CHAPTER FOUR: MULTIPLE LINEAR REGRESSION

4.1 Introduction

In a simple regression, a dependent variable is a function of only one explanatory variable. However, in economics you hardly found that one variable is affected by only one explanatory variable. Hence, a two variable model is often inadequate in practical works. The multiple linear regression is concerned with the relationship between a dependent variable and two or more explanatory variables.

4.2 The three-variable model: specification and assumptions

Consider a regression model with two explanatory variables.

$Y = f(X_1, X_2)$

Example: Demand for a commodity may be influenced not only by the price of the commodity but by the consumers' income. Since the theory does not specify the mathematical form of the demand function, we assume the relationship between Y, and the regressors (X_1, and X_2) is linear. Hence we may write the three variables Population Regression Function (PRF) as follows:

$$Y_i = \beta_o + \beta_1 X_{1i} + \beta_2 X_{2i} + U_i$$

Where Y is the quantity demanded

X_1 and X_2 are the price and income respectively

β_o is the intercept term

β_1 is the coefficient of X_1 and its expected sign is negative (due to the law of demand)

β_2 is the coefficient of X_2 and its expected sign is positive assuming that the good is a normal good.

Interpretation of partial regression coefficients

Given $Y_i = \beta_o + \beta_1 X_{1i} + \beta_2 X_{2i} + U_i$

The meaning of partial regression coefficient is as follows: β_1 measures the change in the mean value of Y, E(Y), per unit change in X_1, holding the value of X_2 constant. Likewise, β_2 measures the change in the mean value of Y per unit change in X_2, holding the value of X_1 constant.

Consider also a model relating a person's wage to observed educ (years of education), exper (years of labor market experience), tenure (years with the current employer) and other unobserved factors.

$Wage = \beta_o + \beta_1 educ + \beta_2 exper + u$

For instance, $\beta_1$measures the change in hourly wage given another year of education, holding experience fixed. Similarly, β_2 measures the change in hourly wage given another year of experience, holding education fixed.

To complete the specification of our simple model we need some assumptions about the random variable U. These assumptions are the same as those assumptions already explained in the simple linear regression model.

Assumptions of the model

1. **Zero mean value of U_i**

   The random variable U has a zero mean value for each $X_{i.}$ that is, $E(U_i/X_{1i}, X_{2i}) = 0$ for each i.

2. **Homoscedasticity:** The variance of each U_i is the same for all the X_i values

   $\text{Var}(U_i) = E(U_i^2) = \sigma_u^2$

3. **Normality:** The values of each U_i are normally distributed. That is, $U_i \sim N(0, \sigma_u^2)$

4. **No serial correlation:** The values of U_i (corresponding to X_i) are independent of the values of any other U_j (corresponding to X_j).

   $\text{Cov}(U_i, U_j) = 0$ for $i \neq j$

5. **Independence of U_i and X_i:** Every disturbance term U_i is independent of the explanatory variables. That is, there is zero covariance between U_i and each X variables.

   $\text{Cov}(U_i, X_{1i}) = \text{Cov}(U_i, X_{2i}) = 0$

   Here the values of the X's are a set of fixed numbers in all hypothetical samples.

6. **No perfect multicollinearity (No collinearity between the X variables):** The explanatory variables are not perfectly linearly correlated. That is, there is no exact linear relationship between X_1 and X_2.

7. **Correct specification of the model:** The model has no specification error in that all the important explanatory variables appear explicitly in the function and the mathematical form is correctly defined.

4.3 Estimation: the method of least squares

Suppose the sample regression function (SRF)

$$\mathrm{Y_i} = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \hat{u}_i$$

where $\hat{\beta}_0, \hat{\beta}_1$ and $\hat{\beta}_2$ are estimates of the true parameters β_0, β_1 and β_2

$\hat{u}_i$ is the residual term.

But since U_i is unobservable the above equation becomes

$\hat{Y}_i = \hat{\beta}_o + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i}$ is the estimated regression line.

In the least squares estimation, the estimates will be obtained by choosing the values of the unknown parameters that will minimize the sum of squares of the residuals. (OLS requires the $\sum \hat{u}_i^{\,2}$ be as small as possible). Symbolically,

$$\text{Min } \sum \hat{u}_i^{\,2} = \sum(\mathrm{Y_i} - \hat{Y}_i)^2 = \sum_{i}^{n} (\mathrm{Y_i} - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i})^2$$

A necessary condition for a minimum value is that the partial derivatives of the above expression with respect to the unknowns (i.e. $\hat{\beta}_0, \hat{\beta}_1$ and $\hat{\beta}_2$) should be set to zero.

$$\frac{\partial \sum \left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i} \right)^2}{\partial\, \hat{\beta}_0} = 0$$

$$\frac{\partial \sum \left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i} \right)^2}{\partial\, \hat{\beta}_1} = 0$$

$$\frac{\partial \sum \left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i} \right)^2}{\partial \hat{\beta}_2} = 0$$

After differentiating, we get the following normal equations:

$$\sum Y_i = n\hat{\beta}_o + \hat{\beta}_1 \sum X_{1i} + \hat{\beta}_2 \sum X_{2i}$$

$$\sum X_{1i} Y_i = \hat{\beta}_o \sum X_{1i} + \hat{\beta}_1 \sum X^2_{1i} + \hat{\beta}_2 \sum X_{1i} X_{2i}$$

$$\sum X_{2i} Y_i = \hat{\beta}_o \sum X_{2i} + \hat{\beta}_1 \sum X_{1i} X_{2i} + \hat{\beta}_2 \sum X^2_{2i}$$

Solving the above normal equations we can obtain values for $\hat{\beta}_o , \hat{\beta}_1$ and $\hat{\beta}_2$

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

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

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

where the variables x and y are in deviation forms.

Derivation

$$\hat{u}_i = y_i - \hat{y}_i$$

$$\sum \hat{u}_i^2 = \sum (y_i - \hat{\beta}_1 x_{1i} - \hat{\beta}_2 x_{2i})^2$$

$$\frac{\partial \sum \hat{u}_i^2}{\partial \hat{\beta}_1} = -2 \sum (y_i - \hat{\beta}_1 x_{1i} - \hat{\beta}_2 x_{2i}) x_{1i} = 0$$

$$\frac{\partial \sum \hat{u}_i^2}{\partial \hat{\beta}_2} = -2 \sum (y_i - \hat{\beta}_1 x_{1i} - \hat{\beta}_2 x_{2i}) x_{2i} = 0$$

After rearranging ,

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

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

Solving these two equations simultaneously we will get the values of $\hat{\beta}_1$ and $\hat{\beta}_2$ given above.

4.4 The Mean and Variance of the Parameter Estimates

The mean of the estimates of the parameters in the three-variable model is derived in the same way as in the two-variable model.

The estimates are unbiased estimates of the true parameters of the relationship between Y, X_1 and X_2.

$$E(\hat{\beta}_0) = \beta_0 \quad E(\hat{\beta}_1) = \beta_1 \quad E(\hat{\beta}_2) = \beta_2$$

The variance of $\hat{\beta}_0, \hat{\beta}_1$ and $\hat{\beta}_2$ are also given as follows

$$\text{Var}(\hat{\beta}_o) = \hat{\sigma}_u^2 \left[\frac{1}{n} + \frac{\overline{X}_1^2 \sum x_2^2 + \overline{X}_2^2 \sum x_1^2 - 2\overline{X}_1 \overline{X}_2 \sum x_1 x_2}{\sum x_1^2 \sum x_2^2 - \left(\sum x_1 x_2\right)^2} \right]$$

$$\text{Var}(\hat{\beta}_1) = \hat{\sigma}^2 \frac{\sum x_2^2}{\sum x_1^2 \sum x_2^2 - \left(\sum x_1 x_2\right)^2}$$

$$\text{Var}(\hat{\beta}_2) = \hat{\sigma}^2 \frac{\sum x_1^2}{\sum x_1^2 \sum x_2^2 - \left(\sum x_1 x_2\right)^2}$$

Where $\hat{\sigma}^2 = \frac{\sum \hat{U}_i^2}{n-k}$, (in the three variable model, k = 3).

Note that x_1 and x_2 are in deviations form.

4.5 The multiple coefficient of determination (R^2) and the Adjusted R^2

In a two variable regression model, the coefficient of determination (r^2) measures the goodness of fit of the regression equation. This notion of r^2 can be easily extended to regression models containing more than two variables.

In the three-variable model, we would like to know the proportion of the variation in Y explained by the variables X_1 and X_2 jointly. The quantity that gives this information is known as the *multiple coefficient of determination*. It is denoted by R^2, with subscripts the variables whose relationships are being studies.

Example: $R^2{}_{y.X_1X_2}$- shows the percentage of the total variation in Y explained by the regression plane, that is, by changes in X_1 and X_2.

$$R^2{}_{y.X_1X_2} = \frac{\sum \hat{y}_i{}^2}{\sum y_i{}^2} = \frac{\sum\left(\hat{Y}_i - \bar{Y}\right)^2}{\sum\left(Y_i - \bar{Y}\right)^2}$$

$$= 1 - \frac{\sum \hat{u}_i{}^2}{\sum y_i{}^2} = 1 - \frac{RSS}{TSS}$$

where: RSS – residual sum of squares

TSS – total sum of squares

Recall that

$\hat{y}_i = \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i}$ (the variables are in deviation forms)

$y_i = \hat{y}_i + \hat{u}_i$

$\sum \hat{u}_i{}^2 = \sum(y_i - \hat{y}_i)^2 = \sum(y_i - \hat{\beta}_1 x_{1i} - \hat{\beta}_2 x_{2i})^2$

or $\sum \hat{u}_i{}^2 = \sum \hat{u}_i \,.\, \hat{u}_i = \sum \hat{u}_i \,(y_i - \hat{\beta}_1 x_{1i} - \hat{\beta}_2 x_{2i})$

$= \sum \hat{u}_i \,.y_i - \hat{\beta}_1 \sum \hat{u}_i \,.x_{1i} - \hat{\beta}_2 \sum \hat{u}_i \,.x_{2i}$

but $\sum \hat{u}_i \,.x_{1i} = \sum \hat{u}_i \,.x_{2i} = 0$

Hence $\sum \hat{u}_i{}^2 = \sum \hat{u}_i \, y_i$

$= \sum (y_i - \hat{y}_i)y_i$ since $\hat{u}_i = y_i - \hat{y}_i$

$= \sum y_i \,(y_i - \hat{\beta}_1 x_{1i} - \hat{\beta}_2 x_{2i})$

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

By substituting the value of $\sum \hat{u}_i{}^2$ in the formula of R^2, we get

$$R^2{}_{y.X_1X_2} = 1 - \frac{\sum \hat{u}_i{}^2}{\sum y_i{}^2} = 1 - \frac{\left[\sum y_i{}^2 - \hat{\beta}_1 \sum x_{1i} y_i - \hat{\beta}_2 \sum x_{2i} y_i\right]}{\sum y_i{}^2}$$

$$= \frac{\hat{\beta}_1 \sum x_{1i} y_i + \hat{\beta}_2 \sum x_{2i} y_i}{\sum y_i^2}$$, where x_{1i}, x_{2i} and y_i are in their deviation forms.

The value of R^2 lies between 0 and 1. The higher R^2 the greater the percentage of the variation of Y explained by the regression plane, that is, the better the goodness of fit of the regression plane to the sample observations. The closer R^2 to zero, the worse the fit is.

An important property of R^2 is that it is a nondecreasing function of the number of explanatory variables or regressors present in the model; 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.

To see this, recall the definition of R^2

$$R^2 = 1 - \frac{\sum \hat{u}_i^2}{\sum y_i^2}$$

It is clear that $\sum y_i^2$ is independent of the number of X variables in the model because it is simply $\sum(Y_i - \bar{Y})^2$. The residual sum of squares (RSS), $\sum \hat{u}_i^2$, however, depends on the number of explanatory variables present in the model. It can be noticed that as the number of X variables increases, $\sum \hat{u}_i^2$ is bound to decrease (at least it will not increase), hence R^2 will increase. Therefore, in comparing two regression models with the same dependent variable but different number of X variables, one should be very wary of choosing the model with the highest R^2. An explanatory variable which is not statistically significant may be retained in the model if one looks at R^2 only.

Therefore, to correct for this defect, we adjust R^2 by taking into account the degrees of freedom, which clearly decreases as new regressors are introduced in the function.

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

$$\text{or } \bar{R}^2 = 1 - (1 - R^2) \frac{(n-1)}{n-k}$$

where k = the number of parameters in the model (including the intercept term)

n = the number of sample observations

It is immediately apparent from the above equation that for $k > 1$, $\bar{R}^2 < R^2$ which implies that as the number of explanatory variables increases, the adjusted R^2 is increasingly less than the unadjusted R^2. The adjusted R^2, i.e. $\overline{R}^2$, can be negative, although R^2 is necessarily non-negative. In this case its value is taken as zero. If n is large, $\overline{R}^2$ and R^2 will not differ much. But with small samples, if the number of regressors (X's) is large in relation to the sample observations, $\overline{R}^2$ will be much smaller than R^2.

4.6 Partial Correlation Coefficients

In the two variable regression model, we have used the simple correlation coefficient, r, as measure of the degree of linear association between two variables. For three variables case, we can compute three correlation coefficients: r_{yx1} (correlation between y and x_1), r_{yx2}, and r_{x1x2} – these are called gross or simple correlation coefficients, or correlation coefficients of zero order. But, for example, r_{yx1} does not likely reflect the true degree of association between Y and X_1 in the presence of X_2. Therefore, what we need is a correlation coefficient that is independent of the influence, if any, of X_2 on X_1 and Y. Such a correlation coefficient is known as the *partial correlation coefficient*. Conceptually, it is similar to the partial regression coefficient.

$r_{yx1.x2}$ = partial correlation coefficient between Y and X_1, holding X_2 constant

$r_{yx2.x1}$ = partial correlation coefficient between Y and X_2, holding X_1 constant

$r_{x1x2.y}$ = partial correlation coefficient between X_1 and X_2, holding Y constant.

We can compute the partial correlations from the simple or zero order correlation coefficients as follows.

$$r_{yx1.x2} = \frac{r_{yx1} - r_{yx2}\, r_{x1x2}}{\sqrt{(1 - r_{yx2}^2)(1 - r_{x1x2}^2)}}$$

$$r_{yx2.x1} = \frac{r_{yx2} - r_{yx1}\, r_{x2x3}}{\sqrt{(1 - r_{yx1}^2)(1 - r_{x1x2}^2)}}$$

$$r_{x1x2.y} = \frac{r_{x1x2} - r_{yx1}\, r_{yx2}}{\sqrt{(1 - r_{yx1}^2)(1 - r_{yx2}^2)}}$$

Note: By order we mean the number of secondary subscripts. For example, $r_{yx1.x2x3}$ is the correlation coefficient of order two, whereas $r_{yx1.x2x3x4}$ represents the correlation coefficient of order three, and so on.

4.7 The Confidence Interval for β_i

The principle involved in constructing the confidence is identical with that of simple regression.

4.8 Test of significance in multiple regression

4.8.1 Hypothesis Testing about Individual Partial Regression Coefficients

We can test whether a particular variable X_1 or X_2 is significant or not holding the other variable constant. The t test is used to test a hypothesis about any individual partial regression coefficient. The partial regression coefficient measures the change in the mean value of Y $E(Y/X_2,X_3)$, per unit change in X_2, holding X_3 constant

$$t = \frac{\hat{\beta}_1 - \beta_i}{se(\hat{\beta}_i)} \sim t(n-k) \quad (i = 0, 1, 2, \ldots., k)$$

This is the observed (or sample) value of the t ratio, which we compare with the theoretical value of t obtainable from the t-table with n – k degrees of freedom.

The theoretical values of t (at the chosen level of significance) are the critical values that define the critical region in a two-tail test, with n – k degrees of freedom.

let us postulate that

H_0: $\beta_i = 0$

H_1: $\beta_i \neq 0$ or one sided ($\beta_i > 0, \beta_i < 0$)

The null hypothesis states that, holding X_2 constant, X_1 has no (linear) influence on y.

If the computed t value exceeds the critical t value at the chosen level of significance, we may reject the hypothesis; otherwise, we may accept it ($\hat{\beta}_1$ is not significant at the chosen level of significance and hence the corresponding regression does not appear to contribute to the explanation of the variations in Y).

Example. Suppose the estimated hourly wage equation is given as follows:

$$\log(\hat{wage}) = .284 + .092\, educ + .0041\, exper + .022\, tenure$$

s.e (.104) (.007) (.0017) (.003)

$$n = 526,\ R^2 = .316,$$

Test whether the return to exper, controlling for educ and tenure, is zero in the population, against the alternative that it is positive. That is, H0: $\beta_{exper} = 0$ against H1: $\beta_{exper} > 0$.

4.8.2 Testing the Overall Significance of a Regression

This test aims at finding out whether the explanatory variables (X_1, X_2, …X_k) do actually have any significant influence on the dependent variable. The test of the overall significance of the regression implies testing the null hypothesis

H_0: $\beta_1 = \beta_2 = \ldots = \beta_k = 0$

Against the alternative hypothesis

H_1: not all β_i's are zero.

If the null hypothesis is true, then there is no linear relationship between y and the regressors.

The above joint hypothesis can be tested by the analysis of variance (ANOVA) technique. The following table summarizes the idea.

Source of variation	Sum of squares (SS)	Degrees of freedom (Df)	Mean square (MSS)
Due to regression (ESS)	$\sum \hat{y}_i^2$	k – 1	$\frac{\sum \hat{y}^2}{k-1}$
Due to Residual (RSS)	$\sum \hat{u}_i^{\ 2}$	n – k	$\frac{\sum \hat{u}_i^{\ 2}}{n-k}$
Total (Total variation)	$\sum y_i^2$	n – 1	

Therefore to undertake the test first find the calculated value of F and compare it with the F tabulated. The calculated value of F can be obtained by using the following formula.

F = $\frac{\sum \hat{y}_i^{\ 2}/(k-1)}{\sum \hat{u}_i^{\ 2}/(n-k)} = \frac{ESS/(k-1)}{RSS/(n-k)}$ follows the F distribution with k – 1 and n – k df.

where k – 1 refers to degrees of freedom of the numerator

n – k refers to degrees of freedom of the denominator k – number of parameters estimated

Decision Rule: If $F_{calculated} > F_{tabulated}$ ($F_\alpha(k - 1, n - k)$), reject H_0: otherwise you may accept it, where $F_\alpha(k - 1, n - k)$ is the critical F value at the α level of significance and (k – 1) numerator df and (n – k) denominator df.

Note that there is a relationship between the coefficient of determination R^2 and the F test used in the analysis of variance.

From the above:

$$F = \frac{(n-k)ESS}{(k-1)RSS}$$

$$F = \frac{(n-k)ESS/TSS}{(k-1)(1-ESS/TSS)}$$

$$F = \frac{(n-k)R^2}{(k-1)(1-R^2)}$$

$$\mathrm{F} = \frac{R^2 / (k-1)}{(1-R^2 / (n-k)}$$

When $R^2 = 0$, F is zero. The larger the R^2, the greater the F value. In the limit, when $R^2 = 1$, F is infinite. Thus the F test, which is a measure of the overall significance of the estimated regression, is also a test of significance of R^2. Testing the null hypothesis is equivalent to testing the null hypothesis that (the population) R^2 is zero.