UNIT 1 INTRODUCTION TO ECONOMETRICS

Structure

1.0 OBJECTIVES

After going through this unit you will be in a position to:

- explain why we should study econometrics;
- appreciate the scope of econometrics; and
- learn certain basic statistical tools used in econometrics.

1.1 INTRODUCTION

Econometrics has emerged as a specialized branch of economics which is concerned mostly with empirical verification of economic theory. Econometrics has both theoretical side as well as applied side. In theoretical econometrics several estimation techniques and test statistics for estimates are developed. Theoretical developments in econometrics have been quite rapid in recent years. It is somewhat difficult to keep track of all these developments and cover the entire lot in a single course.

The present unit serves as a foundation for the course. It reviews some of the background materials considered to be useful for studying econometrics. Thus it covers in very brief the concepts of probability distribution and statistical inference. The Unit begins with a definition of econometrics and how econometric methods are applied to real life situations.

1.2 THE NATURE OF ECONOMETRICS

Econometrics is a branch of economics which deals with economic

economic phenomena based on the concurrent development of theory and observation, related by appropriate methods of inference' [Samuelson, Koopmans and Stone, 1954]. We should notice three issues being emphasized in the above statement. First, econometrics deals with quantitative analysis of economic relationship. Second, it is based on economic theory and logic. Third, it requires appropriate methods to draw inferences.

Through logic and experience economists have attempted to establish relationship between several variables. Movements in certain economic variables are explained in terms of changes in some other variables. For example, the law of demand states that as price increases the quantity demanded of a commodity decreases. You will find numerous examples of such relationships. Some of these are apparent while others are quite complex. As we do not know what exactly is happening we often make conjectures based on logic and our understanding of things. Due to differences in perception, often different persons explain the same phenomenon differently.

Some of the economic relationships have been proved through empirical studies and validated under several situations. Some more are put forth as hypothesis and unequivocal conclusions are yet to be drawn. In both the cases, however, logic needs to be supported by empirical observation. An advantage with the subject matter of economics, unlike many other disciplines in social sciences, is that many of the economic variables can be quantified. This has helped in empirical measurement of economic variables and validation of economic theory.

We will show how econometrics is different from other braches of economics such as mathematical economics or economic statistics. Mathematical economics usually presents economic theory in mathematical form. It does not bother about empirical measurement of economic theory. Econometrics on the other hand is mainly concerned with empirical verification of economic theory. In doing so, econometrics makes use of the mathematical equations suggested in mathematical economics. Economic statistics deals with collection, processing and presentation of data. Statistics provides many theories on the basis of which inferences can be drawn. For example, the probability laws and sampling distribution are quite important in hypothesis testing. The application of these laws to economic theory through empirical research is a part of econometrics.

An important application of econometrics is prediction and forecasting on the basis of econometric models. By taking into account actual data it helps estimation of the parameters of a model and on the basis of those estimates how to make predictions. Thus, econometrics is helpful in policy analysis where we can simulate different values of exogenous variables and interpret their effects on endogenous variables.

Let us point out the important steps involved in undertaking an econometric study.

a) The first and foremost issue in econometrics is the statement of hypothesis. It is drawn from economic theory or some logic built by the researcher.

b) Second step is transformation of the above hypothesis into mathematical equation(s). This is our econometric model that we want to study.

c) Third step is the collection of relevant data required for the study. Data could be from primary or secondary sources.

d) The next step is estimation of the parameters of the econometric model. This requires selection of appropriate estimation method. As you will learn in subsequent units, there are several estimation techniques or methods available to us.

e) Once estimates are obtained, the next tasks testing of the hypothesis put forth in the first step above. Basically it amounts to statistical significance of the estimates obtained by us.

f) Finally, we need to interpret the results. On the basis of inferences drawn through hypothesis testing, we find out the implications of the results for the econometric model.

We mentioned above that statistical theory offers important guidelines to us in formulation and testing of hypothesis. We present below in very brief, some of the relevant statistical theory that we would be using frequently.

1.3 PROBABILITY DISTRIBUTIONS

Let us begin with concept of *random variable*. It is *a variable that takes different values with some probabilities*. Suppose the random variable X under consideration refers to the possible outcomes (called events) of tossing a coin. There are two events in this case (Head and Tail), which are exhaustive and mutually exclusive, and have probability of occurrence $\frac{1}{2}$ each. Similarly for a fair dice there are six outcomes and each event has a probability of $\frac{1}{6}$. We can generalize the concept as follows: A random variable X assumes values $X_1, X_2, \ldots\ldots\ldots X_n$ with corresponding probabilities $p_1, p_2, p_3, \ldots\ldots\ldots p_n$. We express it as $p(X = X_1) = p_1$. When the random variable assumes certain isolated values we call it a discrete random variable. On the other hand, if it assumes *any* value in an interval it is termed continuous random variable.

A *probability distribution* is a statement about the possible values of a random variable along with their respective probabilities. Thus it describes how the probabilities are distributed over the values of the random variable. For a discrete random variable the probability distribution function is called probability mass function whereas for a continuous random variable it is called probability density function.

1.3.1 Discrete Probability Distribution

For a discrete random variable, the *probability mass function* is given by $p(X)$, where p_i is probability and X_i is a value of the random variable, The probability mass function should satisfy the following two conditions:

i) Probability of an event cannot be negative, i.e., for any value X_i we have $p(X_i) \geq 0$

ii) Probabilities of all possible outcomes sum to unity, i.e., $\sum_{all\ x} p(X_i) = 1$

There are quite a few discrete probability distributions. We present below the probability mass function of an important probability distribution, that is, binomial distribution.

Binomial Distribution

The binomial distribution is an example of a discrete probability distribution. It signifies two possible outcomes of an experiment, the occurrence of an event or the non-occurrence of the event. A probability experiment can be termed as a Bernoulli experiment, if it satisfies the following conditions.

1) The experiment consists of a sequence of n repeated trials.

2) Each trial results in an outcome that may be classified either as a *success* or a *failure*.

3) The probability of a success, denoted by p, is known and remains the same in each trial. Consequently, the probability of a failure, denoted by $q = (1 - p)$ is also known and remains the same in each trial.

4) The trials are independent.

The probability of x successes in n trials is given by

$$p(x) = {}^{n}C_{x}\, p^{x} (1-p)^{n-x} \qquad \ldots(1.1)$$

$x = 0, 1, 2, \ldots\ldots, n.$

1.3.2 Continuous Probability Distribution

A continuous random variable X has a zero probability of assuming exactly any of its values. Apparently, this seems to be a surprising statement. Let us try to explain this by considering a random variable say, weight. Obviously weight is a continuous random variable since it can vary continuously. Suppose, we do not know the weight of a person exactly but have a rough idea that her weight falls between 50 kg and 51 kg. Now, there are an infinite number of possible weights between these two limits. As a result, by its definition, the probability of the person assuming a particular weight say, 50.3 kg will be negligibly small; almost equal to zero. But we can definitely attach some probability to the person's weight being between 50 kg and 51 kg. Thus, for a continuous random variable X, we assign a probability to an interval and not to a particular value. Here, we look for a function $p(X)$, called the *probability density function*, such that with the help of this function we can compute the probability $P(a < X < b)$, a and b are the limits of an interval *(a, b)* where, $a < b$.

A probability density function is defined in such a manner that the area under its curve bounded by x-axis is equal to one when computed over the domain of X for which $p(X)$ is defined. The probability density function for a continuous random variable X defined over the entire set of real numbers R should satisfy the following conditions.

1) $p(X_i) \geq 0$ for all $X_i \in R$

2) $\int_{-\infty}^{+\infty} p(X)dX = 1$

3) $p(a < X < b) = \int_a^b p(X)dX$

Normal Distribution

Normal distribution is perhaps the most widely used distribution in Statistics and related subjects. It has found applications in inquiries concerning heights and weights of people, IQ scores, errors in measurement, rainfall studies and so on. The probability density function $p(x)$ of a continuous random variable that follows the normal distribution is given by

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad \ldots(1.2)$$

where $-\infty < x < \infty$, and

$\pi = 3.17141$ (approximately)

$e = 2.71828$ (approximately).

The normal density function is completely determined by the parameters μ and σ. It means that given the values of μ and σ, we can trace out the normal curve by obtaining the values of $p(x)$ for different values of x. In fact, it can be shown that μ and σ are respectively the mean and the standard deviation of the normal distribution. When a random variable X follows normal distribution with mean μ and standard deviation σ we write it in symbols as $X \sim N(\mu, \sigma)$ and read as 'X follows normal distribution with mean μ and standard deviation σ.' The normal curve is a symmetrical bell-shaped curve as shown in Fig. 1.1.

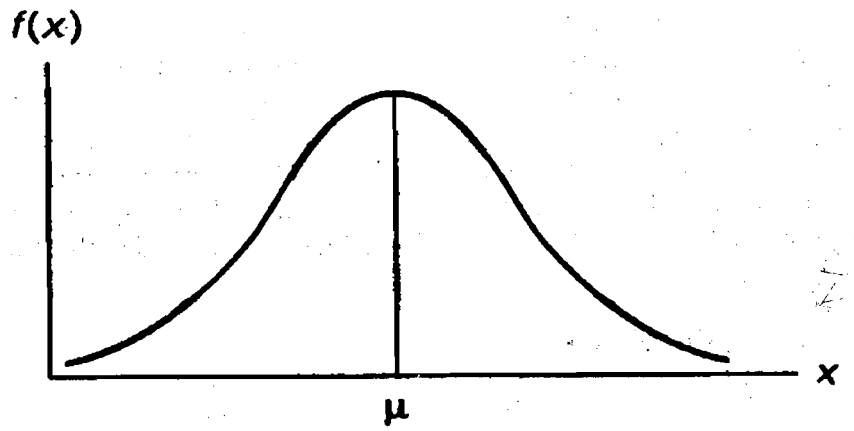

Fig 1.1: Normal Curve

The normal curve stretches from $-\infty$ to $+\infty$. It is symmetric about its mean. The following area-properties hold for a normal distribution. In Fig.1.2 below we plot a normal curve with mean $\mu = 50$ and standard deviation $\sigma = 4$.

a) 68.8 % of the area under the normal curve lies between the ordinates at $\mu - \sigma$ and $\mu + \sigma$. Thus in Fig.1.2, 68.8% area is covered when x ranges between 46 and 54.

b) 95.5% of the area under the normal curve lies between the ordinates at $\mu - 2\sigma$ and $\mu + 2\sigma$. In Fig. 1.2 95.5% area is covered when $42 \leq X \leq 58$.

c) 99.7% of the area (i.e., almost the whole of the distribution) under the normal curve lies between the ordinates at $\mu - 3\sigma$ and $\mu + 3\sigma$. In Fig. 15.2 we find that 99.7% area is covered when $38 \leq X \leq 62$.

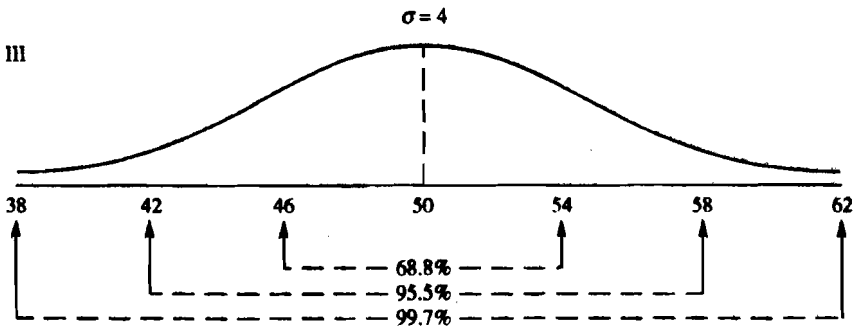

Fig. 1.2: Area Under Normal Curve

A problem encountered here is that μ and σ can take any value and finding out corresponding probability is time consuming. This problem is tackled by subtracting μ from the normal variable and dividing it by σ. This way we obtain the 'standard normal variate', $z = \frac{x-\mu}{\sigma}$, which has mean = 0 and standard deviation = 1.

The probability density function of the *standard normal variate*, z, is given by

$$p(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \qquad -\infty < z < \infty \qquad \ldots(1.3)$$

Once we obtain a standard normal variate, our seemingly hopeless task of obtaining probability areas for different combinations of μ and σ becomes elegantly simple. We should note that a standard normal variate has a unique mean of 0 and a unique standard deviation of 1. It means, if we can construct a table for probability areas of such a unique standard normal variate, it can be used for obtaining probability for any normal variable with any combination of mean and standard deviation. The only thing is that the given normal variable is to be transformed into the standard normal variate. In fact, such a table for areas (or probability) has been compiled for a standard normal variate (see Table A1 at the end of this Block: Areas under the Standard Normal Curve)

The Student's-*t* Distribution

W.S. Gosset presented the *t*-distribution. The interesting story is that Gosset was employed in a brewery in Ireland. The rules of the company did not permit any employee to publish any research finding independently. So, Gosset adopted the pen-name 'student' and published his findings about this distribution anonymously. Since then, the distribution has come to be known as the student's-*t* distribution or simply, the *t* distribution.

If z_1 is a standard normal variant, i.e., $z_1 \sim N(0,1)$ and z_2 is another independent variable that follows the chi-square distribution with k degrees of freedom, i.e. $z_2 \sim \chi^2_k$, then the variable

$$t = \frac{z_1}{\sqrt{(z_2 / k)}} = \frac{z_1 \sqrt{k}}{\sqrt{z_2}}$$

is said to follow student's-*t* distribution with *k* degrees of freedom.

The probability curves for the student's-*t* distribution for different degrees of freedom are presented in Fig. 1.3

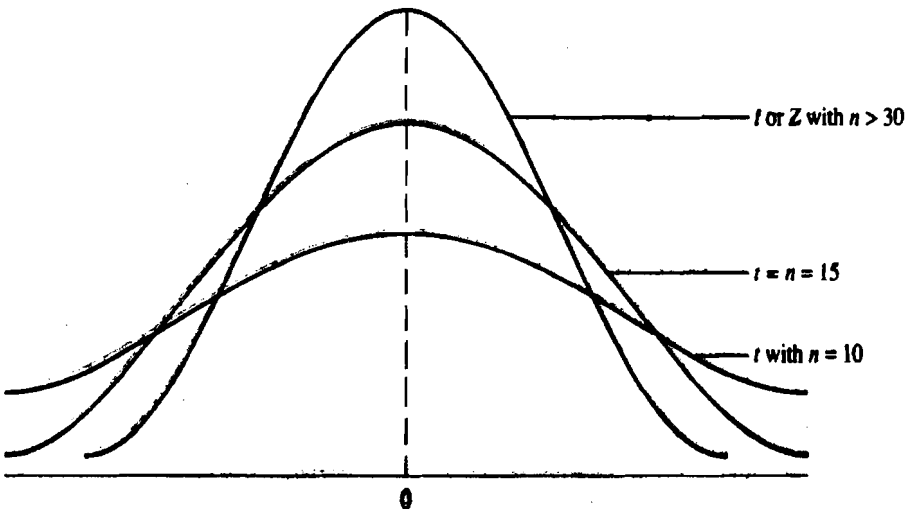

Fig 1.3: Student's-t Probability Curves

We note the important characteristics of this distribution.

1) As we can see in Fig. 1.3, like the normal distribution, the student's-*t* distribution is also symmetric and its range of variation is also from $-\infty$ to $+\infty$; however, it is flatter than the normal distribution. We should note that as the *degrees of freedom* increase, the student's-*t* distribution approaches the normal distribution.

2) The mean of the student's-*t* distribution is zero and its variance is $\frac{k}{(k-2)}$, where, *k* is the degrees of freedom.

Like the normal distribution, the student's-*t* distribution is often used in statistical inferences, particularly when the sample size is small. The task involves the integration of its density function; which may prove to be tedious. As a result, in this case also, like the normal distribution, a table has been constructed for ready-reference purposes (see Table A2 at the end of the Block).

Chi-Square Distribution

If *z* is a standard normal variate, as defined above, then the variable z^2 is said to be distributed as a χ^2 variable with one degree of freedom. Since χ^2 is a squared term, when z ranges from $-\infty$ to $+\infty$, chi-squares ranges from 0 to $+\infty$. Moreover, since z has a mean 0, most of the values taken by χ^2 will be close to 0. As a result, probability density of χ^2 distribution will be the maximum near 0. We can generalize the above. If $z_1, z_2, \ldots z_k$ are *k*

independent chi-square variables then the variable $z = \sum_{i=1}^{k} z_i$ follows χ^2 distribution with k degrees of freedom. We denote this by χ^2_k. The probability density functions of χ^2 for different degrees of freedom are given in Fig. 1.4. Observe that as degree of freedom increases, the distribution approaches normal distribution.

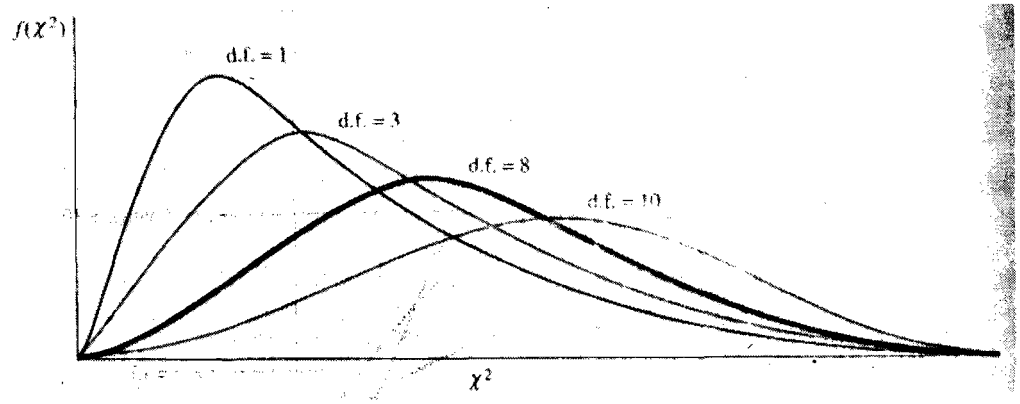

Fig 1.4: Chi-square Probability Curves

The critical values of Chi-square distribution is given in Table A3 at the end of the Block.

F- Distribution

Another continuous probability distribution that finds use in econometrics is the *F* distribution.

If z_1 and z_2 are two chi-square variables that are independently distributed with k_1 and k_2 degrees of freedom respectively, the variable

$$F = \frac{z_1 / k_1}{z_2 / k_2}$$

follows *F* distribution with k_1 and k_2 degrees of freedom respectively. The variable is denoted by F_{k_1,k_2}, where, the subscripts k_1 and k_2 are the degrees of freedom associated with the chi-square variables.

We may note here that k_1 is called the numerator degrees of freedom and in the same way, k_2 is called the denominator degrees of freedom.

Some important properties of the F distribution are mentioned below.

1) The *F* distribution, like the chi-square distribution, is also skewed to the right. But, as k_1 and k_2 increase, the *F* distribution approaches the normal distribution.

2) An *F* distribution with 1 and *k* as the numerator and denominator degrees of freedom respectively is the square of a student's-*t* distribution with *k* degrees of freedom.

   Symbolically,

$$F_{1,k} = t_k^2$$

3) For fairly large denominator degrees of freedom k_2, the product of the numerator degrees of freedom k_1 and the F value is approximately equal to the chi-square value with degrees freedom k_1, i.e., $k_1 F = \chi^2_{k_1}$.

The F distribution is extensively used in *statistical inference*. Again, such uses require obtaining areas under the F probability curve and consequently integrating the F density function. In this case also our task is facilitated by the provision of the F Table. The critical value for F distribution is given in Table A4 at the end of the Block.

1.4 SAMPLING DISTRIBUTION

Due to constraints such as costs, time and manpower it is often not possible to undertake a survey of the entire population (that is, census) and we resort to sampling (that is, we study part of the population). If the sample is drawn in a random manner through appropriate probability attached to each population unit and the sample size is not very small, the sample can be a representative one of the population. It is often argued that sampling results in less of *non-sampling errors* compared to census and thus it is advisable to go for sampling rather than census. Theoretically it is possible to draw a number of samples from a given population and each sample provides us with a sample mean (or any other statistic). Thus the sample means can be arranged in the form of a frequency distribution, called the *sampling distribution* of sample mean.

As sample mean ($\bar{x}$) assumes different values and for each value we can attach certain probability, it is considered as a random variable. In real life situations we have a finite population and the number of samples (and therefore number of sample means) is finite. In this case $\bar{x}$ is a discrete random variable but when there is infinite number of samples, $\bar{x}$ could be a continuous random variable.

Now let us consider another important concept: the *central limit theorem*. It says that sampling distribution of $\bar{x}$ is normal if the parent population from which the sample is drawn is normal. However, sampling distribution of $\bar{x}$ is approximately normal if sample size (n) is large, even if the parent population is not normal. If the parent population is approximately normal then sampling distribution of sample means is approximately normal even when sample size is small. This provides some rationale for assuming that probability distribution of a statistic is normal.

We know that dispersion of sample means is smaller in value than dispersion of the parent population from which the sample is drawn. The standard deviation of the sampling distribution is called *standard error*. Thus if the population has a standard deviation σ then the standard error of sample means is $\frac{\sigma}{\sqrt{n}}$.

Usually we consider a sample to be large in size if $n > 30$. For small samples ($n \leq 30$), sampling distribution of sample means is similar to student's t distribution. Recall that in the case of t distribution the shape of the probability curve changes according to its 'degrees of freedom'. For $n > 30$

the sampling distribution of a statistic is considered to follow normal distribution.

1.5 STATISTICAL INFERENCE

Statistical inference deals with the methods of drawing conclusions about the population characteristics on the basis of information contained in a sample drawn from the population. Population mean is not known to us, but we know the sample mean. In statistical inference we would be interested in answering two types of questions. First, what would be the value of the population mean? The answer lies in making an informed guess about the population mean. This aspect of statistical inference is called *estimation*. The second question pertains to certain assertion made about the population mean. Suppose a manufacturer of electric bulbs claims that the mean life of electric bulbs is equal to 2000 hours. On the basis of the sample information, can we say whether the assertion is correct or not? This aspect of statistical inference is called *hypothesis testing*. We discuss these two aspects below.

1.5.1 Estimation

Estimation could be of two types: *point estimation* and *interval estimation*. In point estimation we estimate the value of the population parameter as a single point. On the other hand, in the case of interval estimation we estimate lower and upper bounds around sample mean within which population mean is likely to remain.

a) Point Estimation

As mentioned earlier we do not know the parameter value and want to guess it by using sample information. Obviously the best guess will be the value of the sample statistic. Here we use a single value or point as 'estimate' of the parameter.

Let us distinguish between the concepts *estimate* and *estimator*. The estimator is the formula and estimate is the particular value obtained by using the formula. For example, if we use sample mean for estimation of population mean, then $\frac{1}{n}\sum_i x_i$ is the estimator. Suppose I collect data on a sample, and put the sampling units to this formula and obtain a particular value for sample mean, say 120. Then 120 is an estimate of population mean. It is possible that you draw another sample from the same population, use the formula for sample mean, that is $\frac{1}{n}\sum_i x_i$, and obtain a different value, say 123. Here both 120 and 123 are estimates of population mean. But in both the cases the estimator is the same, which is $\frac{1}{n}\sum_i x_i$. Remember that the term *statistic*, which is used to mean a function of sample values, is a synonym for estimator.

There may be situations when you would find more than one potential estimator (alternative formulae) for a parameter. In order to choose the best among these estimators, we need to follow certain criteria. These are as follows:

i) *Unbiasednes:* If $\hat{\theta}$ is a statistic based on n observations then it is said to be unbiased if $E(\hat{\theta}) = \theta$. It implies that an estimate may be higher or lower than the unknown value of the parameter but the expected value of the estimate should be equal to the parameter. The extent of bias in a statistic (or estimator) is given by $E(\hat{\theta}) - \theta$. You can find that sample mean is an unbiased estimate of population mean.

ii) *Consistency:* Consistency and asymptotic bias are large sample properties of an estimator. A consistent estimator is defined as follows: an estimator $\hat{\theta}$ is said to be consistent if $plim(\hat{\theta}) = \theta$. The notation *plim* stands for 'probability limit' and it plies that as sample size n increases infinitely the statistic should be equal to the parameter value. Notice the difference between unbiasedness and consistency. In the case unbiasedness the statistic on the average equals parameter value. In the case of consistency, the statistic equals the parameter value when sample size tends to infinity. Consistency is important because some estimators may not be consistent. Let us take an example from Patterson (2000).

Suppose the statistic (a random variable, as we mentioned earlier) assumes two discrete values with corresponding probabilities:

$\hat{\theta} = \theta$ with probability $\frac{n-1}{n}$

$\hat{\theta} = n$ with probability $\frac{1}{n}$

In this case $plim\ (\hat{\theta}) = \theta$. Because as $n \to \infty$ the probability $\frac{1}{n} \to 0$. Thus the statistic is consistent. However, it is not unbiased. For example, if sample size is 10, then $E(\hat{\theta}) = \theta\ (\frac{9}{10}) + \frac{1}{10}\ (10) = 0.9\theta + 1 \neq \theta$.

iii) *Efficiency*: It refers to the variance of an estimator. An estimator with a smaller variance is said to be more efficient. Usually comparison in variance is done among estimators of the same class. For example, we can compare among linear estimators (that is, estimators which are linear combinations of sample observations) and say that OLS estimator is efficient than other linear estimators.

iv) *Asymptotic bias*: A statistic is said to be asymptotically unbiased if its 'asymptotic expectation' equals the parameter value. Asymptotic expectation of a statistic is given by

$$AE(\hat{\theta}) = \lim E(\hat{\theta}) \quad \text{as } n \to \infty .$$

We observe that an unbiased estimator is asymptotically unbiased but the reverse is not essentially true. For example, suppose we modify our previous example as

$\hat{\theta} = \theta$ with probability $\frac{n-1}{n}$

$\hat{\theta} = n^{\frac{1}{2}}$ with probability $\frac{1}{n}$

In this case

$$bias = E(\hat{\theta}) - \theta = \theta\left(\frac{n-1}{n}\right) + n^{\frac{1}{2}}\left(\frac{1}{n}\right) - \theta = \theta - \frac{\theta}{n} + \frac{1}{n^{\frac{1}{2}}} - \theta = -\frac{\theta}{n} + \frac{1}{n^{\frac{1}{2}}}$$

Thus the above statistic is not unbiased. In the limiting case, as $n \to \infty$, however, $\lim E(\hat{\theta}) = \theta$. Thus it is asymptotically unbiased.

Interval Estimation

The point estimate may not be realistic in the sense that the parameter value may not exactly be equal to it. An alternative procedure is to give an interval, which would hold the parameter with certain probability. Here we specify a lower limit and an upper limit within which the parameter value is likely to remain. Also we specify the probability of the parameter remaining in the interval. We call the *interval* as 'confidence interval' and the *probability* of the parameter remaining within this interval as 'confidence level' or 'confidence coefficient'.

'How do we find out the confidence interval and confidence coefficient?' Let us begin with confidence coefficient.

We know that the sampling distribution of $\bar{x}$ for large samples is normally distributed with mean μ and standard error $\frac{\sigma}{\sqrt{n}}$, where n is the size of the sample. By transforming the sampling distribution ($z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$) we obtain *standard normal variate*, which has zero mean and unit variance. The standard normal curve is symmetrical and therefore, the area under the curve for $0 \le z \le \infty$ is 0.5 as we can see from Table A1 given at the end of the Block. If we want our confidence coefficient to be 95 per cent (that is, 0.95), we find out a range for z which will cover 0.95 area of the standard normal curve. Since distribution of z is symmetrical, 0.475 area should remain to the right and 0.475 area should remain to the left of $z = 0$. From Table A1 we find that 0.475 area is covered when z=1.96. Thus the probability that z ranges between −1.96 to 1.96 is 0.95. From this information let us work out backward and find the range within which μ will remain.

We find that

$$P(-1.96 \le z \le 1.96) = 0.95$$

$$or\ P\left(-1.96 \le \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \le 1.96\right) = 0.95$$

$$or\ P\left(-1.96\frac{\sigma}{\sqrt{n}} \le \bar{x} - \mu \le 1.96\frac{\sigma}{\sqrt{n}}\right) = 0.95$$

$$or\ P\left(\bar{x} - 1.96\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}\right) = 0.95 \qquad \ldots (1.6)$$

As each sample would provide us with a different value of $\bar{x}$, the confidence interval would be different. In each case the confidence interval may contain the unknown parameter or it may not. Equation (1.6) means that if a large number of random samples, each of size n, are drawn from the given population and if for each such sample the interval

$\left(\bar{x} - 1.96\frac{\sigma}{\sqrt{n}}, \leq \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}\right)$ is determined, then in about 95% of the cases, the interval will include the population mean μ.

The confidence coefficient is denoted by $(1-\alpha)$ where α is the level of significance. Confidence coefficient could take any value. We can very well ask for a confidence level of say 81 per cent or 97 per cent depending upon how precise our conclusions should be. However, conventionally two confidence levels are frequently used, namely, 95 per cent and 99 per cent.

Let us find out the confidence interval when confidence coefficient $(1-\alpha) = 0.99$. In this case 0.495 area should remain on either side of the standard normal curve. If we look into the normal area table (Table A1 at the end of the Block) we find that 0.495 area is covered when z = 2.58.

Thus

$$P(-2.58 \leq z \leq 2.58) = 0.99 \qquad \ldots(1.7)$$

By rearranging the terms in the above we find that

$$P\left(\bar{x} - 2.58\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + 2.58\frac{\sigma}{\sqrt{n}}\right) = 0.99 \qquad \ldots(1.8)$$

Equation (1.8) implies that 99 per cent confidence interval for μ is given by $\bar{x} \pm 2.58\frac{\sigma}{\sqrt{n}}$.

By looking into the normal area table you can work out the confidence interval for confidence coefficient of 0.90 and find that

$$P\left(\bar{x} - 1.65\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + 1.65\frac{\sigma}{\sqrt{n}}\right) = 0.90 \qquad \ldots(1.9)$$

We observe from (1.7), (1.8) and (1.9) that as the interval widens, the probability of the interval holding a population parameter (in this case μ) increases.

The two limits of the confidence interval are called *confidence limits*. For example, for 95 per cent confidence level we have the lower confidence limit as $\bar{x} - 1.96\frac{\sigma}{\sqrt{n}}$ and upper confidence limit as $\bar{x} + 1.96\frac{\sigma}{\sqrt{n}}$. The confidence coefficient can be interpreted as the confidence or trust that we place in these limits for actually holding μ.

1.5.2 Hypothesis Testing

A hypothesis is a tentative statement about a characteristic of a population. It could be an assertion or a claim also. In hypothesis testing there are four important components: i) null hypothesis, ii) alternative hypothesis, iii) test statistic, and iv) interpretation of results.

Null and Alternative Hypothesis

Usually statistical hypotheses are denoted by the alphabet *H*. There are two types of hypothesis: null hypothesis and alternative hypothesis. A null hypothesis is the statement that we consider to be true about the population

and put to test by a test statistic. We denote the null hypothesis by H_0. Suppose on the basis of our logic we put forth the view that female literacy in a village in Orissa is higher than that for Orissa. We begin with the presumption that female literacy in the village is equal to that of Orissa (which is say 51 per cent). Thus our null hypothesis is that 'sample mean is equal to population mean'.

$$H_0 : \mu = 0.51 \qquad \ldots(1.10)$$

where μ is the parameter, in this case female literacy in Orissa.

There is a possibility that the null hypothesis that we intend to test is not true and female literacy in Orissa is not equal to 51 per cent. Thus there is a need for an alternative hypothesis which holds true in case the null hypothesis is not true. We denote the alternative hypothesis by the symbol H_A and formulate it as follows:

$$H_A : \mu \neq 0.51 \qquad \ldots(1.11)$$

We have to keep in mind that null hypothesis and alternative hypothesis are mutually exclusive, that is, both cannot be true simultaneously. Secondly, both H_0 and H_A exhaust all possible options regarding the parameter, that is, there cannot be a third possibility. For example, in the case of female literacy in the village, there are two possibilities – literacy rate is 51 per cent or it is not 51 per cent; a third possibility is not there.

It is a rare coincidence that sample mean ($\bar{x}$) is equal to population mean (μ). In most cases we find a difference between $\bar{x}$ and μ. Is the difference because of sampling fluctuation or is there a genuine difference between the sample and the population? In order to answer this question we need a test statistic to test the difference between the two. The result that we obtain by using the test statistic needs to be interpreted and a decision needs to be taken regarding whether the null hypothesis be rejected or not.

Rejection Region

While discussing confidence interval we mentioned if confidence level is 95 per cent then 5 per cent area of the standard normal curve remains under the rejection region. Let us look into the standard normal curve presented in Fig. 1.5, where the x-axis represents the variable z and the y-axis represents the probability of z, that is $p(z)$. If the estimate falls under the rejection region then the null hypothesis is rejected. Otherwise, the hypothesis is not rejected.

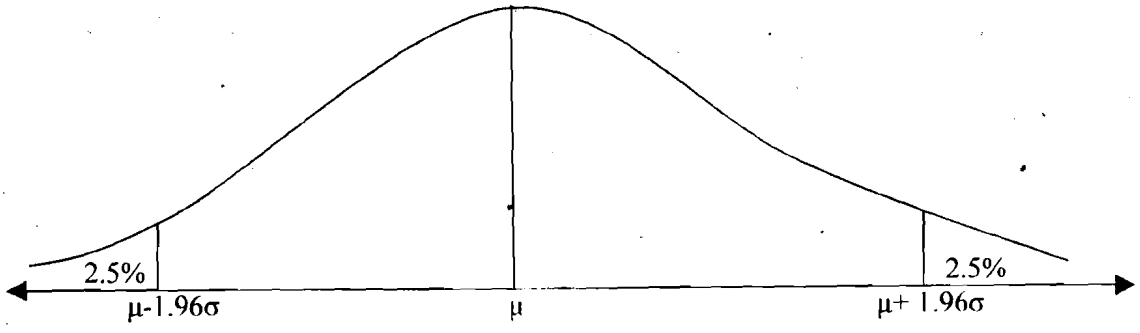

Fig. 1.5: Rejection Region in Standard Normal Curve

We should note the following points.

- When sample mean is equal to population mean (that is, $\bar{x} = \mu$) we find that $z = 0$. When $\bar{x} > \mu$ we find that z is positive and when $\bar{x} < \mu$ we find z to be negative.
- Note that we are concerned with the difference between $\bar{x}$ and μ. Therefore, negative or positive sign of z does not matter much.
- Higher the difference between $\bar{x}$ and μ, higher is the absolute value of z. Thus z-value measures the discrepancy between $\bar{x}$ and μ, and therefore can be used as a test statistic.
- We should find out a critical value of z beyond which the difference between $\bar{x}$ and μ is significant.
- If the absolute value of z is less than the critical value we should not reject the null hypothesis.
- If the absolute value of z exceeds the critical value we should reject the null hypothesis and accept the alternative hypothesis.

Thus in the case of large samples the absolute value of z can be considered as test statistic for hypothesis testing such that

$$z = \frac{|\bar{x} - \mu|}{\sigma/\sqrt{n}} \qquad \ldots(1.12)$$

When we have a significance level of 5 percent, the area covered under the standard normal curve is 95 per cent. Thus 95 per cent area under the curve is bounded by $-1.96 \le z \le 1.96$. The remaining 5 per cent area is covered by $z \le -1.96$ and $z \ge 1.96$. Thus 2.5 per cent of area on both sides of the standard normal curve constitute the rejection region.

For small samples ($n \le 30$), if population standard deviation is known we apply z-statistic for hypothesis testing. On the other hand, if population standard deviation is not known we apply t-statistic. The same criteria apply to hypothesis testing also.

In the case of small samples if population standard deviation is known the test statistic is

$$z = \frac{|\bar{x} - \mu|}{\sigma/\sqrt{n}} \qquad \ldots(1.13)$$

On the other hand, if population standard deviation is not known the test statistic is

$$t = \frac{|\bar{x} - \mu|}{s/\sqrt{n}} \qquad \ldots(1.14)$$

In the case of t-distribution, however, the area under the curve (which implies probability) changes according to degrees of freedom. Thus while finding the critical value of t we should take into account the degrees of freedom. When sample size is n, degrees of freedom is $n-1$. Thus we should remember two

things while finding critical value of t. These are: i) significance level, and ii) degrees of freedom.

One-tail and Two-tail Tests

In Fig. 1.4 we have shown the rejection region on both sides of the standard normal curve. However, in many cases we may place the rejection region on one side (either left or right) of the standard normal curve.

Remember that if α is the level of significance, then for a two-tail test $\frac{\alpha}{2}$ area is placed on both sides of the standard normal curve. But if it is a one-tail test then α area is placed on one-side of the standard normal curve. Thus the critical value for one-tail and two tail test differ.

The selection of one-tail or two-tail test depends upon the formulation of the alternative hypothesis. When the alternative hypothesis is of the type $H_A : \bar{x} \neq \mu$ we have a two-tail test, because $\bar{x}$ could be either greater than or less than μ. On the other hand, if alternative hypothesis is of the type $H_A : \bar{x} < \mu$, then entire rejection is on the left hand side of the standard normal curve. Similarly, if the alternative hypothesis is of the type $H_A : \bar{x} > \mu$, then the entire rejection is on the right hand side of the standard normal curve.

1.6 SOFTWARE PACKAGES FOR ECONOMETRIC ANALYSIS

As econometric analysis deals with empirical data it involves cumbersome computations. We can think of estimating the parameters manually in some of the simpler estimation methods, such as ordinary least squares. In other cases, however, it is quite difficult and time consuming. Some of the econometric methods are iterative in nature and thus require repeated calculations.

In recent years, easy access to computers and availability of Software Packages for Econometric Analysis have made the job simpler. We find that the most difficult computations can be done by the computer in few seconds. There are several software packages available for econometric applications. Most of these software packages include the important econometric applications. Moreover, since they apply similar methods, they provide the same results for a problem. We provide a list of selected software packages below.

SPSS/PC+ by SPP Inc., 444 N. Michigan Avenue, Chicago, IL 60611, USA.

STATA by Computing Resources Centre, 10801 National Blvd., 3rd Floor, Los Angeles, CA 90064, USA.

SHAZAM by Kenneth J. White, department of Economics, university of British Columbia, Vancouver, BC V6T 1Y2, Canada.

LIMDEP by W. H. Greene, Stern Graduate School of Business, New York University, 100 Trinity Place, New York, USA.

SAS/STAT by SAS Institute Inc., PO Box 1818, Evanston., IL, USA.

1.7 LET US SUM UP

In this Unit we provided an outline of some essential statistical tools studied some continuous probability distributions. Among these distributions, the normal distribution is considered to be the most important one.

Besides the normal distribution, we have considered three other continuous probability distributions, viz., the chi-square distribution, the student's-*t* distribution and the *F* distribution. We have already seen from the features of the chi-square, student's-*t* and the *F* distributions that for large degrees of freedom, these distributions approach the normal distribution. This relationship between the chi-square distribution, the student's-*t* distribution and the *F* distribution on one hand and the normal distribution on the other has tremendous practical implications. When the degrees of freedom happen to be fairly large, instead of using the chi-square distribution or the student's-*t* distribution or the *F* distribution separately as the situation may demand; we can uniformly apply the normal distribution. As a result, our task gets considerably simplified.

1.8 KEY WORDS

Chi-square Distribution	:	It is an asymmetric distribution where the range of variation for the random variable is from zero to infinity. For fairly large degrees of freedom, it approaches the normal distribution.
Continuous Probability Distribution	:	It is the probability distribution for a continuous random variable.
Degrees of Freedom	:	It refers to the number of pieces of independent information that are required to compute some characteristic of a given set of observations.
Discrete Probability Distribution	:	It is the probability distribution for a discrete random variable.
Estimation	:	It is the method of prediction about parameter values on the basis of sample statistics.
Estimator	:	It is another name given to statistic in the theory of estimation.
***F* Distribution**	:	It is an asymmetric distribution that is skewed to the right. For fairly large degrees of freedom, it approaches the normal distribution.
Normal Distribution	:	The best known of all the theoretical probability distributions. It traces out a bell-shaped symmetric probability curve.
Normal Variable	:	A random variable that follows the normal distribution.
Parameter	:	It is a measure of some characteristic of the population.

Population : It is the collection of all units of a specified type in a given place and at a particular point of time.

Probability Distribution : It is a statement about the possible values of a random variable along with their probabilities.

Random Sampling : It is a procedure where every member of the population has a definite chance or probability of being selected in the sample. It is also called probability sampling.

Sample : It is a sub-set of the population. It can be drawn from the population in a scientific manner by applying the rules of probability so that personal bias is eliminated. Many samples can be drawn from a population and there are many methods of drawing a sample.

Sampling Distribution : It is the relative frequency or probability distribution of the values of a statistic when the number of samples tends to infinity.

Sampling Error : In the sampling method, we try to approximate some feature of a given population from a sample drawn from it. Now, since in the sample all the members of the population are not included, howsoever close the approximation is, it is not identical to the required population feature and some error is committed. This error is called the sampling error.

Significance Level : There may be certain samples where population mean would not remain within the confidence interval around sample mean. The percentage (probability) of such cases is called significance level. It is usually denoted by α. When $\alpha = 0.05$ (that is, 5 percent) we can say that in 5 per cent cases we are likely to reach an incorrect decision.

Standard Error : A sample statistic varies across samples which can be presented in the form of a probability distribution. Standard error is the standard deviation of the sampling distribution of a statistic.

Standard Normal Variate : A normal variable with mean 0 and standard deviation equal to 1.

Statistic : It is a function of the values of the units that are included in the sample. The basic purpose of a statistic is to estimate some population parameter.

Statistical Inference : It is the process of concluding about an unknown population from a known sample drawn from it.

1.9 SOME USEFUL BOOKS/REFERENCES

Gujarati, D., 1995, Basic Econometrics, McGraw-Hill, Singapore.

IGNOU, 2005, *EEC-13: Elementary Statistical Methods and Survey Techniques*, Blocks 5, 6 and 7.

Nagar, A. L. and R. K. Das, 1989, *Basic Statistics:* Oxford University Press, Delhi.

Patterson, K, 2000, *An Introduction to Applied Econometrics*, Palgrave, New York.

Samuelson, P. A., T. C. Koopmans, and J. R. N. Stone, 'Report of the Evaluative committee for Econometrica', *Econometrica*, vol. 22., no. 2, pp. 141-46.

UNIT 2 ESTIMATION OF TWO-VARIABLE REGRESSION MODEL

Structure

2.0 OBJECTIVES

After going this unit you will be able to:

- explain the concept of two variable model;
- explain the justification of error term;
- estimate beta coefficients;
- explain the concept of coefficient of determination; and
- explain the concept of intrinsically linear models.

2.1 INTRODUCTION

The simplest economic relation is represented through a two-variable model. For example, quantity demanded is a function of price; consumption is a function of income, etc. However, more realistic formulations require the specifications of several variables in each model. To simplify the presentation of the estimation for these relationships statistically, we examine first the two-variable linear equation model, namely

$$Y = a + bX \qquad \ldots(2.1)$$

where a and b are unknown parameters indicating the intercept and the slope of the model respectively. They are also called the regression coefficients. Y, the left-hand variable, is called the dependent variable and X, the right-hand variable, is called the independent variable.

2.2 ERROR TERM

Economic theory specifies exact functional relationships among its variables. In reality the measurement of the functional relationships among its variables is not exact. For a given observed value of X (the independent variable), we may observe many possible values of Y (the dependent variable). As an example, consider the consumption of an individual who receives a certain amount of income each year. Since the amount of money spent on a particularly item, say food, is likely to vary each year, we assume that for each observation X, observations on Y will differ randomly. To describe this situation formally, we add a random "error" term to the model, and writing it as

$$Y_i = \alpha + \beta X_i + \varepsilon_i \qquad \ldots(2.2)$$

where Y is a random variable, X is fixed or non-stochastic, and ε is a random error term whose value is based on an underlying probabilistic distribution. We have switched our notation to use Greek letters α and β to represent the intercept and slope of the line, i.e., the regression parameters, since our model now contains a random error term.

There are several justifications for the introduction of error term into the equation. First, we may have errors of specification. Errors appear because the model is a simplification of reality. It is not always possible to include all relevant variables in the functional relation. For example, we assume that price is a sole determinant of demand for a product. In fact several variables related to demand, like individual tastes, population, income, etc. do influence the price of the product, which are omitted from the function. Some of these variables have positive effects while others have negative effects. The net effects of these omitted variables in the equation are represented by the error term. If the effects of these omitted variables are small, it is reasonable to assume that the error term is random.

A second source of error is associated with the collection and measurement of the data. It is likely that income and consumption information we obtained from a household may not be accurate. If the data are obtained from a government statistical report, the data may not be accurate as a result of clerical handling and rounding error.

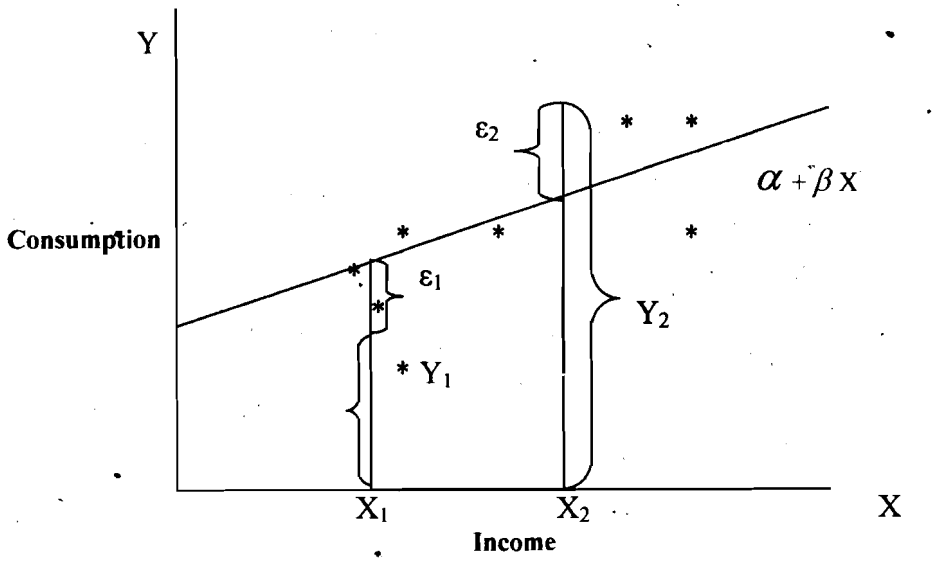

Fig. 2.1: Least Squares Fitting and Random Disturbances

A third source of error is the sampling error. For instance, consider eq. (2) as a household-consumption function, where Y is consumption and X is income. Even if eq. (2.2) is a correct relationship, the sample we randomly choose to examine may turn out to be predominantly poor families. Thus our estimation of α and β from this sample group may not be as good as the estimates from a balanced sample group.

When eq. (2.2) has either one or more of these three sources of error, it is justified for the introduction of an error term. Figure 2.1 illustrates eq. (2.2) using a household consumption as an example.

Given the above sources of error, the representation of the relationship in eq (2.2) as a stochastic one is clear. For every value of X there exists a probability distribution of ε and therefore a probability distribution of the Y's. Thus we say variable Y is stochastic in nature. On the other hand, the variable X is non-stochastic. Its values are kept fixed from sample to sample.

2.3 ESTIMATION

In the specification of the model as described by eq. (2.2), the values of the parameters α and β are not known, as a result the population regression line is not known. When the values of α and β are estimated, we obtain a sample regression line that serves as an estimate of the population regression line. If α and β are estimated by $\hat{\alpha}$ and $\hat{\beta}$ respectively, then the sample regression line is given by,

$$\hat{Y}_i = \hat{\alpha} + \hat{\beta} X_i \qquad \ldots(2.3)$$

where $\hat{Y}_i$ is the fitted value of Y_i.

The theory of estimation can be divided into two parts: point estimation and interval estimation. In point estimation, the aim is to use the prior and the sample information for the purpose of calculating a value which would be, in some sense, our best guess as to the actual value of the parameter of interest. In interval estimation, on the other hand, the same information is used for the purpose of producing an interval which would contain the true value of the parameter with some given level of probability. Since an interval is fully characterised by its limits, estimating an interval is equivalent to estimating its limits. The interval itself is usually called a confidence interval. Confidence intervals can also be viewed as possible measures of the precision of a point estimator.

The problem of point estimation is that of producing an estimate that would represent our best guess about the value of the parameter. To solve this problem we have to do two things. First, to specify what we mean by 'best guess' and second, to devise estimators that would meet this criterion. The first part of the problem amounts to specifying various properties of an estimator that can be considered desirable. The properties of the estimators are: unbiasedness, efficiency, minimum mean square error and consistency. The second part, on the other hand, involves devising estimators that would have at least some of the desirable properties. The estimators are given names

that indicate nature of the principle used in devising the formula. We will discuss here the least-squares method of devising estimators.

Check your Progress 1

1) What do you mean by two-variable regression model?

.....

.....

.....

.....

.....

2) Give the justifications of the error term in regression model.

.....

.....

.....

.....

.....

2.4 LEAST SQUARES METHOD OF ESTIMATION

The dominating and powerful estimating principle is that of least squares. We want to estimate the relationship between Y and X from the sample observations above such that

$$\hat{Y}_i = \hat{\alpha} + \hat{\beta} X_i$$

where $\hat{\alpha}$ and $\hat{\beta}$ are estimates of the unknown parameters α and β and $\hat{Y}$ is the estimated value of Y. The deviations between the observed and estimated values of Y are called residuals ε, i.e.,

$$\varepsilon = Y - \hat{Y}. \qquad \ldots(2.4)$$

The principle of least squares is to choose $\hat{\alpha}$ and $\hat{\beta}$ values that will minimise the sum of squared deviations between the observed and estimated values of Y.

The estimated equation will be the best fitting curve on the least squares criterion. We have therefore

$$\sum \varepsilon_i^2 = \sum \left(Y_i - \hat{Y}_i\right)^2 \quad \text{where} \quad i = 1, 2, \cdots, n$$

$$= \sum \left(Y_i - \hat{\alpha} - \hat{\beta} X_i\right). \qquad \ldots(2.5)$$

Now making deviation of both sides and equating zero, we have

$$\frac{\partial}{\partial \hat{\alpha}} \sum \varepsilon_i^2 = -2 \sum \left(Y_i - \hat{\alpha} - \hat{\beta} X_i \right) = 0$$

$$\frac{\partial}{\partial \hat{\beta}} \sum \varepsilon_i^2 = -2 \sum X_i \left(Y_i - \hat{\alpha} - \hat{\beta} X_i \right) = 0$$

Rearranging these two equations gives the normal equations.

$$\sum Y_i = n\hat{\alpha} + \hat{\beta} \sum X_i \qquad \ldots(2.6)$$

and

$$\sum X_i Y_i = \hat{\alpha} \sum X_i + \hat{\beta} \sum X_i^2 \qquad \ldots (2.7)$$

Now solving the above two normal equations, we can get the values of $\hat{\alpha}$ and $\hat{\beta}$

$$\hat{\beta} = \frac{\left(\sum X_i \right) \left(\sum Y_i \right) - n \sum X_i Y_i}{\left(\sum X_i \right)^2 - n \sum X_i^2} \qquad \ldots(2.8)$$

and

$$\hat{\alpha} = \bar{Y} - \hat{\beta} \bar{X} \qquad \ldots(2.9)$$

where

$$\bar{Y} = \frac{1}{n} \sum Y_i$$

and

$$\bar{X} = \frac{1}{n} \sum X_i .$$

The value of $\hat{\beta}$ can also be shown in deviation of Y and X from their means.

Using $x_i = X_i - \bar{X}$ and $y_i = Y_i - \bar{Y}$, we get

$$\hat{\beta} = \frac{\sum x_i y_i}{\sum x_i^2} \qquad \ldots(2.10)$$

Example 2.1

We can illustrate the application of the two variable models discussed in the earlier section. Following is a numerical illustration of estimating a least-squares equation.

Let us take the relationship between savings and disposable income as follows:

$$Y_i = \hat{\alpha} + \hat{\beta} X_i + \varepsilon_i$$

where Y is savings and X is disposable income. We take the data for the above variables as follows:

Year	Savings (Y) (Rs)	Disposable income (X) (Rs)
1996	16.95	0.84
1997	18.25	1.34
1998	19.56	1.75
1999	20.46	1.55
2000	21.76	1.63
2001	23.39	1.89
2002	25.00	2.13
2003	26.47	2.23
2004	27.52	2.18
2005	29.27	2.20

From the above table we can get the following values

$$\sum X = 228.63 \qquad \sum X^2 = 5381.9041$$

$$\sum Y = 17.74 \qquad \sum Y^2 = 33.2974$$

$$\sum XY = 420.9776$$

Using these aggregates, we can get

$$\sum xy = \sum XY - \frac{1}{N}\sum X \sum Y = 15.3880$$

$$\sum x^2 = \sum X^2 - \frac{1}{N}\left(\sum X\right)^2 = 154.7364$$

N, the number of observations, is equal to 10.

The regression coefficient is estimated as:

$$\hat{\beta} = \frac{\sum xy}{\sum x^2}$$

$$= \frac{15.3880}{154.7364}$$

$$= 0.099$$

The intercept is calculated by using the expression as

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X} = \frac{\sum Y}{N} - \hat{\beta}\frac{\sum X}{N}$$

$$= 1.774 - (0.099 \times 22.863)$$

$$= 1.774 - 2.263$$

$$= -0.489$$

Having estimated the intercept and slope of the regression line, we can now construct the regression equation of savings of income as

$$\hat{Y} = \hat{\alpha} + \hat{\beta}X = -0.489 + 0.099X \qquad \ldots(2.11)$$

Example 2.2

Let us take the relationship between consumption and income as follows:

$$Y_i = \hat{\alpha} + \hat{\beta}X_i + \varepsilon_i$$

where Y is consumption and X is income. Let us take the hypothetical data for the above variables as follows:

Consumption (Y) (Rs)	Income (X) (Rs)
70	80
65	100
90	120
95	140
110	160
115	180
120	200
140	220
145	240
150	260

From the above table we can get the following values

$$\sum X = 1700 \qquad \sum X^2 = 322000$$

$$\sum Y = 1110 \qquad \sum Y^2 = 132100$$

$$\sum XY = 205500$$

Using these aggregates, we can get

$$\sum xy = \sum XY - \frac{1}{N}\sum X \sum Y = 16800$$

$$\sum x^2 = \sum X^2 - \frac{1}{N}\left(\sum X\right)^2 = 33000$$

N, the number of observations, is equal to 10

The regression coefficient is estimated as:

$$\hat{\beta} = \frac{\sum xy}{\sum x^2}$$

$$= \frac{16800}{33000}$$

$$= 0.5091$$

The intercept is calculated by using the expression as

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X} = \frac{\sum Y}{N} - \hat{\beta}\frac{\sum X}{N}$$

$$= 111 - 0.5091(170)$$

$$= 24.4545$$

The estimated regression line therefore is

$$\hat{Y} = \hat{\alpha} + \hat{\beta}X = 24.4545 + 0.5091X \quad \ldots(2.12)$$

2.5 COEFFICIENT OF DETERMINATION

Regression residuals can provide a useful measure of the fit between the estimated regression line and the data. A good regression equation is one which helps explain a large proportion of the variance of Y. To find a measure of goodness of fit, it is reasonable to use the residual variance divided by the variance of Y. The variation in Y is given by

$$\text{Variation}(Y) = \sum(Y_i - \bar{Y})^2$$

Our goal is to divide the variation of Y into two parts, the first accounted for by the regression equation and the second associated with the unexplained portion of the model.

Consider the following identity, which holds for all observations:

$$(Y_i - \bar{Y}) = (Y_i - \hat{Y}_i) + (\hat{Y}_i - \bar{Y}) \quad \ldots(2.13)$$

The term on the left hand side of the equals sign denotes the difference between the sample value of Y and the mean of Y. The first right hand term gives the residual $\hat{\varepsilon}_i$, and the second right-hand term gives the difference between the predicted value of Y and the mean of Y.

To measure variation, we square both sides of eq. (2.13) and sum over all observations (1 to N),

$$\sum(Y_i - \bar{Y})^2 = \sum(Y_i - \hat{Y}_i)^2 + \sum(\hat{Y}_i - \bar{Y})^2 + 2\sum(Y_i - \hat{Y}_i)(\hat{Y}_i - \bar{Y}) \ldots(2.14)$$

The last term in eq. (2.14) can be shown to be 0 by using two properties of least-squares residuals, $\sum \hat{\varepsilon}_i = 0$ and $\sum \hat{\varepsilon}_i X_i = 0$. It follows that

$$\sum(Y_i - \bar{Y})^2 = \sum(Y_i - \hat{Y}_i)^2 + \sum(\hat{Y}_i - \bar{Y})^2$$

Variation in Y	Residual Variation	Explained Variation

or, TSS = ESS + RSS ...(2.15)

Total sum of squares | Error (Residual) sum of squares | Regression sum of squares

From the above[1] we say that the 'total sum of squares' (TSS) is the sum of the 'error sum of squares' and the 'regression sum of squares'.

We divide both sides of eq. (2.15) by the total sum of squares (TSS) to get

$$1 = \frac{ESS}{TSS} + \frac{RSS}{TSS}$$

Then we define r^2 as

$$r^2 = \frac{RSS}{TSS} = \frac{\sum\left(\hat{Y}_i - \bar{Y}\right)^2}{\sum\left(Y_i - \bar{Y}\right)^2} = 1 - \frac{\sum \hat{\varepsilon}_i^2}{\sum\left(Y_i - \bar{Y}\right)^2}.$$

or,

$$r^2 = 1 - \frac{\sum \hat{\varepsilon}_i^2}{\sum y_i^2} \qquad \ldots(2.16)$$

Since $\sum \hat{\varepsilon}_i^2$ has a limit between $\sum y_i^2$ and 0, r^2 will lie between 0 and 1. Here r^2 is called the coefficient of determination. It measures the proportion of the variation in Y. $\sum\left(Y_i - \bar{Y}\right)^2$ is the total variation of the Y values, and $\sum\left(\hat{Y}_i - \bar{Y}\right)^2$ is the variation of Y explained by variations in X. Therefore, this coefficient indicates the proportion of Y variance explained by the variation of X.

In most cases we use the notation R^2 as the coefficient of determination instead of r^2. There is a slight difference between r^2 and R^2. r^2 denotes the coefficient of determination between two variables, while R^2 denotes the coefficient of determination of more than two variable cases. Therefore r^2 is called the simple coefficient of determination, while R^2 is called multiple coefficient of determination. For simplicity, we will call R^2 the coefficient of determination regardless of the number of variables contained in the equation.

Example 2.3

Now to know how much of the sum of the squared variation of Y is explained by the regression equation, we calculate r^2. Following is a numerical example based on the data in Example 2.1.

[1] Notice that in some of the text books on Econometrics the acronym ESS denotes 'explained sum of squares' and RSS denotes 'residual sum of squares' so that R^2 is given by ESS/TSS. Be careful and ascertain how these two acronyms are defined. We in this course uniformly follow the notation that ESS stands for 'error sum of squares' and RSS for 'regression sum of squares'.

$$r^2 = \frac{\hat{\beta}\sum xy}{\sum y^2}$$

We have $\hat{\beta} = 0.099$

$$\sum xy = 15.3880$$

$$\sum y^2 = \sum Y^2 - \frac{1}{N}\left(\sum Y\right)^2$$

$$= 33.2974 - \frac{1}{10}(17.74)^2$$

$$= 1.8266$$

Now $$r^2 = \frac{0.099 \times 15.3880}{1.8266}$$

$$= \frac{1.5234}{1.8266}$$

$$= 0.834$$

With $r^2 = 0.834$, we could say that over 83 per cent of the variation of y (savings) about its mean value is accounted for by the relationship found.

Example 2.4

Let us take an illustration of the relationship between coffee consumption and average retail price of coffee. The hypothetical data is presented in the following. Our purpose is to fit the two-variable linear model.

Year	Coffee consumption Per person per day (no. of cups)(Y)	Price of coffee (Rs)(X)
1980	2.57	0.77
1981	2.50	0.74
1982	2.35	0.72
1983	2.30	0.73
1984	2.25	0.76
1985	2.20	0.75
1986	2.11	1.08
1987	1.94	1.80
1988	1.97	1.39
1989	2.06	1.20
1990	2.02	1.17

The results can be obtained by the formula used in the examples 2.2 and 2.3 as follows:

$$\hat{\beta} = -0.4795$$

$$\hat{\alpha} = 2.6911$$

$$\hat{Y}_i = 2.6911 - 0.4795X_i \qquad \ldots(2.17)$$

and $r^2 = 0.6628$

The interpretation of the estimated regression is as follows: If the average real retail price of coffee goes up by a rupee, the average consumption of coffee per day is expected to decrease by about half a cup. If the price of coffee were to be zero, the average per person consumption of coffee is expected to be about 2.69 cups a day. The r^2 value means that about 66 per cent of the variation in per capita daily coffee consumption is explained by variation in the retail price of coffee.

2.6 INTRINSICALLY LINEAR MODELS

Linear regression model can be applied to a more general class of equations that are intrinsically (inherently) linear. Intrinsically linear models can be expressed in a form that is linear in the parameters by transforming the variables. That means, if a model is non-linear and after transformation of its variables becomes linear then the model can be said to be intrinsically linear. Let us take the case of the following non-linear model

$$Y = \alpha X^{\beta} \qquad \ldots(2.18)$$

This model will be intrinsically linear if it can be transformed into

$$Y^* = \alpha + \beta X^* + \varepsilon \qquad \ldots(2.19)$$

Using the logarithm of each of the variable in eq. (2.18), we get the following transformed equation

$$\log Y = \alpha + \beta \log X + \varepsilon \text{ make it log alpha} \qquad \ldots (2.20)$$

The relationship in eq. (2.20) is intrinsically linear because it is linear with respect to the parameters α and β. We can apply OLS to estimate these parameters. Because of linearity, such models are also called log-log, double-log, or log-linear models.

One attractive feature of this model, which has made it popular in applied work, is that the slope coefficient β measures the elasticity of Y with respect to X, that is the percentage change in Y for a percentage change in X. Thus if Y represents the quantity of a commodity demanded and X its unit price, β measures the price elasticity of demand. If the relationship between quantity demanded and price is shown in Fig. 2.2 (a), the formed equation as shown in Fig. 2.2 (b) will then give the estimate of price elasticity $(-\beta)$.

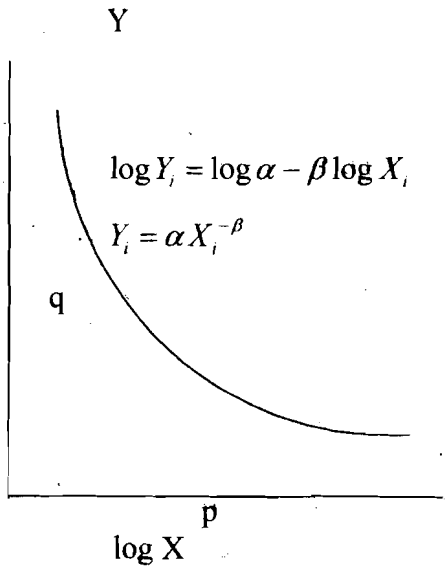

Figure 2.2 (a)

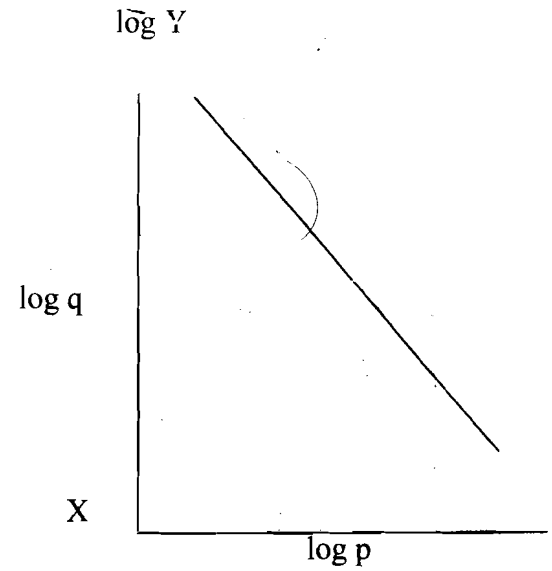

Figure 2.2 (b)

Example 2.5

We can refer to the coffee demand example for the data set and find the logarithmic value of the variables as follows:

Year	Coffee consumption Per person per day (no. of cups) (Y)		Price of coffee (Rs)(X)	Log Y Log X
1990	2.57	0.77	0.409933	-0.11351
1991	2.50	0.74	0.39794	-0.13077
1992	2.35	0.72	0.371068	-0.14267
1993	2.30	0.73	0.361728	-0.13668
1994	2.25	0.76	0.352183	-0.11919
1995	2.20	0.75	0.342423	-0.12494
1996	2.11	1.08	0.324282	0.033424
1997	1.94	1.80	0.287802	0.255273
1998	1.97	1.39	0.294466	0.143015
1999	2.06	1.20	0.313867	0.079181
2000	2.02	1.17	0.305351	0.068186

$\sum \log X = -0.18867$ $\sum \log X^2 = 0.196487857$

$\sum \log Y = 3.761043$ $\sum \log Y^2 = 1.302677016$

$\sum (\log X)(\log Y) = -0.11361$

Using these aggregates, we can get

$$\sum(\log x)(\log y) = \sum(\log X)(\log Y) - \frac{1}{N}\sum \log X \sum \log Y = -0.0491$$

$$\sum \log x^2 = \sum \log X^2 - \frac{1}{N}\left(\sum \log X\right)^2 = 0.1933$$

$$\sum \log y^2 = \sum \log Y^2 - \frac{1}{N}\left(\sum \log Y\right)^2 = 0.0167$$

N, the number of observations, is equal to 11.

The regression coefficient is estimated as:

$$\hat{\beta} = \frac{\sum(\log x)(\log y)}{\sum \log x^2} = -0.2541$$

$$\hat{\alpha} = \log \bar{Y} - \hat{\beta} \log \bar{X} = \frac{\sum \log Y}{N} - \hat{\beta}\frac{\sum \log X}{N}$$

$$= 0.3376$$

$$\log \hat{Y}_i = 0.3376 - 0.2541 \log X_i \qquad (2.2$$

$$r^2 = \frac{\hat{\beta}\sum(\log x)(\log y)}{\sum \log y^2} = 0.7459$$

From this result we see that the price elasticity coefficient is –0.2541, implying that for a 1 per cent increase in the real price of coffee, the demand for coffee (as measured by cups of coffee consumed per day) on the average decreases by about 0.25 per cent. Since the price elasticity value of 0.25 is less than 1 in absolute terms, we can say that the demand for coffee is price-inelastic.

Now considering the results of the linear demand function (Example 2.4) and log-linear demand function (example 2.5), we may question which model is better. Can we say that eq. (2.21) is better than eq. (2.17) because its r^2 value is higher (0.7456 vs. 0.6628)? Unfortunately, we cannot say that, for when the dependent variable of two models is not same (here, log$Y$ vs. Y), the two r^2 values are not directly comparable. We cannot directly compare the slope coefficients either, for in eq. (2.17) the slope coefficient gives the effect of a unit change in the price of coffee on the constant absolute (i.e., not relative) amount of decrease in coffee consumption, which is 0.4795 cups per day. On the other hand, the coefficient of –0.2541 obtained from eq. (2.21) gives the constant percentage decrease in coffee consumption as a result of a 1 per cent increase in the price of coffee.

Check your Progress 2

1) Explain how you would estimate the beta coefficient in a two-variable regression model.

.....

.....

.....

.....

2) Define the coefficient of determination. What does it signify?

.....

.....

.....

.....

3) Find the values of α, β and r^2 using the following data.

X = 2, 3, 1, 5, 9

Y = 4, 7, 3, 9, 17

.....

.....

.....

.....

4) What do you mean by intrinsically linear model? Can you compare the results of an intrinsically linear model with that of a linear model? Why or why not?

.....

.....

.....

.....

2.7 LET US SUM UP

In this unit we studied the basic framework of regression analysis. The reason for inclusion of error term in the regression model is described vividly. We also studied the estimation of beta coefficient, which serves as an estimate of the population regression line. The overall goodness of fit of the regression model is measured by the coefficient of determination, r^2. It tells what proportion of the variation in the dependent variable is explained by the explanatory (independent) variable. The unit also discussed about the intrinsically linear models that can be expressed in a form that is linear in the parameters by transforming the variables. A number of illustrative applications are presented in this unit so as to make the students easy for understanding.

2.8 KEY WORDS

Two-variable model : An equation, in which there are two variables – one dependent variable and the other independent variable is called a two variable model.

Error Term : The measurement of the functional relationships among variables in reality is not exact. For a given independent variable X, we may observe many possible dependent variable values of Y. To describe this situation formally, we add a random "error" term to the model.

Stochastic Nature : The representation of the relationship in an equation can be said to be stochastic in nature if for every value of X (independent variable) there exists a probability distribution of ε and therefore a probability distribution of the Y's (dependent variable).

Estimation : In the specification of the two variable model, the values of the parameters α and β are not known, as a result the population regression line is not known. When the values of α and β are estimated, we obtain a sample regression line that serves as an estimate of the population regression line. This is known as estimation.

Least-Squares : The principle of least squares is to choose $\hat{\alpha}$ and $\hat{\beta}$ values that will minimise the sum of squared deviations between the observed and estimated values of Y.

Coefficient of Determination : The coefficient of determination indicates the proportion of Y variance explained by the variation of X.

Intrinsically Linear Models : If a model is non-linear and after transformation of its variables becomes linear then the model can be said to be intrinsically linear.

2.9 SOME USEFUL BOOKS/REFERENCES

Gujarati, Damodar N., 1995, *Basic Econometrics*, McGraw-Hill Inc., Singapore.

Johnston, Jack and John Dinardo, 1997, *Econometric Methods*, The McGraw-Hill Companies Inc., Singapore.

Pindyck, Robert S. and Daniel L. Rubinfeld, 1998, *Econometric Models and Economic Forecasts*, Irwin/McGraw-Hill, Singapore.

2.10 ANSWERS/HINTS TO CHECK YOUR PROGRESS EXERCISES

Check Your Progress 1

1) Go through Sections 2.1.

2) Go through Section 2.2.

Check Your Progress 2

1) Go through Section 2.5.

2) Go through Section 2.6.

3) Go through Sections 2.5 and 2.6. The results are 1, 1.75 and 0.9879 respectively.

4) Go through Sections 2.5, 2.6 and 2.7.

UNIT 3 STATISTICAL INFERENCE IN SIMPLE REGRESSION MODELS

Structure

3.0 OBJECTIVES

After going this unit you will be able to:

- describe the assumptions of ordinary least squares;
- find the best linear unbiased estimator;
- explain the concept of confidence interval;
- test the significance of the estimators; and
- make prediction of two-variable models

3.1 INTRODUCTION

We have discussed in Unit 2 about the specification of the two-variable linear regression model,

$$Y_i = \alpha + \beta X_i + \varepsilon_i \qquad \ldots(3.1)$$

where we have estimated the relationship between Y and X from the sample observations such that

$$\hat{Y}_i = \hat{\alpha} + \hat{\beta} X_i .$$

Our objective was to find the estimated values of the unknown parameters α and β, and the estimated value of Y_i. However, our objective is not only to estimate the coefficients but also to draw inferences about their true values. In order to estimate the parameters of the model statistically, the relationship of the above equation requires some assumptions about the probability distribution of the error term.

3.2 CLASSICAL ASSUMPTIONS FOR OLS

The assumptions of the ordinary least squares model are as follows:

1) The relationship between Y and X is linear.

2) The error has zero expected value, i.e.

$$E(\varepsilon) = 0 \quad \ldots (3.2)$$

3) The error term has constant variance for all observations, i.e.

$$E(\varepsilon^2) = \sigma^2. \quad \ldots (3.3)$$

4) The random variables ε_i are statistically independent. Thus,

$$E(\varepsilon_i \varepsilon_j) = 0 \text{ for all } i \neq j. \quad \ldots (3.4)$$

5) The X's are non-stochastic variables whose values are fixed. Therefore the X's are uncorrelated to the error term, i. e.

$$E(X_i \varepsilon_i) = 0, \text{ for } i = 1, 2, \cdots, n. \quad \ldots (3.5)$$

These assumptions imply that error terms ε has 0 mean and constant variance σ^2 and that error terms are uncorrelated to each other. Further, X_i and ε_i are also uncorrelated.

3.3 BEST LINEAR UNBIASED ESTIMATOR (BLUE)

To examine the characteristics of the least-squares parameter estimates, recall that they result from a specific sample of observations of the dependent and independent variables. Since the sample may vary, the estimates may vary as well and thus are associated with a random variable. Because the model is stochastic, we have shown the formulas for regression intercept and slope as $\hat{\alpha}$ and $\hat{\beta}$, where the 'hats' over α and β represent estimated values. Here $\hat{\beta}$ refers to the slope estimate resulting from a specific sample as well as to the estimator (a formula that applies to any sample) that follows a probability distribution.

The ordinary least-squares (OLS) estimators are unbiased and consistent. In fact, one of the properties of the ordinary least-squares estimator is that of all the estimators which are linear and which yield unbiased estimates, the estimates resulting from the OLS estimator have the minimum variance. This

therefore, states that given the assumptions of OLS, the estimators α and β are the best (most efficient) linear unbiased estimators (BLUE) of α and β in the sense that they have the minimum variance of all linear unbiased estimators. If some of the assumptions of the Gauss-Markov theorem do not hold, the least-squares estimators will no longer be BLUE. It is important to realize that the Gauss-Markov theorem does not apply to nonlinear estimators. In the following we will prove the Gauss-Markov theorem. Since the properties of $\hat{\alpha}$ and $\hat{\beta}$ are exactly the same we will discuss the properties of $\hat{\beta}$ only.

In order to prove that $\hat{\beta}$ is BLUE, we will first prove that $\hat{\beta}$ is an unbiased linear estimator of β and then show that it has minimum variance.

We note that $\hat{\beta}$ is a linear estimator, since $\hat{\beta}$ can be written as a weighted average of the individual observations on Y. The unbiased property of $\hat{\beta}$ can be seen from the following.

From eq. (2.10) of Unit 2 we can write

$$\hat{\beta} = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{\sum x_i(\beta x_i + \varepsilon_i)}{\sum x_i^2} \qquad \text{since } y_i = \beta x_i + \varepsilon_i$$

$$= \beta \frac{\sum x_i^2}{\sum x_i^2} + \frac{\sum x_i \varepsilon_i}{\sum x_i^2}$$

or

$$\hat{\beta} = \beta + \frac{\sum x_i \varepsilon_i}{\sum x_i^2} \qquad \ldots(3.6)$$

If we take the expected value of (3.6) we find that

$$E(\hat{\beta}) = \beta \qquad \text{since } E(\varepsilon_i) = 0.$$

We know that the parameter β is a constant; thus $E(\beta) = \beta$. Since X_i is fixed in nature we have

$$E\left(\frac{\sum x_i \varepsilon_i}{\sum x_i^2}\right) = \frac{\sum x_i E(\varepsilon_i)}{\sum x_i^2} = 0.$$

Hence $\hat{\beta}$ is an unbiased estimator of β. This along with the linear properties shows that $\hat{\beta}$ is an unbiased linear estimator of β.

Next we establish that the variance of $\hat{\beta}$ is the best among a class of linear unbiased estimators, i.e. $\hat{\beta}$ has the smallest variance. Let us first obtain the variance of $\hat{\beta}$, $\sigma_{\hat{\beta}}^2$. By using eq. (3.6), we can find out

$$\sigma_{\hat{\beta}}^2 = E\left(\hat{\beta} - \beta\right)^2 = E\left(\frac{\sum x_i \varepsilon_i}{\sum x_i^2}\right)^2$$

$$= E\left[\frac{1}{\left(\sum x_i^2\right)^2}(x_1^2\varepsilon_1^2 + x_2^2\varepsilon_2^2 + \cdots + x_n^2\varepsilon_n^2 + 2x_1x_2\varepsilon_1\varepsilon_2 + \cdots\right]$$

$$= \frac{1}{\left(\sum x_i^2\right)^2}\left[x_1^2E(\varepsilon_1^2) + x_2^2E(\varepsilon_2^2) + \cdots + x_n^2E(\varepsilon_n^2)\right]$$

$$\frac{\sum x_i^2}{(\sum x_i^2)^2}\,E(\varepsilon_i^2) = \frac{\sum x_i^2}{\left(\sum x_i^2\right)^2}\sigma^2$$

Thus, the variance of β is given by

$$\sigma_{\hat{\beta}}^2 = \frac{\sigma^2}{\sum x_i^2} \qquad \ldots (3.7)$$

To prove that $\hat{\beta}$ has the smallest variance of all linear unbiased estimates, we define

$$\beta^* = \sum c_i y_i$$

where $c_i = \frac{x_i}{\sum x_i^2} + d_i$,

and where d_i are arbitrary constants. β^* can be written as

$$\beta^* = \sum c_i\left(\beta x_i + \varepsilon_i\right).$$

Therefore,

$$E\left(\beta^*\right) = \beta\sum c_i x_i + \sum c_i\varepsilon_i\,. \qquad \ldots (3.8)$$

To show that $E\left(\beta^*\right) = \beta$, we must assume that $\sum c_i x_i = 1$.

Since $c_i = \frac{x_i}{\sum x_i^2} + d_i$ we have

$c_i x_i = \frac{x_i^2}{\sum x_i^2} + d_i x_i$, and therefore, $\sum c_i x_i = \frac{\sum x_i^2}{\sum x_i^2} + \sum d_i x_i = 1 + \sum d_i x_i$

Thus $E\left(\beta^*\right) = \beta$ only if the following condition is fulfilled:

$$\sum d_i x_i = 0\,.$$

Secondly, let us look into the variance of β^*. It is given by

$$\sigma_{\beta^*}^2 = E\left(\beta^* - \beta\right)^2$$

$$= E\left(\sum c_i\varepsilon_i\right)^2$$

$$= \sigma^2\sum c_i^2$$

$$= \sigma^2\left[\sum\left(\frac{x_i}{\sum x_i^2}\right)^2 + \sum d_i^2 + 2\sum \frac{x_i}{x_i^2} d_i\right]$$

$$= \frac{\sigma^2}{\sum x_i^2} + \sigma^2 \sum d_i^2$$

$$= \sigma_{\hat{\beta}}^2 + \sigma^2 \sum d_i^2 \qquad \ldots(3.9)$$

Therefore, σ_{β}^2 is at least equal to or larger than $\sigma_{\hat{\beta}}^2$, since $\sum d_i^2$ is at least 0 or larger. This shows that $\hat{\beta}$ has the smallest variance of all linear unbiased estimates. Hence, we conclude that $\hat{\beta}$ is a best linear unbiased estimator (BLUE).

3.4 PROBABILITY DISTRIBUTION OF ERROR TERMS

It has been shown that the OLS estimators $\hat{\alpha}$ and $\hat{\beta}$ are both linear function of the error term, which is random by assumption. Therefore, the sampling or probability distributions of the OLS estimators will depend upon the assumptions made about the probability distribution of the error term. Since the probability distributions of these estimators are necessary to draw inferences about their population values, the nature of the probability distribution of the error term assumes an extremely important role in hypothesis testing.

Since the method of OLS does not make any assumption about the probabilistic nature of the error term, it is of little help for the purpose of drawing inferences about population from sample. This void can be filled if we are willing to assume that the error term follow some probability distribution. In the regression context, it is usually assumed that error terms follow the normal distribution.

The normal distribution is a bell-shaped probability distribution and can be fully described by its mean and variance. The classical normal linear regression assumes that each error term is distributed normally with zero mean and constant variance, i.e. $\varepsilon_i \sim N(0, \sigma^2)$.

We know from (3.1) that $\varepsilon_i = (Y_i - \alpha - \beta X_i)$. Thus the probability distribution of ε_i would be

$$p(\varepsilon_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(Y_i - \alpha - \beta X_i)^2\right] \qquad \ldots(3.10)$$

We would apply the above property of ε_i in hypothesis testing which is given in Section 3.6.

Check Your Progress 1

1) What are the assumptions of two-variable linear regression model?

2) Prove that the estimated residuals from the linear regression and the corresponding sample values of X are uncorrelated; that is, $\sum X_i \hat{\varepsilon}_i = 0$.

3) Prove that $\hat{\beta}$ is the best linear unbiased estimator of β.

3.5 MAXIMUM LIKELIHOOD METHOD OF ESTIMATION

Maximum likelihood estimation focuses on the fact that different populations generate different samples; any one sample being scrutinized is more likely to have come from some populations than from others.

We define the maximum-likelihood estimator of a parameter β as the value of $\hat{\beta}$, which would most likely generate the observed sample observations $Y_1, Y_2, \cdots, Y_N$. In general Y_i is normally distributed and each of the Y_i's is drawn independently, the maximum likelihood estimator maximizes

$$p(Y_1)\,p(Y_2)\cdots\cdots p(Y_N) \tag{3.13}$$

where each p represents a probability associated with the normal distribution.

Thus, the calculated maximum likelihood estimate is a function of the particular sample of Y's chosen. A different sample would result in a different maximum likelihood estimate. The maximum likelihood, or we can say likelihood, function depends not only on the sample values but also on the unknown parameters of the problem. Maximum likelihood estimation involves a search over alternative parameter estimators to find those estimators which most likely would generate the sample.

We will now show how this can be applied to the estimation of econometric models. We will see that a great advantage of maximum likelihood estimation is that under a broad set of conditions parameters are both consistent and (for large samples) asymptotically efficient.

We begin the analysis with the linear regression model

$$Y_i = \alpha + \beta X_i + \varepsilon_i$$

We know that each Y_i is normally distributed with mean $\alpha + \beta X_i$ and variance σ^2. The probability can be written as

$$p(Y_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(Y_i - \alpha - \beta X_i)^2\right]$$

The likelihood function is the product of the individual probabilities taken over all N observations. In this case the likelihood function is

$$L(Y_1, Y_2, \cdots, Y_N, \alpha, \beta, \sigma^2) = p(Y_1)\,p(Y_2)\cdots p(Y_N)$$

$$= \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left[-\sum\left(\frac{Y_i - \alpha - \beta X_i}{2\sigma^2}\right)^2\right] \quad \ldots(3.11)$$

With maximum likelihood estimation our goal is to find the values of the parameters α, β and σ^2 which are most likely to generate the sample observations $Y_1, Y_2, \cdots, Y_N$. This is achieved by maximizing the likelihood function given above with respect to each of the parameters. To do this it is more convenient to work with the logarithm of the likelihood function. The log-likelihood function is given by

$$LogL = -\frac{N}{2}\log(\sigma^2) - \frac{N}{2}\log(2\pi) - \frac{1}{2\sigma^2}\sum(Y_i - \alpha - \beta X_i)^2 \quad \ldots(3.12)$$

To find the maximum we differentiate the log-likelihood function with respect to each of the three unknown parameters, equate the derivatives to zero, and solve.

Differentiating eq. (3.12) partially with respect to α, β and σ^2 and setting the derivatives equal to zero yields

$$\frac{\partial(\log L)}{\partial\alpha} = \frac{1}{\sigma^2}\sum(Y_i - \alpha - \beta X_i) = 0 \quad \ldots(3.13)$$

$$\frac{\partial(\log L)}{\partial\beta} = \frac{1}{\sigma^2}\sum\left[X_i(Y_i - \alpha - \beta X_i)\right] = 0 \quad \ldots(3.14)$$

$$\frac{\partial(\log L)}{\partial\sigma^2} = -\frac{N}{2\sigma^2} + \frac{1}{2\sigma^4}\sum(Y_i - \alpha - \beta X_i)^2 = 0 \quad \ldots(3.15)$$

From (3.11) we get

$$\sum Y_i = N\hat{\alpha} + \hat{\beta}\sum X_i \quad \ldots(3.16)$$

From (3.12) we get

$$\sum Y_i X_i = \hat{\alpha}\sum X_i + \hat{\beta}\sum X_i^2 \qquad \ldots(3.17)$$

From (3.13) we get

$$\frac{\sum\left(Y_i - \hat{\alpha} - \hat{\beta}X_i\right)^2}{2} = \frac{N\hat{\sigma}^4}{2\hat{\sigma}^2}$$

Or

$$\sum\left(Y_i - \hat{\alpha} - \hat{\beta}X_i\right)^2 = N\hat{\sigma}^2$$

Or

$$\hat{\sigma}^2 = \frac{1}{N}\sum\left(Y_i - \hat{\alpha} - \hat{\beta}X_i\right)^2 \qquad \ldots(3.18)$$

Note that eq. (3.14) and eq. (3.15) are the same as the normal equations obtained in OLS method. Solving eq. (3.14) and eq. (3.15) we get $\hat{\alpha}$ and $\hat{\beta}$.

So

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X} \qquad \ldots(3.19)$$

$$\hat{\beta} = \frac{\sum x_i y_i}{\sum x_i^2} \qquad \ldots(3.20)$$

$$\hat{\sigma}^2 = \frac{1}{N}\sum\left(Y_i - \hat{\alpha} - \hat{\beta}X_i\right)^2 \qquad \ldots(3.21)$$

Clearly, the maximum likelihood estimators of α and β are identical to the least-squares estimators when the error terms of OLS follow the normal distribution. It follows therefore that $\hat{\alpha}$ and $\hat{\beta}$ are best linear unbiased estimators (BLUE). $\hat{\sigma}^2$ is, however, a biased (although consistent) estimator of σ^2. To obtain an unbiased estimator, we need to divide the numerator by $N-2$, adjusting for the degrees of freedom.

Since the method of least squares with the added assumption of normality of ε_i provides us with all the tools necessary for both estimation and hypothesis testing of the linear regression models, there is no loss for readers who may not want to pursue the maximum likelihood method because of its slight mathematical complexity.

3.6 STATISTICAL INFERENCE IN REGRESSION MODEL

3.6.1 Estimation of Standard Error

To make statistical inferences about $\hat{\beta}$ or to test its statistical significance, we have to know the variance of $\hat{\beta}$. According to eq. (3.7), the variance of $\hat{\beta}$ is a function of the disturbance variance σ^2. Therefore, an estimator of the

disturbance variance σ^2 should first be obtained from the sample observations in order to estimate the variance of $\hat{\beta}$.

The estimation of σ^2 can be based on the error sum of squares $\sum \hat{\varepsilon}_i^2$:

$$\hat{\varepsilon}_i = Y_i - \hat{Y}_i$$

$$= \left(y_i + \bar{Y}\right) - \left(\hat{y}_i + \bar{Y}\right) \text{ since } y_i = Y_i - \bar{Y} \;\&\; \hat{y}_i = \hat{Y}_i - \bar{Y} = y_i - \hat{y}_i$$

Averaging $Y_i = \alpha + \beta X_i + \varepsilon_i$ over the N sample observations, we obtain $\bar{Y} = \alpha + \beta \bar{X} + \bar{\varepsilon}$, so that

$$y_i = \beta x_i + \left(\varepsilon_i - \bar{\varepsilon}\right)$$

Also recall that

$$\hat{Y}_i = \hat{\alpha} + \hat{\beta} X_i \text{ and } \bar{Y} = \hat{\alpha} + \hat{\beta} \bar{X}$$

We have, therefore, $\hat{y}_i = \hat{\beta} x_i$.

Now

$$\hat{\varepsilon}_i = \beta x_i + \left(\varepsilon_i - \bar{\varepsilon}\right) - \beta x_i = -\left(\hat{\beta} - \beta\right) x_i + \left(\varepsilon_i - \bar{\varepsilon}\right)$$

Squaring and summing both the sides, we obtain

$$\sum \hat{\varepsilon}_i^2 = \left(\hat{\beta} - \beta\right)^2 \sum x_i^2 + \sum \left(\varepsilon_i - \bar{\varepsilon}\right)^2 - 2\left(\hat{\beta} - \beta\right) \sum x_i \left(\varepsilon_i - \bar{\varepsilon}\right)$$

Taking expectations on both sides gives

$$E\left(\sum \hat{\varepsilon}_i^2\right) = \sum x_i^2 E\left(\hat{\beta} - \beta\right)^2 + E\left[\sum \left(\varepsilon_i - \bar{\varepsilon}\right)^2\right] - 2E\left[\left(\hat{\beta} - \beta\right) \sum x_i \left(\varepsilon_i - \bar{\varepsilon}\right)\right]$$

$$= \sum x_i^2 E\left(\hat{\beta} - \beta\right) + E\left[\sum \varepsilon_i^2 - \frac{\sum \left(\varepsilon_i\right)^2}{N}\right] - 2E\left[\frac{\sum x_i \varepsilon_i}{\sum x_i^2} \left(\sum x_i \varepsilon_i - \sum x_i \bar{\varepsilon}\right)\right]$$

$$= \sigma^2 + (N-1)\sigma^2 - 2E\left[\frac{\left(\sum x_i \varepsilon_i\right)^2}{\sum x_i^2}\right] \quad \text{since } \sum x_i = 0$$

$$= \sigma^2 + (N-1)\sigma^2 - 2\sigma^2$$

$$= (N-2)\sigma^2$$

Therefore, if we define

$$s^2 = \hat{\sigma}^2 = \frac{\sum \hat{\varepsilon}_i^2}{N-2} \qquad \ldots(3.22)$$

Its expected value is

$$E\left(\hat{\sigma}^2\right) = \frac{1}{N-2} E\left(\sum \hat{\varepsilon}_i^2\right) = \sigma^2 \qquad \ldots(3.23)$$

which shows that $\hat{\sigma}^2$ (or s^2) is an unbiased estimator of true σ^2.

$\hat{\sigma}$ (or s) is a square root of $\hat{\sigma}^2$ (or s^2), and it is called the standard error of the estimate.

An unbiased estimator of $\sigma_{\hat{\beta}}^2$ can be obtained as follows:

$$\hat{\sigma}_{\hat{\beta}}^2 = \frac{\hat{\sigma}^2}{\sum x_i^2} = \frac{\sum \hat{\varepsilon}_i^2}{(N-2)\sum x_i^2} \quad \ldots(3.24)$$

The square root of $\hat{\sigma}_{\hat{\beta}}^2$, $\hat{\sigma}_{\hat{\beta}}$, is called the standard error of the coefficient.

3.6.1 Confidence Interval

Confidence intervals provide a range of values, which are likely to contain the true regression parameters. With every confidence interval we associate a level of statistical significance. The confidence intervals are constructed so that the probability that the interval contains the true parameter is 1 minus the level of significance.

Confidence intervals are particularly useful for testing statistical hypotheses about the estimated regression parameters. We begin with a null hypothesis, which usually states that a certain effect is not present. Because we often hope to accept the model, the null hypothesis is constructed in a way that makes its rejection possible. To test the validity of a model we set up the null hypothesis that β equals 0. We hope to reject the null hypothesis by obtaining a value of $\hat{\beta}$ which is sufficiently different from 0 to cast significant doubt on the hypothesis that β equals 0. Assume, for example, that $\hat{\beta}$ is 0.9. If we choose a level of significance of 10 per cent, the 90 per cent confidence interval for β might be

$$0.6 < \beta < 1.2$$

This means that the probability that β is within the range 0.6 to 1.2 is 0.90. In addition it means that we can reject the null hypothesis that β equals 0 with 90 per cent confidence.

3.6.2 Tests of Significance

Here we will discuss the tests of significance of the regression coefficient and that of the coefficient of determination. The former is done through t test, while the latter is made through F test. Both these tests are presented in the following.

Tests of Regression Coefficients

The statistical test for rejecting null hypotheses associated with a regression coefficient is usually based on the t distribution. The t distribution is relevant because for statistical testing we need to utilize a sample estimate of the error variance rather than its true value. To use the t distribution to construct 95 per cent confidence intervals for the estimated parameters, we first standardize the estimated regression parameter, say $\hat{\beta}$, by subtracting its hypothesized true value β_0 and dividing by the estimate of its standard error

(see Unit 1). Thus, $t = \frac{\hat{\beta} - B_0}{s_{\hat{\beta}}}$ where $s_{\hat{\beta}}$ is the standard error of $\hat{\beta}$. We consider the null hypothesis that $\beta = 0$ or equivalently, that there is no relationship between the variables X and Y in the two-variable model. In this case the t statistic is given by

$$t_{N-2} = \frac{\hat{\beta}}{s_{\hat{\beta}}} \qquad \ldots(3.25)$$

The subscript (N – 2) indicates the degree of freedom for which t-test should be undertaken. If the t statistic (calculated value) is greater than t_c, the critical value (tabulated value), we reject the null hypothesis. Here the subscript 'c' indicates the level of significance.

More generally, we can test the null hypothesis that $\beta = \beta_0$. To do so we calculate the t statistic:

$$t_{N-2} = \frac{\hat{\beta} - \beta_0}{s_{\hat{\beta}}} \qquad \ldots(3.26)$$

the standardized variable t_{N-2} also follows that a t distribution with $N-2$ degrees of freedom. With a 5 per cent test, the critical value is defined so that

$$\text{Prob}\left(-t_c < t_{N-2} < t_c\right) = .95 \qquad \ldots(3.27)$$

where Prob denotes probability.

Now, by substituting from eq. (3.2 6) we obtain

$$\text{Prob}\left(-t_c < \frac{\hat{\beta} - \beta_0}{s_{\hat{\beta}}} < t_c\right) = .95 \qquad \ldots(3.28)$$

Modifying eq. (3.28) we get

$$\text{Prob}\left(\hat{\beta} - t_c s_{\hat{\beta}} < \beta_0 < \hat{\beta} + t_c s_{\hat{\beta}}\right) = .95 \qquad \ldots(3.29)$$

From eq. (3.29) we obtain a 95 per cent confidence interval for β:

$$\hat{\beta} \pm t_c s_{\hat{\beta}} \qquad \ldots(3.30)$$

Using a similar procedure, we obtain a 95 per cent confidence interval for α:

$$\hat{\alpha} \pm t_c s_{\hat{\alpha}} \qquad \ldots(3.31)$$

It is possible to determine confidence intervals for any level of significance as long as the critical value of t distribution is correctly chosen. Eq. (3.30) tells us that an interval of t_c standard deviations on either side of the estimated slope parameter has a probability of .95 of containing the true parameter.

Example 3.1

Following is a numerical illustration of testing the regression coefficient.

Y	X	$y = Y - \bar{Y}$	$x = X - \bar{X}$
40	4	-22	-4
60	6	-2	-2
50	7	-12	-1
70	10	8	2
90	13	28	5

From the above data we can calculate

$$\sum Y = 310 \qquad \sum X = 40$$

$$\sum X^2 = 370 \qquad \sum XY = 2740$$

$$\sum x_i y_i = 260 \qquad \sum x_i^2 = 50$$

$$\sum \hat{\varepsilon}_i^2 = 128$$

We have

$$\hat{\beta} = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{260}{50} = 5.2$$

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}$$

$$= 62 - 5.2(8) = 62 - 41.6 = 20.4$$

The estimated equation is written as

$$\hat{Y} = 20.4 + 5.2X$$

Now we can calculate

$$\hat{\sigma}_{\hat{\beta}}^2 = \frac{\sum \hat{\varepsilon}_i^2}{(N-2)\sum x_i^2} = \frac{128}{3(50)} = \frac{128}{150} = 0.85$$

$$\hat{\sigma}_{\hat{\beta}} = \sqrt{0.85} = 0.92$$

$$t_{N-2} = \frac{\hat{\beta}}{s_{\hat{\beta}}} = \frac{0.52}{0.92} = 5.63$$

For the 5% level of significance (i.e., 95% level of confidence), the value of t is 3.182 for a two-tailed test with degree of freedom, 3. The calculated t value is 5.63, which is higher than tabulated t value, 3.182. Therefore, we reject the null hypothesis and conclude that X is a statistically significant variable in explaining the variation of Y.

Example 3.2

Let us return to our consumption-income example in Unit 2.

Consumption (Y) (Rs)	Income (X) (Rs)
70	80
65	100
90	120
95	140
110	160
115	180
120	200
140	220
155	240
150	260

We found that

$$\hat{\beta} = 0.5091$$

$$\hat{\alpha} = 24.4545$$

$$\sum \hat{\varepsilon}_i^2 = 337.2728$$

The estimated regression line is

$$\hat{Y} = 24.4545 + 0.5091X$$

We can calculate

$$\hat{\sigma}_{\hat{\beta}}^2 = \frac{\sum \hat{\varepsilon}_i^2}{(N-2)\sum x_i^2} = 0.0013$$

$$\hat{\sigma}_{\hat{\beta}} = \sqrt{0.0013} = 0.0357$$

$$t_{N-2} = \frac{\hat{\beta}}{s_{\hat{\beta}}} = \frac{0.5091}{0.0357} = 14.26$$

For the 1% level of significance (i.e., 99% level of confidence), the value of t is 3.355 for a two-tailed test with degree of freedom, 8. The calculated t value is 14.26, which is higher than tabulated t value, 3.355. Therefore, we reject the null hypothesis and conclude that X is a statistically significant variable in explaining the variation of Y.

The 99% confidence interval for $\hat{\beta}$ would be

$$\hat{\beta} \pm t_c s_{\hat{\beta}} = 0.5091 \pm (3.355)(0.0357) = 0.5091 \pm 0.1198$$

or

$$0.389 < \beta < 0.629$$

We observe that 0 lies outside the 99 percent confidence interval for β, allowing us to reject at the 1 per cent level of significance the null hypothesis that $\beta = 0$.

Similarly,

$$\hat{\sigma}_{\hat{\alpha}}^2 = \frac{\sum \hat{\varepsilon}_i^2}{(N-2)}\left(\frac{\sum X_i^2}{N\sum x_i^2}\right)$$

$$= \frac{337.2728}{8}\left(\frac{322000}{10\times 33000}\right)$$

$$= 41.1370$$

$$\hat{\sigma}_{\hat{\alpha}} = \sqrt{41.1370} = 6.4138$$

$$t_{N-2} = \frac{\hat{\alpha}}{s_{\hat{\alpha}}} = \frac{24.4545}{6.4138} = 3.813$$

The 99% confidence interval for α would be

$$\hat{\alpha} \pm t_c s_{\hat{\alpha}} = 24.4545 \pm (3.355)(6.4138) = 24.4545 \pm 21.5183$$

or

$$2.9362 < \alpha < 45.9728$$

We observe that 0 lies outside the 99 percent confidence interval for α, allowing us to reject at the 1 per cent level of significance the null hypothesis that $\alpha = 0$. Equivalently, we observe that the calculated value of t (3.813) is greater than the critical value of 3.355 and again reject the null hypothesis.

Tests for the Coefficient of Determination

To test the significance of R^2 that R^2 is not different from 0, that Y is not explained by the explanatory variable in the equation, we can make use of F test. This is an example of the analysis of variance. In the case of two-variable models with the null hypothesis $\beta = 0$,

$$F = \frac{\hat{\beta}^2 \sum x_i^2/(k-1)}{\sum \hat{\varepsilon}_i^2/(N-k)} = \frac{\sum \hat{y}_i^2/(2-1)}{\sum \hat{\varepsilon}_i^2/(N-2)} = \frac{RSS/(2-1)}{ESS/(N-2)} \quad \ldots(3.32)$$

where F has $(1, N-2)$ degrees of freedom. The high value of F leads to the rejection of the null hypothesis. Therefore, it is quite natural that rejection of null hypothesis occurs when RSS is high relative to ESS, that is, when R^2 is high.

The F test is designed to test the significance of all variables or a set of variables in the equation. However, in the two-variable case the F test is used for testing the single explanatory variable and at the same time is equivalent to test the significance of R^2.

Example 3.3

Following is a numerical illustration of testing the significance of R^2. We use the data provided in the Example 3.1.

From the data provided in example 3.1 we calculate

$$\sum x_i y_i = 260 \qquad \sum x_i^2 = 50 \qquad \sum y_i^2 = 1480$$

$$\sum \hat{y}_i^2 = 1352 \qquad \sum \hat{\varepsilon}_i^2 = 128$$

$$r = \frac{260}{\sqrt{50}\sqrt{1480}} = 0.97$$

$$R^2 = \frac{1352}{1480} = 0.93$$

The F value is calculated based on the information above as

$$F = \frac{1352/(2-1)}{128/(5-2)} = 31.68$$

At a 5% level of significance $F(1,3)$ is 10.13, which is less than 31.68. Therefore, R^2 is statistically significant.

3.7 PREDICTION

One of the objectives of econometrics is the prediction of economic phenomena. The estimated regression equation $\hat{Y} = \hat{\alpha} + \hat{\beta}X$ is used for predicting the value of Y for given values of X. Let X_0 be the given value of X. Then we predict the corresponding value of Y_0 of Y by

$$\hat{Y}_0 = \hat{\alpha} + \hat{\beta}X_0$$

where the true value Y_0 is given by

$$Y_0 = \alpha + \beta X_0 + \varepsilon_0$$

The prediction error will be

$$\varepsilon = Y_0 - \hat{Y}_0$$

$$= (\alpha + \beta X_0 + \varepsilon_0) - (\hat{\alpha} + \hat{\beta}X_0)$$

$$= (\alpha - \hat{\alpha}) + (\beta - \hat{\beta})X_0 + \varepsilon_0$$

The expectation of the prediction error will be

$$E(\varepsilon) = E(\alpha - \hat{\alpha}) + E(\beta - \hat{\beta})X_0 + E(\varepsilon_0)$$

$$= 0 \qquad \text{since } E(\hat{\alpha}) = \alpha,\ E(\hat{\beta}) = \beta, \text{ and } E(\varepsilon_0) = 0.$$

Therefore, the predicted value of Y is an unbiased estimate. In fact, it can be shown that Y_0 is the best linear unbiased estimate (BLUE).

Now the variance of the prediction error will be

$$E\left[\varepsilon - E(\varepsilon)\right]^2 = E\left[(\alpha - \hat{\alpha}) + (\beta - \hat{\beta})X_0 + (\varepsilon_0)\right]^2$$

$$= E\left[(\alpha - \hat{\alpha}) + (\beta - \hat{\beta})X_0\right]^2 + E(\varepsilon_0)^2$$

$$= \sigma_{\hat{Y}_0}^2 + \sigma^2 \qquad \ldots(3.33)$$

where $\sigma_{\hat{Y}_0}^2$ is $E\left[(\alpha - \hat{\alpha}) + (\beta - \hat{\beta})X_0\right]^2$. Eq. (3.33) implies that the error made in predicting an individual value is the sum of two uncorrelated errors: the error in estimating the parameters of the regression equation and the error of the random disturbance during the forecasting period.

Check Your Progress 2

1) Find the maximum likelihood estimators. Are they same as least-squares estimators?

.....

.....

.....

.....

2) Prove that $\hat{\sigma}^2$ is not an unbiased estimator of σ^2.

.....

.....

.....

.....

3) Define the concept of confidence interval.

.....

.....

.....

.....

4) Test the hypothesis that there is no relationship between the variables X and Y in the two-variable model.

.....

.....

.....

.....

5) Find the values of α, β and r^2, and conduct t and F tests from the following data.

$Y = 4.0, 3.0, 3.5, 2.0, 3.0, 3.5, 2.5, 2.5$

$X = 21, 15, 15, 9, 12, 18, 6, 12$

.....

.....

.....

.....

3.8 LET US SUM UP

In this unit we first discussed about the classical assumptions of ordinary least squares. We presented five assumptions of least squares, viz., linear relationship of the variables, zero expected value of error term, constant variance, independent random variables, and non-stochastic independent variable. The sixth assumption on normal distribution was discussed later. We discussed the estimation of best linear unbiased estimator. The maximum likelihood (ML) method of estimation is also discussed. It is found that the ML estimators are the same as those of least squares estimators. We discussed the tests of significance of regression model like t distribution and F test. Finally, we made prediction about the dependent variable for known values of explanatory variable.

3.9 KEY WORDS

Best Linear Unbiased Estimator : Given the assumptions of OLS, the estimators are said to be the best linear unbiased estimators (BLUE) when they have the minimum variance of all linear unbiased estimators.

Confidence Interval : Confidence intervals provide a range of values, which are likely to contain the true regression parameters. With every confidence interval we associate a level of statistical significance. The confidence intervals are constructed so that the probability that the interval contains the true parameter is 1 minus the level of significance. These are particularly useful for testing statistical hypotheses about the estimated regression parameters.

Maximum Likelihood : Maximum likelihood estimation focuses on the fact that different populations generate different samples; any one sample being

scrutinized is more likely to have come from some populations than from others. It involves a search over alternative parameter estimators to find those estimators which most likely would generate the sample.

Normal Distribution : The normal distribution is a bell-shaped probability distribution. It can be fully described by its mean and variance.

Probability Distribution : A random variable is examined by the process which generates its values. This process is called probability distribution, which lists all possible outcomes and the probability that each will occur.

3.10 SOME USEFUL BOOKS/REFERENCES

Gujarati, Damodar N., 1995, *Basic Econometrics*, McGraw-Hill Inc., Singapore.

Johnston, Jack and John Dinardo, 1997, *Econometric Methods*, The McGraw-Hill Companies Inc., Singapore.

Pindyck, Robert S. and Daniel L. Rubinfeld, 1998, *Econometric Models and Economic Forecasts*, Irwin/McGraw-Hill, Singapore.

3.11 ANSWERS/HINTS FOR CHECK YOUR PROGRESS EXERCISES

Check Your Progress 1

1) Go through Sections 3.1.

2) The problem will be easier if you work with the data in deviations form.

3) Go through Sections 3.2.

Check Your Progress 2

1) Go through Section 3.4.

2) Go through Section 3.5.1.

3) Go through Section 3.5.2.

4) Go through Section 3.5.3.

5) Go through Section 3.5.3.

UNIT 4 MULTIPLE REGRESSION MODEL

Structure

4.0 OBJECTIVES

After going through this unit, you will be in a position to:

- explain the concept of multiple regression;
- explain the concept of coefficient of determination;
- test the structural stability of regression models; and
- make prediction with multiple regression.

4.1 INTRODUCTION

In Units 2 and 3, we studied the basic statistical tools and procedures for analysing relationships between two variables. However, the two-variable framework is too restrictive for realistic analyses of economic phenomena. Economic models generally contain more than two variables. When a regression equation has three or more than three variables, we call it a multiple regression model. The statistical formulas for estimating the parameters, variance, and testing the parameters are very similar, or in some cases identical, to the two-variable regression model.

The simplest possible multiple regression model (without using matrix algebra) is three-variable regression, with one dependent variable and two explanatory variables. If you can understand the analysis of a relationship between three variables, you should be able to generalize the concept to a multiple regression model. A conventional example of a three variable equation is a demand equation in which quantity demanded depends not only on the price of the commodity but also on the income of a consumer.

4.2 MULTIPLE REGRESSION MODEL

We extend the two-variable model by assuming that the dependent variable Y is a linear function of a series of independent variables $X_1, X_2, \cdots\cdots, X_k$ and an error term. This model is a natural extension of the two-variable model. We write the multiple regression model as

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} + \varepsilon_i \qquad \ldots(4.1)$$

where Y is the dependent variable, X's are the independent variables, and ε is the error term. X_{2i} represents the ith observation on explanatory variable X_2. Explanatory variable X_1 is taken here as 1. β_1 is the constant term or intercept and $\beta_2, \cdots, \beta_k$ are slopes of the equation.

For simplicity we have used a three-variable regression model. The model is written as follows:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \varepsilon_i \ , \qquad i = 1, 2, \cdots, n \qquad \ldots (4.2)$$

The least-squares procedure is equivalent to searching for parameter estimates which minimize the error sum of squares, defined as

$$ESS = \sum \hat{\varepsilon}_i^{\,2} = \sum \left(Y_i - \hat{Y}_i\right)^2, \text{ where } \hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i}$$

Just as we did in Unit 2, we can find the values of β_1, β_2 and β_3 which minimize ESS. To minimize ESS, we partially differentiate it with respect to the three unknown parameters β_1, β_2 and β_3, and set the resulting expression to zero and solve them simultaneously.

To simplify we use the model in deviation form, so that

$$ESS = \sum \left(y_i - \hat{\beta}_2 x_{2i} - \hat{\beta}_3 x_{3i}\right)^2$$

then

$$\frac{\partial \sum \hat{\varepsilon}_i^{\,2}}{\partial \beta_2} = 0, \text{ or } \sum x_{2i} y_i = \beta_2 \sum x_{2i}^{\,2} + \beta_3 \sum x_{2i} x_{3i} \qquad \ldots (4.3)$$

$$\frac{\partial \sum \hat{\varepsilon}_i^{\,2}}{\partial \beta_3} = 0, \text{ or } \sum x_{3i} y_i = \beta_2 \sum x_{2i} x_{3i} + \beta_3 \sum x_{3i}^{\,2} \qquad \ldots (4.4)$$

To solve, we multiply eq.(4.3) by $\sum x_{3i}^{\,2}$ and multiply eq.(4.4) by $\sum x_{2i} x_{3i}$, and then subtract the latter from the former:

$$\sum x_{2i} y_i \sum x_{3i}^{\,2} - \sum x_{3i} y_i \sum x_{2i} x_{3i} = \beta_2 \left[\sum x_{2i}^{\,2} \sum x_{3i}^{\,2} - \left(\sum x_{2i} x_{3i}\right)^2\right]$$

or,

$$\hat{\beta}_2 = \frac{\left(\sum x_{2i} y_i\right)\left(\sum x_{3i}^{\,2}\right) - \left(\sum x_{3i} y_i\right)\left(\sum x_{2i} x_{3i}\right)}{\left(\sum x_{2i}^{\,2}\right)\left(\sum x_{3i}^{\,2}\right) - \left(\sum x_{2i} x_{3i}\right)^2} \qquad \ldots (4.5)$$

Similarly, multiplying eq.(4.3) by $\sum x_{2i}^2$ and eq.(4.2) by $\sum x_{2i}x_{3i}$ and subtracting the latter by the former and solving them, we get

$$\hat{\beta}_3 = \frac{\left(\sum x_{3i}y_i\right)\left(\sum x_{2i}^2\right) - \left(\sum x_{2i}y_i\right)\left(\sum x_{2i}x_{3i}\right)}{\left(\sum x_{2i}^2\right)\left(\sum x_{3i}^2\right) - \left(\sum x_{2i}x_{3i}\right)^2} \quad \ldots (4.6)$$

Finally, if we set the derivative of ESS with respect to β_1 equals to zero, we find that

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2\bar{X}_2 - \hat{\beta}_3\bar{X}_3 \quad \ldots (4.7)$$

In the three-variable model, the coefficient β_2 measures the change in Y associated with a unit change in X_2 on the assumption that the variable X_3 is held constant. Likewise, the coefficient β_3 measures the change in Y associated with a unit change in X_3 while X_2 is held constant. In both cases the assumption that the values of the remaining explanatory variables are constant is crucial to our interpretation of the coefficients.

4.3 ADDITIONAL ASSUMPTIONS

The assumptions of the multiple regression model are quite similar to those of the two-variable model. Besides the assumptions of two-variable regression model, the multiple regression model has another assumption, viz. no exact linear relationship exists between two or more independent variables, or we can say there is no multicollinearity. Suppose in eq.(4.2), Y, X_2, and X_3 represent consumption expenditure, income and wealth of the consumer respectively. In postulating that consumption expenditure is linearly related to income and wealth, economic theory presumes that wealth and income may have some independent influence on consumption. If not, there is no sense in including both income and wealth variables in the model. In the extreme, if there is an exact linear relationship between income and wealth, we have only one independent variable, not two, and there is no need to include both the variables. In short, the assumption of no multicollinearity requires that in the regression model we include only those variables which are not linear functions of some of the variables in the model.

4.4 TESTS OF SIGNIFICANCE

To test the statistical significance of individual regression coefficients, we need to ask whether the Gauss-Markov theorem extends to the multiple regression models and whether one can obtain an unbiased estimate of the variance σ^2 as well as information about the distribution of the estimated regression parameters. The statistical properties of the multiple regression models are presented in the following:

1) The Gauss-Markov theorem applies to the multiple regression models. That is, the ordinary least-squares estimator of each coefficient β_j. $j = 1, 2, \cdots, k$, is BLUE.

2) An unbiased and consistent estimate of σ^2 is provided by

$$s^2 = \frac{\sum \hat{\varepsilon}_i^2}{N-k} \qquad \ldots (4.8)$$

3) When the error is normally distributed, t tests can be applied because

$$\frac{\hat{\beta}_j - \beta_j}{s_{\hat{\beta}j}} \sim t_{N-k} \quad \text{for } j = 1,2,\cdots,k \qquad \ldots (4.9)$$

In other words, the estimated regression parameters, which are normalised by subtracting the mean and dividing by the estimated standard error, follow the t distribution with $N-k$ degrees of freedom[1].

4) The F-statistic calculated by most regression programmes can be used in multiple regression models to test the significance of the R^2 statistic. The F-statistic with $k-1$ and $N-k$ degrees of freedom allows us to test the hypothesis that none of the explanatory variables helps explain the variation of Y about its mean. In other words, the F-statistic tests the joint hypothesis that $\beta_2 = \beta_3 = \cdots \beta_k = 0$. It can be shown that

$$F_{k-1,N-k} = \frac{RSS/k-1}{ESS/N-k}$$

$$= \frac{R^2/k-1}{(1-R^2)/N-k} = \frac{R^2}{1-R^2}\frac{N-k}{k-1} \qquad \ldots (4.10)$$

If the null hypothesis is true, then we would expect RSS, R^2, and therefore F to be close to 0. Thus, a high value of the F-statistic is a rationale for rejecting the null hypothesis. As F-statistic is not significantly different from 0 lets us conclude that the explanatory variables do little to explain the variation of Y about its mean. In the two-variable model, the F statistic tests whether the regression line is horizontal. In such a case, R^2 = 0 and the regression explains none of the variation in the dependent variable. Note that we do not test whether the regression passes through the origin; our objective is simply to see whether we can explain any variation around the mean of Y.

The F-tests of the significance of a regression equation may allow for rejection of the null hypothesis even though none of the regression coefficients are found to be significant according to individual t tests. This situation may arise, for example, if the independent variables are highly correlated with each other. The result may be high standard errors for the coefficients and low t values, yet the model as a whole may fit the data well.

4.5 COEFFICIENT OF DETERMINATION

R^2 is called the coefficient of determination. To use R^2 as a measure of goodness of fit in the multiple regression model, we extend the discussion in Unit 2 about the decomposition of the variation in the dependent variable Y.

[1] Recall that in a two-variable model we tested for the significance of $\hat{\beta}$ at (N – 2) degree of freedom since the model included two explanatory variables, viz., constant term and X.

We can break down the difference between Y_i and its mean $\bar{Y}$ as follows:

$$\left(Y_i - \bar{Y}\right) = \left(Y_i - \hat{Y}_i\right) + \left(\hat{Y}_i - \bar{Y}\right)$$

Squaring both sides and summing over all observations (1 to N), we obtain

$$\sum\left(Y_i - \bar{Y}\right)^2 = \sum\left(Y_i - \hat{Y}_i\right)^2 + \sum\left(\hat{Y}_i - \bar{Y}\right)^2$$

Variation in Y | Residual Variation | Explained Variation

or, TSS = ESS + RSS

Total sum of squares | Residual (Error) sum of squares | Regression sum of squares

Then we define R^2 as

$$R^2 = \frac{RSS}{TSS} = \frac{\sum\left(\hat{Y}_i - \bar{Y}\right)^2}{\sum\left(Y_i - \bar{Y}\right)^2} = 1 - \frac{\sum \hat{\varepsilon}_i^2}{\sum\left(Y_i - \bar{Y}\right)^2}. \qquad \ldots(4.11)$$

R^2 measures the proportion of the variation in Y which is explained by the multiple regression equation. R^2 is often used informally as a goodness-of-fit statistic and to compare the validity of regression results under alternative specifications of the independent variables in the model.

The difficulty with R^2 as a measure of goodness of fit is that R^2 pertains only to explained and unexplained variation in Y and therefore does not account for the number of degrees of freedom. A natural solution is to use *variances* (which equals variation divided by degrees of freedom), not *variations*, thus eliminating the dependence of goodness of fit on the number of independent variables in the model. We define $\bar{R}^2$, or corrected/adjusted R^2, as

$$\bar{R}^2 = 1 - \left(1 - R^2\right)\frac{N-1}{N-k} \qquad \ldots (4.12)$$

where N is number of observations in the equation and k is number of parameters in the equation. It can be seen that as N becomes large, the differences between $\bar{R}^2$ and R^2 tend to be small.

$\bar{R}^2$ has a number of properties which make it more desirable goodness of fit measure than R^2. When new variables are added to a regression model, R^2 always increases, while $\bar{R}^2$ may rise or fall. The use of $\bar{R}^2$ eliminates at least some of the incentives for researchers to include numerous variables in a model without much thought about why they should appear.

Example 4.1

Following is a numerical example for a three variable case.

Y	X_1	X_2	y	x_1	x_2
40	4	8	− 22	− 4	− 0.4
60	6	12	− 2	− 2	3.6
50	7	10	− 12	− 1	1.6
70	10	5	8	2	− 3.4

The above information gives us

$$\sum Y = 310 \qquad \sum X_1 = 40 \qquad \sum X_2 = 42$$

$$\sum x_1 y = 260 \qquad \sum x_2 y = -84 \qquad \sum x_1 x_2 = -21$$

$$\sum x_1^2 = 50 \qquad \sum x_2^2 = 29.70 \qquad \sum y^2 = 1480$$

$$N = 5$$

Therefore,

$$\hat{\beta}_1 = 5.72$$

$$\hat{\beta}_2 = 1.24$$

$$\hat{\alpha} = 5.86$$

$$\text{ESS} = \sum \hat{\varepsilon}^2 = 96.84$$

so that

$$s = 6.96$$

$$s_{\hat{\beta}_1} = 1.18$$

$$s_{\hat{\beta}_2} = 1.54$$

Thus the estimated equation will be

$$\hat{Y} = 5.86 + \underset{(1.18)}{5.72}\ X_1 + \underset{(1.54)}{1.24}\ X_2$$

where the value in parentheses is the standard error of the coefficient. From this equation we can compute

$$\text{RSS} = \sum \hat{y}^2 = 1383.16$$

$$\text{TSS} = \sum y_2 = 1480.00$$

and

$$R^2 = \frac{RSS}{TSS} = 0.93$$

Finally, the test of significance of R^2 is provided by F ratio

$$F_{(2,2)} = \frac{RSS/(k-1)}{ESS/(n-k)} = 14.28$$

This F ratio is a test of significance of both explanatory variables simultaneously. Here it is significant at 10 per cent level of significance as the computed value (14.28) exceeds the critical value (9.0) at 10 % level of significance.

Example 4.2

For illustrating the ideas introduced so far, we consider the following information, where Y = actual rate of inflation (%) at time t, X_2 = unemployment rate (%) at time t, and X_3 = expected or anticipated inflation rate (%) at time t. This model is known as the expectations-augmented Philips curve.

Year	Y	X_2	X_3
1970	5.92	4.9	4.78
1971	4.30	5.9	3.84
1972	3.30	5.6	3.13
1973	6.23	4.9	3.44
1974	10.97	5.6	6.84
1975	9.14	8.5	9.47
1976	5.77	7.7	6.51
1977	6.45	7.1	5.92
1978	7.60	6.1	6.08
1979	11.47	5.8	8.09
1980	13.46	7.1	10.01
1981	10.24	7.6	10.81
1982	5.99	9.7	8.00

Based on the above data, the OLS method gives the following result.

$$\hat{Y} = \underset{(1.5948)}{7.1933} - \underset{(0.3050)}{1.3925} X_2 + \underset{(0.1758)}{1.4700} X_3 \qquad R^2 = 0.8766$$

where figures in the parentheses are the estimated standard errors. The interpretation of this regression is as follows: For the sample period if both X_2 and X_3 were fixed at zero, the average rate of actual inflation would have been about 7.19 per cent. The partial regression coefficient of –1.3925 means that by holding X_3 (the expected inflation rate) constant the actual inflation rate on the average increased (decreased) by about 1.4 per cent for every one unit (here one percentage point) decrease (increase) in the unemployment rate over the period 1970-1982. Likewise, by holding the unemployment rate constant, the coefficient value of 1.4700 implies that over the same time period the actual inflation rate on the average increased by about 1.47 per cent for every one percentage point increase in the anticipated or expected rate of inflation. The R^2 value of 0.88 means that the two explanatory variables together account for about 88 per cent of the variation in the actual inflation rate, a fairly high amount of explanatory power since R^2 can at best be one.

Example 4.3

With the information given in example – 4.2, we will now test whether the coefficients significantly determine the rate of inflation. We can postulate the following hypotheses.

1) $H_0 : \beta_2 = 0$, and $H_1 : \beta_2 \neq 0$

2) $H_0 : \beta_3 = 0$ and $H_1 : \beta_3 \neq 0$

The null hypothesis states that, holding X_3 constant, unemployment has no (linear) influence on the actual rate of inflation. Similarly, holding X_2 constant, expected inflation has no influence on actual inflation. To test the null hypotheses, we use the t test. If the computed t value exceeds the critical t value at the chosen level of significance, we may reject the null hypothesis.

1) Noting that $\beta_2 = 0$ under the null hypothesis, we obtain

$$t = \frac{-1.3925}{0.3050} = -4.566$$

Using the two-tailed t test, the critical t value is 3.169. Since the computed t value of 4.566 exceeds the critical value of 3.169, we may reject the null hypothesis and say that $\hat{\beta}_2$ is statistically significant, that is, significantly different from zero.

2) Similarly, for testing the significance of β_3, we obtain the t value as

$$t = \frac{1.4700}{0.1758} = 8.362$$

In this case, the computed t value exceeds the critical t value of 3.169 at 1 per cent level of significance. Therefore, we reject the null hypothesis and find that the coefficient is significantly different from zero.

The above two t - tests show that the independent variables have significant influence on the dependent variables. In other words, the increase (decrease) in actual inflation is due to both the decrease (increase) in unemployment rate as well as the increase (decrease) in expected inflation.

Check Your Progress 1

1) The following sums were obtained from 10 sets of observations on Y, X_1 and X_2:

$$\sum Y = 20 \qquad \sum X_1 = 30 \qquad \sum X_2 = 40$$

$$\sum Y^2 = 88.2 \qquad \sum X_1^2 = 92 \qquad \sum X_2^2 = 163$$

$$\sum YX_1 = 59 \qquad \sum YX_2 = 88 \qquad \sum X_1X_2 = 119$$

Estimate the regression of Y on X_1 and X_2, including an intercept term, and test the hypothesis that the coefficient of X_2 is zero.

.....

.....

.....

2) Consider a multiple regression model for which all classical assumptions hold, but in which there is no constant term. Suppose you wish to test the null hypothesis that there is no relationship between Y and X, that is,

$$H_0 : \beta_2 = \beta_3 = \cdots \beta_k = 0$$

against the alternative that at least one of the β's is nonzero. Present the appropriate test and state its distributions (including the numbers of degrees of freedom).

.....

.....

.....

4.6 MATRIX FORM OF MULTIPLE REGRESSION MODEL

The multiple regression model can be represented in matrix form. We begin by representing the linear model in matrix form. We can see from eq. (4.1) that the regression model includes $k+1$ variables – a dependent variable and k independent variables (including the constant term). Since there are N observations, we can summarise the regression model by writing a series of N equations, as follows:

$$\begin{aligned} Y_1 &= \beta_1 + \beta_2 X_{21} + \beta_3 X_{31} + \cdots + \beta_k X_{k1} + \varepsilon_1 \\ Y_2 &= \beta_1 + \beta_2 X_{22} + \beta_3 X_{32} + \cdots + \beta_k X_{k2} + \varepsilon_2 \\ Y_N &= \beta_1 + \beta_2 X_{2N} + \beta_3 X_{3N} + \cdots + \beta_k X_{kN} + \varepsilon_N \end{aligned} \quad \ldots (4.13)$$

The matrix formulation of the above model is

$$Y = X\beta + \varepsilon \quad \ldots (4.14)$$

in which

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \cdots \\ Y_N \end{bmatrix} \quad X = \begin{bmatrix} 1 & X_{21} & \cdots & X_{k1} \\ 1 & X_{22} & \cdots & X_{k2} \\ \cdots & \cdots & \cdots & \cdots \\ 1 & X_{2N} & \cdots & X_{kN} \end{bmatrix} \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \cdots \\ \beta_N \end{bmatrix} \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdots \\ \varepsilon_N \end{bmatrix} \quad \ldots (4.15)$$

where

$Y = N \times 1$ column vector of dependent variable observations

$X = N \times k$ matrix of independent variable observations

$\beta = k \times 1$ column vector of unknown parameters

$\varepsilon = N \times 1$ column vector of errors

In our representation of the matrix X, each component X_{ji} has two subscripts, the first satisfying the appropriate column (variable) and the second signifying row (observation). Each column of X represents a vector of N observations on a given variable, with all observations associated with the intercept equal to 1.

The assumptions of the classical linear regression model can be represented as follows:

(i) The elements of X are fixed and have finite variance. In addition X has rank k, which is less than the number of observations N.

(ii) ε is normally distributed with $E(\varepsilon) = 0$ and $E(\varepsilon\varepsilon') = \sigma^2 I$, where I is an $N \times N$ identity matrix.

The *variance-covariance matrix* $\sigma^2 I$ appears as follows

$$E(\varepsilon\varepsilon') = E\left\{\begin{bmatrix}\varepsilon_1\\ \varepsilon_2\\ \cdots\\ \varepsilon_N\end{bmatrix}\begin{bmatrix}\varepsilon_1 & \varepsilon_2 & \cdots & \varepsilon_N\end{bmatrix}\right\}$$

$$= E\begin{bmatrix}(\varepsilon_1^2) & (\varepsilon_1\varepsilon_2) & \cdots & (\varepsilon_1\varepsilon_N)\\ (\varepsilon_2\varepsilon_1) & (\varepsilon^2) & \cdots & (\varepsilon_2\varepsilon_N)\\ \cdots & \cdots & \cdots & \cdots\\ (\varepsilon_N\varepsilon_1) & (\varepsilon_N\varepsilon_2) & \cdots & (\varepsilon_N^2)\end{bmatrix}$$

$$= \begin{bmatrix}E(\varepsilon_1^2) & E(\varepsilon_1\varepsilon_2) & \cdots & E(\varepsilon_1\varepsilon_N)\\ E(\varepsilon_2\varepsilon_1) & E(\varepsilon^2) & \cdots & E(\varepsilon_2\varepsilon_N)\\ \cdots & \cdots & \cdots & \cdots\\ E(\varepsilon_N\varepsilon_1) & E(\varepsilon_N\varepsilon_2) & \cdots & E(\varepsilon_N^2)\end{bmatrix}$$

$$= \begin{bmatrix}Var(\varepsilon_1^2) & Cov(\varepsilon_1,\varepsilon_2) & \cdots & Cov(\varepsilon_1,\varepsilon_N)\\ Cov(\varepsilon_2,\varepsilon_1) & Var(\varepsilon^2) & \cdots & Cov(\varepsilon_2,\varepsilon_N)\\ \cdots & \cdots & \cdots & \cdots\\ Cov(\varepsilon_N,\varepsilon_1) & Cov(\varepsilon_N,\varepsilon_2) & \cdots & Var(\varepsilon_N^2)\end{bmatrix}$$

Using the assumptions of homoscedasticity and no serial correlation, the above matrix reduces to

$$E(\varepsilon\varepsilon') = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix}$$

$$= \sigma^2 \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

$$= \sigma^2 I$$

where I is an $N \times N$ matrix.

Our objective is to find a vector of parameters $\hat{\beta}$ which minimise ESS.

$$\text{ESS} = \sum_{i=1}^{N} \hat{\varepsilon}_i^2 = \hat{\varepsilon}'\hat{\varepsilon} \qquad \ldots (4.16)$$

Where

$$\hat{\varepsilon} = Y - \hat{Y} \qquad \ldots (4.17)$$

And

$$\hat{Y} = X\hat{\beta} \qquad \ldots (4.18)$$

$\hat{\varepsilon}$ represents the $N \times 1$ vector of regression residuals, while $\hat{Y}$ represents the $N \times 1$ vector of fitted values for Y. Substituting eqs. (4.17) and (4.18) into eq. (4.16), we get

$$\hat{\varepsilon}'\hat{\varepsilon} = \left(Y - X\hat{\beta}\right)'\left(Y - X\hat{\beta}\right)$$

$$= Y'Y - \hat{\beta}'X'Y - Y'X\hat{\beta} + \hat{\beta}'X'X\hat{\beta}$$

$$= Y'Y - 2\hat{\beta}'X'Y + \hat{\beta}'X'X\hat{\beta}$$

The last steps follows because $\hat{\beta}'X'Y$ and $Y'X\hat{\beta}$ are both scalars and are equal to each other.

From Unit 2 we know that the least-squares estimators can be obtained by minimizing ESS. In order to minimise ESS we take partial derivatives with respect to β so that :

$$\frac{\partial ESS}{\partial \hat{\beta}} = 2X'Y + 2X'X\hat{\beta} = 0$$

By solving for the above we obtain

$$\hat{\beta} = (X'X)^{-1}(X'Y) \qquad \ldots (4.19)$$

The matrix $X'X$, called the *cross-product matrix* is guaranteed to have an inverse because of our assumption that X has rank k. The cross product matrix can be shown as

$$X'X = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ X_{21} & X_{22} & \cdots & X_{2N} \\ \cdots & \cdots & \cdots & \cdots \\ X_{k1} & X_{k1} & \cdots & X_{kN} \end{bmatrix} \cdot \begin{bmatrix} 1 & X_{21} & \cdots & X_{k1} \\ 1 & X_{22} & \cdots & X_{k2} \\ \cdots & \cdots & \cdots & \cdots \\ 1 & X_{2N} & \cdots & X_{kN} \end{bmatrix}$$

$$= \begin{bmatrix} N & \sum X_{2i} & \cdots & \sum X_{ki} \\ \sum X_{2i} & \sum X_{2i}^2 & \cdots & \sum X_{2i}X_{ki} \\ \cdots & \cdots & \cdots & \cdots \\ \sum X_{ki} & \sum X_{ki}X_{2i} & \cdots & \sum X_{ki}^2 \end{bmatrix}$$

Now consider the properties of the least-squares estimator $\hat{\beta}$. First, we can prove that $\hat{\beta}$ is an unbiased estimator of β.

Substituting $Y = X\beta + \varepsilon$ into eq. (4.19) gives,

$$\begin{aligned} \hat{\beta} &= (X'X)^{-1} X'(X\beta + \varepsilon) \\ &= (X'X)^{-1} X'X\beta + (X'X)^{-1} X'\varepsilon \\ &= \beta + (X'X)^{-1} X'\varepsilon \end{aligned} \qquad \ldots (4.20)$$

Now

$$\begin{aligned} \mathrm{E}(\hat{\beta}) &= E\left[\beta + (X'X)^{-1} X'\varepsilon\right] \\ &= \beta + (X'X)^{-1} X'E(\varepsilon) \\ &= \beta \qquad \text{since } E(\varepsilon) = 0 \end{aligned}$$

Hence, the least-squares estimators are unbiased.

The least squares estimator will be normally distributed, since $\hat{\beta}$ is a linear function of ε and ε is normally distributed.

From eq. (4.20) we get

$$\hat{\beta} - \beta = (X'X)^{-1} X'\varepsilon \qquad \ldots (4.21)$$

By definition

$$\begin{aligned} \mathrm{Var}(\hat{\beta}) &= E\left[\left(\hat{\beta} - \beta\right)\left(\hat{\beta} - \beta\right)'\right] \\ &= E\left\{\left[(X'X)^{-1} X'\varepsilon\right]\left[(X'X)^{-1} X'\varepsilon\right]'\right\} \\ &= E\left[(X'X)^{-1} X'\varepsilon\varepsilon' X (X'X)^{-1}\right] \\ &= (X'X)^{-1} X'E(\varepsilon\varepsilon')X(X'X)^{-1}, \end{aligned}$$

since X's are non-stochastic

$$= (X'X)^{-1} X'\sigma^2 IX(X'X)^{-1}$$

$$= \sigma^2 (X'X)^{-1} \quad \ldots (4.22)$$

In deriving the above result use is made of the assumption that $E(\varepsilon'\varepsilon) = \sigma^2 I$.

We have already proved that the least-squares estimator is linear and unbiased. In fact, $\hat{\beta}$ is the best linear unbiased estimator of β in the sense that it has the minimum variance of all unbiased estimators. To complete the proof of the Gauss-Markov theorem, we need to show that any other unbiased linear estimator has greater variance than $\hat{\beta}$.

We have

$$\hat{\beta} = (X'X)^{-1}(X'Y)$$

and

$$Var(\hat{\beta}) = \sigma^2 (X'X)^{-1}$$

Let we take

$(X'X)^{-1} X' = w'$, so that $\hat{\beta} = w'Y$

Let us take an arbitrary linear estimator

$\beta^* = c'Y$, where $c' = w' + d'$

$$= c'(X\beta + \varepsilon)$$

$$= c'X\beta + c'\varepsilon$$

$$E(\beta^*) = c'X\beta + c'E(\varepsilon)$$

$$= c'X\beta + c'(0)$$

$$= c'X\beta$$

β^* will be an unbiased estimator, if and only if,

$$c'X = I_k$$

So

$$E(\beta^*) = c'X\beta = I_k\beta = \beta$$

We have

$$\beta^* - c'Y = c'(X\beta + \varepsilon) = c'X\beta + c'\varepsilon = \beta + c'\varepsilon$$

Or

$$\beta^* - \beta = c'\varepsilon$$

Now

$$Var(\beta^*) = E[(\beta^*-\beta)(\beta^*-\beta)']$$
$$= E[(c'\varepsilon)(\varepsilon'c)]$$
$$= c'E(\varepsilon\varepsilon')c$$
$$= c'\sigma^2 I_N c$$
$$= \sigma^2 I_N c'c$$
$$= \sigma^2 c'c$$

We have

$$Var(\hat{\beta}) = \sigma^2 (X'X)^{-1}$$
$$= \sigma^2 w'w$$

We have taken

$$c' = w'+d'$$

Or

$$c'c = (w'+d')(w+d)$$
$$= w'w + d'd + w'd + d'w$$
$$= w'w + d'd$$

since $w'd = 0$ and $d'w = 0$

Now

$$\sigma^2 c'c = \sigma^2 w'w + \sigma^2 d'd$$

or

$$Var(\beta^*) = Var(\hat{\beta}) + \sigma^2 d'd$$

Since

$$d'd \geq 0, \; Var(\hat{\beta}) \leq Var(\beta^*)$$

Hence, $\hat{\beta}$ is the best linear unbiased estimator.

Now for calculating R^2, we decompose the total variation of Y as follows.

$$Y = X\hat{\beta} + \hat{\varepsilon}$$

Then

$$Y'Y = (X\hat{\beta} + \hat{\varepsilon})'(X\hat{\beta} + \hat{\varepsilon})$$
$$= \hat{\beta}'X'X\hat{\beta} + \hat{\varepsilon}'X\hat{\beta} + \hat{\beta}'X'\hat{\varepsilon} + \hat{\varepsilon}'\hat{\varepsilon}$$
$$= \hat{\beta}'X'X\hat{\beta} + \hat{\varepsilon}'\hat{\varepsilon} \text{ (since } X'\hat{\varepsilon} = 0 \text{ and } \hat{\varepsilon}'X = 0) \quad \ldots (4.23)$$

or, $TSS = RSS + ESS.$

Then

$$R^2 = 1 - \frac{ESS}{TSS} = 1 - \frac{\hat{\varepsilon}'\hat{\varepsilon}}{Y'Y} = \frac{\hat{\beta}'X'X\hat{\beta}}{Y'Y} \qquad \ldots(4.24)$$

To correct for the dependence of goodness of fit on degrees of freedom, we define $\bar{R}^2$ as

$$\bar{R}^2 = 1 - \frac{\hat{\varepsilon}'\hat{\varepsilon}/(N-k)}{Y'Y/(N-1)} \qquad \ldots (4.25)$$

Example 4.4

From a sample of 10 observations corresponding to the regression model, the following quantities were calculated:

$$\sum Y = 30 \qquad \sum X_1 = 10 \qquad \sum X_2 = 40$$

$$\sum Y^2 = 100 \qquad \sum X_1^2 = 51 \qquad \sum X_2^2 = 170$$

$$\sum YX_1 = 45 \qquad \sum YX_2 = 128 \qquad \sum X_1X_2 = 60$$

Find the least squares estimates of β_0, β_1, β_2 using the data in the deviation form. Calculate R^2 and interpret the estimated relation.

From the above data we construct the following deviation forms:

$$\sum y^2 = \sum Y^2 - \frac{1}{N}\left(\sum Y^2\right) = 10$$

$$\sum x_1^2 = \sum X_1^2 - \frac{1}{N}\left(\sum X_1^2\right) = 41$$

$$\sum x_2^2 = \sum X_2^2 - \frac{1}{N}\left(\sum X_2^2\right) = 10$$

$$\sum x_1 y = \sum X_1 Y - \frac{1}{N}\left(\sum X_1\right)\left(\sum Y\right) = 15$$

$$\sum x_2 y = \sum X_2 Y - \frac{1}{N}\left(\sum X_2\right)\left(\sum Y\right) = 8$$

$$\sum x_1 x_2 = \sum X_1 X_2 - \frac{1}{N}\left(\sum X_1\right)\left(\sum X_2\right) = 20$$

Now we hav

$$X'X = \begin{bmatrix} \sum x_1^2 & \sum x_1 x_2 \\ \sum x_1 x_2 & \sum x_2^2 \end{bmatrix} = \begin{bmatrix} 41 & 20 \\ 20 & 10 \end{bmatrix}$$

$$X'Y = \begin{bmatrix} \sum x_1 y \\ \sum x_2 y \end{bmatrix} = \begin{bmatrix} 15 \\ 8 \end{bmatrix}$$

We can find the value of $\hat{\beta}_1$ and $\hat{\beta}_2$ from the following

$$\begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = (X'X)^{-1} X'Y$$

$$= \begin{bmatrix} 41 & 20 \\ 20 & 10 \end{bmatrix}^{-1} \begin{bmatrix} 15 \\ 8 \end{bmatrix}$$

Since

$$\begin{vmatrix} 41 & 20 \\ 20 & 10 \end{vmatrix} = 10$$

We have

$$\hat{\beta}_1 = \frac{\begin{vmatrix} 15 & 20 \\ 8 & 10 \end{vmatrix}}{10} = -1$$

$$\hat{\beta}_2 = \frac{\begin{vmatrix} 41 & 15 \\ 20 & 8 \end{vmatrix}}{10} = 2.8$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}_1 - \hat{\beta}_2 \bar{X}_2$$

$$= \frac{30}{10} - (-1)\left(\frac{10}{10}\right) - 2.8\left(\frac{40}{10}\right)$$

$$= 3 + 1 - 11.2$$

$$= -7.2$$

$$R^2 = \frac{RSS}{TSS} = \frac{\sum y^2}{\hat{\beta}' X'Y} = \frac{\sum y^2}{\hat{\beta}_1 \sum x_1 y + \hat{\beta}_2 \sum x_2 y}$$

$$= \frac{7.4}{10} = 0.74$$

The R^2 value of 0.74 means that the two explanatory variables together account for about 74 per cent of the variation in the dependent variable.

4.7 STRUCTURAL STABILITY OF REGRESSION MODELS – CHOW TEST

So far we have estimated the model on the assumption that the parameters (βs) remain constant throughout the model. In reality, however, the parameters

may undergo change so that there is a structural change in the model. In order test for structural stability of a model there are several tests.

The Chow forecast test leads naturally to more general tests of structural change. A structural change or break occurs if the parameters underlying a relationship differ from one subset of the data to another. There may, of course, be several relevant subsets of data, with the possibility of several structural breaks. For the moment we will consider just two subsets of n_1 and n_2 observations making up the total sample of $n = n_1 + n_2$ observations. Suppose; for example, that we wish to investigate whether savings in a country differs between pre-liberalisation and post-liberalisation periods. Suppose we have observations on the relevant variables for n_1 pre-liberalisation years and n_2 post-liberalisation years. A Chow test could be performed by using the estimated pre-liberalisation function to forecast savings in the post-liberalisation years. However, provided $n_2 > k$, one might alternatively use the estimated post-liberalisation function to forecast savings in pre-liberalisation years. It is, however, not clear which choice should be made, and the two procedures might well yield different answers. If the subsets are large enough it is better to estimate both functions and test for common parameters.

To see the structural change, we formulate the savings functions for the two periods as follows:

Period I:

$$Y_t = \alpha_1 + \alpha_2 X_t + \varepsilon_{1t} \qquad \ldots (4.26)$$

$$t = 1, 2, \cdots, n_1$$

Period II:

$$Y_t = \beta_1 + \beta_2 X_t + \varepsilon_{2t} \qquad \ldots (4.27)$$

$$t = 1, 2, \cdots, n_2$$

where Y is personal savings, X is personal income, the ε's are the disturbance terms in the two equations, n_1 and n_2 are the number of observations in the two periods.

A structural change may mean that the two intercepts are different, or the two slopes are different, or both the intercept and the slopes are different, or any other suitable combination of the parameters. If there is no structural change (i.e., structural stability), we can combine all the n_1 and n_2 observations and just estimate one savings function as

$$Y_t = \lambda_1 + \lambda_2 X_t + \varepsilon_t \qquad \ldots (4.28)$$

The assumptions of Chow test are two fold:

(a) $\varepsilon_{1t} \sim N(0, \sigma^2)$ and $\varepsilon_{2t} \sim N(0, \sigma^2)$, i.e., the two error terms are normally distributed with the same variance, and

(b) ε_{1t} and ε_{2t} are independently distributed.

The Chow test proceeds as follows:

Step I: Combining all the n_1 and n_2 observations, we estimate (4.28) and obtain restricted residual (error) sum of squares, say ESS_R, with DF = $(n_1 + n_2 - k)$, where k is the number of estimated parameters estimated.

Step II: Estimate (4.26) and (4.27) individually and obtain their ESS, say ESS_1 and ESS_2, with DF = $(n_1 - k)$ and $(n_2 - k)$, respectively.

Step III: Add these two residual sum of squares, say ESS_{UR} (unrestricted residual sum of squares), with DF = $(n_1 + n_2 - 2k)$. Find $ESS_R - ESS_{UR}$, with DF = k.

Step IV: Given the assumptions of Chow test, it can be shown that

$$F = \frac{(ESS_R - ESS_{UR})/k}{ESS_{UR}/(n_1 + n_2 - 2k)} \quad \ldots (4.29)$$

follows the F distribution with DF = $(k, n_1 + n_2 - 2k)$. If the F computed from (4.29) exceeds the critical value at the chosen level of confidence level, reject the hypothesis that the regressions (4.26) and (4.27) are the same, that is, reject the hypothesis of structural stability.

Example 4.5

We present the data on personal savings and personal income in the United Kingdom for the period 1946-1963 in the following. We want to test whether the savings function is same in the two time periods.

	Period I			Period II	
1946-1954	Savings	Income	1955-1963	Savings	Income
1946	0.36	8.8	1955	0.59	15.5
1947	0.21	9.4	1956	0.90	16.7
1948	0.08	10.0	1957	0.95	17.7
1949	0.20	10.6	1958	0.82	18.6
1950	0.10	11.0	1959	1.04	19.7
1951	0.12	11.9	1960	1.53	21.1
1952	0.41	12.7	1961	1.94	22.8
1953	0.50	13.5	1962	1.75	23.9
1954	0.43	14.3	1963	1.99	25.2

In the above data, we have $n_1 = n_2 = 9$.

Step I:

$$\hat{Y}_t = -1.0821 + 0.1178\ X_t$$

$$(0.1452) \quad (0.0088)$$

$$t = \quad (-7.4548) \quad (13.4316)$$

$$R^2 = 0.9185, \qquad ESS_R = 0.5722, \qquad DF = 16$$

Step II: Period 1946 – 1954

$$\hat{Y}_t = \quad -0.2622 \quad + \quad 0.0470 \; X_t$$

$$(0.3054) \quad (0.0266)$$

$$t = \quad (-0.8719) \quad (1.7700)$$

$$R^2 = 0.3092, \qquad ESS_1 = 0.1396, \qquad DF = 7$$

Period 1955 – 1963

$$\hat{Y}_t = \quad -1.7502 \quad + \quad 0.1504 \; X_t$$

$$(0.3576) \quad (0.0175)$$

$$t = \quad (-4.8948) \quad (8.5749)$$

$$R^2 = 0.9131, \qquad ESS_2 = 0.1931, \qquad DF = 7$$

Step III:

$$ESS_{UR} = 0.3327$$

$$ESS_R - ESS_{UR} = 0.2395$$

Step IV:

$$F = \frac{0.2395/2}{0.3327/14} = 5.04$$

At the 5% level of significance the critical value of $F_{2,14} = 3.74$. Since the observed F value of 5.04 exceeds this critical value, we can reject the hypothesis that the savings function in the two time periods is the same.

Though we accept the conclusion that the savings functions in the two time periods are different, this difference is due to either a difference in the intercept values, or slope values, or both. Although the Chow test can be adapted to show this, dummy variables can show this more readily.

4.8 PREDICTION WITH MULTIPLE REGRESSION MODEL

The formulas for prediction in multiple regression are similar to those in the case of simple regression except that to compute the standard error of the predicted value we need the variances and covariances of all the regression coefficients.

Let the estimated regression equation be

$$\hat{Y} = \hat{\beta}_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3 .$$

Now consider the prediction of the value Y_0 of Y given values X_{20} of X_2 and X_{30} of X_3, respectively.

Then we have

$$Y_0 = \beta_1 + \beta_2 X_{20} + \beta_3 X_{30} + \varepsilon_0$$

Consider

$$\hat{Y}_0 = \hat{\beta}_1 + \hat{\beta}_2 X_{20} + \hat{\beta}_3 X_{30}$$

The prediction error is

$$\hat{Y}_0 - Y_0 = \hat{\beta}_1 - \beta_1 + \left(\hat{\beta}_2 - \beta_2\right) x_{20} + \left(\hat{\beta}_3 - \beta_3\right) X_{30} - \varepsilon_0$$

Since $E\left(\hat{\beta}_1 - \beta_1\right)$, $E\left(\hat{\beta}_2 - \beta_2\right)$, $E\left(\hat{\beta}_3 - \beta_3\right)$, and $E(\varepsilon_0)$ are all equal to zero, we have $E\left(\hat{Y}_0 - Y_0\right) = 0$. Thus the predictor $\hat{Y}_0$ is unbiased. Since $\hat{Y}_0$ and Y_0 are random variables $E\left(\hat{Y}_0\right) = E(Y_0)$.

The variance of the prediction error is

$$\sigma^2\left(1 + \frac{1}{n}\right) + \left(X_{20} - \bar{X}_2\right)^2 Var\left(\hat{\beta}_2\right)$$

$$+2\left(X_{20} - \bar{X}_2\right)\left(X_{30} - \bar{X}_3\right) Cov\left(\hat{\beta}_2, \hat{\beta}_3\right) + \left(X_{30} - \bar{X}_3\right)^2 Var\left(\hat{\beta}_3\right)$$

We estimate σ^2 by $RSS/(n-3)$ in the case of two explanatory variables and by $RSS/(n-k-1)$ in the general case.

Check Your Progress 2

1) Data on a three-variable problem yield the following results:

$$X'X = \begin{bmatrix} 33 & 0 & 0 \\ 0 & 40 & 20 \\ 0 & 20 & 60 \end{bmatrix} \qquad X'Y = \begin{bmatrix} 132 \\ 24 \\ 92 \end{bmatrix}$$

$$\sum\left(Y - \bar{Y}\right)^2 = 150$$

(a) What is the sample size?

(b) Compute the regression equation.

(c) Estimate the standard error of $\hat{\beta}_2$ and test the hypothesis β_2 is zero.

.....

.....

.....

.....

2) Data on expenditure and income of a State in India for two periods of time is presented in the following. Test whether the expenditure function is same in the two time periods.

.....

.....

.....

.....

.....

	Period I			Period II	
1981-91	Expenditure (Rs. Cr.)	NSDP (Rs. Cr.)	1991-2001	Expenditure (Rs. Cr.)	NSDP (Rs. Cr.)
1981-82	741	3844	1991-92	3291	12505
1982-83	1007	4070	1992-93	3636	13416
1983-84	967	5253	1993-94	4067	15481
1984-85	1133	5191	1994-95	4662	18156
1985-86	1258	6226	1995-96	5145	21463
1986-87	1571	6748	1996-97	5996	22669
1987-88	1798	6853	1997-98	6393	28000
1988-89	2075	8681	1998-99	7733	31211
1989-90	2273	9917	1999-00	9258	34223
1990-91	2742	9664	2000-01	9668	33906

4.9 LET US SUM UP

This unit introduced the simplest possible multiple linear regression model, namely, the three-variable regression model. This has also been generalised to a multiple linear regression involving k number of variables model with the help of matrix. This unit also discussed testing the individual significance of regression coefficient and testing the overall significance of regression. The coefficient of determination, which measures the proportion of the variation in Y is often used informally as a goodness-of-fit statistic and to compare the validity of regression results under alternative specifications of the

independent variables in the model. This has been dealt with in this unit. Besides, this unit discussed testing for structural stability of regression models and prediction of dependent variables, given the values of the independent variables.

4.10 KEY WORDS

Adjusted R-square : It is a measure of goodness of fit which accounts for not only explained and unexplained variation in Y but also for the number of degrees of freedom. It is therefore corrected R^2 which accounts for only the former.

Multicollinearity : When linear relationship exists between two or more independent variables, we say there is multicollinearity.

Multiple regression model : When a regression equation has three or more than three variables, we call it a multiple regression model. The simplest possible multiple regression model is three-variable regression, with one dependent variable and two explanatory variables.

Regression sum of squares : The variation of Y is divided into two parts – The first part, which accounted for by the regression equation, i.e. explained portion, is called regression sum of square.

Residual sum of squares : The variation of Y is divided into two portions. The unexplained portion of the model is called the residual (or error) sum of square.

Structural stability : When there is no structural change or break, i.e., the parameters underlying a relationship do not differ from one subset of the data to another, we say there is structural stability of a function.

4.11 SOME USEFUL BOOKS

Gujarati, Damodar N., 1995, *Basic Econometrics*, McGraw-Hill Inc., Singapore.

Johnston, Jack and John Dinardo,1997, *Econometric Methods*, The McGraw-Hill Companies Inc., Singapore.

Pindyck, Robert S. and Daniel L. Rubinfeld, 1998, *Econometric Models and Economic Forecasts*, Irwin/McGraw-Hill, Singapore.

4.12 ANSWERS/HINTS TO CHECK YOUR PROGRESS EXERCISES

Check Your Progress 1

1) Go through sections 4.2, 4.4 and 4.5.

2) Go through section 4.4.

Check Your Progress 2

1) Go through sections 4.4 and 4.6.

2) Go through section 4.7.

UNIT 5 GENERALISED LEAST SQUARES

Structure

5.0 OBJECTIVES

After going this unit, you will be able to:

- specify the generalised least squares model;
- estimate the parameters by the generalised least squares method; and
- show that GLS covers classical regression, heteroscedasticity and autocorrelation as special cases.

5.1 INTRODUCTION

The multiple regression model discussed in the previous Unit is based on a number of assumptions concerning the stochastic error term. Two important assumptions regarding the error term ε_i are that, it has a constant variance (that is, homoscedasticity) and the error terms are not related to one another (no autocorrelation). These two assumptions may not hold in real life and actual data that we take for analysis may not fulfil them. When the error term does not have a constant variance, that is, error term varies according to changes in certain variable we have the problem of heteroscedasticity. On the other hand, when the error terms are related to one another we encounter the problem of autocorrelation. We will learn about the consequences, detection and remedial measures for these two problems in Units 7 and 8 respectively. Here in this Unit we discuss an estimation method, that is, generalised least squares, which can solve both these problems.

5.2 ASSUMPTIONS OF CLASSICAL REGRESSION MODEL

The classical regression is based on rather restrictive assumptions concerning the behaviour of the error term. An alternate model, known as 'generalised linear regression model', is considerably less restrictive in this respect. If we do not make these assumptions, but retain all other assumptions of the classical normal linear regression model, we have the so-called generalised linear regression model. Now the full description of the model is given below.

1) $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} + \varepsilon_i$.

2) The joint distribution of $\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_N$ is multivariate normal.

3) $E(\varepsilon_i) = 0$, $\quad i = 1,2,\cdots,N$.

4) $E(\varepsilon_i \varepsilon_j) = \sigma_{ij}$, $\quad i, j = 1,2,\cdots,N$.

5) Each of the explanatory variables is non-stochastic and such that, for any sample size

$$\frac{1}{N}\sum_{i=1}^{N}\left(X_{ki} - \bar{X}_k\right)^2$$

is a finite number different from zero for every $k = 2,\cdots,k$.

6) The number of observations exceeds the number of explanatory variables plus one, i.e., $N > k+1$.

7) No exact linear relation exists between any of the explanatory variables.

According to assumptions (3) and (4) above, σ_i^2 is the variance of ε_i and σ_{ij} $(i \neq j)$ is the covariance of ε_i and ε_j. Using the matrix notation, we can restate the model as

$$Y = X\beta + \varepsilon \qquad \ldots(5.1)$$

where

$Y = N \times 1$ column vector of dependent variable observations

$X = N \times k$ matrix of independent variable observations

$\beta = k \times 1$ column vector of unknown parameters

$\varepsilon = N \times 1$ column vector of errors

The assumption (4) given above can be written as

$$E(\varepsilon\varepsilon') = V$$

where

$$V = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1N} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2N} \\ \cdots & \cdots & \cdots & \cdots \\ \sigma_{N1} & \sigma_{N2} & \cdots & \sigma_N^2 \end{bmatrix} \quad \ldots(5.2)$$

This model is called 'generalised' because it includes other models as special cases. The classical normal linear regression model is one such special case, in which V is a diagonal matrix with σ^2 in place of each of the diagonal elements and zero at other places. Thus for the classical regression model we have

$$V = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix} \quad \ldots(5.3)$$

Another special case is the heteroscedasticitic model, where V is again diagonal, but all the diagonal elements are not necessarily the same. Thus we have

$$V = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \sigma_N^2 \end{bmatrix} \quad \ldots(5.4)$$

For the model in which the disturbances follow a first-order autoregressive scheme, the error variance-covariance matrix V becomes

$$V = \sigma^2 \begin{bmatrix} 1 & \rho & \cdots & \rho^{n-1} \\ \rho & 1 & \cdots & \rho^{n-2} \\ \cdots & \cdots & \cdots & \cdots \\ \rho^{n-1} & \rho^{n-2} & \cdots & 1 \end{bmatrix} \quad \ldots(5.5)$$

In Unit 4, under OLS estimation, we assumed that

$$E(\varepsilon\varepsilon') = \sigma^2 I$$

where I is the identity matrix.

Under this assumption we obtained the OLS estimator as (see equation (4.19))

$$\hat{\beta} = (X'X)^{-1}(X'Y)$$

It is easy to show that $\hat{\beta}_{OLS}$ is unbiased so long as $E(\varepsilon) = 0$. We have seen from (4.20) that

$$\hat{\beta} = (X'X)^{-1}(X'Y) = \beta + (X'X)^{-1}X'\varepsilon$$

Therefore, we have

$$E(\hat{\beta}) = \beta.$$

However, if $E(\varepsilon\varepsilon') \neq \sigma^2 I$

then $\hat{\beta}_{OLS}$ is unbiased but not efficient.

5.3 ESTIMATION PROCEDURE

The problem now is to estimate β in (5.1), that is,

$$Y = X\beta + \varepsilon$$

when

$$E(\varepsilon) = 0 \text{ and } E(\varepsilon\varepsilon') = V$$

where V is assumed to be known, symmetric, positive-definite matrix.

This estimation problem may be approached in a number of ways. We present the simplest method in the following:

We know that a positive definite matrix can be expressed in the form PP', where P is non-singular. Note that P' is the transpose of the matrix P. Thus, we write

$$V = PP'$$

so that

$$(P^{-1}VP^{-1})' = I$$

and

$$(P^{-1})'(P^{-1}) = V^{-1}$$

Now pre-multiplying the model $Y = X\beta + \varepsilon$ by P^{-1}, we get

$$P^{-1}Y = P^{-1}X\beta + P^{-1}\varepsilon$$

or

$$M = N\beta + u \qquad \ldots(5.6)$$

where

$$P^{-1}Y = M,\ P^{-1}X = N,\ P^{-1}\varepsilon = u$$

Applying least squares to the transformed model (5.6), we get

$$\begin{aligned}\hat{\beta}_{gls} &= (N'N)^{-1}N'M \\ &= \{(P^{-1}X)'(P^{-1}X)\}^{-1}\{(P^{-1}X)'(P^{-1}Y)\} \\ &= (X'P^{-1'}P^{-1}X)^{-1}(X'P^{-1'}P^{-1}Y)\end{aligned}$$

$$= (X'V^{-1}X)^{-1}(X'V^{-1}Y) \qquad \ldots(5.7)$$

This is called the generalised least squares estimator of β.

5.3.1 Properties of GLS Estimators

Now putting the value of Y in the above equation, we get

$$\begin{aligned}\hat{\beta}_{gls} &= (X'V^{-1}X)^{-1}\{X'V^{-1}(X\beta+\varepsilon)\}\\ &= (X'V^{-1}X)^{-1}(X'V^{-1}X)\beta+(X'V^{-1}X)^{-1}(X'V^{-1}\varepsilon)\\ &= \beta+(X'V^{-1}X)^{-1}X'V^{-1}\varepsilon \qquad \ldots(5.8)\end{aligned}$$

Since $E(\hat{\beta})=\beta$ in (5.8) above we can say that $\hat{\beta}_{GLS}$ is unbiased estimator of β.

By re-arranging terms in (5.8) we obtain

$$\hat{\beta}-\beta = (X'V^{-1}X)^{-1}X'V^{-1}\varepsilon$$

The variance of $\hat{\beta}$ is given by

$$\begin{aligned}Var(\hat{\beta}) &= E(\hat{\beta}-\beta)(\hat{\beta}-\beta)'\\ &= E\left[\{(X'V^{-1}X)^{-1}X'V^{-1}\varepsilon\}\{\varepsilon'V^{-1}X(X'V^{-1}X)^{-1}\}\right]\\ &= E\{(X'V^{-1}X)^{-1}X'V^{-1}\varepsilon\varepsilon'V^{-1}X(X'V^{-1}X)^{-1}\}\\ &= (X'V^{-1}X)^{-1}X'V^{-1}E(\varepsilon\varepsilon')V^{-1}X(X'V^{-1}X)^{-1}\\ &= (X'V^{-1}X)^{-1}X'V^{-1}VV^{-1}X(X'V^{-1}X)^{-1}\\ &= (X'V^{-1}X)^{-1}(X'V^{-1}X)(X'V^{-1}X)^{-1}\\ &= (X'V^{-1}X)^{-1} \qquad \ldots(5.9)\end{aligned}$$

The estimator $\hat{\beta}_{gls} = (X'V^{-1}X)^{-1}(X'V^{-1}Y)$ given at (5.7) is the Aitken's generalised least-squares estimator. If we apply OLS to $Y=X\beta+\varepsilon$ the resultant estimator would be unbiased linear but it would not be minimum variance unbiased linear estimator. The presence of assumption $E(\varepsilon\varepsilon')=V$ rather than the simple $E(\varepsilon\varepsilon')=\sigma^2 I_n$ destroys the minimum variance property of OLS.

5.3.2 Testing of Hypothesis

For testing of the hypothesis that GLS estimators are statistically significant we proceed in the same manner as for multiple regression in Unit 4. As you know, for hypothesis testing we need two values: value of the estimator to be tested, and its standard error. Given V, the generalised least squares (GLS) estimator can be computed from eq. (5.7) and standard error from eq. (5.9) so that the usual significance tests and confidence intervals for the β_i's can be constructed. Thus the test statistic can designed and its observed value can be compared with the tabulated value.

Alternatively, we can work in an analysis of variance framework by establishing the appropriate sum of squares from eq. (5.1).

Defining $\hat{\varepsilon} = Y - X\hat{\beta}$ and pre-multiplying by P^{-1}, we get

$$P^{-1}\hat{\varepsilon} = P^{-1}Y - P^{-1}X\hat{\beta}$$

$$= M - N\hat{\beta},$$

where

$$M = P^{-1}Y \text{ and } N = P^{-1}X$$

or

$$M = N\hat{\beta} + P^{-1}\hat{\varepsilon}$$

Now

$$M'M = (N\hat{\beta} + P^{-1}\hat{\varepsilon})'(N\hat{\beta} + P^{-1}\hat{\varepsilon})$$

$$= (\hat{\beta}'N' + \hat{\varepsilon}'P^{-1\prime})(N\hat{\beta} + P^{-1}\hat{\varepsilon})$$

$$= \hat{\beta}'N'N\hat{\beta} + \hat{\beta}'N'P^{-1}\hat{\varepsilon} + \hat{\varepsilon}'P^{-1\prime}N\hat{\beta} + \hat{\varepsilon}'P^{-1\prime}P^{-1}\hat{\varepsilon}$$

$$= \hat{\beta}'(P^{-1}X)'(P^{-1}X)\hat{\beta} + \hat{\beta}'(P^{-1}X)'(P^{-1}\hat{\varepsilon}) + \hat{\varepsilon}'P^{-1\prime}(P^{-1}X)\hat{\beta} + \hat{\varepsilon}'P^{-1\prime}P^{-1}\hat{\varepsilon}$$

$$= \hat{\beta}'(X'P^{-1\prime}P^{-1}X)\hat{\beta} + \hat{\beta}'(X'P^{-1\prime}P^{-1}\hat{\varepsilon}) + \hat{\varepsilon}'(P^{-1\prime}P^{-1}X)\hat{\beta} + \hat{\varepsilon}'(P^{-1\prime}P^{-1})\hat{\varepsilon}$$

$$= \hat{\beta}'(X'V^{-1}X)\hat{\beta} + \hat{\varepsilon}'V^{-1}\hat{\varepsilon},$$

since $X'\hat{\varepsilon} = \hat{\varepsilon}'X = 0$

or

$$(P^{-1}Y)'(P^{-1}Y) = \hat{\beta}'(X'V^{-1}X)\hat{\beta} + \hat{\varepsilon}'V^{-1}\hat{\varepsilon}$$

or

$$Y'P^{-1\prime}P^{-1}Y = \hat{\beta}'X'V^{-1}X\hat{\beta} + \hat{\varepsilon}'V^{-1}\hat{\varepsilon}$$

or

$$Y'V^{-1}Y = \hat{\beta}'X'V^{-1}Y + \hat{\varepsilon}'V^{-1}\hat{\varepsilon} \qquad \ldots(5.10)$$

Equation (5.10) can be viewed as a decomposition of the total sum of squares (TSS) into an regression sum of squares (RSS) and error (residual) sum of squares (ESS). Eq. (5.10) can serve as a basis for the usual analysis of variance table but since the analysis is derived from the transformed variables N and M we see that we are weighting sums of squares in the original variables Y and X by the matrix V^{-1}

5.4 SPECIAL CASES OF GLS

Let us now consider some special cases.

First, if

$$V = \sigma^2 I_n = E(\varepsilon\varepsilon'),$$

that is, if the generalised model reduces to the classical model, we have

$$V^{-1} = \frac{1}{\sigma^2} I_n$$

and Aitken's generalised estimator is the same as the ordinary least-squares estimator.

Second, if

$$V = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \sigma_n^2 \end{bmatrix}$$

that is, if the generalised model reduces to a purely heteroscedastic model, then

$$V^{-1} = \begin{bmatrix} \frac{1}{\sigma_1^2} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sigma_2^2} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \frac{1}{\sigma_n^2} \end{bmatrix}$$

and Aitken's generalised estimator is the same as the best linear unbiased estimator developed for the heteroscedastic model (See Unit 8).

Finally, if

$$V = \sigma^2 \begin{bmatrix} 1 & \rho & \cdots & \rho^{n-1} \\ \rho & 1 & \cdots & \rho^{n-2} \\ \cdots & \cdots & \cdots & \cdots \\ \rho^{n-1} & \rho^{n-2} & \cdots & 1 \end{bmatrix}$$

that is, if the error term follows a first-order autoregressive scheme, we obtain

$$V^{-1} = \frac{1}{\sigma^2(1-\rho^2)} \begin{bmatrix} 1 & -\rho & \cdots & 0 \\ -\rho & (1+\rho^2) & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

In this case Aitken's generalised estimator is the same as the best linear unbiased estimator (BLUE) derived for the regression model with autoregressive disturbances (see Unit 7).

Check Your Progress 1

1) Explain the procedure of obtaining GLS estimator.

.....

.....

.....

.....

2) Find the estimate for variance of the GLS estimator.

.....

.....

.....

.....

.....

5.5 LET US SUM UP

In this unit we discussed the method of generalised least squares which applies to several situations. In reality we observe that economic data used for estimation of multiple regression models do not fulfil some of the assumptions made under classical regression model. In order to obtain estimates which are unbiased as well as efficient we apply the method of generalised least squares.

There are several procedures of obtaining estimates for GLS models. We discussed an important method, that is, Aitken's method in the Unit. You will find that the GLS method provides efficient estimators in the presence of autocorrelation and heteroscedasticity. These two issues will be taken up for discussion in Units 7 and 8.

5.6 KEY WORDS

Autoregressive scheme : Here the error terms are related to one another. The first order autoregressive scheme is given by $\varepsilon_t = \rho\varepsilon_{t-1} + u_t$.

Generalised least squares : It is called generalised least squares as several models such as classical regression, heteroscedasticity and autocorrelation can be considered as its special cases.

Heteroscedasticity : A situation where the variance of the error terms changes within the model across observations.

Multicollinearity : A situation where the explanatory variables are not independent of one another.

5.7 SOME USEFUL BOOKS/REFERENCES

Johnston, Jack and John Dinardo, 1997, *Econometric Methods*, The McGraw-Hill Companies Inc., Singapore.

Pindyck, Robert S. and Daniel L. Rubinfeld,1998, *Econometric Models and Economic Forecasts*, Irwin/McGraw-Hill, Singapore.

5.8 ANSWERS/HINTS TO CHECK YOUR PROGRESS EXERCISES

Check Your Progress 1

1) See Section 5.3 and answer.

2) See Section 5.3 and answer.

TABLE A1: NORMAL AREA TABLE

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990

TABLE A2: CRITICAL VALUE OF T-DISTRIBUTION

df\p	0.40	0.25	0.10	0.05	0.025	0.01	0.005	0.0005
1	0.324920	1.000000	3.077684	6.313752	12.70620	31.82052	63.65674	636.6192
2	0.288675	0.816497	1.885618	2.919986	4.30265	6.96456	9.92484	31.5991
3	0.276671	0.764892	1.637744	2.353363	3.18245	4.54070	5.84091	12.9240
4	0.270722	0.740697	1.533206	2.131847	2.77645	3.74695	4.60409	8.6103
5	0.267181	0.726687	1.475884	2.015048	2.57058	3.36493	4.03214	6.8688
6	0.264835	0.717558	1.439756	1.943180	2.44691	3.14267	3.70743	5.9588
7	0.263167	0.711142	1.414924	1.894579	2.36462	2.99795	3.49948	5.4079
8	0.261921	0.706387	1.396815	1.859548	2.30600	2.89646	3.35539	5.0413
9	0.260955	0.702722	1.383029	1.833113	2.26216	2.82144	3.24984	4.7809
10	0.260185	0.699812	1.372184	1.812461	2.22814	2.76377	3.16927	4.5869
11	0.259556	0.697445	1.363430	1.795885	2.20099	2.71808	3.10581	4.4370
12	0.259033	0.695483	1.356217	1.782288	2.17881	2.68100	3.05454	4.3178
13	0.258591	0.693829	1.350171	1.770933	2.16037	2.65031	3.01228	4.2208
14	0.258213	0.692417	1.345030	1.761310	2.14479	2.62449	2.97684	4.1405
15	0.257885	0.691197	1.340606	1.753050	2.13145	2.60248	2.94671	4.0728
16	0.257599	0.690132	1.336757	1.745884	2.11991	2.58349	2.92078	4.0150
17	0.257347	0.689195	1.333379	1.739607	2.10982	2.56693	2.89823	3.9651
18	0.257123	0.688364	1.330391	1.734064	2.10092	2.55238	2.87844	3.9216
19	0.256923	0.687621	1.327728	1.729133	2.09302	2.53948	2.86093	3.8834
20	0.256743	0.686954	1.325341	1.724718	2.08596	2.52798	2.84534	3.8495
21	0.256580	0.686352	1.323188	1.720743	2.07961	2.51765	2.83136	3.8193
22	0.256432	0.685805	1.321237	1.717144	2.07387	2.50832	2.81876	3.7921
23	0.256297	0.685306	1.319460	1.713872	2.06866	2.49987	2.80734	3.7676
24	0.256173	0.684850	1.317836	1.710882	2.06390	2.49216	2.79694	3.7454
25	0.256060	0.684430	1.316345	1.708141	2.05954	2.48511	2.78744	3.7251
26	0.255955	0.684043	1.314972	1.705618	2.05553	2.47863	2.77871	3.7066
27	0.255858	0.683685	1.313703	1.703288	2.05183	2.47266	2.77068	3.6896
28	0.255768	0.683353	1.312527	1.701131	2.04841	2.46714	2.76326	3.6739
29	0.255684	0.683044	1.311434	1.699127	2.04523	2.46202	2.75639	3.6594
30	0.255605	0.682756	1.310415	1.697261	2.04227	2.45726	2.75000	3.6460
inf	0.253347	0.674490	1.281552	1.644854	1.95996	2.32635	2.57583	3.2905

TABLE A3: CRITICAL VALUES OF CHI-SQUARE DISTRIBUTION

df\area	0.1	0.05	0.025	0.01	0.005
1.00	2.71	3.84	5.02	6.63	7.88
2.00	4.61	5.99	7.38	9.21	10.60
3.00	6.25	7.81	9.35	11.34	12.84
4.00	7.78	9.49	11.14	13.28	14.86
5.00	9.24	11.07	12.83	15.09	16.75
6.00	10.64	12.59	14.45	16.81	18.55
7.00	12.02	14.07	16.01	18.48	20.28
8.00	13.36	15.51	17.53	20.09	21.95
9.00	14.68	16.92	19.02	21.67	23.59
10.00	15.99	18.31	20.48	23.21	25.19
11.00	17.28	19.68	21.92	24.72	26.76
12.00	18.55	21.03	23.34	26.22	28.30
13.00	19.81	22.36	24.74	27.69	29.82
14.00	21.06	23.68	26.12	29.14	31.32
15.00	22.31	25.00	27.49	30.58	32.80
16.00	23.54	26.30	28.85	32.00	34.27
17.00	24.77	27.59	30.19	33.41	35.72
18.00	25.99	28.87	31.53	34.81	37.16
19.00	27.20	30.14	32.85	36.19	38.58
20.00	28.41	31.41	34.17	37.57	40.00
21.00	29.62	32.67	35.48	38.93	41.40
22.00	30.81	33.92	36.78	40.29	42.80
23.00	32.01	35.17	38.08	41.64	44.18
24.00	33.20	36.42	39.36	42.98	45.56
25.00	34.38	37.65	40.65	44.31	46.93
26.00	35.56	38.89	41.92	45.64	48.29
27.00	36.74	40.11	43.19	46.96	49.64
28.00	37.92	41.34	44.46	48.28	50.99
29.00	39.09	42.56	45.72	49.59	52.34
30.00	40.26	43.77	46.98	50.89	53.67

TABLE A4: CRITICAL VALUE OF F-DISTRIBUTION

(Level of significance = 0.05)

df2/df1	1	2	3	4	5	6	7	8	9	10
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91
inf	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83

TABLE A4: CRITICAL VALUE OF F-DISTRIBUTION (CONTD.)

(Level of significance = 0.05)

df2/df1	12	15	20	24	30	40	60	120	INF
1	243.91	245.95	248.01	249.05	250.10	251.14	252.20	253.25	254.31
2	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
3	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.37
6	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
26	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
27	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
28	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
29	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
30	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
inf	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

TABLE A4: CRITICAL VALUE OF F-DISTRIBUTION (CONTD.)

(Level of significance = 0.01)

df2/df1	1	2	3	4	5	6	7	8	9	10
1	4052.2	4999.5	5403.4	5624.6	5763.7	5859.0	5928.4	5981.1	6022.5	6055.9
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05
6	13.75	10.93	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10
14	8.86	6.52	5.56	5.04	4.70	4.46	4.28	4.14	4.03	3.94
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.90	3.81
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69
17	8.40	6.11	5.19	4.67	4.34	4.10	3.93	3.79	3.68	3.59
18	8.29	6.01	5.09	4.58	4.25	4.02	3.84	3.71	3.60	3.51
19	8.19	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26
23	7.88	5.66	4.77	4.26	3.94	3.71	3.54	3.41	3.30	3.21
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17
25	7.77	5.57	4.68	4.18	3.86	3.63	3.46	3.32	3.22	3.13
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09
27	7.68	5.49	4.60	4.11	3.79	3.56	3.39	3.26	3.15	3.06
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03
29	7.60	5.42	4.54	4.05	3.73	3.50	3.33	3.20	3.09	3.01
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47
inf	6.64	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32

TABLE A4: CRITICAL VALUE OF F-DISTRIBUTION (CONTD.)

(Level of significance = 0.01)

df2/df1	12	15	20	24	30	40	60	120	INF
1	6106.3	6157.3	6208.7	6234.6	6260.6	6286.8	6313.0	6339.4	6365.9
2	99.42	99.43	99.45	99.46	99.47	99.47	99.48	99.49	99.50
3	27.05	26.87	26.69	26.60	26.51	26.41	26.32	26.22	26.13
4	14.37	14.20	14.02	13.93	13.84	13.75	13.65	13.56	13.46
5	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	9.02
6	7.72	7.56	7.40	7.31	7.23	7.14	7.06	6.97	6.88
7	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	5.65
8	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	4.86
9	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40	4.31
10	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4.00	3.91
11	4.40	4.25	4.10	4.02	3.94	3.86	3.78	3.69	3.60
12	4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.45	3.36
13	3.96	3.82	3.67	3.59	3.51	3.43	3.34	3.26	3.17
14	3.80	3.66	3.51	3.43	3.35	3.27	3.18	3.09	3.00
15	3.67	3.52	3.37	3.29	3.21	3.13	3.05	2.96	2.87
16	3.55	3.41	3.26	3.18	3.10	3.02	2.93	2.85	2.75
17	3.46	3.31	3.16	3.08	3.00	2.92	2.84	2.75	2.65
18	3.37	3.23	3.08	3.00	2.92	2.84	2.75	2.66	2.57
19	3.30	3.15	3.00	2.93	2.84	2.76	2.67	2.58	2.49
20	3.23	3.09	2.94	2.86	2.78	2.70	2.61	2.52	2.42
21	3.17	3.03	2.88	2.80	2.72	2.64	2.55	2.46	2.36
22	3.12	2.98	2.83	2.75	2.67	2.58	2.50	2.40	2.31
23	3.07	2.93	2.78	2.70	2.62	2.54	2.45	2.35	2.26
24	3.03	2.89	2.74	2.66	2.58	2.49	2.40	2.31	2.21
25	2.99	2.85	2.70	2.62	2.54	2.45	2.36	2.27	2.17
26	2.96	2.82	2.66	2.59	2.50	2.42	2.33	2.23	2.13
27	2.93	2.78	2.63	2.55	2.47	2.38	2.29	2.20	2.10
28	2.90	2.75	2.60	2.52	2.44	2.35	2.26	2.17	2.06
29	2.87	2.73	2.57	2.50	2.41	2.33	2.23	2.14	2.03
30	2.84	2.70	2.55	2.47	2.39	2.30	2.21	2.11	2.01
40	2.67	2.52	2.37	2.29	2.20	2.11	2.02	1.92	1.81
60	2.50	2.35	2.20	2.12	2.03	1.94	1.84	1.73	1.60
120	2.34	2.19	2.04	1.95	1.86	1.76	1.66	1.53	1.38
inf	2.19	2.04	1.88	1.79	1.70	1.59	1.47	1.33	1.00