Chapter 3
Theory of Single and Multiple Regressions
Theory of Single and Multiple Regressions

Regression analysis attempts to establish the nature of the relationship between dependent variable and explanatory or independent variables and thereby provide a basis for prediction, or forecasting of the value of dependent variable. The variable, which is used to predict the variable of interest, is called the independent variable or explanatory variable and the variable we are trying to predict is called the dependent variable or explained variable. While doing regression analysis we are mainly concerned with the statistical and not deterministic or functional relationship between variables. Once such a relationship is arrived at, it can be utilized to predict the values of the dependent variable.

Regression analysis relevant to the present study is presented in the following sections and is based on the standard theoretical treatment. (Bethea et al., 1985; Gunst and Manson, 1980; Bhattacharya, et al., 1977).

[3.1] Simple Regression Model

The regression measures the cause and effect relationship where the independent variables cause the dependent variable to change. The most simple regression model is the linear regression, which involves two variables i.e. one dependent and other independent.

\[ Y_i = \beta_1 + \beta_2 X_i + \varepsilon_i \]  \hspace{1cm} (3.1)

Where \( Y_i \) is a random variable, \( X_i \) is fixed or non-stochastic and \( \varepsilon_i \) is a stochastic error term where values are based on an underlying probability distribution. \( \beta_1 \) and \( \beta_2 \) are constant and regression coefficients respectively. Because the regression model is not a mathematical model, which gives the exact relation between
dependent and independent variable. therefore to make this relationship stochastic, the random variable \( \varepsilon_i \) is included. \( \varepsilon_i \) contains all the variables from the model which also influence the dependent variable directly or indirectly. To get the efficient estimates the stochastic regression model has to satisfy some assumptions on \( \varepsilon_i \) and the explanatory variables as well. The important assumptions made to specify fully the two variables linear regression model are:

1. The relationship between \( Y \) and \( X \) is linear: The assumption of linearity means that it does not consider the quadratic or other nonlinear form of relationship.

2. The \( X_i \)'s are non stochastic variables whose values are fixed.

3. (a) The error term has zero expected value and constant variance for all observations i.e., \( E(\varepsilon_i) = 0 \) and \( E(\varepsilon_i^2) = \sigma^2 \)

The expected value of random term (\( \varepsilon_i \)) will be zero because it is assumed that out of the excluded variable from the model some will give negative impact while other will give positive impact on dependent variable so that on an average the influence will be cancel out while the variance i.e. assumed to be constant means that there is no tendency to expand or contract in the variance with respect to observation. In other words, the variance is assumed to be homoscedestistic.

(b) The random variable \( \varepsilon_i \) are statistically independent-

\[
E(\varepsilon_i \varepsilon_j) = 0 \text{ for } i \neq j
\]

This assumption implies that the successive values of the random term (\( \varepsilon_i \)) will be independent to each other. If there is any relationship between/among successive values of term then it can influence the estimates of regression parameters.
(c) The error term is normally distributed.

The values of $\beta_1$ and $\beta_2$ are obtained by the method of least square which states that the line should be drawn through the plotted points $(Y_i, V_s, X_i)$ in such a manner that the sum of the squares of the deviation of the actual $y$ value from the computed $y$ value is the least, or in other words in order to obtain a line which fits and data points best i.e., error sum of squares (ESS) should be minimum.

$$\text{ESS} = \sum_{i=1}^{N} (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^{N} \varepsilon_i^2$$ \quad (3.2)

Where $\varepsilon_i = Y_i - \hat{Y}_i$. $Y_i$ and $\varepsilon_i$ are the estimate values of dependent variable and error, respectively.

When the simple regression model is expanded to include more explanatory variables, the resultant model is called linear multiple regression model. In the following, we present a detailed analysis of multiple regression theory.

[3.2] Multiple Regression Model

We consider a model consisting of a dependent variable $Y$ and $K$ independent variables ($X$'s) which include the constant term. Assuming that there are $N$ observations, we can write the regression model in the form of $N$ equations.

$$Y_1 = \beta_1 + \beta_2 X_{21} + \beta_3 X_{31} + \beta_4 X_{41} + \ldots + \beta_K X_{K1} + \varepsilon_1$$

$$Y_2 = \beta_1 + \beta_2 X_{22} + \beta_3 X_{32} + \beta_4 X_{42} + \ldots + \beta_K X_{K2} + \varepsilon_2$$

$$Y_3 = \beta_1 + \beta_2 X_{23} + \beta_3 X_{33} + \beta_4 X_{43} + \ldots + \beta_K X_{K3} + \varepsilon_3$$
In matrix form, the above equation can be written as

\[
Y_N = \beta_1 + \beta_2 X_{2N} + \beta_3 X_{3N} + \beta_4 X_{4N} + \cdots + \beta_K X_{KN} + \varepsilon_N
\]

In matrix form, the above equation can be written as

\[
Y = \beta X + \varepsilon \quad \text{(3.3)}
\]

Where

\[
Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_N \end{bmatrix}, \quad X = \begin{bmatrix} 1 & X_{21} & \cdots & X_{1K1} \\ 1 & X_{22} & \cdots & X_{1K2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{2N} & \cdots & X_{1KN} \end{bmatrix}
\]

\[
\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_N \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{bmatrix}
\]

Where \( Y = N \times 1 \) Column vector of observations on the dependent variable \( Y \).

\( X = N \times K \) Matrix of independent variable observations.

\( \beta = K \times 1 \) Column vector of regression parameters.

\( \varepsilon = N \times 1 \) Column vector of errors.

In case of multiple regression model, there is one more assumption that model has to follow to give efficient estimates. The assumption is-

\[
\text{Rank of } X = K < N
\]
It means that the independent variables are not strongly correlated to each other and the numbers of parameters to be estimated are less than the number of observations. We can write the vector $Y$ as the sum of its predicted values $\hat{Y} = X\hat{\beta}$ and the residual vector $\hat{\varepsilon}$

$$Y = \hat{\beta} X + \hat{\varepsilon} \quad (3.4)$$

Where $\hat{\beta}$ and $\hat{\varepsilon}$ are matrices of estimated regression coefficient and error term respectively.

The total sum of squares

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

Since $X'\hat{\varepsilon} = 0$ and $\hat{\varepsilon}'X = 0$

$$Y'Y = \hat{\beta}'X'X\hat{\beta} + \hat{\varepsilon}'\hat{\varepsilon} \quad (3.5)$$

or

$$TSS = RSS + ESS \quad (3.6)$$

The first term of R.H.S. of Eq (3.5) $\hat{\beta}'X'X\hat{\beta}$ is the regression sum of squares (RSS) which represents explained variation of $Y$. The second term on R.H.S. of Eq (3.5), $\hat{\varepsilon}'\hat{\varepsilon}$ denotes the error sum of squares (ESS) which represents the unexplained variation of $Y$.

$X' = \text{Transpose of matrix } X$

To determine the least squares estimator, we minimize ESS. Writing ESS as

$$\hat{\varepsilon}'\hat{\varepsilon} = (Y - \hat{\beta}'X)(Y - \hat{\beta}X) = Y'Y - 2\hat{\beta}'X'Y + \hat{\beta}'X'X\hat{\beta} \quad (3.7)$$
and equating the first derivative of ESS w.r.t. $\beta$ to zero, we obtain,

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

$$\hat{\beta} = (X'X)^{-1}(X'Y) \quad (3.8)$$

The matrix $(X'X)$ is called the cross product matrix. Having estimated matrix $\hat{\beta}$, we now determine variance of $\hat{\beta}$ and denote it by vector $V \equiv \text{Var}(\hat{\beta})$.

We can then write $V$ as

$$V = E\left[ (\hat{\beta} - \beta)(\hat{\beta} - \beta)' \right]$$

$$= \left[ \frac{E(\hat{\beta}_k - \beta_k)^2}{E(\hat{\beta}_k - \beta_k)(\hat{\beta}_l - \beta_l)} \right] - \frac{E(\hat{\beta}_k - \beta_k)^2}{E(\hat{\beta}_k - \beta_k) - \beta_k}$$

$$= \left[ \frac{\text{Var}(\hat{\beta}_k)}{\text{Cov}(\hat{\beta}_k, \hat{\beta}_l)} \right] - \frac{\text{Cov}(\hat{\beta}_k, \hat{\beta}_l)}{\text{Var}(\hat{\beta}_k)} \quad (3.9)$$

The diagonal elements of $V$ represent the variance of the estimated parameters, while the off-diagonal terms represent the covariances.

$$\text{Var}(\hat{\beta}) = E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'] = E[(A\epsilon)(A\epsilon)']$$

$$= E(A\epsilon\epsilon' A')$$

Chapter - 3

The matrix $(X'X)$ is called the cross product matrix. Having estimated matrix $\hat{\beta}$, we now determine variance of $\hat{\beta}$ and denote it by vector $V \equiv \text{Var}(\hat{\beta})$.

We can then write $V$ as

$$V = E\left[ (\hat{\beta} - \beta)(\hat{\beta} - \beta)' \right]$$

$$= \left[ \frac{E(\hat{\beta}_k - \beta_k)^2}{E(\hat{\beta}_k - \beta_k)(\hat{\beta}_l - \beta_l)} \right] - \frac{E(\hat{\beta}_k - \beta_k)^2}{E(\hat{\beta}_k - \beta_k) - \beta_k}$$

$$= \left[ \frac{\text{Var}(\hat{\beta}_k)}{\text{Cov}(\hat{\beta}_k, \hat{\beta}_l)} \right] - \frac{\text{Cov}(\hat{\beta}_k, \hat{\beta}_l)}{\text{Var}(\hat{\beta}_k)} \quad (3.9)$$

The diagonal elements of $V$ represent the variance of the estimated parameters, while the off-diagonal terms represent the covariances.

$$\text{Var}(\hat{\beta}) = E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'] = E[(A\epsilon)(A\epsilon)']$$

$$= E(A\epsilon\epsilon' A')$$
\[ AE(\varepsilon \varepsilon')A' = A(\sigma^2 I)A' = \sigma^2 AA' \]  \hspace{1cm} (3.10)

Since \( A \) and \( A' \) are matrices of fixed numbers.

\[ AA' = [(X'X)^{-1}X'][(X'X)^{-1}X']' \]
\[ = (X'X)^{-1}(X'X)(X'X)^{-1} = (X'X)^{-1} \]  \hspace{1cm} (3.11)

From equation (3.10) and (3.11), we get

\[ \text{Var}(\hat{\beta}) = E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'] = \sigma^2 (X'X)^{-1} \]  \hspace{1cm} (3.12)

[3.2.1] t-Statistics

To test the significance levels of the individual estimated regression coefficients, t-statistics is used. In order to compute the t-value, the best estimate of \( \text{Var}(\hat{\beta}) \) i.e., \( \sigma^2 (X'X)^{-1} \) is required. This implies that we need best estimate of \( \sigma^2 \). It can be shown as (Gujarati, 2004)

\[ s^2 = \frac{\tilde{\varepsilon}'\tilde{\varepsilon}}{N-K} \]  \hspace{1cm} (3.13)

is the best unbiased estimator of \( \sigma^2 \). It follows that \( s^2 (X'X)^{-1} \) yields an unbiased estimator of \( \text{Var}(\tilde{\beta}) \) we rely on the use of the t test when \( s^2 \) is used to approximate \( \sigma^2 \). To do so, we use the following statistical results:

1) \( \tilde{\varepsilon}'\tilde{\varepsilon} / \sigma^2 \) is distributed as chi square with N-K degrees of freedom.

2) \((N - K) s^2 / \sigma^2 \) is distributed as chi square with N-K degrees of freedom.
3) \((\beta_i' - \beta_i)\) for \(i = 1, 2 \ldots \ldots \ldots k\), is normally distributed with mean 0 and variance \(\sigma^2 V_i\), where \(V_i\) is the \(i\)th diagonal element of \((X'X)^{-1}\).

4) \((N-K) s^2 / \sigma^2\) and \((\beta_i' - \beta_i)\) are independently distributed.

\[
\frac{t_{n-k}}{s \sqrt{V_i}} = \frac{(\hat{\beta}_i - \beta_i)}{s \sqrt{V_i}}
\]

(3.14)

To test a hypothesis about a particular value of \(\beta_i\), we substitute that value into Eq (3.14). If the \(t\)-value is great enough in absolute value, we reject the null hypothesis at the appropriately chosen level of confidence. A 95% confidence interval for \(\beta_i\) is given by

\[
\hat{\beta}_i \pm t_c (s \sqrt{V_i})
\]

Where, \(t_c\) is the critical value of the \(t\) distribution associated with a 5% level of significance.

[3.2.2] F Statistics

\(F\) statistics is used in the multiple regression model to test the significance of the \(R^2\) statistic.

We now define the coefficient of determination \((R^2)\) of the regression as

\[
R^2 = 1 - \frac{ESS}{TSS}
\]

\[
= 1 - \frac{\hat{\epsilon}'\hat{\epsilon}}{Y'Y}
\]
From the eq (2.5) we get,

\[ R^2 = \frac{\hat{\beta}' X' X \hat{\beta}}{Y'Y} \]  

(3.15)

In order to correct for the dependence of goodness of fit on degree of freedom, we define \( R^2 (\text{Adj}) \).

\[ R^2 (\text{Adj}) = 1 - \frac{\hat{\varepsilon}' \hat{\varepsilon} / (N - K)}{Y'Y / (N - 1)} = 1 - \frac{\hat{\varepsilon}' \hat{\varepsilon} \ N - 1}{Y'Y \ N - K} \]  

(3.16)

The most frequently used test for goodness of fit involves the test of the joint null hypothesis that \( \beta_2 = \beta_3 = \cdots \beta_k = 0 \) i.e., all the regression coefficients are simultaneously zero. This is done by computing F-statistics for (K-1) and (N-K) degrees of freedom (Gujarati, 2004)

\[ F_{k-1, N-K} = \frac{R^2 \ N - K}{1 - R^2 \ K - 1} \]  

(3.17)

and comparing it with critical value of \( F_{k-1, N-K} \) for 5% level of significance. If the computed F statistics is greater than critical F value, the null hypothesis is rejected.

As a practical matter, the F test of the significance of a regression equation is of limited use, because it is likely that the F statistic will allow for rejection of the null hypothesis, whether or not the model explains the structure under study. In fact, it is possible for \( R^2 \) to be significant at a given level, even though none of the regression coefficients are found to be significant according to individual t-tests. If the value of the F statistic is not significantly different from 0, we must conclude that the explanatory variables do little to explain the variation of about its mean. (Gujarati, 2004).  

39