a_2 t^2 + a_3 t^3)$$

t	t^2	t^3	t^4	t^5	t^6
-3	9	-27	81	-243	729
-2	4	-8	16	-32	64
-1	1	-1	1	-1	1
0	0	0	0	0	0
1	1	1	1	1	1
2	4	8	16	32	64
3	9	27	81	243	729
$\Sigma t = 0$	$\Sigma\, t^2 = 28$	$\Sigma t^3 = 0$	$\Sigma\, t^4 = 196$	$\Sigma\, t^5 = 0$	$\Sigma t^6 = 1588$

Substituting these values in the normal equation we have

$\sum_{t=-3}^{3} X_t = 7\, a_0 + 28\, a_2$ (1)

$\sum_{t=-3}^{3} X_t = a_0 \Sigma t + a_1\, \Sigma\, t^2 + a_2 \Sigma t^3 + a_3 \Sigma\, t^4$

$= 28\, a_1 + 196\, a_3$ (2)

Similarly,

$\sum_{t=-3}^{3} t^2\, X_t = 28\, a_0 + 196\, a_2$ (3)

Finally,

$\sum_{t=-3}^{3} t^3\, X_t = 196\, a_1 + 1558\, a_3$ (4)

Equation 1-3 gives

$\sum_{t=-3}^{3} X_t = 7\, a_0 + 28\, a_2$ (5)

$\sum_{t=-3}^{3} t^2\, X_t = 28\, a_0 + 196\, a_2$ (6)

Multiply equation (5) by seven then subtract it from equation (6)

$\sum_{t=-3}^{3} 7X_t - \sum_{t=-3}^{3} t^2\, X_t = 21 a_0$

$a_0 = \frac{1}{21}\left(\sum_{t=-3}^{3} 7X_t - \sum_{t=-3}^{3} t^2\, X_t\right)$ substituting the values we have

$$a_0 = \frac{1}{21}\,[7[X_{-3} + X_{-2} + X_{-1} + X_1 + X_2 + X_3] - 9X_{-3} - 4X_{-2} - X_{-1} - X_1 - 4X_2 - 9X_{-3}]$$

$$= \frac{1}{21}[-2X_{-3} + 3X_{-2} + 6X_{-1} + 7X_0 + 6X_1 + 3X_2 - 2X_3]$$

Hence the required coefficients are $\frac{1}{21}[-2, 3, 6\,, 3, -2]$

Which are the coefficients or weight to determine the trend?

Exercise

Using the method above, confirm the weight of a cubic polynomial by the method of moving average to a set of 5 points.

Ans. $\frac{1}{35}(-3, 12, 17, 12, -3)$

SOME USEFUL STOCHASTIC TIME SERIES MODELS

This method describes several different types of stochastic processes which are sometimes useful in setting up a model for a time series.

i) A purely random process (white noise)

A discrete time process is called a purely random process if it consists of a sequence of random variables X_t which are mutually independent and identically distributed. From the definition it follows that the process has a constant mean and variance and that

$$\sigma(h) = cov(X_t, X_{t+h}) = \begin{cases} \sigma^2 & h = 0 \\ 0 & h = \pm 1, \pm 2, \pm 3 \ldots \end{cases}$$

A purely random process is sometimes called a white noise particularly by engineers. Process of this type is useful in many situations as building blocks for more complicated process such as MA process.

ii) Random walk process

Suppose that $\{e_t\}$ is a discrete random process with mean μ and variance σ^2. A process $\{X_t\}$ is said to be a random walk if

$X_t = X_{t-1} + e_t, X_0 = 0$

The process is customarily started at zero when $t = 0$ so that

$X_1 \;\; = X_0 + e_1 = e_1$

$X_2 \;\; = X_1 + e_2 = e_1 + e_2$

$X_3 \;\; = X_2 + e_3 = e_1 + e_2 + e_3$

$X_t \; = X_{t-1} + e_t = e_1 + e_2 + e_3 + \cdots + e_t.$

$X_t = \sum_{i=1}^{t} e_i$

$E(X_t) = \sum_{i-1}^{t} E(e_i) = \sum_{i-1}^{t} \mu$

$E(X_t) = t\mu$

$var(X_t) = \sum_{i-1}^{t} Var(e_i) = \sum_{i=1}^{t} \sigma^2 = t\sigma^2$

Hence as the mean and variance changes with time, this process is non-stationary. However, it is interesting to note that, first difference of random walk is given by

Change in $X_t = X_t - X_{t-1} = (e_1 + e_2 + \cdots + e_t) - (e_1 + e_2 + \cdots + e_{t-1})$

Change in $X_t = e_t$

This is therefore stationary.

iii) Moving average process

Suppose that $\{e_t\}$ is a white noise with mean zero and variance σ^2, then the process $\{e_t\}$ is said to be a moving average process of order q i.e MA(q) if

$X_t = \beta_0 e_t + \beta_1 e_{t-1} + \cdots + \beta_q e_{t-q}$ which is written in summation form as

$X_t = \sum_{j=0}^{q} \beta_j e_{t-j}$ where β_j are constants.

Therefore the mean and variance are given by

$E(X_t) = \sum_{j-0}^{q} \beta_j E(e_{t-j}) = 0$ and $var(X_t) = \sum_{j-0}^{q} {\beta_j}^2 Var(e_{t-j}) = \sum_{j-0}^{q} {\beta_j}^2 \sigma^2$

$var(X_t) = \sigma^2 \sum_{j-0}^{q} \beta_j{}^2$

iv) <u>Autoregressive process</u>

Let $\{e_t\}$ be a purely random process with mean zero and variance σ^2. Then a process $\{X_t\}$ is said to be an autoregressive process of order p i.e. AR(P) if

$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \alpha_p X_{t-p} + e_t$

$X_t = \sum_{j=1}^{p} \alpha_j X_{t-j} + e_t$

<u>1st order process (AR(1))</u>

AR(1) process is given by $X_t = \alpha_1 X_{t-1} + e_t \ |\alpha| < 1$

Successively replacing the $X_t s$ we obtain

$X_t = e_1 + \alpha X_{t-1}$

$= e_t + \alpha(\alpha X_{t-2} + e_{t-1}) = e_t + \alpha e_{t-1} + \alpha^2 X_{t-2}$

$= e_t + \alpha e_{t-1} + \alpha^2(\alpha X_{t-3} + e_{t-2})$

$= e_t + \alpha e_{t-1} + \alpha^2 e_{t-2} + \alpha^2 X_{t-3}$

$X_t = e_t + \alpha e_{t-1} + \alpha^2 e_{t-2}$ +…+$\alpha^s e_{t-(s+1)} + \alpha^s e_{t-s} + \alpha^{s+1} X_{t-(s+1)}$.

Hence we have

$X_t = \sum_{j=0}^{s} \alpha^j e_{t-j} + \alpha^{s+1} X_{t-(s+1)}$

$X_t - \sum_{j=0}^{s} \alpha^j e_{t-j} = \alpha^{s+1} X_{t-(s+1)}$

Squaring both sides and taking expectation we get

$E\left(X_t - \sum_{j=0}^{s} \alpha^j e_{t-j}\right)^2 = \alpha^{2s+2} E\left(X_{t-(s+1)}^2\right)$

Note that $E(X_t) = \alpha E(X_{t-1})$ since $E(e_t) = 0$

$E(X_t) - \alpha E(X_{t-1}) = 0$

Assuming the process is stationary.

$E(X_t) = E(X_{t-1})$ then $E(X_t) - \alpha E(X_{t-1}) = 0$

$(1-\alpha)E(X_t) = 0 \qquad |\alpha| < 1$

$E(X_t) = 0$

But $\alpha \neq 1$, hence $E(X_t) = 0$

Therefore

$$E\left(X_t - \sum_{j=0}^{s} \alpha^j e_{t-j}\right)^2 = \alpha^{2s+2} Var\left(X_{t-(s+1)}\right)$$
$$= \alpha^{2s+2} Var(X_t)$$

Since X_t is stationary $Var(X_t)$ is a constant because $|\alpha| < 1$, $\alpha^{2s+2} Var(X_t)$ tends to zero as $s \to \infty$

This implies that when $|\alpha| < 1$, X_t converges in probability to $\sum_{j=0}^{\infty} \alpha^s e_{t-j}$

Hence $X_t = \sum_{j=0}^{\infty} \alpha^s e_{t-j}$ of an MA process.

$E(X_t) = \sum_{j=0}^{\infty} \alpha^j E\left(e_{t-j}\right) = 0$

$Var(X_t) = \sum_{j=0}^{\infty} \alpha^{2j} Var\left(e_{t-j}\right) = \sum_{j=0}^{\infty} \alpha^{2j} \alpha^2$

$Var(X_t) = \alpha^2 \sum_{j=0}^{\infty} \alpha^{2j} = \alpha^2(1 + \alpha + \alpha^2 + \alpha^4 + \cdots)$

$Var(X_t) = \frac{\alpha^2}{1-\alpha^2}$

$\sigma(h) = E(X_t, X_{t+h}) - \left(E(X_t)\right)^2 = cov(X_t, X_{t+h})$

$\sigma(h) = var\ (X_t) = \frac{\alpha^2}{1-\alpha^2}$.

FORECASTING

Introduction.

Forecasting the values of the future time series is an important problem in many areas including economic, periodic planning, sales forecasting and stock control.

Suppose we have in observed series $X_1, X_2 \ldots\ldots X_n$ then the basic problem is to estimate future values such as $X_n + k$ where the integer k is called the lead time.

The forecast of $X_n + k$ made at time n for n steps ahead will be denoted by $\hat{x}(n, k)$.

A wide variety of forecasting procedures are available and its important to realize that no single method is universally applicable rather the analyst must choose the procedure which is more applicable for a given set of conditions. We only consider in this procedure the Box-Jenkins approach.

The Box-Jenkins procedure

This is based on ARIMA models and is usually called the Box- Jenkins procedure. The main stages in setting up a Box-Jenkins forecasting models are as follows;

a) Model identification

The data is examined to see which member of a class of ARIMA processes appear to be appropriate

b) Estimation

After an appropriate model has been chosen, the parameters of the model are estimated.

c) Diagnostic checking

The residues from the fitted values are examined to see if the chosen model is adequate.

d) Consideration of other models if necessary

If the first model appears inadequate for some reasons, then the ARIMA model may be tied until a satisfactory model is found.

When a satisfactory model is found forecasting may readily be computed given data upto time n, this focus will involve the observations and the fitted residuals upto an including time n.

THE BOX-JENKINS APPROACH TO FORECASTING

Box and Jenkins built a general forecasting methodology from the assumption that time series generated by a stationary autoregressive moving average (ARMA) process.

Let an ARMA process (X_t) be defined by the set of equations.

$$X_t = \sum_{i=1}^{p} \alpha_i X_{t-i} + e_{t-i} + \sum_{j=1}^{p} \beta_j e_{t-j} \qquad (1)$$

Where $\{e_t\}$ is a white noise using the operator notation we write this as

$$\theta(B)X_t = \phi(B)e_t \qquad (2)$$

Where $\theta(B)$ and $\phi(B)$ are polynomials of degree p and q respectively with $\theta(0) = \phi(0) = 1$ and with all roots of the polynomial equation.

$\theta(2) = 0$ and $\phi(2) = 0$ outside the unit circle. Let also that X_t has a representation as a several linear process.

$$X_t = \sum_{i=0}^{\infty} \theta_i e_{t-i} \qquad (3)$$

Now suppose that we wish to construct a forecast X_t of the form $X_t(k)$

$X_t(k) = \sum_{j=0}^{t-1} w_j\, X_{t-j}$ (4)

From equation (3)

$X_{t-j} = \sum_{i=0}^{\infty} \theta_i\, e_{k-j-i}$

Substituting this in equation (4) we obtain

$X_t(k) = \sum_{j=0}^{t-1} w_j \sum_{i=0}^{\infty} \theta_i\, e_{t-j-i}$ This simplifies to

$X_t(k) = \sum_{j=0}^{\infty} w\, e_{t-j}$ (5)

In equation (5) we are not concern with the précised relationship between the w_j and θ_i but only with the fact that

$X_t(k)$ has been defined by equation (4) can also be in terms of the underline white noise process $\{e_t\}$ as a linear combination of current and past values we shall now evaluate the mean square forecasting error from equation (3) and (5)

$M = E\{[X_{t+k} - X_t(k)]^2\}$

$= E\left[\left\{\sum_{i=0}^{\infty} \theta_i\, e_{t+k-i} - \sum_{j=0}^{\infty} w_j\, e_{t-j}\right\}^2\right]$

$= E\left[\left\{\sum_{i=0}^{k-1} \theta_i\, e_{t+k-i} + \sum_{i=k}^{\infty}(\theta_i - w_{i-k})\, e_{t+k-i}\right\}^2\right]$

$= var\left[\sum_{i=0}^{k-1} \theta_i\, e_{t+k-i} + \sum_{i=k}^{\infty}(\theta_i - w_{i-k})\, e_{t+k-i}\right]$

$= \sum_{i=0}^{k-1} {\theta_i}^2\, var(e_{t+k-i}) + \sum_{i=k}^{\infty}(\theta_i - w_{i-k})^2\, var(e_{t+k-i})$

$= \sigma^2 \sum_{i=1}^{k-1} {\theta_i}^2 + \sigma^2 \sum_{i=k}^{\infty}(\theta_i - w_{i-k})^2$

$= \sigma^2\left[\sum_{i=1}^{k-1} {\theta_i}^2 + \sum_{i=k}^{\infty}(\theta_i - w_{i-k})^2\right]$ (6)

Using the fact that the e_t are mutually exclusive with mean μ and variance σ^2

Equation (6) is minimized by taking $w_i = \theta_i + k$ $\quad i = 0, 1, 2, \ldots$

The resulting forecast is

$$X_t(k) = \sum_{j=0}^{\infty} w_j\, e_{t-j}$$

$$= \sum_{j=0}^{\infty} \theta_j + k \quad e_{t-j}$$

$$X_t(k) = \sum_{i=k}^{\infty} \theta_i \quad e_{t+k-i} \qquad (7)$$

Comparing equation (7) and (3), the latter can be written as

$$X_{t+k} = \sum_{i=k}^{\infty} \theta_i \quad e_{t+k-i} \qquad (8)$$

We see that the optimal forecast $X_t(k)$ has the following very simple interpretations.

$X_t(k)$ is evaluated from the definition of X_{t+k} except that the future values $e_{t+1}, e_{t+2}, \ldots e_{t+k}$ are set equal to zero. The forecast errors $X_{t+k} - X_t(k)$ have equally simple interpretations in terms of the underline white noise process from equation (7) and (8).

$X_{t+k} - X_t(k) = \sum_{i=0}^{k-1} \theta_i\, e_{t+k-i}$ (9)

In particular it follows from the conditions if the polynomial $\theta(B)$ and $\phi(B)$ in (2) that $\phi(0) = 1$ and therefore that

$X_{t+k} - X_t(1) = e_{t+1}$ i.e the one-step ahead forecast errors are the elements of the underline white noise sequence $\{e_t\}$.

To apply this result in practice, it is simpler to work directly with the explicit formula (1) rather than to evaluate θ_i in terms of the original ARMA parameters α_i andβ_j.

Example

$X_t = \alpha\, X_{t-1} + e_t$

For this AR (1) process

$X_{t+1} = \alpha\, X_t + e_{t+1}$

So the optimal k step ahead forecast for k greater than 1 i.e $k > 1$ can be evaluated different but equivalent ways. The 1st is to use appropriate forecast in place of future values of X_t. Thus for the AR(1) process.

$X_{t+2} = \alpha\, X_{t+1} + e_{t+2}$ Leads to

$X_t(2) = \alpha X_t(1) = \alpha^2 X_t$.

The 2nd is to back substitute from the defining equation of the process, so as to eliminate future values of X_t thus,

$X_{t+2} = \alpha\, X_{t+1} + e_{t+2}$

$= \alpha\, (\alpha X_t e_{t+1}) + e_{t+2}$

$= \alpha^2 X_t + \alpha\, e_{t+1} + e_{t+2}$

$X_t(2) = \alpha^2 X_t$ as before.