Business Forecasting Models (Time Series and Regression Analysis)

14

SLOB Mapped against the Module

To equip oneself with application-oriented knowledge of Business Forecasting techniques to facilitate management decisions for optimisation through resource allocation, managing competition, work scheduling and managing cost overrun, demand estimation, production and cost analysis etc.

Module Learning Objectives

After studying this module, the students will be able to:

- Acquire knowledge on the basics of Business Forecasting
- Understand importance of Business Forecasting
- Understand the pros and cons of Business Forecasting as well as its limitations.
- Have idea about different types as well as methods of Business Forecasting
- Understand how concept of Time Series is used for Business Forecasting
- Understand how concept of Regression Analysis is used for Business Forecasting.

Business Forecasting Models (Time Series and Regression Analysis)

14

Business forecasting is very popular in today's world because any management is interested to know if the organisation to which they belong is financially on track. The more clarity they have over financial situation and future sales, the more empowered are they to take intelligent business decisions and ensure success.

There are many ways to judge the financial situation of a business, but forecasting is the best one for predicting the future events and trends. If the management of a company wants to know the sales figures of a future date, forecasting technique can help them analyse past and present data to generate answers. Actually Forecasting is a technique of a business's future developments based on trends, patterns and analysis of data – both current and historical. Business forecasting allows a company to make long term plans and prepare for any changes in the market. Based on the unique needs of a business, the approach to business forecasting must be targeted and tweaked to be useful. The business forecasting models typically involves the following steps –

- Collecting primary and secondary sources of data
- Analysing the datasets
- Creating strategies for projections
- Comparing forecasting model with the realised outcomes

An accurate business forecast helps to create business budgets, allocate funds, make decisions about cash flow and credit needs and to create timelines for new initiatives or acquisitions. There are different forecasting methods one can use to make predictions. What will work best depends on the circumstances. A start up business with almost no past data would not use the same method as an already established organisation.

One should know why forecasting is important and what approaches are worthwhile for his company before putting enough efforts for forecasting.

Importance of Business Forecasting in different areas of an organisation

Business forecasting is one of the most vital business tools been used by organisations towards future projections. Its importance is manifested in the following areas:

- Facilitates to gain valuable insight – Management can set smarter goals, make informed business decisions and create strategies that reflects the insights they can gain from forecasted data.
- Facilitates allocation of resources among the functional areas of the organisation as well as control the operations of the entity.
- Helps purchasing and supply department towards planning and procurement of raw materials as well as scheduling their delivery time.
- Useful for the finance department towards planning for fund sourcing to forestall cash flow constraint.
- Helps human resources department to plan the manpower requirement as the time it is required.

- Plays a deciding role for a manufacturing entity to schedule its production – when to produce at full capacity and when not to.
- Useful to determine well in advance expected return on investment fairly correct.
- Facilitates learning from past mistakes causing eradication of weaknesses in the current system.
- Useful for cost reduction and profit increase. The management can stick to an accurate budget based on current market conditions and expected future outcomes.
- Helps to foresee market trends causing the management to take measures proactively.

Beyond inter business benefits, any management can also use forecasting to keep customers happy by providing services and the products they are looking for. This leads to greater brand loyalty and of course better profits in the future.

A forecast can also help businesses apply for loans or new lines of credit since many financial institutions will request a report before approving a commercial application.

Pros and Cons of Business Forecasting

Forecasting relies heavily on the past data. Although past events are analysed as a guide to the future, a question is always raised as to the accuracy of these recorded events. In fact the predictions are never 100% accurate, even if they are based on actual numbers of the past.

Any management strives for accuracy which needs more thorough collection and analysis of data. But that can be time consuming, expensive and resource intensive which every business cannot handle. That’s why many companies don’t use forecasting as regularly or consistently as they should.

However, the results that come from committing to forecasting as a business activity are usually worth the effort because one can gain a lot of clarity or even debunk some of the assumptions which are holding back from realizing greater profits. The bottom line is that predicting future trends and events enables informed business decisions.

Limitations of Business Forecasting

Business Forecasting, though a very useful tool for present day businesses, has the following limitations.

- The past data used for forecasting cannot be relied upon to a great extent.
- Accuracy in judgement is almost impossible due to presence of bias.
- Inconsistencies exist in the measurement of forecasts.
- Some methods of forecasting are tedious and cumbersome.
- In most cases forecasting is expensive. It increases the price of the products while failing to reduce the uncertainty attached to the business.
- The veracity of data may be doubtful as they can be wrongly compiled and recorded.

Types of Business Forecasts

There can be various types of Business forecasts depending on what the company wants to know or predict. It can range from the general (Sales next quarter) to the incredibly specific (Demand for a specific product during rainy season). Most common type of business forecasts are as follows.

1. **General Business Forecasting** is used to determine the overall business climate for a future date and can be widely applied for many different businesses and industries. This is used for determining overall market conditions and the impact of the environmental factors in which the business operates. It is best suited for businesses operating in influential environments, such as countries experiencing political upheavals, major technological advancements, dramatic seasonal shifts etc.

2. **Financial Forecasting** is about getting a clear picture of where a company is headed for. It includes weighing assets and liabilities, accounts payable and receivable, operating costs, capital structure and cash flow and general market conditions. It is used for tracking the future trajectory of a company as a whole. This is best suited for any business looking to stay on top of its business's health through financial projections.
3. **Accounting Forecast** is the practice of predicting the future costs which will be incurred by a company, using past and present data to estimate how much the business will pay for raw materials, inventory, man hours, utilities and rent, insurance and more. This is used for determining the future operating costs of a business and suitable for any business concerned with covering future costs. Estimating cyclical changes in a seasonal product's cost (such as the cost of mangos) could be taken as an example.
4. **Demand Forecasting** goes hand in hand with a sales forecast. Demand forecast will predict what the market needs and the sales forecast will predict how a business will be able to capitalize with those needs. This is used for determining market and customer demand for any goods or services in the future and best suited for planning how much to invest in raw materials or inventory, deciding if a new product will perform well.
5. **Sales Forecasting** is used to estimate future sales of a specific product or service within the rage of offerings of a business, using the available sales data. It allows the management to anticipate the future needs of workforce, resources, cash flow, inventory and capital investment. It will provide the management with the figures of expected revenue over the next month or next quarter or next year of a sales cycle. This is useful for predicting sales of a future period and estimating growth as well as cash flow. This is best suited for businesses relying solely on sales history or looking to project sales for investors and funding.
6. **Capital Forecasting** is based on current and future assets and liabilities as well as predictions for liquid capital and cash flow estimates. It is tricky and not as reliable as the other types of forecasts because it involves a lot of factors like Cash and Savings, Assets, Accounts receivable, Revenue, Investment funding, Lines of credit etc. It is mainly used for predicting available capital for a future date or event and best suited for companies preparing for investment, growth, hiring, acquisitions or other changes that will require cash.

Methods or models of Business Forecasting

A Forecasting Type identifies the target a company is pursuing (like sales, cash flow etc.) whereas Forecasting.

Method deals with the method of gathering and identifying data (qualitative versus quantitative). In fact the Forecasting Method is the tool a company use to gather and evaluate the relevant data for its Forecasting Type.

While there are several forecasting methods or models, they all fall within two general categories – QUANTITATIVE and QUALITATIVE. Depending on the available data and the age of the business, one approach will be more beneficial than the other. Quantitative Forecasting focuses on structured data, statistical analysis and experiments while Qualitative Forecasting uses unstructured data since it relies on interviews, surveys and observations.

Quantitative Methods of Forecasting – As the name implies Quantitative Method of Forecasting is all about numbers and measureable data. These models focus on existing data, numbers and formulae and there is very little human interfacing. The analysis is statistical, the outcomes are conclusive and the patterns observed provide a straightforward course of action. When cause-effect relationships are discovered or suspected, a business can leverage the variables for maximum benefit. These models help to get answers like "How many" and "How often" which are helpful when businesses are adding new services or products or making adjustments to existing ones. It's also beneficial for predicting sales figures from one year to the next.

Some of the most frequently used Quantitative Methods of Forecasting are –

- Regression Analysis
- Time Series Analysis

- Exponential Smoothing
- Input – Output Models

Qualitative Methods of Forecasting – These methods aim to gain a qualitative understanding of a given subject, problem or point of interest. Instead of numbers and formulae, qualitative forecasting focuses on human feedback from experts and customers. It is an unstructured, broad, non-statistical approach with subjective interpretations. The outcome of these forecasts is there to help management develop a further understanding of a previously posed question. For example, a business might want to know it's customers' thoughts about a new product. Forecasters will use a qualitative research to measure their opinions. This approach is best for businesses that don't have enough raw data to reach an accurate quantitative forecast. It may also be used by companies who don't need numbers to lead them in the right direction but want to know what brings the most value to their customers.

Under the umbrella of qualitative forecasting there are useful techniques given as –

- Historical Analogy Method
- Executive Opinion Method
- Survey Techniques
- Barometric Techniques
- Delphi Technique

Above mentioned Methods or Models of Business Forecasting are discussed in detail in the following paragraphs

Regression Analysis

In one of the methods of Quantitative Forecasting (also known as Casual Method) a study on relationship between two (or more) variables is conducted where one variable affects the other/s. In other words two or more variables have a cause-effect relationship.

In general the dependent variable represents the prediction. An example is the Sales Growth of certain products. The independent variable might describe the issue or point of interest, for example, short staffed warehouse crew. The Casual Model then asks how the shortage of staff going to affect future sales of the product.

The answer to the question mentioned above can be obtained if it is possible to establish certain relationship between the variables by means of mathematical or statistical equations. Simple Correlation technique is useful for establishing associative relationship between two variables, while the Regression technique (Simple or Multiple) is meant to isolate the casual relationship between them. Several Regression Models are available to test and establish a statistically satisfactory fit between the dependent variable and a specific range of independent variables. Forecasts are made by substituting values of the independent variables in the equation and then computing the value of the dependent variable. These methods are useful for long term forecasting and are relatively more sophisticated and expensive to use.

Prediction or estimation is one of the major issues in almost all spheres of human activity. The estimation or prediction of future production, consumption, prices, investments, sales, profits, income etc. are of paramount importance to a businessman or an economist. Regression Analysis is one of the very scientific techniques for making such predictions. In the words of M.M.Blair "Regression Analysis is a mathematical measure of the average relationship between two or more variables in terms of the original units of data"

In the day to day life we experience numbers of inter-related events. For example, the weight of a human being is dependent on the height of that person, the demand for a particular product is dependent on its price and so on. The Regression Analysis confined to the study of only two variables at a time is called Simple Regression and it is called Multiple Regression when the study involves more than two variables.

In each of the examples given above there is a dependent variable whose value changes depending on the value of the independent variable. In Regression Analysis, the independent variable is called *Regressor* or *Explanator* and the dependent one is called *Regressed* or *Explained* variable.

Simple Linear Regression (SLR)

We confine our study to the linear models of Regression. Simple Regression mentioned above is also called *Simple Linear Regression* or *SLR*. It involves two variables – one Explanatory (or Independent) and one Response (or Dependent). The Explanatory variable is manipulated by the investigator while the Response variable responds to the manipulation of the Explanatory one. When "Time spent studying for an examination" is the Explanatory variable then the Response variable could be "Examination score". SLR Analysis yields a regression model between two variables that can be used to make predictions or estimations about observations inside and outside the range of the data used to make the regression model. One important component of conducting a SLR Analysis is Ordinary Least Squares (which is also termed OLS in short). OLS is a method for estimating the unknown parameters of a linear regression model. In every regression analysis there are independent and dependent variables that are being studied from a dataset. In the majority of the cases this dataset is a list of observations for the dependent and independent variables from individuals that make up a sample. This sample is taken from the population because it is extremely rare to have the data from the entire population. The population is made up of the individuals, objects or ideas we want to study. The measurements of the population are called Parameters while each measurement of a sample is called a Statistic. When it comes to linear regression there are many different ways we can try to estimate these population parameters with the help of statistic of the samples. OLS is so vital in case of linear regression because it minimizes the difference between the observations in the dataset and the predicted responses of the linear regression model.

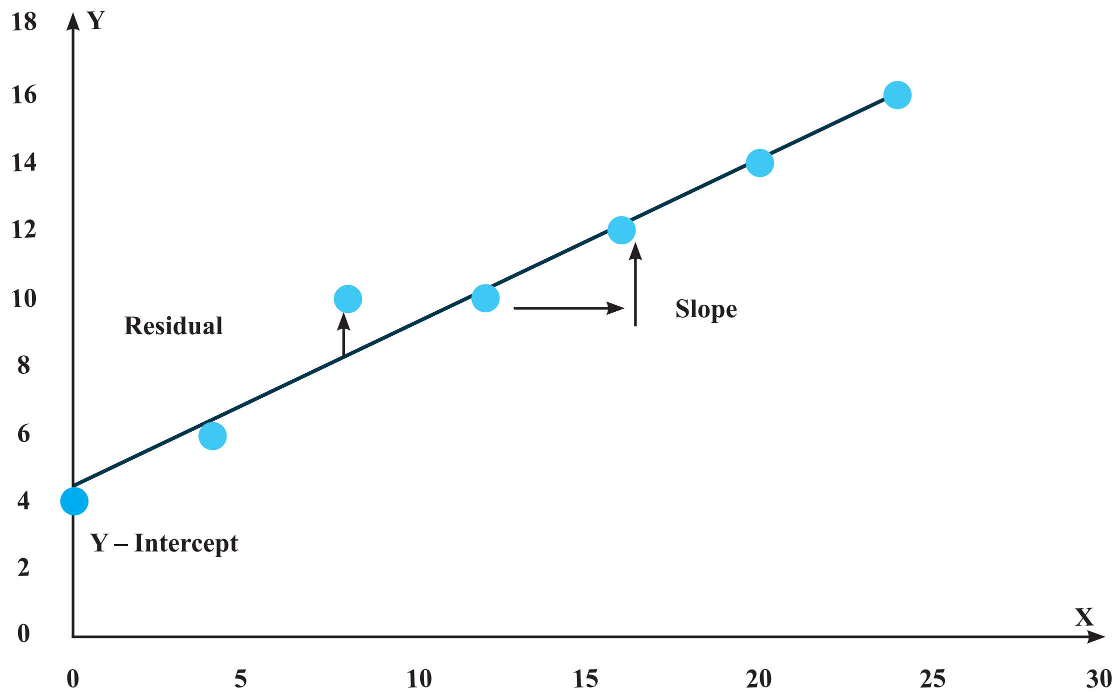

Figure No. 14.1: Linear Regression Line

The formulae for the Regression Line, Estimators and Residuals of OLS are given as follows:

Model : $y = \alpha + \beta x + u$ Here u indicate the amount of Error

Slope : $b = r(\sigma_y / \sigma_x)$ This is the slope of the Regression model (r = Correlation Coefficient & σ = S.D)

y intercept : $a = \bar{y} - b\bar{x}$ This is the y intercept of the Regression model

Residual : $[(y_i) - (y_i)']$ This is the difference between the observed and the predicted values of y

The reason behind the use of OLS Regression model is the fact that the line obtained by this method has least value of the sum of the squares of all the residuals. In other words it is the regression line for which sum of the squares of all residuals or the difference between the dataset and the predicted data points is minimum.

It can be seen that the Linear Regression line includes Residuals which are the estimates of Error. Rearranging the formula for the Model we see the Residuals given as $u_i = y_i - \alpha - \beta x_i$ because the best Linear Regression Line is one that reduces the distance between its predictions and the actual observed values from the sample, the goal would naturally be to want to reduce the residuals for all points in the sample. As residuals can be positive or negative, depending on whether the line under or overestimates y, they are squared so that equal comparison of the positive residuals with the negative ones is possible. This concept is shown in the diagram below.

Since the aim is to reduce the square of each of these residuals it is called Ordinary Least Squares or OLS.

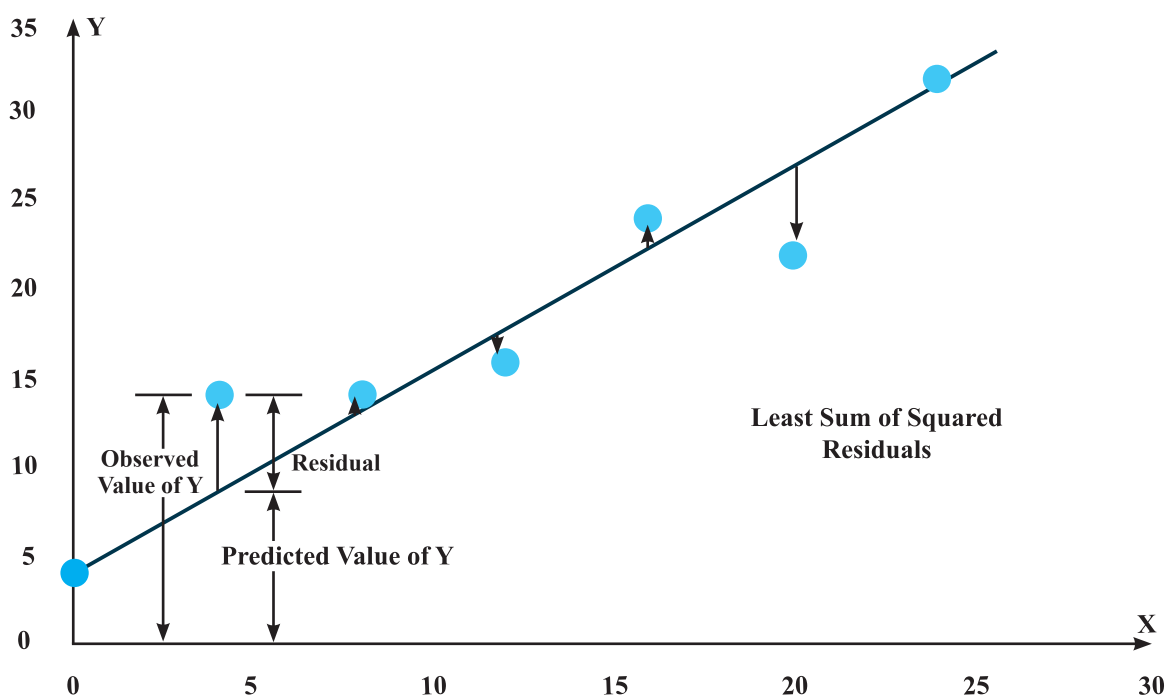

Figure No. 14.2: Concept of Residuals

Understanding the meaning of the Estimators within the formula of Linear Regression Line is important. The slope of the Regression Model is called Regression Coefficient while y intercept is known as y intercept only.

Suppose there is a Regression relation between the variables Sales (y) and Price (x) given as y = 40 -5x, where Price is in Rupees and Sales is in Units. As the slope (b) is negative, this means that for an unit increase of price, there would be a decrease of 5 units in Sales.

Illustration 1

There are two variables that need to be studied – Exports of raw cotton and Imports of manufactured goods into India. Following dataset for 7 years is provided. What kind of regression model should be used here? What are the results of this regression? Interpret the model estimators.

	₹ in Crores						
Exports	42	44	58	55	89	98	60
Imports	56	49	53	58	67	76	58

Solution:

As only two variables are involved, Simple Linear Regression or SLR model should be used.

Let the Regression model be $y = a + bx$ where x = Exports and y = Imports

Using the concept of Least Squares the values of a and b can be calculated. For that the Normal Equations are -

$\Sigma y = a.N + b.\ \Sigma x$ (where N = No. of pairs of observations) and $\Sigma xy = a.\Sigma x + b.\Sigma x^2$

Sl. No.	x in ₹ Crores	y in ₹ Crores	x^2	xy
1	42	56	1764	2352
2	44	49	1936	2156
3	58	53	3364	3074
4	55	58	3025	3190
5	89	67	7921	5963
6	98	76	9604	7448
7	60	58	3600	3480
Total	446	417	31214	27663

Putting the values of Σy, Σx and N in the 1st Normal Equation we get $417 = a.7 + b.\ 446$..........................(1)

Similarly putting the values of Σxy, Σx and Σx^2 in the 2nd Normal Equation we get

$27663 = a.446 + b.31214$..(2)

From (1) we get $a = (417 - 446b)/7$

Substituting the above expression for 'a' in (2) we get $27663 = 446.(417 - 446b)/7 + 31214b$

Or, $27663 = 26568.86 - 28416.57b + 31214b$

Or, $1094.14 = 2797.43b$ Or, $b = 0.39$

Putting $b = 0.39$ in the expression for 'a' we have $a = (417 - 446 \times 0.39)/7 = 34.72$

Substituting the values of a and b in the assumed model we get, $y = 34.72 + 0.39x$

Thus, the given data can be fit to the SLR Model $y = 34.72 + 0.39x$

From this, Estimators Slope and y intercept can be interpreted as follows –

Slope = b = 0.39 indicates the amount of change in y (Imports) per unit change in the value of x (Exports)

y intercept = value of y when x = 0 i.e 34.72 indicates that even for no Exports there will be Imports worth 34.72 Crores of Rupees.

Important notes on SLR

1. SLR deals with only two variables x and y of which x is conventionally considered as the Independent one, but on a number of occasions y is considered as Independent, too. Thus there can be two types of SLR models known as – Regression line y on x (where x is Independent) and Regression line x on y (where y is Independent).
2. General form of Regression line y on x is y = a +bx. Using statistical measures this equation is also written as $y - \bar{y} = b_{yx}(x - \bar{x})$ where $\bar{x}$ and $\bar{y}$ are Means of x & y and b_{yx} = Regression Coefficient y on x = $r(\sigma_y / \sigma_x)$

   r = Correlation Coefficient and σ_y & σ_x are S.Ds of y and x respectively
3. Similarly General form of Regression line x on y is x = a + by. This equation is also expressed as

   $x - \bar{x} = b_{xy}(y - \bar{y})$ where $\bar{x}$ and $\bar{y}$ are Mean values of x & y and b_{xy} = Regression Coefficient x on y = $r(\sigma_x / \sigma_y)$ r = Correlation Coefficient and σ_y & σ_x are S.Ds of y and x respectively
4. Two Regression lines intersect at the point $(\bar{x}, \bar{y})$.
5. Sign of b_{yx} and b_{xy} should be same always – either both positive or both negative.
6. $b_{yx}.b_{xy} = r^2$ Or, $r = \pm\sqrt{b_{yx}.b_{xy}}$ [r will take the sign same as b_{yx} and b_{xy}]
7. Slope of Regression line y on x is b_{yx} and that of x on y is $1/b_{xy}$
8. Two Regression lines coincide if and only if there is a perfect correlation between x and y i.e r = ±1 In such situation slopes of the two lines will be equal i.e $b_{yx} = 1/b_{xy}$ Or, $b_{yx}.b_{xy} = 1$
9. Two Regression lines will be perpendicular to each other if $b_{yx}.\ 1/b_{xy} = -1$ Or, $b_{yx} = -\ b_{xy}$

   [Correlation Coefficient = r = Cov (x,y)/ $\sigma_x.\sigma_y$ where Cov (x,y) = $\Sigma xy/N - (\Sigma x/N).(\Sigma y/N)$

   & $\sigma_x = \sqrt{[\Sigma x^2/N - (\Sigma x/N)^2]}$

Multiple Linear Regression (MLR)

Multiple Linear Regression or MLR is the extension of Simple Linear Regression. Basic concepts behind MLR are same as those of SLR. The main difference however is that MLR has one response variable with more than one explanatory variable. In fact majority of the phenomenon around us - the demand for goods, growth of plants etc. typically have more than just one variable related to them. Mathematically this also makes sense the more variables are added to the model, the predictions from the model become better.

However introducing more variables means extra precaution should be taken during analysis. Having a high value of r^2 doesn't always mean the best regression model is in use. Often too high value of r^2 can signal towards underlying problems with the model. Some of the common problems one can encounter while building MLR Model are as given in the following table:

Concept	Definition	Resulting problem
Over-fitting	Adding too many predictors	The model is too closely related or fit to the sample dataset to the point that it introduces a lot of variability.
Under-fitting	Adding too few predictors	The model does not fit the data well enough because it is not complex enough to the point that it introduces bias
Multi-collinearity	Pairs of explanatory variables are too highly correlated.	Reduces the reliability of the model because it affects the variance.

The first two concepts are often referred to as the bias – variance trade-off. The more complex a model is, higher is the risk of over-fitting the data and therefore having higher variance. The less complex is the model, higher is the risk of under-fitting the data and therefore having higher bias. The best model finds the sweet spot between the over-fitted and under-fitted models which can be visualized in the graph below.

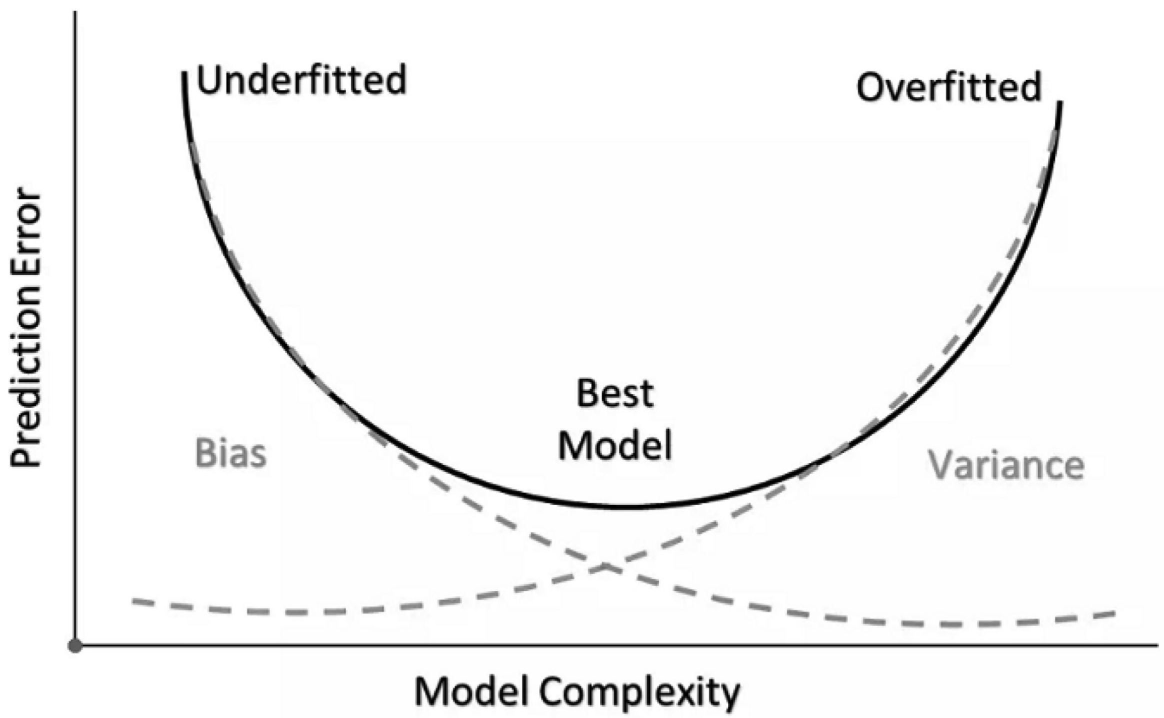

Figure No. 14.3: Model Complexity vs Prediction Error

The formula of Multiple Regression model is given as $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_n x_n + \varepsilon$ where x_1, x_2 ...... x_n are n number of linear Independent (or Explanatory) Variables and β_1, β_2 .... β_n are the respective coefficients of the Independent variables. β_0 is the Intercept and ε is the Error. The Error term is based on the real population parameters. While using the MLR equation to estimate a value of the Response (y), this Error term is assumed to be zero because true population parameters are not known.

For the MLR Model $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$ calculation of the coefficients β_1 and β_2 & the intercept β_0 are done by using the formulae given below.

$\beta_0 = \bar{y} - \beta_1\bar{x}_1 - \beta_2\bar{x}_2$ where $\bar{y}$ = Mean value of y, $\bar{x}_1$ = Mean value of x_1 and $\bar{x}_2$ = Mean value of x_2

$\beta_1 = [\Sigma(X_2)^2.\Sigma(X_1y) - \Sigma(X_1X_2).\Sigma(X_2y)] \div [\Sigma(X_1)^2.\Sigma(X_2)^2 - (\Sigma X_1X_2)^2]$

$\beta_2 = [\Sigma(X_1)^2.\Sigma(X_2y) - \Sigma(X_1X_2).\Sigma(X_1y)]\ \ [\Sigma(X_1)^2.\Sigma(X_2)^2 - (\Sigma X_1X_2)^2]$

where $\Sigma(X_1)^2 = \Sigma(x_1)^2 - (\Sigma x_1)^2/N$, $\Sigma(X_2)^2 = \Sigma(x_2)^2 - (\Sigma x_2)^2/N$

$\Sigma X_1y = \Sigma x_1y - (\Sigma x_1.\Sigma y)/N$, $\Sigma X_2y = \Sigma x_2y - (\Sigma x_2.\Sigma y)/N$ and $\Sigma X_1X_2 = \Sigma x_1x_2 - (\Sigma x_1.\Sigma x_2)/N$

N = Number of observation sets

Illustration 2

In order to find out the effect of Educational Qualification and Experience on the Earnings of the workers of a CNC Machine Shop, a study has been conducted on 10 workers and the dataset below is provided to you.

Worker	1	2	3	4	5	6	7	8	9	10
Education	11th Std	11th Std	12th Std	12th Std	1st Yr. Bach.	2nd Yr Bach.	2nd Yr Bach.	1st Yr. Masters	1st Yr. Masters	1st Yr. Masters
Experience (Years)	10	6	10	5	5	6	5	8	7	2
Salary (₹ per month)	30000	27000	20000	25000	29000	35000	38000	40000	45000	28000

Conduct a Multiple Regression Analysis considering the following numerical equivalents for Educational Qualification – 11th Std. = 11, 12th Std. = 12, 1st Yr Bachelors = 13, 2nd Yr. Bachelors = 14, 1st Yr. Masters = 16 Also interpret the meaning of the coefficients of the variables and the constant term.

Solution:

Let the Multiple Regression Model be $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$

For calculation of β_0, β_1 and β_2 following formulae to be used.

$\beta_0 = \bar{y} - \beta_1\bar{x}_1 - \beta_2\bar{x}_2$ where $\bar{y}$ = Mean value of y, $\bar{x}_1$ = Mean value of x_1 and $\bar{x}_2$ = Mean value of x_2

$\beta_1 = [\Sigma(X_2)^2.\Sigma(X_1y) - \Sigma(X_1X_2).\Sigma(X_2y)] \div [\Sigma(X_1)^2.\Sigma(X_2)^2 - (\Sigma X_1X_2)^2]$

$\beta_2 = [\Sigma(X_1)^2.\Sigma(X_2y) - \Sigma(X_1X_2).\Sigma(X_1y)] \div [\Sigma(X_1)^2.\Sigma(X_2)^2 - (\Sigma X_1X_2)^2]$

where $\Sigma(X_1)^2 = \Sigma(x_1)^2 - (\Sigma x_1)^2/N$, $\Sigma(X_2)^2 = \Sigma(x_2)^2 - (\Sigma x_2)^2/N$

$\Sigma X_1y = \Sigma x_1y - (\Sigma x_1.\Sigma y)/N$, $\Sigma X_2y = \Sigma x_2y - (\Sigma x_2.\Sigma y)/N$ and $\Sigma X_1X_2 = \Sigma x_1x_2 - (\Sigma x_1.\Sigma x_2)/N$

N = Number of observation sets = 10 (In this case)

Worker	Salary (₹ y)	Education (x_1)	Experience (x_2 years)	$(x_1)^2$	$(x_2)^2$	x_1y	x_2y	x_1x_2
1	30000	11	10	121	100	330000	300000	110
2	27000	11	6	121	36	297000	162000	66
3	20000	12	10	144	100	240000	200000	120
4	25000	12	5	144	25	300000	125000	60
5	29000	13	5	169	25	377000	145000	65
6	35000	14	6	196	36	490000	210000	84
7	38000	14	5	196	25	532000	190000	70
8	40000	16	8	256	64	640000	320000	128
9	45000	16	7	256	49	720000	315000	112
10	28000	16	2	256	4	448000	56000	32
Total	317000	135	64	1859	464	4374000	2023000	847

Putting the values of $\Sigma(x_1)^2$ and Σx_1 in the above mentioned formula we have $\Sigma(X_1)^2 = 1859 – (135)^2/10 = 36.5$

Putting the values of $\Sigma(x_2)^2$ and Σx_2 in the above mentioned formula we have $\Sigma(X_2)^2 = 464 – (64)^2/10 = 54.4$

Putting the values of $\Sigma(x_1y)$, Σx_1 & Σy in the above mentioned formula we have

$\Sigma(X_1y) = 4374000 – (135)(317000)/10 = 94500$

Putting the values of $\Sigma(x_2y)$, Σx_2 & Σy in the above mentioned formula we have

$\Sigma(X_2y) = 2023000 – (64)(317000)/10 = - 5800$

Putting the values of $\Sigma(x_1x_2)$, Σx_1 & Σx_2 in the above mentioned formula we have

$\Sigma(X_1X_2) = 847 – (135)(64)/10 = - 17$

Substituting the values calculated above for $\Sigma(X_2)^2$, $\Sigma(X_1y)$, $\Sigma(X_1X_2)$, $\Sigma(X_2y)$ and $\Sigma(X_1)^2$ in the formula of β_1 we get

$\beta_1 = [54.5x94500 – (-17)x(-5800)] / [36.5x54.4 – (-17)^2] = 2977.51$

Substituting the values calculated above for $\Sigma(X_2)^2$, $\Sigma(X_1y)$, $\Sigma(X_1X_2)$, $\Sigma(X_2y)$ and $\Sigma(X_1)^2$ in the formula of β_2 we get

$\beta_2 = [36.5x(-5800) – (-17)x94500] / [36.5x54.4 – (-17)^2] = 822.11$

Also we have $\bar{y} = \Sigma y /N = 317000/10 = 31700$, $\bar{x}_1 = \Sigma x_1/N = 135/10 = 13.5$ & $\bar{x}_2 = \Sigma x_2/N = 64/10 = 6.4$

So $\beta_0 = \bar{y} – \beta_1\bar{x}_1 – \beta_2\bar{x}_2 = 31700 – 2977.51x13.5 – 822.11x6.4 = – 13757.75$

Required Multiple Regression Model is **$y = -13757.75 + 2977.51x_1 + 822.11x_2$**

Meaning of β_1**:-** 1 unit increase in Education (x_1) will cause increase in Salary(y) by ₹ 2977.51 when x_2 = Const.

Meaning of β_2 **:-** 1 unit increase in Experience (x_2) will cause an increase in Salary (y) by ₹ 822.11 when Educational Qualification (x_1) is held constant.

Meaning of β_0 **:-** When both the predictors Educational Qualification (x_1) and Experience (x_2) are zero then the value of y is (– 13757.75)

Illustration 3

The CNC Machine Shop mentioned in the problem above is going to bid for few very critical machining items required for a Nuclear Power project. For that they have to recruit a dedicated programmer for the machines. Besides they need to procure two new CNC Vertical Machining centres which have to be operated by qualified operators. The profile they are looking for the programmer is Graduate Engineer with 5 years of experience and that for the operators of the new machining centres is 2nd year Bachelors' degree holder with 8 years of working experience. If 1000 nos. of the critical items are to be supplied within a period of 3 years then how much extra the company should bid for each item to accommodate the recruitments of a programmer and two operators. Use the dataset of the previous illustration. [Assume educational qualification of a Graduate Engineer = 20 and that of an operator with 2nd year Bachelors' degree = 14]

Solution:

As the dataset of the previous illustration is to be used, the Regression model should be as find out there.

Therefore $y = -13757.75 + 2977.51x_1 + 822.11x_2$ with y = Salary/month**,** x_1= Education & x_2= Experience

Estimated monthly salary of the Graduate Engineer having x_1= 20 and x_2 = 5 years is

$y = -13757.75 + 2977.51x20 + 822.11x5$ = ₹ 49903

Similarly estimated monthly salary of each of the Operator having x_1= 14 and x_2= 8 years is

$y = -13757.75 + 2977.51x14 + 822.11x8$ = ₹ 34504.27

Hence total estimated monthly outlay of the company for accommodating the programmer and 2 operators is

49903 + 2x34504.27 = ₹ 118911.54

As the supply has to be completed within 3 years, the total estimated outlay = 118911.54x3x12 = ₹ 4280815.44

So for 1000 items, the company has to bid 4280815.44/1000 = ₹ 4280.81 extra per piece to accommodate the new recruits.

Time Series Analysis

The term "*Time Series*" refers to a series of observations recorded in accordance with the time of occurrence. A study of time series data discloses that observed values of the variable are always fluctuating from time to time. The fluctuations are the result of the joint action of various forces, like changes in tastes and habits of people, increase in population, development of new technology resulting in lower cost of production, changes in environmental conditions etc. The forces are ever-changing, and due to the interaction among them, values of the variable undergo change with the passage of time. The objects of time series analysis are to isolate and measure the effects of various components. Such an analysis helps understand the past behaviour, so that the future tendency may be predicted. To business executives who have to plan their production much ahead of sales, the analysis of time series is of great importance in planning for the hiring of personnel for peak periods, to accumulate an inventory of raw materials, to keep the equipment ready and finally in forecasting the future demand of their product.

Following are few examples of Time Series data –

- Profits earned by a company for each of the past ten years
- Number of students registered for CMA Examination for the past decade.
- Percentage change in quarterly Consumer Price Index (CPI) of a country for the last 45 years
- Number of employees hired by a company for each of the last five years

Components of Time Series data

Graphical representation of the data of Percentage change in quarterly Consumer Price Index (CPI) of a country during the period 1960 to 2005 reveals the changes over time. This is shown in the graph below However these changes are not totally haphazard and a part of it at least can be accounted for.

The four components of Time Series are –

- Secular Trend or Trend (T)
- Seasonal Variation (S)
- Cyclical Fluctuation (C)
- Irregular or Random Movement (I)

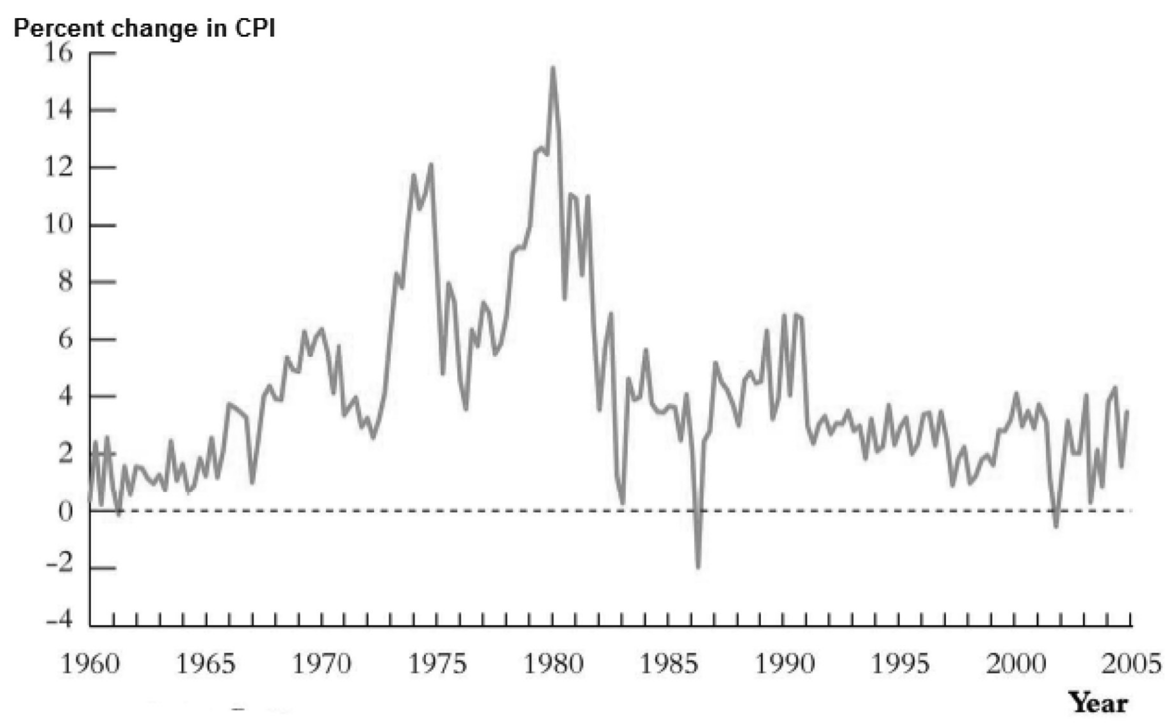

Figure No. 14.4: Percentage change in CPI of a country during the period 1960 to 2005

In the traditional approach, it is assumed that there is a multiplicative relationship among the four components. In other words any particular observation of a time series data is considered to be the product of the effects of these components. Thus, $y_t = T \times S \times C \times I$ as per Multiplicative Model

Another approach is to assume an additive relationship among them. Thus $y_t = T+S+C+I$ as per *Additive Model*.

Although the additive model facilitates easier calculations, it has been found inappropriate in many practical situations and hence is not generally used.

Secular Trend or simply Trend of time series is the smooth, regular and long term movement exhibiting the tendency of growth or decline over a period of time. In a time series, observations are taken over successive periods. Though most of these data generally display some random fluctuations, the time series may still show gradual shifts to relatively higher or lower values over an extended period. This gradual shifting is referred to as the trend of the time series. A trend emerges due to one or more long term factors, such as changes in population size, changes in the demographic characteristics of the population, changes in tastes and preferences of customers etc. For example manufacturers of automobiles in India may see that there are substantial variations in automobile sales from one month to the next. But in reviewing auto sales in the past 15 to 20 years, the manufacturers may discover a gradual increase in annual sales volume. In this case trend for auto sales is increasing over time. It is a standard practice to describe an increasing trend by an upward sloping line and a decreasing trend by a downward sloping line. The line mentioned here may be a straight line or a curved line representing linear or non-linear trends respectively.

Seasonal Variation is a type of periodic movement where the period is at the most one year. Business activities are found to have brisk and slack periods at different parts of the year. This ups and downs recurring with remarkable regularity year after year, is attributable to the presence of seasonal variation. For example usage of swimming pools experience low values in the fall and winter months, but witness peak values during summer and springs. On the other hand manufacturers of woollen garments experience exactly opposite behaviour.

Cyclical Fluctuation is another type of periodic movement where the period is more than a year. Such movements are fairly regular and oscillatory in nature. One complete period is called a cycle. Although a time series may often exhibit a trend over a long period, it may also display alternating sequences of points that lie above and below the trend line and last more than a year. The time series of the aggregate output in the economy (called real Gross Domestic Product or GDP) provides a good example of cyclical fluctuation in Time Series. While the trend line for GDP is sloping upwards, the output growth displays a cyclical behaviour around the trend line. This cyclical behaviour of GDP has been dubbed as business cycle by the economists.

Irregular or Random movements are such variations which are caused by factors of erratic nature. These are completely unpredictable or caused by such unforeseen events as war, natural calamities, strike, lockout etc. Random Movements do not reveal any pattern of the repetitive tendency and may be considered as residual variation.

A very common practice for the purpose of forecasting is to segregate the above mentioned components from the observed dataset of Time Series. Segregation of the components facilitates to increase accuracy of the prediction being made for a future date.

Measurement of Trend

There are four methods of isolating Secular Trend from the Time Series data.

- Free Hand Method
- Semi Average Method
- Moving Average Method
- Fitting Mathematical Curves

In **Free Hand Method**, the given data are plotted as points on a graph paper. The time series data (y_t) are taken along vertical axis and time (t) along horizontal. Then a smooth free hand curve is drawn through the scatter of the plotted points which appears to represent their pattern of movement over time. The drawn curve is the Trend line and the value of y_t against any value of t can be obtained from this. This method not much used for actual forecasting because too much of individual judgement is involved in it. It is used only to get a preliminary idea about the possible nature of the trend line.

Semi Average Method deals with dividing the dataset into two equal parts and then finding average of each. These averages are plotted as points on a graph paper against the mid-point of the time interval covered by each part. The straight line joining the two points is considered to be the trend line. Although the method is simple to apply, it may lead to poor results when used indiscriminately.

Moving Average Method is very commonly used for the isolation of trend and in smoothing out fluctuations in time series. In this method a series of arithmetic means of successive observations known as Moving Averages are calculated from overlapping groups of successive elements of the given data and these moving averages are used as trend values. Each moving average is based on values covering a fixed time interval called *Period of Moving Average* and is shown against centre of the first. The composition of items is adjusted successively by replacing the first observation of the previously averaged group by the next observation below that group. Thus the moving average for period k is a series of successive averages of observations at a time starting with 1st, 2nd, 3rd **........** to k terms. Thus the first average is the mean of the 1st to k terms, the second is the mean of the k terms from the 2nd to the (k+1)th term, the third is the mean of the 3rd to the (k+2)th term and so on.

When **the period of moving average is ODD** then each moving average calculated is shown against the middle most value of the concerned group of observations. It is illustrated below.

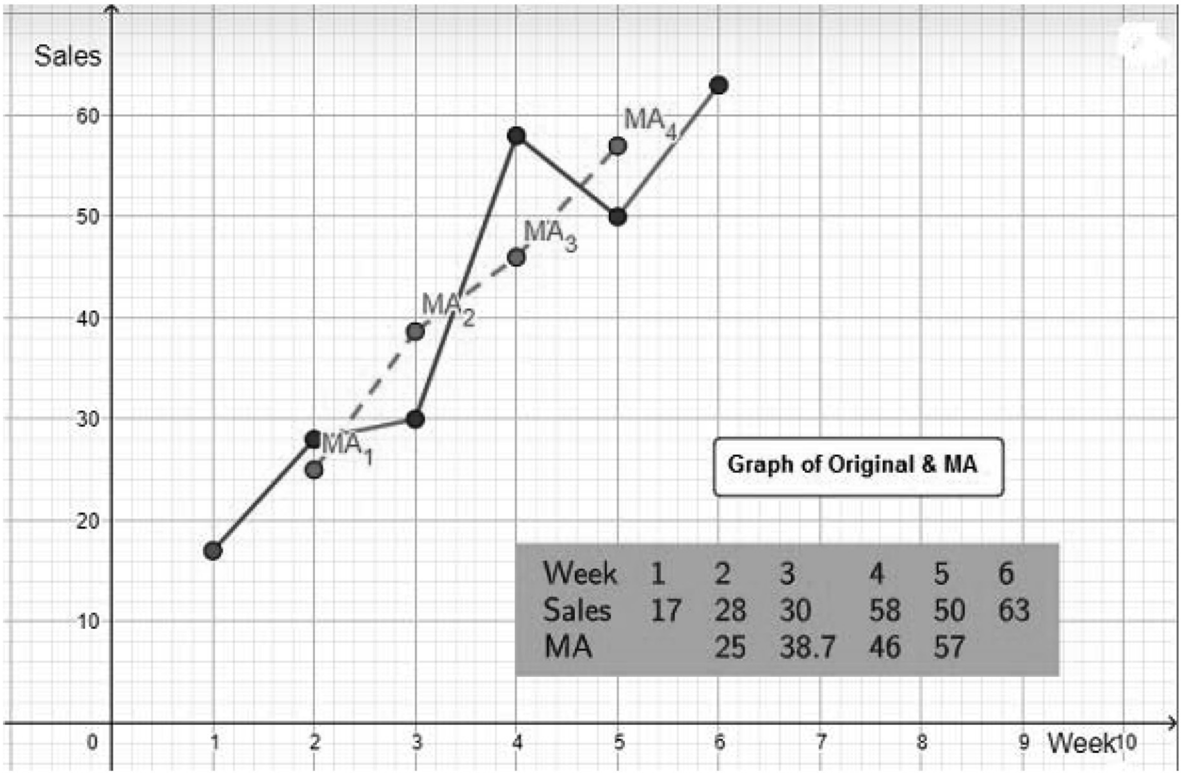

Figure No. 14.5: Graph showing Original and Moving Average Values

With respect to the data provided along with the diagram above the calculation of 3 weekly Moving Average (MA) is explained here and shown below.

1st MA = (17+28+30)/3 = 25 appearing against week 2**,** 2nd MA = (28+30+58)/3 = 38**.**7 appearing against week 3

Similarly 3rd MA = (30+58+50)/3 = 46 appearing against week 4 & 4th MA = (58+50+63)/3 = 57 against week 5

Week	Sales	3 weekly Moving Total	3 weekly Moving Average
1	17	-	-
2	28	75	25
3	30	116	38.7
4	58	138	46
5	50	171	57
6	63	-	-

Both the Original values as well as the Moving Average values are plotted on a graph and joined using firm lines and dotted lines as shown in the diagram above. It is clear from the diagram that the fluctuations in the Original data have been smoothened a lot in the Moving Average graph. This is due to the fact that MA contains only the Trend component of the time series.

When **the period of moving average is EVEN** then there are two middle periods and the moving average value is placed in between the two middle terms of the time interval it covers. So in this case the moving average value will not coincide with a period of the given time series. To synchronize the moving average values with the original data, an average of two already calculated moving averages is computed and placed them corresponding to the given time set. This method is known as Centring and the corresponding moving averages are called Centred Moving Averages. This is explained in the Illustration No. 4 below.

The logic behind moving average method is that if the period of moving average is exactly equal to the period of cycle (or its multiple) present in the time series then the method will completely eliminate cyclical fluctuations. The method is very simple and involves no complicated mathematical calculation. But the main disadvantage is that some of the value/s at the beginning and at the end of the series cannot be obtained.

Method of Fitting Mathematical Curves is perhaps the best and most objective method of determining trend. In this method an appropriate type of mathematical equation is selected for trend and the constants appearing in the trend equation are determined on the basis of the given time series data. The choice of the appropriate type of equation is facilitated by a graphical representation of the data.

If the plotted data show approximately a straight line tendency on a graph paper, the equation used is $y = a + bx$

If they show a straight line tendency on a semi logarithmic graph paper, the equation used is $\log y = a + bx$ i.e an Exponential Curve.

If the graphical representation of data shows a parabola then the equation used is $y = a + bx + cx^2$. Sometimes the equation may also be of higher degree polynomial.

The constants appearing in the equations mentioned above are obtained by applying Principle of Least Square. In fact it is a special category of Least Square Regression in which the independent variable is always time. Though this method involves considerable numerical calculations, but found to be most suitable for forecasting the trend.

Illustration 4 (Moving Average trend with EVEN Period)

Find trend values of the following year wise data of Goods carried by a fleet of trucks of a Transport Company having pan India network using the Moving Average Method. [Assume a 4 yearly cycle]

Year	2012	2013	2014	2015
Goods carried (Tons)	2204	2500	2360	2680
Year	2016	2017	2018	2019
Goods carried (Tons)	2424	2634	2904	3098
Year	2020	2021	2022	2023
Goods carried (Tons)	3172	2952	3248	3172

Solution:

Calculations for 4 Yearly Moving Average Trend values

Year	Goods carried (Tons)	4 Yearly Moving Total	4 Yearly Moving Average (Not centred)	2 item Moving Total (Centred)	4 Yearly Moving Average (Centred)
(1)	(2)	(3)	(4) = (3) 4	(5)	(6) = (5) 2
2012	2204	-	-	-	-
2013	2500	-	-	-	-
		9744	2436		
2014	2360			4927	**2463.50**
		9964	2491		
2015	2680			5015.5	**2507.75**
		10098	2524.5		
2016	2424			5185	**2592.50**
		10642	2660.5		
2017	2634			5425.5	**2712.75**
		11060	2765		
2018	2904			5717	**2858.50**
		11808	2952		
2019	3098			5983.5	**2991.75**
		12126	3031.5		
2020	3172			6149	**3074.5**
		12470	3117.5		
2021	2952			6253.5	**3126.75**
		12544	3136		
2022	3248	-	-	-	-
2023	3172	-	-	-	-

Method of calculation –

1st entry of Column (3) = Sum of the entries in Column (2) for the period 2012 to 2015 = 2204+2500+2360+2680 and it is placed in between the years 2013 & 2014 i.e at the middle of the first 4 year period under consideration.

2nd entry of Column (3) = Sum of the entries in Column (2) for the period 2013 to 2016 = 2500+2360+2680+2424 and it is placed in between the years 2014 & 2015 i.e at the middle of the second 4 year period under consideration.

3rd entry of Column (3) = Sum of the entries in Column (2) for the period 2014 to 2017 = 2360+2680+2424+2634 and it is placed in between the years 2015 & 2016 i.e at the middle of the third 4 year period under consideration.

Thus it is clear that except the first entry, each of the other entries are made by omitting the first value of the previous period and adding the value against the new year taken into consideration. This way all the other entries of column (3) are made. As none of these entries appear against a particular year, they are called "Not centred" values.

Similarly in case of column (5) except the first entry, each of the other entries are made by omitting the first value considered for the previous calculation and adding a new value from the column (4) values. It is shown below.

1st entry of Column (5) = Sum of the first two entries of column (4) = 2436 + 2491 = 4927

2nd entry of Column (5) = Sum of the second & third entries of column (4) = 2491 + 2524.5 = 5015.5

3rd entry of Column (5) = Sum of the third & fourth entries of column (4) = 2524.5 + 2660.5 = 5185

Other entries of Column (5) are made using the same method.

[**N. B –** The calculations can be reduced to some extent as follows

Find two item Moving total (Centred) by taking 2 values at a time from column (3) and place it in column (4) by following the similar method as described in case of column (5) in the illustration above.

Next find 4 Yearly Moving Average (Centred) by dividing each entry of this new column (4) by 8]

Thus the results can be obtained by making 5 columns instead of 6 as shown in the illustration above.

Illustration 5 (Mathematical Curve fitting with data for ODD number of years given)

The following table relates to the tourist arrivals in India during 2017 to 2023.

Year	2017	2018	2019	2020	2021	2022	2023
Tourist arrivals (lakhs)	18	20	23	25	24	28	30

Fit a Straight Line trend by the Method of Least Squares and estimate the number of tourists that would arrive in the year 2027.

Solution:

Let the best fit Trend line to the given data be y = a + bx (Origin at the year 2020 and x unit = 1 year)

Normal equations are $\Sigma y = a.n + b.\Sigma x$ **......** (1) and $\Sigma xy = a.\Sigma x + b.\Sigma x^2$ **........** (2) where n = No. of years = 7 (here)

Calculations for fitting Straight Line Trend

Year	Tourist arrivals (y in lakhs)	x	x^2	xy
2017	18	–3	9	– 54
2018	20	–2	4	– 40
2019	23	–1	1	– 23
2020	25	0	0	0
2021	24	1	1	24
2022	28	2	4	56
2023	30	3	9	90
Total	168	0	28	53

Putting the values of Σy, Σx and n in equation (1) we get 168 = a.7 + b.0 Or, a = 24

Also putting the values of Σxy, Σx and Σx^2 in equation (2) we get, 53 = a.0 + b.28 Or, b = 1.893

So the required equation of Straight Line Trend is y = 24 + 1.893x (Origin = 2020, x unit = 1 year)

For the year 2027, x = 7. So the estimated number of tourists in the year 2027 = 24 + 1.893.7 = 37.25 lakhs

Illustration 6 (Mathematical Curve fitting with data for EVEN number of years given)

From the following past data of Sales (in lakhs Rupees) of a company estimate the same for the year 2029.

Year	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
Sales	15.3	14.6	16.8	17.3	17.2	20.9	22.3	20	23.1	24.5

Assume the trend line to be linear. What is the monthly rate of increase of Sales?

Solution:

Let the best fit Linear Trend line to the given data be y = a + bx

(Origin at the middle of the years 2018 & 2019 and x unit = 6 months)

Normal equations are $\Sigma y = a.n + b.\Sigma x$ .... (1) where n = No. of years = 10 (here)

$\Sigma xy = a.\Sigma x + b.\Sigma x^2$.................... (2)

Using the values (from calculations below) of Σy, Σx and n in equation (1) we get 192 = a.10 + b.0 Or, a = 19.2

Also using the values (from calculations below) of Σxy, Σx and Σx^2 and putting in the equation (2) we get, 177 = a.0 + b.330 Or, b = 0.536

Calculations for fitting Straight Line Trend

Year	Sales (y in ₹ Millions)	x	x^2	xy
2014	15.3	-9	81	- 137.7
2015	14.6	-7	49	- 102.2
2016	16.8	-5	25	- 84
2017	17.3	-3	9	- 51.9
2018	17.2	-1	1	- 17.2
2019	20.9	1	1	20.9
2020	22.3	3	9	66.9
2021	20.0	5	25	100
2022	23.1	7	49	161.7
2023	24.5	9	81	220.5
Total	192	0	330	177

So the required equation of Straight Line Trend is y = 19.2 + 0.536x

(Origin = At the middle of 2018 & 2019, x unit = 6 months)

For the year 2029, x = 21. So the **estimated sales for the year 2029 = 19.2 + 0.536 × 21 = 30.456 Million ₹**

Yearly rate of increase in Sales = b = 0.536. So **monthly rate of increase in Sales = b/12 = 0.0467 Million ₹**

Measurement of Seasonal Variation

Four methods are generally used for measuring Seasonal Variation or Short term fluctuation of time series.

- Method of Simple averages (monthly or quarterly)
- Moving Average Method (Normally Quarterly type with Monthly type at times)
- Trend Ratio Method
- Link Relative Method

Method of Simple Averages is generally applied to the time series data which do not contain trend or cyclical fluctuation to any appreciable extent. From the given quarterly data the averages (A_1, A_2, A_3 & A_4) for the four quarters are calculated and also Grand Average [$G = (A_1 + A_2 + A_3 + A_4)/4$] is calculated.

If the Additive Model is used, the deviations of the Quarterly Averages from the Grand Average give the measures of Seasonal Variation of the four quarters as $S_1 = A_1 - G$, $S_2 = A_2 - G$, $S_3 = A_3 - G$ and $S_4 = A_4 - G$

If the Multiplicative Model is used, each Quarterly Average is expressed as percentage of Grand Average to give the Seasonal Indices for the 4 quarters as $S_1 = (A_1/G)x100$, $S_2 = (A_2/G)x100$, $S_3 = (A_3/G)x100$, $S_4 = (A_4/G)x100$

When monthly figures are given, then instead of Quarterly Averages we need to calculate Monthly Averages.

Total of Seasonal Variations = 0 for Additive Model and Average Seasonal Index = 100 for Multiplicative Model

In *Moving Average Method* firstly quarterly or monthly (as the case may be) Trend values are calculated using the concept of Moving Average. Thereafter the effect of Trend is eliminated from the original data.

For Additive Model, this is done by subtracting Trend values from the original data to give "Deviations from Trend". Thereafter with these values the technique used in the Method of Simple Averages (as explained above) is applied to get the measures of Seasonal Variations. The total of these Seasonal Variations should be 0. In case that does not happen then some adjustment in the calculated values are done.

For Multiplicative Model, the effect of Trend is eliminated by finding "Ratio to Moving Average" expressed as percentage which is actually (Original data for a quarter or month /Moving Average value of that period) x 100. Thereafter the technique used in the Method of Simple Averages (as explained above) is used to find out Seasonal Indices. The total of these Seasonal Indices should be 400 for quarterly data and 1200 for monthly data. In case that does not happen then some adjustment in the calculated values are done.

Trend Ratio Method is applicable only for the Multiplicative type Time Series data. Here trend values are ob tained by fitting a mathematical curve and the original data are expressed as percentages of corresponding trend. Thereafter the technique used for Multiplicative Model cases of Simple Average Method is utilised to get the values of Seasonal Indices of the different quarters or months.

Link Relative Method – If quarterly data are given, each value is expressed as a percentage of the value for the immediately preceding period. These are called Link Relatives (L.R). For the first quarter it cannot be calculated because there is no data before this. Then the L.Rs are arranged by quarters and the average L.R for each quarter is found. From these L.Rs we find Chain Relatives (C.R) by relating them to a common base. For the first quarter C.R is taken as 100. The C.R for any quarter is then obtained as illustrated below.

Second C.R for 1st Quarter = (C.R for 1st Quarter) × (L.R for 1st Quarter) /100

Usually the second C.R for 1st Quarter will differ from the originally assumed value of 100 due to the presence of trend. Therefore some adjustments to the C.Rs are necessary.

Let c be the average quarterly deviation of the 2nd C.R of 1st Quarter from 100 i.e c = (2nd C.R for Q_1 – 100) /4

Subtracting c, 2c, 3c and 4c from the C.Rs for Q_2, Q_3, Q_4 and the 2nd C.R for Q_1, we find that both the C.Rs for Q_1 are now equal to 100. The adjusted C.Rs for Q_1, Q_2, Q_3 and Q_4 are now expressed as percentages of their A.M to give the Seasonal Indices. The total of these Seasonal Indices should be 400.

Illustration 7

Calculate the Seasonal Indices for the following quarterly data in certain units. Appropriate method for finding the Indices has to be decided by you with due explanation.

Year	1st Quarter	2nd Quarter	3rd Quarter	4th Quarter
2020	39	21	52	81
2021	45	23	63	76
2022	44	26	69	75
2023	53	23	64	84

Solution:

The values in any quarter do not reveal any definite tendency to change. Thus there is no appreciable trend in the given dataset. So it is decided to use Method of Simple Average (Quarterly) to find out the Seasonal Indices. Also a Multiplicative Model is assumed for the data.

Calculations for Seasonal Index

Year	1st Quarter	2nd Quarter	3rd Quarter	4th Quarter	Total
2020	39	21	52	81	-
2021	45	23	63	76	-
2022	44	26	69	75	-
2023	53	23	64	84	-
Total	181	93	248	316	838
Arithmetic Mean	45.25	23.25	62	79	209.5
Seasonal Index	**86.4**	**44.4**	**118.4**	**150.8**	400

Calculations

Arithmetic Mean for any Quarter = Toal for that quarter /4, Grand Average = Total of the Arithmetic Means /4

Seasonal Index for any Quarter = (Arithmetic Mean of that Quarter / Grand Average)x100

Illustration 8

Calculate Seasonal Fluctuation from the following Time Series data obtained from a Mini Steel Plant

	Quarterly Output of Steel in '000 Tons			
Year	1st Quarter	2nd Quarter	3rd Quarter	4th Quarter
2020	65	58	56	61
2021	68	63	63	67
2022	70	59	56	52
2023	60	55	51	58

Use 4 Quarter Moving Average Method. Consider Additive Model for the Time Series data.

Deseasonalise the given dataset.

Solution:

Calculations for 4 Quarter Moving Averages & Deviations from Trend

Year	Quarter	Output in '000 Tons	4 Quarter Moving Total (Not centred)	2 Period Moving Total (Centred)	4 Quarter Moving Average (Centred)	Deviation from Trend
(1)	(2)	(3)	(4)	(5)	(6) = (5) ÷ 8	(7) = (6) – (3)
2020	1st	65	-	-	-	-
	2nd	58	-	-	-	-
			240			
	3rd	56		483	60.38	- 4.38
			243			
	4th	61		491	61.38	- 0.38
			248			
2021	1st	68		503	62.88	5.12
			255			
	2nd	63		516	64.50	- 1.50
			261			
	3rd	63		524	65.50	- 2.50
			263			
	4th	67		522	65.25	1.75
			259			
2022	1st	70		511	63.88	6.12
			252			
	2nd	59		489	61.12	- 2.12
			237			
	3rd	56		464	58.00	- 2.00
			227			
	4th	52		450	56.25	- 4.25
			223			
2023	1st	60		441	55.12	4.88
			218			
	2nd	55		442	55.25	- 0.25
			224			
	3rd	51	-	-	-	-
	4th	58	-	-	-	-

Calculations

1st entry of Column (4) = 65+58+56+61 = 240,

2nd entry of Column (4) = 58+56+61+68 = 243

3rd entry of Column (4) = 56+61+68+63 = 248 and so on.

1st entry of Column (5) = 240+243 = 483

2nd entry of Column (5) = 243+248 = 491 and so on

Calculations for Seasonal Fluctuations

Year	Deviation from Trend for				Total
	1st Quarter	2nd Quarter	3rd Quarter	4th Quarter	
2020	-	-	- 4.38	- 0.38	-
2021	5.12	- 1.50	- 2.50	1.75	-
2022	6.12	- 2.12	- 2.00	- 4.25	-
2023	4.88	- 0.25	-	-	-
Total	16.12	- 3.87	- 8.88	- 2.88	0.49
Arithmetic Mean	5.37	- 1.29	- 2.96	- 0.96	0.16
Adjustment	- 0.04	- 0.04	- 0.04	- 0.04	- 0.16
Seasonal Fluctuation	**5.33**	**- 1.33**	**- 3.00**	**- 1.00**	0

Calculations

Arithmetic Mean = (Quarterly total value of Deviation from trend) /3 ,

Grand Average = (Total of Arithmetic Mean values) /4 = 0.16/4 = 0.04

Adjustment for each quarter = - (Grand Average)

Seasonal Fluctuation for any Quarter = Arithmetic Mean for that Quarter + Adjustment _

[As explained in the theory, Adjustment has been done to get a total of zero value for the Seasonal Fluctuations]

Deseasonalising given Time Series data

Year	Quarter	Output (y_t in '000 Tons)	Seasonal Fluctuation (S)	Deseasonalised Data (y_t – S)
2020	1st	65	5.33	**59.67**
	2nd	58	– 1.33	**59.33**
	3rd	56	– 3.00	**59.00**
	4th	61	– 1.00	**62.00**
2021	1st	68	5.33	**62.67**
	2nd	63	– 1.33	**64.33**
	3rd	63	– 3.00	**66.00**
	4th	67	– 1.00	**68.00**
2022	1st	70	5.33	**64.67**
	2nd	59	– 1.33	**60.33**
	3rd	56	– 3.00	**59.00**
	4th	52	– 1.00	**53.00**
2023	1st	60	5.33	**54.67**
	2nd	55	– 1.33	**56.33**
	3rd	51	– 3.00	**54.00**
	4th	58	– 1.00	**59.00**

[**N.B** – Deseasonalisation means elimination of Seasonal Variation from the Time Series data. For data in Additive Model it is done by subtracting Seasonal Variation from the original data. When Multiplicative Model is used then it is done by dividing the original data by Seasonal Effect which is Seasonal Index/100]

Illustration 9

On the basis of quarterly Sales (in ₹ Lakhs) of a certain commodity for the period 2019 to 2023 the following calculations were made.

Trend:- Straight Line trend equation is y = 25 + 0.6t, Origin – 1st Quarter of 2001, t unit – 1 Quarter, y – Sales

Seasonal Variations:-

Quarter	I	II	III	IV
Seasonal Index	90	95	110	105

Estimate the Quarterly Sales figures for the year 2028

Solution:

Given Trend equation is y = 25 + 0.6t [Origin at the 1st Quarter of 2001, t unit = 1 Quarter]

For the 1st Quarter of the year 2028, the value of t = 36, For the 2nd Quarter of the same year, t = 37

For the 3rd Quarter of the same year, t = 38 and for the 4th Quarter, t = 39

Putting these values of t in the Trend equation we find –

Trend for the 1st Quarter of 2028 = 25 + 0.6 × 36 = 46.6, Trend for the 2nd Quarter of 2028 = 25+0.6×37 = 47.2

Trend for the 3rd Quarter of 2028 = 25+0.6×38 = 47.8 & Trend for the 4th Quarter of 2028 = 25+0.6×39 = 48.4

Quarter of the year 2028	Trend (T) in ₹ Lakhs	Seasonal Index	Seasonal Effect (S) = Seasonal Index/100	Estimated Sales in ₹ Lakhs (TxS)
I	46.6	90	0.90	**41.94**
II	47.2	95	0.95	**44.84**
III	47.8	110	1.10	**52.58**
IV	48.4	105	1.05	**50.82**

So the estimated Sales is calculated by multiplying Trend and Seasonal values. But for the cases where the data follows Additive Model, the estimated Value should be calculated by adding Trend and Seasonal Variation.

Exponential Smoothing

This method can be considered as a more sophisticated extension of Moving Average Method of forecasting technique which weighs past data in an exponential manner so that the most recent data carries more weight in the moving average. Unlike moving average method, where equal weightage is given to all the past observations, this method assigns decreasing weightages to the past observations. This is definitely more reasonable since older the observation the less relevance it holds for the future.

In Exponential Smoothing, the weights used are – α for the current observation, $\alpha(1-\alpha)$ for the immediately preceding observation, $\alpha(1-\alpha)^2$ for the still preceding observation and so on, where α is a constant known as *Smoothing Constant* and has a value lying between 0 and 1. It can be mentioned that the weights of observations from the current period backwards are diminishing with a common ratio $(1 - \alpha)$. In fact they form an infinite G.P having Sum of all the terms equal to 1.

The exponentially smoothed average which is used as a "Forecast" for time t is thus given as –

$u_t = \alpha y_t + \alpha(1-\alpha)y_{t-1} + \alpha(1-\alpha)^2 y_{t-2} + \alpha(1-\alpha)^3 y_{t-3} + \ldots\ldots\ldots\ldots$ to infinity

This may also be written as a recurrence relation $\mathbf{u_t = \alpha y_t + (1-\alpha)u_{t-1}}$ where u_t and u_{t-1} are the Forecasts at time t and (t-1) respectively, y_t is the observation for the time t of the given Time Series data and α is the Smoothing Coefficient.

Above relation may also be expressed as $u_t = u_{t-1} + \alpha(y_t - u_{t-1})$ Or, $\mathbf{u_t = u_{t-1} + \alpha e_t}$ where $e_t = y_t - u_{t-1}$ is called *Error* or *Discrepancy* of the latest observation from the forecast in the previous period.

For numerical computations, above relation may be used with the following steps

- Find the Error or Discrepancy in the latest observation from the previous forecast using $e_t = y_t - u_{t-1}$
- Multiply the Error by the Smoothing Coefficient to obtain the Correction (αe_t)
- Add the Correction (αe_t) to the previous forecast (u_{t-1}) to get forecast (u_t) for the current period t.

In Exponential Smoothing the first forecast has to be obtained by some subjective method or by taking the average of the first few time periods. Subsequent forecasts are then obtained by repeatedly using the relationship

$u_t = u_{t-1} + \alpha e_t$

When $\alpha = 0$, no Correction is necessary to the previous forecast. When $\alpha = 1$, the new forecast will always be the latest observation. Smoothing Coefficient α lies between 0 and 1 and it is selected on the basis of experiments with different values. Higher the value of α, sooner it discounts the effects of old observations. So when old observations are also of some significance, it is wise to take smaller values of α. For all practical purposes it is kept in the range 0.005 to 0.3 to smooth the forecast.

Illustration 10

M/S B.P.Leathers, a shoe manufacturer has modern outlook and they depend heavily on Business Forecasting methodology to plan their business activities like manufacturing, marketing, finance etc. At the beginning of the year 2023 they have forecasted data of demand of their shoes for the beginning of the month of March as 1000 pairs. But the actual demand turned out to be 900 pairs. Using a Smoothing Coefficient of 0.1 forecast the demand at the beginning of the 2nd week of March 2023.

Also forecast the demands using Exponential Smoothing technique at the beginning of each week till mid April 2023 when the actual demands are as follows –

At the beginning of the 2nd week of March – 1010 pairs, At the beginning of the 3rd week of March – 1032 pairs, At the beginning of the 4th week of March – 976 pairs, At the beginning of the 1st week of April – 934 pairs, At the beginning of 2nd week of April – 1008 pairs & At the end of the 2nd week of April – 1020 pairs.

Solution:

As per the concept of Exponential Smoothing we have $u_t = u_{t-1} + \alpha e_t$ where $e_t = y_t - u_{t-1}$ = Forecast Error & $\alpha = 0.1$

Calculations for Exponential Smoothing

Beginning of	Demand of shoe (y_t in Pairs)	Previous Forecast (u_{t-1})	Forecast Error ($e_t = y_t - u_{t-1}$)	Correction (αe_t) ($\alpha = 0.1$)	New Forecast ($u_t = u_{t-1} + \alpha e_t$)
March 1st week	900	1000	- 100	- 10	**990**
March 2nd week	1010	990	20	2	**992**
March 3rd week	1032	992	40	4	**996**
March 4th week	976	996	- 20	- 2	**994**
April 1st week	934	994	- 60	- 6	**988**
April 2nd week	1008	988	20	2	**990**
April 2nd week end or Mid April	1020	990	30	3.0	**993**

[**Note** – Except the 1st entry of 3rd column, all the other entries are taken from the last column].

Input – Output Model

Under this model a forecast of output is based on given inputs if the coefficients of input – output relationship are known. Similarly, input requirements can be forecast on the basis of final output with a given input – output relationship. Due to this mechanics of forecasting the model is named as Input – Output Model.

Assumptions on which the model is established are as follows –

1. The economy consists of a number of interacting industries.
2. Each industry produce only one item and use only one process of production.
3. To produce an item, the industry requires as input the goods made by other industries, labour and perhaps imports. An industry may use some of its own goods. Such use may be taken as sale to itself.
4. The output of any industry becomes either the input to another industry or the final demand.
5. In any productive process all inputs are used in fixed proportions and increase in input is in proportion with the level of output. Production takes place through processes with constant technical coefficients. Technical Coefficient shows the number of units of any industry's output needed to produce one unit of another industry's output.
6. All transactions may be taken in terms of money values because it is a suitable common unit for aggregating inputs and outputs of industries.

Uniqueness of Input – Output Model lies in the fact that the same is the only forecasting technique which is based on inter industry flow of goods and services, given technical data on input usage by various industries. However the product quality, technology and industrial organisation undergo changes over time causing inter industry relationship to change, too. The degree of accuracy of the forecasts through this model, therefore depends on the extent to which the technical coefficients truly project the latest inter industry flow relationship.

Historical Analogy Method

Under this method forecast in regard to a particular phenomenon is based on some analogous conditions elsewhere in the past. For example, the forecast of demand for a product in India can be based on the demand for the same in some developed country in the past if it is found that the present conditions in India are very much like the same prevailing in that country in the past.

In fact such a method is more useful for indicating qualitative changes in society. It is said social analogies have helped in indicating the trends of changes in the norms of corporate behaviour in terms of attitude of the worker against inequality etc. find similarities in various countries at different stages of the history of industrial growth. But it is difficult to quantify most of these phenomena and therefore this method does not have much relevance for statistical analysis.

Jury or Executive Opinion Method

This method attempts to pool the knowledge, experience and judgement of managers inside the organisation, by asking about their opinion on the likely sales of a product in future. This is particularly applicable for a new product which does not have and past data to facilitate forecasting.

Survey Technique

First hand data on the behaviour of certain desired variables can be obtained through selective surveys by means of questionnaires and interviews. The data so collected can be processed and analysed for purposes of testing some predetermined hypotheses and making predictions and future estimates of the behaviour of the variables under study.

Barometric Technique

The behaviour of certain economic or business variables can have an important effect on some other variables. For example, a shortfall in supply of a commodity may lead to a rise in its price. Here the shortfall is an indicator or a barometer of the likely increase in price. Indicators can be of three types – Lead, Lag and Concurrent. Issuance of Industrial License is a Lead Indicator of future industrial activity. Similarly demand for Household Furniture in a housing complex is a Lag Indicator because it becomes active only after the construction of houses are completed i.e after a certain time lag. Concurrent Indicators move together i.e no lead or lag time is involved.

Delphi Technique

This technique is used to make more realistic judgemental forecasts by minimizing bias. In this method a panel of experts (including decision makers, staff personnel and respondents) is asked sequential questions. It is a step by step procedure and final forecast is obtained by the common opinion of all the experts.

EXERCISE

A. Theoretical Questions:

Multiple Choice Questions

1. In Exponential Smoothing Method which one of the following is true?

   (a) $0 \leq \alpha \leq 1$ and high value of α is used for stable demand.

   (b) $0 \leq \alpha \leq 1$ and high value of α is used for unstable demand.

   (c) $\alpha \geq 1$ and high value of α is used for stable demand.

   (d) $\alpha \leq 0$ and high value of α is used for unstable demand.

2. Which of the following is not a Casual Forecasting Method?

   (a) Trend adjusted Exponential Smoothing

   (b) Econometric models

   (c) Linear Regression

   (d) Multiple Regression

3. Which of the following is a Forecasting technique?

   (a) PERT / CPM

   (b) Exponential Smoothing

   (c) Gantt Chart

   (d) Control Chart

4. The number of averaging period in the Simple Moving Average Method of forecasting is increased for greater smoothing but at the cost of –

   (a) Accuracy

   (b) Stability

   (c) Visibility

   (d) Responsiveness to changes

5. In a Time Series forecasting model, the demands for five time periods are 10, 13, 15, 18 and 22. A linear regression fit resulted in the equation $y_t = 6.9 + 2.9t$, where y_t is the forecast for the period t. The sum of the absolute deviations for the five data with respect to their corresponding forecasts (taking t = 1 for the first one) is

   (a) 2.3

   (b) 0.2

   (c) 1.2

   (d) 2.2

6. Which of the following is not a part of Quantitative type of Forecasting Model

   (a) Moving Average

   (b) Simple Average

(c) Delphi Method

(d) Exponential Smoothing

7. Which of the following Forecasting technique uses three types of participants: Decision Makers, Staff personnel and Respondents?

   (a) Expert's Opinion

   (b) Sales Force Survey

   (c) Consumer Survey

   (d) Delphi Method

8. Sales data for the numbers sold for a particular product during January to May 2007 shows the values 10, 11, 16, 19 and 25. Regarding forecast for the month of June which one of the following statement is true?

   (a) Moving Average will forecast a higher value compared to regression.

   (b) Exponential Smoothing will forecast a higher value compared to regression

   (c) Regression will forecast a higher value compared to moving average.

   (d) None of the above.

9. The Time Series forecasting method that gives equal weightage to each of the N most recent observations is –

   (a) Moving Average Method

   (b) Exponential Smoothing with linear Trend

   (c) Triple Exponential Smoothing

   (d) None of the above

10. Which of the following is not a forecasting technique?

    (a) Trend line estimate

    (b) Delphi Method

    (c) Hungarian Method

    (d) Judgemental technique

11. In Simple Exponential Smoothing forecast, to give higher weightage to recent demand information, the smoothing constant must be close to –

    (a) -1

    (b) 0

    (c) 0.5

    (d) 1

12. Which of the following is not true for forecasting?

    (a) Forecasts are rarely perfect.

    (b) The underlying casual system will remain same in the future.

    (c) Forecast for group of items is accurate than individual item

    (d) Short range forecasts are less accurate than long range forecasts.

13. In which of the following forecasting technique, data obtained from past experience is analysed?
    (a) Judgemental forecast
    (b) Time Series forecast
    (c) Associative model
    (d) All of the above
14. Delphi Method is used for –
    (a) Judgemental forecast
    (b) Time Series forecast
    (c) Associative model
    (d) All of the above
15. Short term regular variations related to the calendar or time of the day is known as –
    (a) Trend
    (b) Seasonality
    (c) Cycles
    (d) Random variations
16. A linear Trend equation has the form –
    (a) F = a - bt
    (b) F = a + bt
    (c) F = 2a – bt
    (d) F = 2a + bt
17. The actual demand for a period is 100 units. But forecast demand was 90 units. The forecast error is –
    (a) – 10
    b) 10
    c) 5
    d) None of the above
18. Which of the following is not a forecasting technique?
    a) Judgemental
    b) Time Series
    c) Time Horizon
    d) Associative
19. Which of the following is not a Qualitative Forecasting technique?
    (a) Surveys of consumer expenditure plans
    (b) Perspective of foreign advisory councils
    (c) Consumer intention polling
    (d) Time Series analysis

20. Which of the following is not one of the four types of variation that is estimated in the Time Series analysis?

    (a) Predictable

    (b) Trend

    (c) Cyclical

    (d) Irregular

21. In Time Series Analysis which source of variation can be estimated by the ratio to trend method?

    (a) Cyclical

    (b) Trend

    (c) Seasonal

    (d) Irregular

22. A qualitative forecast

    (a) Predicts the quality of a new product

    (b) Predicts the direction but not the magnitude of change in a variable.

    (c) Is a forecast that is classified on a numerical scale from 1 (poor quality) to 10 (perfect quality).

    (d) Is a forecast that is based on econometric methods.

23. The first step in Time Series analysis is to –

    (a) Perform preliminary Regression calculations.

    (b) Calculate a moving average.

    (c) Plot the data on a graph

    (d) Identify relevant correlated variables.

24. If the estimate of the Trend Component is 158.2, the estimate of Seasonal Component is 94%, the estimate of the Cyclical Component is 105% and the estimate of the Irregular Component is 98%, then the multiplicative model will produce a forecast of –

    (a) 1.53

    (b) 1.53%

    (c) 153.02

    (d) 153,020,532

25. From the first two supplied values of a Time Series and its corresponding Exponential Smoothing forecast as given below, the forecast for the time period 3, assuming Smoothing Constant = 0.3, will be

Time Period (t)	Value (y_t)	Exponential Smoothing Forecast (u_t)
1	18	18
2	22	18

    (a) 18

    (b) 19.2

    (c) 20

    (d) 40

26. For the Regression Equations y = 0.516x + 33.73 and x = 0.512y + 32.52, the Arithmetic Mean values of x and y are nearly

(a) 67.6 and 68.6

(b) 68.6 and 68.6

(c) 67.6 and 58.6

(d) 68.6 and 58.6

27. For a bivariate dataset (x,y), if the Means, Standard Deviations and Correlation Coefficient are respectively $\overline{x} = 1$, $\overline{y} = 2$, $\sigma_x = 3$, $\sigma_y = 9$ and r = 0.8. The regression line y on x is –

(a) y = 1 + 2.4(x -1)

(b) y = 2 + 0.27(x – 1)

(c) y = 2 + 2.4(x – 1)

(d) y = 1 + 0.27(x – 2)

28. The data about Sales and Advertisement Expenditure of a firm is given below

	Sales (₹ Crores)	**Advertisement Expenditure (₹ Crores)**
Mean	40	6
S.D	10	1.5

The Correlation Coefficient between Sales and Advertisement Expenditure is 0.9. The likely Sales for a proposed Advertisement Expenditure of ₹ 10 Crores is –

(a) ₹ 64 Crores

(b) ₹ 67 Crores

(c) ₹ 70 Crores

(d) ₹ 58 Crores

29. Given the Regression Lines x + 2y – 5 = 0 and 2x + 3y – 8 = 0 and Var (X) = 12. The value of Var (Y) is –

(a) 3/4

(b) 4/3

(c) 16

(d) 4

30. The equations of the two lines of Regression are 4x + 3y + 7 = 0 and 3x + 4y + 8 = 0. The Coefficient of Correlation between x and y is –

(a) 1.25

(b) 0.25

(c) – 0.75

(d) 0.92

Answers:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
b	a	b	d	d	c	d	c	a	c	d	d	b	a	b
16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
b	b	c	d	a	c	b	c	c	b	a	c	a	d	c

State True or False

1. Surveys and Opinion Polls are Qualitative Techniques
2. The Delphi Method generates forecasts by surveying consumers to determine their opinions.
3. Irregular or random influences on time series data give rise to secular trend.
4. The linear trend equation can be estimated by ordinary least squares Regression Analysis.
5. The ratio to trend method is used to find a linear trend equation.
6. Forecasts based on lead indicators are qualitative.
7. The long run increase or decrease in time series data is referred to as Cyclical Fluctuation.
8. The choice of a forecasting method should be based on an assessment of costs and benefits of each method in a specific application.
9. Barometric forecasting methods are most useful for long term forecasts.
10. Qualitative forecasts based on surveys tend to perform particularly well during periods of unexpected international political upheaval.
11. Councils of distinguished foreign dignitaries and business people are used to obtain qualitative forecasts with a foreign perspective.
12. Time analysis generates forecasts by identifying cause and effect relationship between variables.
13. Time Series data are observations on a variable at different points of time.
14. A time series that displays regular seasonal variation is said to exhibit cyclical fluctuation.
15. Smoothing techniques are most useful for time series data that is primarily influenced by irregular variation.

Answers:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
T	F	F	T	F	T	F	T	F	F	T	T	T	F	T

Fill in the blanks

1. Shortfall in supply of a commodity may lead to a rise in its price. Here the shortfall is a _______ of the likely increase in price.
2. Multiple linear regression or ___ is an extension to the Simple linear regression.
3. Deseasonalisation is _____ of Seasonal Variation component from Time Series data.

4. Two regression lines _______ if $b_{yx}.b_{xy} = 1$.
5. Signs of Regression Coefficients and Correlation Coefficient should be ____ always.
6. In the regression line x on y the independent variable is ___.
7. Coordinates of the point of intersection of two regression lines give ___ values of the variables.
8. $r(\sigma_y/\sigma_x)$ is known as the ____ of the regression line y = a + bx.
9. ____Forecasting is based on current and future assets and liabilities as well as predictions for liquid capital and cash flow estimates.
10. Business _____ allows a company to make long term plans and prepare for any changes in the market.

Answers:

1.	Barometer	2.	MLR
3.	Removal	4.	Coincide
5.	Same	6.	y
7.	Mean	8.	Slope
9.	Capital	10.	Forecasting

Short essay type questions

1. What is Secular Trend in Time Series? Name the methods by which it can be measured.
2. Define Regression. What are the estimators in a linear regression equation?
3. What do you mean by Demand Forecasting?
4. Name the different models of Qualitative Forecasting.
5. What is the speciality of Delphi technique?
6. "To deseasonalise time series data firstly trend component of it needs to be removed" – discuss.

Essay type questions

1. What are the limitations of Business Forecasting?
2. What are the assumptions on which Input – Output Model is established?
3. Briefly describe the different models of time series data.
4. Write short notes on SLR and MLR.
5. Describe the importance of Business Forecasting in different areas of an organisation.

B. Numerical Questions

Comprehensive Numerical Problems

1. A company dealing in logistics business has a business wing named Ship Unloading. During the years 2022 and 2023 they have the following quarterly figures (in tonnage) of material unloaded from Ships.

Year	2022				2023			
Quarter	I	II	III	IV	I	II	III	IV
Material Unloaded	180	168	159	175	190	205	180	182

If the forecast for the first quarter of 2022 is 175 tons then what is the forecast figure for the first quarter of 2024? Use a Smoothing coefficient of 0.1

2. The number of quarterly traffic accidents in a Metro city during 2021-2023 are as below –

Year / Quarter	1st	2nd	3rd	4th
2021	165	135	140	180
2022	152	121	127	163
2023	140	100	105	158

Find the Seasonal Indices by Trend Ratio Method, assuming a linear trend for the data.

3. For the following series of observations verify that the 4 yearly Centred Moving Average is equivalent to the 5 yearly Weighted Moving Average with weights 1, 2, 2, 2, 1 respectively.

Year	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
Sales (‘0000 ₹)	2	6	1	5	3	7	2	6	4	8	3

4. In the following dataset consider y as the Response Variable and x_1 and x_2 as the Predictor Variables to find the Linear Regression model. Interpret the estimators of the model.

y	140	155	159	179	192	200	212	215
x_1	60	62	67	70	71	72	75	78
x_2	22	25	24	20	15	14	14	11

5. While calculating the Coefficient of Correlation between two variables x and y, the following results were obtained: $N = 25, \Sigma x = 125, \Sigma y = 100, \Sigma x^2 = 650, \Sigma y^2 = 460, \Sigma xy = 508$. It was however discovered later on that two pairs of observation (x,y) were copied as (6,14) and (8,6) instead of (8,12) and (6,8) respectively. Determine the correct equations of the two Regression Lines.

Answers:

1. 178.6 Tons
2. 105 for 1st quarter, 83 for 2nd Quarter, 89 for 3rd Quarter and 123 for 4th Quarter.
4. $y = -6.867 + 3.148\,x_1 - 1.656\,x_2$
5. $9x - 5y - 25 = 0$ and $4x - 5y = 0$

References:

- Hamke & Wichern – Business Forecasting, Pearson Publications
- Das N. G. – Statistical Methods in Commerce (Part II), Accountancy and Economics, H. Das & Co.
- Kapoor V. K. – Operations Research (For Managerial Decision Making), Sultanchand & Sons.