Economics Honours (Semester VI)

Basic Econometrics

Violation of the Assumptions of CLSM

(Concepts Only)

Multicollinearity

One of the assumptions of the Classical Ordinary Least Square Method is that in a multiple regression model of the form,

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

there is no exact collinearity between the X variables, or no exact linear relationship between X_2 and X_3. The violation of this assumption gives rise to the problem of *multicollinearity*. Stated differently, multicollinearity refers to the case in which two or more explanatory variables in the regression are highly correlated, thus making it difficult or impossible to isolate their individual effects on the dependent variable.

A test for multicollinearity is that the estimated coefficients may be statistically insignificant (and even have the wrong sign) even though R^2 may be high.

Example:

Suppose we obtain the following multiple regression function based on a sample size n= 15 on Y = growth rate of imports, X_2 = GDP and X_3 = inflation. It is expected that the level of imports will be greater as GDP and domestic prices increase.

$$\hat{Y}_i = 0.0015 + \underset{(1.46)}{1.39\, X_2} + \underset{(1.85)}{0.09 X_3}$$

where $\hat{\beta}_1 = 0.0015$

$\hat{\beta}_2 = 1.39$ and $t = 1.46$

$\hat{\beta}_3 = 0.09$ and $t = 1.85$

also $R^2 = 0.42$

and $r_{23} = 0.38$

- On observation it is found that $\hat{\beta}_2$ and $\hat{\beta}_3$ both have positive coefficients. This means that there is a positive relationship between GDP and level of imports and between domestic increase in prices and level of imports.
- **Comparing the determined values of *t* for $\hat{\beta}_2$ and $\hat{\beta}_3$ i.e. *t*= 1.46 and *t* = 1.85 respectively with the critical values of *t* for 12 df at 5% level of significance, it is found that neither $\hat{\beta}_2$ nor $\hat{\beta}_3$ is statistically significant at 5% level.**
- R^2 indicates that 42% of the variation in Y is explained even though none of the variables stand out individually.

Thus we may conclude that the above model in our example suffers from multicollinearity.

Multicollinearity can sometimes be overcome or reduced by collecting more data, by utilizing a priori information, by transforming the functional relationship or by dropping one of the highly collinear variables.

Heteroscedasticity

If the OLS assumption the variance of the error term is constant for all observations do not hold, we face the problem of *heteroscedasticity*. This leads to unbiased by inefficient (that is larger than minimum variance) estimates of coefficients, as well as biased estimated of standard error as the latter is determined from the former. This in turn gives rise to incorrect statistical tests and confidence intervals.

One test of heteroscedasticity involves arranging the data from small to large values of the independent variable X. Here we apply two regression analysis, one for small values of X and one for large values of X omitting one fifth of the middle observations. Then we test the ratio of the error sum of squares (ESS) of the second regression to the first regression is significantly different from zero, using F table with (n - d -2k)/2 degrees of freedom, where n is the total number of observations, d is the number of omitted observations and k is the number of estimated parameters.

If the error term variance is proportional to X^2, heteroscedasticity can be overcome by dividing every term of the model by X and then reestimating the regression using the transformed variables.

Autocorrelation

If the OLS assumption, $\text{cov}\,(u_i, u_j) = 0,\ i \neq j$ gets violated we face the problem of *autocorrelation* or serial correlation. Stately, differently when the error term in one time period is positively correlated with the error term in the previous time period, we face the problem of (positive first-order) autocorrelation. This is common in time series analysis and leads to downward-biased standard errors and in turn this gives rise to incorrect statistical tests and confidence intervals.

The presence of first-order autocorrelation is tested by utilizing the table of Durbin-Watson statistic at the 5% level of significance for n observations and k explanatory variables. If the calculated value of d from the formula given below is smaller than the tabular value of d_L (lower limit), the hypotheses of positive first-order autocorrelation is accepted that is cannot be rejected.

$$d = \frac{\Sigma\,(e_t - e_{t-1})^2}{\Sigma\, e_t^2}$$ where t = 2…n in the numerator and t =1,…n in the denominator.

The hypothesis is rejected if $d > d_U$ (upper limit), and the test is inconclusive if $d_L < d < d_U$.

*N.B. Explanations of all the theories and examples have been taken from the following references:

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.