SEM - 6 (Econometrics) 14/04/2020

Problem of multicollinearity

One of the important assumptions of OLS is the two explanatory variables are not perfectly related

i.e. $r_{12} \neq \pm 1$

i.e. $cov(x_1, x_2)$ ~~[illegible]~~ exists

but x_1 and x_2 are not perfectly related

There are two types of multi-collinearity —

(1) Exact multicollinearity or Perfect MC

(2) Near - exact MC

In case of exact MC

$$X_1 = KX_2, \quad K \neq 0$$

$$var(X_1) = K^2 \, Var(X_2)$$

$$cov(X_1, X_2) = cov[KX_2, X_2]$$

$$= K \, Var(X_2)$$

$$\mathrm{cov}(x_1, x_2)$$

$$= 1/n \sum_{i=1}^{n} (x_1 - \bar{x}_1)(x_2 - \bar{x}_2)$$

$$= 1/n \sum_{i=1}^{n} k (x_2 - \bar{x}_2)(x_2 - \bar{x}_2)$$

$$= k \cdot 1/n \sum_{i=1}^{n} (x_2 - \bar{x}_2)^2$$

$$= k \cdot \mathrm{Var}(x_2)$$

Multicollinearity arises when there are more than one explanatory variables. That is in case of 2 variable model, there is no multicollinearity But in case of 3 variable model there is the problem of multicollinearity

The 3 variable multiple regression model is,

$$Y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$$

By applying OLS when there is no MC,

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

$$\hat{\beta}_1 = \frac{\mathrm{cov}(Y, X_1)\,\mathrm{Var}(X_2) - \mathrm{cov}(Y, X_2) \cdot \mathrm{cov}(X_1, X_2)}{\mathrm{Var}(X_1) \cdot \mathrm{Var}(X_2) - \mathrm{cov}(X_1, X_2) \cdot \mathrm{cov}(X_1, X_2)}$$

$$\hat{\beta}_2 = \frac{cov(Y, X_2)\, Var(X_1) - cov(Y, X_1)\, cov(X_1, X_2)}{Var(X_1)\, Var(X_2) - cov(X_1, X_2)\, cov(X_1, X_2)}$$

Now, when we apply OLS with the presence of perfect MC. then we get,

$$Var(X_1) \cdot Var(X_2) - \{cov(X_1, X_2)\}^2$$

$$= K^2 \{Var(X_2)\}^2 - K \cdot Var(X_2) \cdot K \cdot Var(X_2)$$

$$= K^2 \cancel{\{Var(X_2)\}^2} - K^2 \cancel{\{Var(X_2)\}^2}$$

$$= 0$$

As denominator of $\hat{\beta}_1$ and $\hat{\beta}_2$ are zero, therefore we are unable to estimate the parameters with exact MC

<u>In case of near exact MC</u>

In case of near exact MC that is r_{12} close to ± 1. When $|r_{12}| > 0.5$, then the problem of multicollinearity arises in the model. Then if we apply OLS in the presence of near - exact MC then, parameters are estimated. But, their variances increases i.e they are not remain BLUE.

$$Var(\hat{\beta}_1) = \frac{\sigma_u^2}{\Sigma x_1^2 (1 - r_{12}^2)}$$

$$Var(\hat{\beta}_2) = \frac{\sigma_u^2}{\Sigma x_2^2 (1 - r_{12}^2)}$$

The term $\frac{1}{1-r_{12}^2}$ is known as variance inflationary factor (VIF)

$$Var(\hat{\beta}_1) = \frac{\sigma_u^2}{\Sigma x_1^2} \cdot (VIF)$$

$$Var(\hat{\beta}_2) = \frac{\sigma_u^2}{\Sigma x_2^2} (VIF)$$

(1) when $r_{12} = 0$ then $VIF = 1$

(2) when $r_{12} = \pm 1$ then $VIF = \infty$

VIF ranges from 1 to ∞

The value of VIF determines the existence of the model with application of OLS. When VIF ↑es then variances and SEs are over-estimated and 't' statistics are under-estimated. That we accept the false H_0. Therefore, we commit Type II error. The whole inference analysis will be erroneous. The model is not good fitted.

Scanned with CamScanner