31 January 2024 08:04 PM

Specification Bias

$E(UU') = \sigma^2 I$ (homoscedasticity)

Digression → Matrix approach to Linear Regression Model:

$$Y_1 = \beta_0 + \beta_1 X_{11} + \beta_2 X_{21} + \beta_3 X_{31} + \cdots + \beta_K X_{K1} + U_1$$

$$Y_2 = \beta_0 + \beta_1 X_{12} + \beta_2 X_{22} + \beta_3 X_{32} + \cdots + \beta_K X_{K2} + U_2$$

$$Y_3 = \cdots$$

$$\vdots$$

$$Y_n = \beta_0 + \beta_1 X_{1n} + \beta_2 X_{2n} + \beta_3 X_{3n} + \cdots + \beta_K X_{Kn} + U_n$$

Putting this equation in matrix form,

$$Y = X\beta + U$$

where

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \qquad X = \begin{bmatrix} 1 & X_{11} & \cdots & X_{K1} \\ 1 & X_{12} & & \vdots \\ \vdots & \vdots & & \vdots \\ 1 & X_{1n} & & X_{Kn} \end{bmatrix}$$

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_K \end{bmatrix} \quad \text{and } U = \begin{bmatrix} U_1 \\ U_2 \\ \vdots \\ U_n \end{bmatrix}$$

Assumptions:

1. $E(U) = 0$
2. $U = \begin{bmatrix} U_1 \\ U_2 \\ \vdots \\ U_n \end{bmatrix}$ $\qquad U' = \begin{bmatrix} U_1 & U_2 & \cdots & U_n \end{bmatrix}$

$$\begin{bmatrix} U_1^2 & U_1U_2 & U_1U_2 & \cdots & U_1U_n \end{bmatrix}$$

$$\begin{bmatrix} u_n \end{bmatrix}$$

$$UU' = \begin{bmatrix} u_1^2 & u_1u_2 & u_1u_3 & \cdots & u_1u_n \\ u_2u_1 & u_2^2 & u_2u_3 & \cdots & u_2u_n \\ \vdots & \vdots & & & \vdots \\ u_nu_1 & u_nu_2 & \cdots & \cdots & u_n^2 \end{bmatrix}$$

$$E(UU') = \begin{bmatrix} E(u_1^2) & E(u_1u_2) & E(u_1u_3) & \cdots & E(u_1u_n) \\ E(u_2u_1) & E(u_2^2) & E(u_2u_3) & \cdots & E(u_2u_n) \\ \vdots & & \ddots & & \vdots \\ E(u_nu_1) & \cdots & \cdots & \cdots & E(u_n^2) \end{bmatrix}$$

$$= \begin{bmatrix} \sigma^2 & cov(u_1u_2) & cov(u_1u_3) & \cdots & cov(u_1u_n) \\ \vdots & \sigma^2 & cov & & \\ & & \sigma^2 & & \\ cov) & cov) & & & \sigma^2 \end{bmatrix}$$

CLRM: $E(u^2) = \sigma^2$

$E(u_iu_j) = 0$

$$= \begin{bmatrix} \sigma^2 & 0 & 0 & 0 & \cdots & 0 \\ 0 & \sigma^2 & 0 & 0 & \cdots & 0 \\ \vdots & & & & & \\ 0 & 0 & 0 & 0 & \cdots & \sigma^2 \end{bmatrix}$$

$$= \sigma^2 \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

$$\boxed{E(UU') = \sigma^2 I_n}$$

— * —

Specification error : <u>Case of Explanatory Variable</u>

The use of ordinary least squares when some variables are left out may introduce bias into the estimates. Bias ... called specification bias.

The use of ordinary ... Bias

left out may introduce bias into the estimates. Bias that originates in this way is called specification bias.

For instance, the true function explaining variation in y is given as: $\boxed{y = \beta_1 x_1 + \beta_2 x_2 + u}$

However either due to ignorance of true relation or because of non-availability of data on x_2 following regression equation is estimated:

$$\boxed{y = \beta_1^* x_1 + v}$$

It can be shown that β_1^* is different from β_1.

On applying OLS to $y = \beta_1^* x_1 + v$

we obtain $\beta_1^* = \dfrac{\sum x_1 y}{\sum x_1^2}$.

On the other hand, the normal equations of the true function: $y = \beta_1 x_1 + \beta_2 x_2 + u$ are

$$\sum x_1 y = \beta_1 \sum x_1^2 + \beta_2 \sum x_1 x_2 \quad \text{— divide by } \sum x_1^2$$

$$\sum x_2 y = \beta_1 \sum x_1 x_2 + \beta_2 \sum x_2^2$$

$$\frac{\sum x_1 y}{\sum x_1^2} = \beta_1 + \beta_2 \left(\frac{\sum x_1 x_2}{\sum x_1^2}\right)$$

since $\beta_1^* = \dfrac{\sum x_1 y}{\sum x_1^2} \Rightarrow \beta_1^* = \beta_1 + \beta_2 \left(\dfrac{\sum x_1 x_2}{\sum x_1^2}\right)$

$\beta^* = \beta$ only if $\sum x_1 x_2 = 0$

$$\beta_1^* = \beta_1 \text{ only if } \frac{\Sigma x_1 x_2}{\Sigma x_1^2} = 0$$

$$\text{if } \Sigma x_1 x_2 = 0$$

$$\therefore \text{ Specification error} = \underbrace{(\beta_1^* - \beta_1)}_{} = \beta_2 \left(\frac{\Sigma x_1 x_2}{\Sigma x_1^2}\right)$$

It is hence proved that β_1^* from incorrect specification (through OLS procedure) is a biased estimate of parameter β_1

The bias, which is equal to $\left[\beta_2 \left\{\frac{\Sigma x_1 x_2}{\Sigma x_1^2}\right\}\right]$ depends on two terms, namely, the regression coefficient of ommited variable in the true relation (β_2) and the covariance of the omitted variable with the included variable $\frac{\Sigma x_1 x_2}{\Sigma x_1^2}$.

Variance of β_1^*

for $y = \beta_1^* x_1 + v$

$$\text{var}(\beta_1^*) = \frac{\sigma_u^2}{\Sigma x_1^2} = \frac{\Sigma e_i^2/(n-2)}{\Sigma x_1^2}$$

$$\text{var}(\beta_1^*) = \frac{\Sigma (y - \beta_1^* x_1)^2}{}$$

$$\operatorname{var}(\beta_1^*) = \frac{\sum (y - \beta_1^* x_1)^2}{(n-2)\sum x_1^2}$$

$$y = \beta_1 x_1 + \beta_2 x_2 + u \Rightarrow \operatorname{var}(\beta_1^*) = \frac{\sum \left(\beta_1 x_1 + \beta_2 x_2 - \beta_1^* x_1\right)^2}{(n-2)\sum x_1^2}$$

(true equation

$$= \frac{\sum \left[-(\beta_1^* - \beta_1)x_1 + \beta_2 x_2\right]^2}{(n-2)\sum x_1^2}$$

$$= \frac{\sum \left\{(\beta_1^* - \beta_1)x_1\right\}^2}{(n-2)\sum x_1^2} + \frac{\sum \beta_2^2 x_2^2}{(n-2)\sum x_1^2}$$

$$= \operatorname{var}(\beta_1^*) + \frac{\sum \beta_2^2 x_2^2}{n-2 \sum x_1^2}$$

This implies that estimator of β_1^* is positively biased.

Therefore, the usual tests of significance concerning β_1 shall be invalid in present circumstances.

The estimator of the constant intercept of the incorrectly specified function turns out to be biased

$$E(\beta_0^*) = E(\bar{y} - \beta_1 \bar{x}_1)$$

$$= E(\beta_0 + \cancel{\beta_1 \bar{x}_1} + \beta_2 \bar{x}_2 - \cancel{\beta_1 \bar{x}_1})$$

$$= E(\beta_0 + \beta_2 \bar{x}_2)$$

$$E(\beta_0^*) = \beta_0 + \beta_2 \bar{x}_2$$

↳ biasedness

β_0^* will be unbiased if $\bar{x}_2 = 0$.

The above discussion can now be summarised as follows:

The omitted or left out explanatory variable is

(i) If the ommitted or left out explanatory variable is correlated with the included explanatory variable, the OLS estimator of β_1 will be biased and inconsistent.

(ii) If the ommitted variable is not correlated with the included variable, the estimator of constant term will still be biased and inconsistent, but estimator of β_1 will be unbiased.

(iii) The variance of β_1 will contain an upward bias. Therefore, the test on significance of the estimator would not lead to correct conclusions.