Economics Honours (Semester VI)
Basic Econometrics
Multiple Regression Analysis

*Meaning

Multiple Regression analysis studies the influence of two or more independent variables on a dependent variable. It is said to be an extended form of the simple regression model. The simplest form of multiple regression model is the three-variable model with one dependent variable, that is the regressand and two independent or explanatory variables also called the predictors or regressors.

The three-variable Population Regression Function can be written as

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$$

where Y is the dependent variable, X_2 and X_3, the explanatory variables, u, the disturbance or error term, the coefficients β_1, the intercept term, $\beta 2$ and $\beta 3$ denote the partial regression coefficients or the partial slope coefficients and i denotes the ith observation in the data series.

Assumptions

1. Zero mean value of u_i, for each i
2. No serial correlation, or cov (u_i, u_j) = 0 $i \neq j$
3. Homoscedasticity, or var (u_i) = σ^2
4. Zero covariance between u_i and each X variable, or cov (ui, X_{2i}) = cov (ui, X_{3i}) = 0
5. No specification bias, or the model is correctly specified
6. No exact collinearity between the X variables, or no exact linear relationship between X_2 and X_3
7. Linear in the parameters, that the values of the regressors are fixed in repeated sampling, and that there is sufficient variability in the values of the regressors.

*The Meaning of Partial Regression Coefficients

The regression coefficients β_2 and β_3 are known as partial regression or partial slope coefficients. β_2 measures the change in the mean value of Y, per unit change in X_2, holding the value of X_3 constant. Similarly, $\beta 3$ measures the change in the mean value of Y per unit change in X3, holding the value of X2 constant.

OLS Estimators

Let us write the sample regression function as follows:

$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \hat{u}_i$ ……(1)

To find the OLS estimators, we need to

Min $\Sigma \hat{u}_i^2 = \Sigma(Yi - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i})^2$ ……(2)

Differentiate $\Sigma \hat{u}_i^2$ with respect to $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\beta}_3$ and setting the results equal to zero, the following three normal equations are obtained.

$\bar{Y} = \hat{\beta}_1 + \hat{\beta}_2 \bar{X}_2 + \hat{\beta}_3 \bar{X}_3$ ……(3)

$\Sigma Y_i X_{2i} = \hat{\beta}_1 \Sigma X_{2i} + \hat{\beta}_2 \Sigma X_{2i}^2 + \hat{\beta}_3 \Sigma X_{2i} X_{3i}$ ……..(4)

$\Sigma Y_i X_{3i} = \hat{\beta}_1 \Sigma X_{3i} + \hat{\beta}_2 \Sigma X_{2i} X_{3i} + \hat{\beta}_3 \Sigma X_{3i}^2$ ……(5)

Solving the normal equations gives the formulae for regression coefficients

$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}_2 - \hat{\beta}_3 \bar{X}_3$ ……(6)

$$\hat{\beta}_2 = \frac{(\Sigma y_i x_{2i})(\Sigma x_{3i}^2) - (\Sigma y_i x_{3i})(\Sigma x_{2i} x_{3i})}{(\Sigma x_{2i}^2)(\Sigma x_{3i}^2) - (\Sigma x_{2i} x_{3i})^2} \quad \text{……(7)}$$

$$\hat{\beta}_3 = \frac{(\Sigma y_i x_{3i})(\Sigma x_{2i}^2) - (\Sigma y_i x_{2i})(\Sigma x_{2i} x_{3i})}{(\Sigma x_{2i}^2)(\Sigma x_{3i}^2) - (\Sigma x_{2i} x_{3i})^2} \quad \text{……..(8)}$$

R^2 AND THE ADJUSTED R^2

In the three variable model we would like to know the proportion of the variation in Y explained by the variables X_2 and X_3 jointly. The quantity that gives this information is known as the multiple coefficient of determination and is denoted by R^2.

Given TSS = ESS + RSS

i.e. Total sum of Squares = Explained sum of squares + Residual sum of squares,

we have, $R^2 = ESS / TSS$

$$= \frac{\hat{\beta}_2 \Sigma y_i x_{2i} + \hat{\beta}_3 \Sigma y_i x_{3i}}{\Sigma y_i^2}$$

Or, $= 1 - RSS/TSS = 1 - (\Sigma \hat{u}_i^2)/\Sigma y_i^2$ ……(9)

The value of R^2 lies between) and 1. The coefficient of multiple correlation is denoted by R and it is a measure of the degree of association between Y and the explanatory variables.

An important property of R^2 as the number of regressors increases, R^2 almost invariably increases and never decreases. Stated differently, an additional X variable will not decrease R^2. Thus, in comparing two regression models with the same dependent variable but differing number of X variables, one should be very wary of choosing the model with the highest R^2. To compare two R^2 terms, one must take into account the number of X variables present in the model. This can be done readily if we consider an alternative coefficient of determination, which is known as the adjusted R^2, denoted by $\bar{R}^2$, (pronounced as R bar square). The term adjusted means adjusted for the degrees of freedom (df) associated with the sums of squares entering into $\Sigma \hat{u}_i^2$ having (n – k) df in a model involving ‘n’ observations and k parameters including the intercept term and Σy_i^2 having (n - 1) df. So

$$\bar{R}^2 = 1 - \frac{\Sigma \hat{u}_i^2 / (n-k)}{\Sigma y_i^2 / (n-1)} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (10)$$

$$\text{Or, } \bar{R}^2 = 1 - (1 - R^2)\frac{n-1}{n-k} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (11)$$

*N.B. Explanations of all the theories have been taken from the following reference:

References:

Gujarati, D. N. Porter, D.C., Gunasekar, S. (2009), *Basic econometrics*. (Fifth ed.) McGraw-Hill Education (India).