2.5 PROPERTIES OF ESTIMATORS

To choose between estimating principles, we look into the properties satisfied by them. These properties are classified into two groups—small sample properties and large sample properties.[7]

Small Sample Properties

- *Unbiasedness*: An estimator $\hat{\beta}$ is said to be an unbiased estimator of β if its mean or expected value is equal to the value of true population parameter, β, i.e., $E(\hat{\beta}) = \beta$.[8] In everyday terms, this means that if repeated samples of a given size are drawn, and $\hat{\beta}$ computed for each sample, the average of such $\hat{\beta}$ values would be equal to β. However, if $E(\hat{\beta}) \neq \beta$ or $E(\hat{\beta}) - \beta \neq 0$, then $\hat{\beta}$ is said to be biased and the extent of bias for $\hat{\beta}$ is measured by $E(\hat{\beta}) - \beta$.
- *Minimum Variance or Bestness*: An estimator $\hat{\beta}$ is said to be a minimum variance or best estimator of β if its variance is less than the variance of any other estimator, say β^*. Thus, when $Var(\hat{\beta}) < Var(\beta^*)$, $\hat{\beta}$ is called the minimum variance or best estimator of β.[9]
- *Efficiency*: $\hat{\beta}$ is an efficient estimator if the following two conditions are satisfied together:

  (i) $\hat{\beta}$ is unbiased and

  (ii) $Var(\hat{\beta}) \leq Var(\beta^*)$

  An efficient estimator is also called as a minimum variance unbiased estimator (MVUE) or best-unbiased estimator.
- *Linearity*: An estimator is said to have the property of linearity if it is possible to express it as a linear combination of sample observations.[10]
- *Mean-Squared Error* (MSE): Sometimes a difficult choice problem arises while comparing two estimators. Suppose we have two different estimators of which one has lower bias, but higher variance, compared with the other. In other words, when

$$(bias\,\hat{\beta}) > (bias\,\beta^*)$$

[7] There is no hard and fast rule to distinguish between small and large samples. However, a working definition is that a small sample has 30 or less observations while the large sample has more than 30 observations.

[8] The unbiasedness property reflects on the *accuracy* of the estimator. Thus, when an estimator is unbiased, we say that it is able to provide an accurate estimate of its true population parameter.

[9] Minimum variance is associated with *reliability* dimension of the estimator. Obviously, the estimator that has lower variance is more reliable than the estimator which has higher variance.

[10] Put differently, linearity is associated with linear (i.e., additive) calculation rather than multiplicative or non-linear calculation.

but

$$Var(\hat{\beta}) < Var(\beta^*)$$

then how to choose between the two estimators? In this situation, where one estimator has a larger bias but a smaller variance than the other estimator, it is intuitively plausible to consider a trade-off between the two characteristics. This notion is given a precise, formal, expression in the mean-squared error.

The mean-squared error for $\hat{\beta}$ is defined as

$$\begin{aligned} MSE(\hat{\beta}) &= E[\hat{\beta}-\beta]^2 \\ &= E[\{\hat{\beta}-E(\hat{\beta})\}+\{E(\hat{\beta})-\beta\}]^2 \\ &= E\{\hat{\beta}-E(\hat{\beta})\}^2+E\{E(\hat{\beta})-\beta\}^2+2E[\{\hat{\beta}-E(\hat{\beta})\}\{E(\hat{\beta})-\beta\}] \\ &= Var(\hat{\beta})+(bias\hat{\beta})^2 \end{aligned}$$

This is so because the cross-product term vanishes, as shown below.

$$\begin{aligned} E[\{\hat{\beta}-E(\hat{\beta})\}\{E(\hat{\beta})-\beta\}] &= E\{\hat{\beta}E(\hat{\beta})-\hat{\beta}\beta-E(\hat{\beta})E(\hat{\beta})+E(\hat{\beta})\beta\} \\ &= E(\hat{\beta})E(\hat{\beta})-E(\hat{\beta})\beta-E(\hat{\beta})E(\hat{\beta})+E(\hat{\beta})\beta \\ &= 0 \end{aligned}$$

Now, according to the mean-squared error property, if $MSE(\hat{\beta}) < MSE(\beta^*)$, we say that $\hat{\beta}$ has lower mean-squared error, and accept it as an estimator of β.

Large Sample or Asymptotic Properties

These properties relate to the distribution of an estimator when the sample size is large, and approaches to infinity. The important properties here are the following.

- *Asymptotic Unbiasedness*: $\hat{\beta}$ is an asymptotically unbiased estimator of β if

$$\lim_{n\to\infty} E(\hat{\beta}) = \beta$$

  This means that the estimator $\hat{\beta}$, which is otherwise biased, becomes unbiased as the sample size approaches infinity. It is to be noted that if an estimator is unbiased, it is also asymptotically unbiased, but the reverse is not necessarily true.
- *Consistency*: Whether or not an estimator is consistent is understood by looking at the behaviour of its bias and variance as the sample size approaches infinity. If the increase in sample size reduces bias (if there were one) and variance of the estimator, and this

continues until both bias and variance become zero, as $n \to \infty$, then the estimator is said to be consistent. Thus, $\hat{\beta}$ is a consistent estimator if

$$\lim_{n \to \infty} [E(\hat{\beta}) - \beta] = 0$$

and

$$\lim_{n \to \infty} Var(\hat{\beta}) = 0$$

2.6 PROPERTIES OF OLS ESTIMATORS

The ordinary least-squares estimators are best, linear, and unbiased. In brief, we say that they are BLUE.[11] The BLUE properties for the least-squares estimators are proved below.

Recall the two-variable linear model

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

where ε_i satisfies the OLS assumptions:

$$\left.\begin{array}{l} E(\varepsilon_i) = 0 \\ E(\varepsilon_i^2) = \sigma_\varepsilon^2 \end{array}\right\} \text{for all } i$$

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

We know that

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

and

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

where

$$x_i = X_i - \bar{X}$$

and

$$y_i = Y_i - \bar{Y}$$

[11] BLUE—best, linear and unbiased estimator. The BLUE property of the OLS estimators is also known as the Gauss-Markov theorem.

Unbiasedness of $\hat{\beta}$

$$\begin{aligned}\hat{\beta} &= \frac{\sum x_i y_i}{\sum x_i^2} \\ &= \frac{\sum x_i (Y_i - \overline{Y})}{\sum x_i^2} \\ &= \frac{\sum x_i Y_i - \overline{Y}\sum x_i}{\sum x_i^2} \\ &= \frac{\sum x_i Y_i}{\sum x_i^2} \qquad [\because \sum x_i = \sum (X_i - \overline{X}) = 0] \\ &= \sum w_i Y_i \end{aligned} \tag{2.12}$$

where

$$w_i = \frac{x_i}{\sum x_i^2} \tag{2.13}$$

It follows from (2.13) that

$$\sum w_i = \frac{\sum x_i}{\sum x_i^2} = 0 \tag{2.14}$$

$$\begin{aligned}\sum w_i X_i &= \frac{\sum x_i X_i}{\sum x_i^2} \\ &= \frac{\sum (X_i - \overline{X}) X_i}{\sum (X_i - \overline{X})^2} \\ &= \frac{\sum X_i^2 - \overline{X}\sum X_i}{\sum X_i^2 - 2\overline{X}\sum X_i + n\overline{X}^2} \\ &= \frac{\sum X_i^2 - n\overline{X}^2}{\sum X_i^2 - 2n\overline{X}^2 + n\overline{X}^2} \\ &= \frac{\sum X_i^2 - n\overline{X}^2}{\sum X_i^2 - n\overline{X}^2} \\ &= 1 \end{aligned} \tag{2.15a}$$

$$\sum w_i^2 = \frac{\sum x_i^2}{(\sum x_i^2)^2} = \frac{1}{\sum x_i^2} \tag{2.15b}$$

From (2.12), we have

$$\hat{\beta}=\sum w_i Y_i$$
$$=\sum w_i(\alpha+\beta X_i+\varepsilon_i)$$
$$=\alpha\sum w_i+\beta\sum w_i X_i+\sum w_i\varepsilon_i$$
$$=\beta+\sum w_i\varepsilon_i \qquad (\because \sum w_i=0 \text{ and } \sum w_i X_i=1) \tag{2.16}$$

Taking expectations,

$$E(\hat{\beta})=E(\beta+\sum w_i\varepsilon_i)$$
$$=\beta+\sum w_i E(\varepsilon_i)$$
$$=\beta \qquad [\because E(\varepsilon_i)=0]$$

This proves that $\hat{\beta}$ is unbiased.

Linearity of $\hat{\beta}$

From (2.12),

$$\hat{\beta}=\sum w_i Y_i$$

Since w_i's are a set of fixed values, we may write

$$\hat{\beta}=w_1Y_1+w_2Y_2+....+w_nY_n$$

This shows that $\hat{\beta}$ is a linear combination of sample values of Y_i, the dependent variable. Thus, $\hat{\beta}$ has the property of linearity.

Minimum Variance or Bestness for $\hat{\beta}$

In order to prove minimum variance or bestness property for $\hat{\beta}$, we shall compute the variance of $\hat{\beta}$ and show that it is lower than the variance of some other estimator.

From (2.16), we have

$$\hat{\beta}=\beta+\sum w_i\varepsilon_i$$
$$\Rightarrow \hat{\beta}-\beta=\sum w_i\varepsilon_i$$
$$\Rightarrow \hat{\beta}-E(\hat{\beta})=\sum w_i\varepsilon_i \qquad [\because E(\hat{\beta})=\beta]$$

Thus,

$$Var(\hat{\beta})=E[\hat{\beta}-E(\hat{\beta})]^2$$
$$=E\left(\sum w_i\varepsilon_i\right)^2$$

$$
\begin{aligned}
&= E\left(\sum w_i^2 \varepsilon_i^2 + 2\sum_{i<j} w_i w_j \varepsilon_i \varepsilon_j\right) \\
&= \sum w_i^2 E(\varepsilon_i^2) + 2\sum_{i<j} w_i w_j E(\varepsilon_i \varepsilon_j) \\
&= \sigma^2 \sum w_i^2 \qquad [\because E(\varepsilon_i^2) = \sigma^2 \text{ and } E(\varepsilon_i \varepsilon_j) = 0 \text{ for } i \neq j] \\
&= \frac{\sigma^2}{\sum x_i^2} \qquad \left(\because \sum w_i^2 = \frac{1}{\sum x_i^2}\right) \qquad (2.17)
\end{aligned}
$$

Let us now consider some other estimator, say β^*, such that

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

where c_i $(i = 1, 2, \ldots, n)$ represents a set of weights

Then,

$$
\begin{aligned}
\beta^* &= \sum c_i(\alpha + \beta X_i + \varepsilon_i) \qquad (2.18) \\
&= \alpha \sum c_i + \beta \sum c_i X_i + \sum c_i \varepsilon_i
\end{aligned}
$$

Taking expectations,

$$E(\beta^*) = \alpha \sum c_i + \beta \sum c_i X_i \quad [\because E(\varepsilon_i) = 0]$$

It is clear that we require the weights to be such that β^* is an unbiased estimator. Th imposes the conditions

$$\sum c_i = 0 \text{ and } \sum c_i X_i = 1 \Rightarrow \sum c_i x_i = 1 \qquad (2.1$$

Let us now compute $Var(\beta^*)$ accepting these conditions. Under these conditi equation (2.18) reduces to

$$\beta^* = \beta + \sum c_i \varepsilon_i \Rightarrow (\beta^* - \beta) = (\beta^* - E(\beta^*)) = \sum c_i \varepsilon_i$$

Thus,

$$
\begin{aligned}
Var(\beta^*) &= E[\beta^* - E(\beta^*)]^2 \\
&= E\left(\sum c_i \varepsilon_i\right)^2 \\
&= E\left(\sum c_i^2 \varepsilon_i^2 + 2\sum_{i<j} c_i c_j \varepsilon_i \varepsilon_j\right) \\
&= \sum c_i^2 E(\varepsilon_i^2) + 2\sum_{i<j} c_i c_j E(\varepsilon_i \varepsilon_j) \\
&= \sigma^2 \sum c_i^2 \qquad [\because E(\varepsilon_i^2) = \sigma^2 \text{ and } E(\varepsilon_i \varepsilon_j) = 0 \text{ for } i \neq j]
\end{aligned}
$$

To compare $Var(\hat{\beta})$ with $Var(\beta^*)$, consider the expression

$$c_i = w_i + (c_i - w_i)$$
$$\Rightarrow \sum c_i^2 = \sum w_i^2 + \sum (c_i - w_i)^2 + 2\sum w_i (c_i - w_i) \qquad (2.21)$$

Note that

$$\begin{aligned}
&\sum w_i (c_i - w_i) \\
&= \sum w_i c_i - \sum w_i^2 \\
&= \frac{\sum c_i x_i}{\sum x_i^2} - \frac{1}{\sum x_i^2} \qquad \left(\because w_i = \frac{x_i}{\sum x_i^2}\right) \\
&= \frac{1}{\sum x_i^2} - \frac{1}{\sum x_i^2} \qquad \left[\because \sum c_i x_i = 1, \text{as shown in } (2.19) \text{ above}\right] \\
&= 0
\end{aligned}$$

Thus,

$$\begin{aligned}
Var(\beta^*) &= \sigma^2 [\sum w_i^2 + \sum (c_i - w_i)^2] \\
&= \frac{\sigma^2}{\sum x_i^2} + \sigma^2 \sum (c_i - w_i)^2 \qquad \left(\because \sum w_i^2 = \frac{1}{\sum x_i^2}\right) \\
&= Var(\hat{\beta}) + \sigma^2 \sum (c_i - w_i)^2
\end{aligned}$$

Since $\sum (c_i - w_i)^2 > 0$ unless $c_i = w_i$ for all i, $Var(\hat{\beta}) < Var(\beta^*)$, and we conclude that $\hat{\beta}$ is a minimum variance or best estimator.

2.7 STATISTICAL INFERENCE IN SLRM

After estimating the population parameters (α and β) of our regression model, our next task is to examine statistical significance of the estimated coefficients ($\hat{\alpha}$ and $\hat{\beta}$) by applying our knowledge of statistical inference.[12] Examination of statistical significance of the estimated coefficients specifically requires the knowledge about their sampling distributions. In this regard, it may be noted that

$$\hat{\alpha} \sim N\left[\alpha, \left(\frac{1}{n} + \frac{\bar{X}^2}{\sum x_i^2}\right)\right] \qquad (2.22)$$

[12] In layman's language, testing statistical significance of an estimated coefficient (say $\hat{\beta}$) enables us to understand the usefulness of X_i as a predictor of Y_i.

$$\hat{\beta} \sim N\left[\beta, \frac{\sigma^2}{\sum x_i^2}\right] \tag{2.23}$$

Expression (2.22) states that $\hat{\alpha}$ has a normal distribution with mean equal to α and variance $\frac{1}{n}+\frac{\overline{X}^2}{\sum x_i^2}$. Similarly, (2.23) states that $\hat{\beta}$ also follows a normal distribution with mean equal to β and variance $\sigma^2/\sum x_i^2$. However, these results are useful when the variance of the disturbance term (σ^2) is known. Unfortunately, in practice, σ^2 is not known and has to be estimated as

$$\hat{\sigma}^2 = \frac{RSS}{n-2} = \frac{\sum e_i^2}{n-2} \tag{2.24}$$

where $\hat{\sigma}^2$ is the estimate of σ^2.[13]

Hypothesis Testing

We formalize the object of testing statistical significance of $\hat{\beta}$ (also $\hat{\alpha}$) by stating that we want to test the validity of the null hypothesis (H_N)[14] that the value of true population parameter β is zero against the alternative hypothesis (H_A)[15] that it is different from zero. In the present context, we set our hypotheses as

$H_N: \beta = 0$
$H_A: \beta \neq 0$ (under two-tailed test)

However, if we have any prior knowledge about the sign of β (say positive), then the hypotheses are set as

$H_N: \beta = 0$
$H_A: \beta > 0$ (one-tailed test)

[13] Note that $\sqrt{\hat{\sigma}^2} = \hat{\sigma}$ is called the *standard error of regression*. It provides an estimate of standard deviation of the regression error (or disturbance term) ε_i.

[14] In simple language, null hypothesis is what we are going to test.

[15] The alternative hypothesis represents our conclusion if the experimental test indicates that the null hypothesis is false.

2.10 SOME IMPORTANT RELATIONS IN THE CONTEXT OF SLRM

Relation between Regression Slope and Correlation Coefficient

There is a relation between the regression slope ($\hat{\beta}$) and correlation coefficient (r) between X_i and Y_i which is demonstrated as follows.

$$r = \frac{\Sigma x_i y_i}{\sqrt{\Sigma x_i^2}\sqrt{\Sigma y_i^2}} \qquad (x_i = X_i - \overline{X} \text{ and } y_i = Y_i - \overline{Y})$$

$$= \frac{\Sigma x_i y_i}{\Sigma x_i^2}\frac{\sqrt{\Sigma x_i^2}}{\sqrt{\Sigma y_i^2}}$$

$$= \frac{\Sigma x_i y_i}{\Sigma x_i^2}\frac{\sqrt{\Sigma x_i^2/n}}{\sqrt{\Sigma y_i^2/n}}$$

$$= \hat{\beta}\frac{S_x}{S_y} \qquad (S_x \text{ and } S_y \text{ are standard deviations of } X_i \text{ and } Y_i \text{ respectively})$$

Thus, the relation between the regression slope and correlation coefficient is given by

$$\hat{\beta} = r\frac{S_y}{S_x} \tag{2.28}$$

Relation between *F*-statistic and r^2

We know that

$$TSS = ESS + RSS$$

$$TSS = \Sigma y_i^2$$

$$ESS = \hat{\beta}\Sigma x_i y_i = r^2 \Sigma y_i^2 \qquad \left(\because r^2 = \hat{\beta}\frac{\Sigma x_i y_i}{\Sigma y_i^2}\right)$$

$$RSS = TSS - ESS = \Sigma y_i^2 - r^2\Sigma y_i^2 = (1 - r^2)\Sigma y_i^2$$

3 The Multiple Linear Regression Model

This chapter extends the discussion of the previous chapter. It is concerned with issues relevant to multiple regression analysis. Specifically, we discuss specification and assumptions of multiple regression model, its estimation, goodness of fit measures, and various problems of inference in the context of multiple regression models. We have added a brief discussion on the LR, Wald, and LM tests which are nowadays widely applied to handle a variety of inference and other problems in multiple regressions. Empirical applications of these tools and techniques have been explained using data set and EViews software package.

3.1 DEFINITION

Sometimes the two-variable regression model may appear to be inadequate as one independent/explanatory variable alone may not adequately explain variation in the dependent variable. In other words, it may appear that there are more than one determinants of the dependent variable. Thus, when we consider more than one determinants or independent variables, it becomes the case of multiple regression models. For instance, if we hypothesise that monthly consumption expenditure of the people is determined by their income, age, education, sex, etc., we have to specify a multiple regression model. In brief, a multiple regression model is the one where two or more independent variables are considered to explain variation in the dependent variable.[1]

Obviously, the easiest example of a multiple regression model is where only two independent variables or regressors are considered. In this chapter, we consider such a model while in the appendix to this chapter we present the multiple regression model involving more than two independent variables.

[1] Geweke et al. (2008, 610) observed that R. Benini, the Italian statistician, was the first to make use of the method of multiple regression in economics in the decade beginning 1900. However, Henry Moore was the first to place the statistical estimation of economic relations at the centre of quantitative analysis of economics in the 1910s. Moore is also credited for laying the foundation of 'statistical economics', the precursor of econometrics.

3.2 SPECIFICATION AND ASSUMPTIONS

The three-variable population regression model involving the dependent variable Y_i and independent/explanatory variables X_{1i} and X_{2i} is specified as:

$$Y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i \tag{3.1}$$

Here ε_i is the stochastic disturbance term and the subscript i denotes the i^{th} observation. As in the case of two-variable model, we make the following assumptions in context of the above multiple regression model.

(i) Zero mean of ε_i: $E(\varepsilon_i | X_{1i}, X_{2i}) = 0$ for each i
(ii) Homoskedasticity: $Var(\varepsilon_i) = \sigma^2$ constant
(iii) Non-autocorrelation: $Cov(\varepsilon_i, \varepsilon_j) = 0$ where $i \neq j$
(iv) Normality: ε_i is normally distributed.
(v) Non-stochastic Xs, which implies that the values of the X-variables are same in repeated samples.
(vi) Zero covariance between ε_i and X variables, i.e., $Cov(\varepsilon_i, X_{1i}) = Cov(\varepsilon_i, X_{2i}) = 0$.
(vii) No exact linear relationship exists between the X variables, i.e., Xs are not correlated.

3.3 OLS ESTIMATION

To obtain the OLS estimates of parameters of the population regression model (3.1), let us write the corresponding sample regression model as

$$Y_i = \hat{\alpha} + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + e_i \tag{3.2}$$

where e_i represents estimated residual values, and $\hat{\alpha}$, $\hat{\beta}_1$, and $\hat{\beta}_2$ are estimates of population parameters α, β_1, and β_2, respectively. As in the two-variable model, we apply the 'least-squares criterion' to obtain these estimates. Following this criterion, we select the values of $\hat{\alpha}$, $\hat{\beta}_1$, and $\hat{\beta}_2$ which minimize $\sum e_i^2$.

Here

$$\sum e_i^2 = \sum (Y_i - \hat{\alpha} - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i})^2$$

The necessary conditions of minimization of $\sum e_i^2$ are

$$\frac{\partial \sum e_i^2}{\partial \hat{\alpha}} = \frac{\partial \sum e_i^2}{\partial \hat{\beta}_1} = \frac{\partial \sum e_i^2}{\partial \hat{\beta}_2} = 0$$

Applying these conditions, the following 'normal equations' are obtained.

$$\sum Y_i = n\hat{\alpha} + \hat{\beta}_1 \sum X_{1i} + \hat{\beta}_2 \sum X_{2i} \tag{3.3a}$$

$$\sum X_{1i}Y_i = \hat{\alpha}\sum X_{1i} + \hat{\beta}_1\sum X_{1i}^2 + \hat{\beta}_2\sum X_{1i}X_{2i} \quad (3.3b)$$

$$\sum X_{2i}Y_i = \hat{\alpha}\sum X_{2i} + \hat{\beta}_1\sum X_{1i}X_{2i} + \hat{\beta}_2\sum X_{2i}^2 \quad (3.3c)$$

It is clear that with the given data on Y_i, X_{1i}, and X_{2i}, we have to compute the following quantities to obtain the values of the estimates.

$$n, \sum Y_i, \sum X_{1i}, \sum X_{2i}, \sum X_{1i}Y_i, \sum X_{2i}Y_i, \sum X_{1i}X_{2i}, \sum X_{1i}^2, \text{ and } \sum X_{2i}^2$$

Putting these values in the aforementioned 'normal equations' and solving, we have solutions for values of $\hat{\alpha}$, $\hat{\beta}_1$, and $\hat{\beta}_2$. Using these values, we write the estimated three-variable multiple regression model as

$$\hat{Y}_i = \hat{\alpha} + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i}$$

An alternative way to compute the estimates is to use the following formulas that can be derived by solving the 'normal equations'.

$$\hat{\alpha} = \bar{Y} - \hat{\beta}_1\bar{X}_1 - \hat{\beta}_2\bar{X}_2 \quad (3.4)$$

$$\hat{\beta}_1 = \frac{\sum x_{1i}y_i \sum x_{2i}^2 - \sum x_{2i}y_i \sum x_{1i}x_{2i}}{\sum x_{1i}^2 \sum x_{2i}^2 - (\sum x_{1i}x_{2i})^2} \quad (3.5)$$

$$\hat{\beta}_2 = \frac{\sum x_{2i}y_i \sum x_{1i}^2 - \sum x_{1i}y_i \sum x_{1i}x_{2i}}{\sum x_{1i}^2 \sum x_{2i}^2 - (\sum x_{1i}x_{2i})^2} \quad (3.6)$$

Here $\bar{Y}, \bar{X}_1$, and $\bar{X}_2$ denote sample mean values for the three variables and the lowercase letters denote deviation from these sample means.

It is also easy to compute the variances of $\hat{\alpha}$, $\hat{\beta}_1$, and $\hat{\beta}_2$ by using the following formulas.

$$Var(\hat{\alpha}) = \left[\frac{1}{n} + \frac{\bar{X}_1^2\sum x_{2i}^2 + \bar{X}_2^2\sum x_{1i}^2 - 2\bar{X}_1\bar{X}_2\sum x_{1i}x_{2i}}{\sum x_{1i}^2\sum x_{2i}^2 - (\sum x_{1i}x_{2i})^2}\right]\sigma^2 \quad (3.7)$$

$$Var(\hat{\beta}_1) = \left[\frac{\sum x_{2i}^2}{(\sum x_{1i}^2)(\sum x_{2i}^2) - (\sum x_{1i}x_{2i})^2}\right]\sigma^2 = \frac{\sigma^2}{\sum x_{1i}^2(1 - r_{12}^2)} \quad (3.8)$$

$$Var(\hat{\beta}_2) = \left[\frac{\sum x_{1i}^2}{(\sum x_{1i}^2)(\sum x_{2i}^2) - (\sum x_{1i}x_{2i})^2}\right]\sigma^2 = \frac{\sigma^2}{\sum x_{2i}^2(1 - r_{12}^2)} \quad (3.9)$$

In the above formulae, r_{12} is the sample coefficient of correlation between X_{1i} and X_{2i} or variance of the disturbance term ε_i which is estimated as

$$\hat{\sigma}^2 = \frac{\sum e_i^2}{n-3} \quad (3.10)$$

where 3 is the number of parameters in the population regression equation estimated in the model.

3.4 PROPERTIES OF OLS ESTIMATORS

As in two-variable linear model, the least-squares estimators ($\hat{\alpha}$, $\hat{\beta}_1$, and $\hat{\beta}_2$) in the three-variable model are also BLUE, i.e., best, linear and unbiased estimators of population parameters (α, β_1, and β_2). It is easy to prove these properties. However, we skip this exercise here as Appendix 3.1 provides proof of BLUE properties in context of the general linear multiple regression model.

3.5 MEASURING GOODNESS OF FIT

After estimating the multiple regression model, we may be interested to assess the goodness or quality of fit of our estimated model. In other words, our objective is to know how well the estimated line fits the sample observations.

The goodness of fit of the estimated model in the context of a two-variable model is understood in terms of the value of r^2-statistic. To recapitulate, r^2-statistic provides a measure of proportion of total variation in the dependent variable that is explained by the independent/explanatory variable of the model. We can extend this concept further to obtain a measure of goodness of fit of the estimated model in the context of estimated multiple regression model. This is done as follows.

In the two-variable model,

$$r^2 = \frac{(\sum x_i y_i)^2}{\sum x_i^2 \sum y_i^2}$$

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

$$= \frac{ESS}{TSS}$$

Let us rewrite the above relation supposing that the variables considered are Y_i and X_{1i}. Then,

$$r^2 = \frac{\hat{\beta}_1 \sum x_{1i} y_i}{\sum y_i^2}$$

Now if we suppose that there are two explanatory variables, X_{1i} and X_{2i}, then

$$R^2 = \frac{\hat{\beta}_1 \sum x_{1i} y_i + \hat{\beta}_2 \sum x_{2i} y_i}{\sum y_i^2}$$

The above formula can be extended further by adding terms in the numerator, when we have more than two explanatory variables. If we have k number of explanatory variables in the model, then the R^2 formula becomes

$$R^2 = \frac{\hat{\beta}_1 \sum x_{1i} y_i + \hat{\beta}_2 \sum x_{2i} y_i + \dots + \hat{\beta}_k \sum x_{ki} y_i}{\sum y_i^2} \tag{3.11}$$

Usefulness of R^2-statistic

R^2-statistic (in brief, R^2) provides a measure of goodness of fit of the estimated multiple regression model to sample data. It also helps to understand the relevance of explanatory variables in the estimated model. The value of R^2 lies between 0 and 1. When the value of R^2 is close to 0, the explanatory variables have not explained much of the variation in the dependent variable of the model and we have a 'bad fit' estimated equation. In other words, we have not considered the explanatory variables that are relevant to explain variation in the dependent variable. On the other hand, when the value of R^2 is high and close to 1, we have a 'good fit' estimated equation, which explains a large part of variation in the dependent variable and the explanatory variables considered in the model are quite relevant.

Misuse of R^2-statistic

In spite of above-mentioned usefulness of the R^2-statistic, one must be cautious about its possible misuses. In particular, it is to be remembered that it is dangerous to play the game of maximizing the value of R^2. Some researchers do this by gradually increasing the number of explanatory variables in the model. However, in empirical research, quite often we come across a situation where the value of R^2 is high but very few of the estimated coefficients are statistically significant and/or they have expected signs. Therefore, the researchers should be more concerned about the logical/theoretical relevance of the explanatory variables to the dependent variable and also their statistical significance. If in this process, a high value of R^2 is obtained, well and good. On the other hand, if R^2 is low, it does not mean that the model is necessarily bad, particularly when a good number of the estimated coefficients have expected signs and are statistically significant.

To illustrate the above point further, let us consider an interesting example given by Rao and Miller (1972, 14–16) which clarifies the difficulty of choosing between two different models solely on the basis of their computed R^2 values. Rao and Miller estimated both the consumption and savings functions using the same time series data on consumption (C_t) and income (Y_t).[2] The estimated equations were

Consumption function:

$$\hat{C}_t = -0.34 + 0.76Y_t + 0.30Y_{t-1} \qquad R^2 = 0.99$$

Savings function:

$$\hat{S}_t = 0.34 + 0.24Y_t - 0.30Y_{t-1} \qquad R^2 = 0.64$$

It is found that estimated consumption model explains 99% of total variation in the dependent variable while the savings model explains 64% of the same. On the basis of this result, it would be incorrect to conclude that the consumption model provides a better causal relation. The reason for this is as follows.

We know that

$$R^2 = 1 - \frac{RSS}{TSS} = 1 - \frac{\sum e_t^2}{\sum y_t^2}$$

In the case of consumption model,

$$R_c^2 = 1 - \frac{\sum e_t^2}{\sum c_t^2} \text{ where } c_t = C_t - \bar{C}.$$

For the savings model,

$$R_s^2 = 1 - \frac{\sum e_t^2}{\sum s_t^2} \text{ where } s_t = S_t - \bar{S}.$$

Now, since marginal propensity to consume (MPC) is greater than the marginal propensity to save (MPS),[3] $\sum c_t^2 > \sum s_t^2 \Rightarrow R_c^2 > R_s^2$. Thus, as long as MPC $\neq$ MPS, the two R^2 values ought to differ. So, it would not be appropriate to conclude that the estimated consumption model with higher R^2 provides a better-fit estimated equation compared with the estimated savings model.

[2] The data series on savings (S_t) has been generated by taking a difference of income and consumption.

[3] This is shown by the values of estimated coefficients for income variable (Y_t) in the consumption and savings models, which are 0.76 and 0.24, respectively.