Economics Honours (Semester VI)
Basic Econometrics
More on Multiple Regression

**Variances and Standard Errors of OLS Estimators*

As in the simple regression model with two-variable case, in the multiple regression model concerning three variable we also require the variances and standard errors of the estimators to establish the statistical significance of the parameter estimates. We have already derived the OLS estimators of the partial regression coefficients in the multiple regression analysis. We now derive the variances and standard errors of these estimators.

Given the sample regression model as follows:

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

where Y_i = dependent variable, $\hat{\beta}_1$ = intercept coefficient, $\hat{\beta}_2$ and $\hat{\beta}_3$ are the partial regression coefficients, X_{2i} and X_{3i} are the two explanatory variables and $\hat{u}_i$ the residual term. Here k which represents the total number of parameters is equal to three.

The OLS estimators are

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}_2 - \hat{\beta}_3 \bar{X}_3$$

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

$$\hat{\beta}_3 = \frac{(\Sigma y_i x_{3i})(\Sigma x_{2i}^2) - (\Sigma y_i x_{2i})(\Sigma x_{2i} x_{3i})}{(\Sigma x_{2i}^2)(\Sigma x_{3i}^2) - (\Sigma x_{2i} x_{3i})^2}$$

We get the following results for Variances and Standard Errors

$$\mathrm{Var}(\hat{\beta}_2) = \sigma^2 . \frac{\Sigma x_{3i}^2}{\Sigma x_{2i}^2 \Sigma x_{3i}^2 - (\Sigma x_{2i} x_{3i})^2}$$

$$\mathrm{se}(\hat{\beta}_2) = \sqrt{\mathrm{Var}(\hat{\beta}_2)}$$

$$\mathrm{Var}(\hat{\beta}_3) = \sigma^2 . \frac{\Sigma x_{2i}^2}{\Sigma x_{2i}^2 \Sigma x_{3i}^2 - (\Sigma x_{2i} x_{3i})^2}$$

$$\mathrm{se}(\hat{\beta}_3) = \sqrt{\mathrm{Var}(\hat{\beta}_3)}$$

$$\sigma^2 = \hat{\sigma}^2 = \frac{\Sigma \hat{u}_i^2}{(n-k)}$$

**Procedure for Testing Partial Regression Coefficients*

On the basis of the assumption that $u_i \sim N(0,\sigma^2)$, we use the *t* test to test for any individual partial regression coefficient. Suppose we decide to test the significance of the partial regression coefficient β_2 and set the null hypothesis against the alternative hypothesis as

$H_0 : \beta_2 = 0$,

against $H_1 : \beta_2 \neq 0$.

Here the null hypothesis states that with X_3 held constant, X_2 has no influence on Y. To test the null hypothesis we compute the value of *t*.

$$t = \frac{\hat{\beta}_2 - \beta_2}{se(\hat{\beta}_2)}$$

For a two tailed test, if the computed *t* value exceeds the critical value of *t* at the chosen level of significance for the given degrees of freedom (denoted by n), we reject the null hypothesis and conclude that the partial regression coefficient is statistically significant. Otherwise if the computed *t* value does not exceed the critical value of *t* at the chosen level of significance for the given degrees of freedom, we may reject the null hypothesis and conclude that the partial regression coefficient is not statistically significant. We test the significance of the partial regression coefficient β_3 in a similar manner.

**Procedure for Testing the Overall Significance of the Multiple Regression*

We use the F test to the overall significance of the multiple regression. Given a k-variable regression model (for our case three variable model, i.e. k=3) of the form

$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \hat{u}_i$.

We set the null hypothesis against the alternative hypothesis as

$H_0 : \beta_2 = \beta_3 = 0$,

(all slope coefficients are simultaneously zero)

against H_1 : Not all slope coefficients are simultaneously zero

To test the null hypothesis we compute the value of F.

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

where n = number of observations, k = total number of parameters in the multiple regression model and R^2 = Coefficient of determination of the multiple regression model.

For a two tailed test, if the computed $F > F_\alpha$ (k-1, n-k), reject H_0, otherwise do not reject it, where F_α (k-1, n-k) is the critical value of F at α level of significance and (k-1) number df and (n – k) denominator df.

**Partial Correlation Coefficients*

The partial correlation coefficients measure the net correlation between the dependent variable and one independent variable after excluding the common influence of (i.e. holding constant) the other independent variables in the model. The partial correlation coefficients for the above three variable model is defined as follows:

$r_{12.3}$ = partial correlation coefficient between Y and X_2, holding X_3 constant

$r_{13.2}$ = partial correlation coefficient between Y and X_3, holding X_1 constant

and computed as

$$r_{12.3} = \frac{r_{12} - r_{13}\, r_{23}}{\sqrt{(1 - r_{13}^2)\,(1 - r_{23}^2)}}$$

$$r_{13.2} = \frac{r_{13} - r_{12}\, r_{23}}{\sqrt{(1 - r_{12}^2)\,(1 - r_{23}^2)}}$$

Here r_{12} = simple correlation between Y and X_2

r_{13} = simple correlation between Y and X_3

r_{23} = simple correlation between X_2 and X_3

The partial correlation coefficients range in value from (-1 to +1), have the sign of the corresponding estimated parameter and are used to determine the relative importance of the different explanatory variables in a multiple regression.

Suppose in any example, $r_{12.3} = 0.7023$ or 70.23% and $r_{13.2} = 0.8434$ or 84.34%. Then it is concluded that as $r_{13.2} > r_{12.3}$, X_3 is more important than X_2 in explaining the variation in Y.

*****N.B. Explanations of all the theories have been taken from the references mentioned.

References:

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

Salvatore, D., & Reagle, D. (2011). *Schaum's outline of statistics and econometrics* (2nd ed.). McGraw-Hill Education.