Omitted Variable Bias

Deterministic Regressors:

True: $y_i = \beta_1 + \beta_2 x_i + \beta_3 w_i + \varepsilon_i$

Est: $y_i = \hat{\beta}_1 + \hat{\beta}_2 x_i + \hat{u}_i$

$u_i = \beta_3 w_i + \varepsilon_i$

$E(u_i) = \beta_3 w_i$

TGM:

$E(\varepsilon_i) = 0$

$E(\varepsilon_i^2) = \sigma_\varepsilon^2$

$E(\varepsilon_i \varepsilon_j) = 0$

$$\hat{\beta}_2 = \frac{\widehat{Cov}(x, y)}{\widehat{Var}(x)} = \frac{\widehat{Cov}(X, \beta_1 + \beta_2 X + \beta_3 W + \varepsilon)}{\widehat{Var}(X)} =$$

$$\frac{\overset{0}{\widehat{Cov}(X, \beta_1)} + \beta_2 \widehat{Cov}(X, X) + \beta_3 \widehat{Cov}(X, W) + \widehat{Cov}(X, \varepsilon)}{\widehat{Var}(X)} =$$

$$= \beta_2 + \beta_3 \frac{\widehat{Cov}(X, W)}{\widehat{Var}(X)} + \frac{\widehat{Cov}(X, \varepsilon)}{\widehat{Var}(X)}$$

$\bar{\varepsilon} \neq 0$

$\bar{\varepsilon} \xrightarrow{P} E(\varepsilon) = 0$

$$E(\hat{\beta}_2) = \beta_2 + \beta_3 E\left(\frac{\widehat{Cov}(X, W)}{\widehat{Var}(X)}\right) + E\left(\frac{\widehat{Cov}(X, \varepsilon)}{\widehat{Var}(X)}\right) =$$

$\hat{\bar{\varepsilon}} = 0$

$$= \beta_2 + \beta_3 \frac{\widehat{Cov}(X, W)}{\widehat{Var}(X)} + \frac{E(\widehat{Cov}(X, \varepsilon))}{\widehat{Var}(X)} =$$

$E(\widehat{Var}(x)) = E\left(\frac{1}{n-1} \sum (x_i - \bar{x})^2\right)$

$$= \beta_2 + \beta_3 \cdot \underbrace{\frac{\widehat{Cov}(X, W)}{\widehat{Var}(X)}}_{\text{bias}}$$

$E(\widehat{Cov}(X, \varepsilon)) =$

$E\left(\frac{1}{n-1} \sum (X_i - \bar{X})(\varepsilon_i - \bar{\varepsilon})\right) =$

bias $= 0$ $\quad \beta_3 = 0$ $\quad \widehat{Cov}(X, W) = 0$ or $X \perp W \,/\, \sum x_i w_i = 0$ $\quad \frac{1}{n-1} \sum (X_i - \bar{X})(E \varepsilon_i - E(\bar{\varepsilon})) = 0$

Stochastic Regressors:

True: $y_i = \beta_1 + \beta_2 x_i + \beta_3 w_i + \varepsilon_i$

Est: $y_i = \hat{\beta}_1 + \hat{\beta}_2 x_i + \hat{e}_i$

$[E(\varepsilon_i | X) = 0] \Rightarrow \mathrm{cov}(\varepsilon, X)$

① Deterministic (t)

$E(\varepsilon_i^2 | X) = \sigma^2$

$y_t = \beta_1 + \beta_2 t + \varepsilon_t$

$E(\varepsilon_i \varepsilon_j | X) = 0$

② Stochastic

$[\pi_t = \beta_1 + \beta_2 \pi_{t-1} + \varepsilon_t]$

$\pi_{t-1} = \beta_1 + \beta_2 \pi_{t-2} + \varepsilon_{t-1}$

$$\operatorname{plim} \hat{\beta}_2 = \operatorname*{plim}_{n \to \infty} \frac{\widehat{\mathrm{Cov}}(x, y)}{\widehat{\mathrm{Var}}(x)} =$$

$$= \frac{\mathrm{Cov}(x_1, y_1)}{\mathrm{Var}(X_1)} = \frac{\mathrm{Cov}(X_1, \beta_1 + \beta_2 X_1 + \beta_3 W_1 + \varepsilon_1)}{\mathrm{Var}(X_1)} =$$

$$\underbrace{\beta_2 + \beta_3 \frac{\mathrm{Cov}(X_1, W_1)}{\mathrm{Var}(X_1)}}_{\text{bias}} + \frac{\mathrm{Cov}(X_1, \varepsilon_1)^{=0}}{\mathrm{Var}(X_1)}$$

$$P(X, Y) = P(X)P(Y)$$
$$\Downarrow$$
$$E(X|Y) = E(X)$$
$$\Downarrow$$
$$\mathrm{cov}(X, Y) = 0$$

$$\mathrm{Cov}(X, \varepsilon) = E(X\varepsilon) - E(X) \cdot E(\varepsilon) = 0$$

$E(\varepsilon) = E(E(\varepsilon | X)) = 0$

$$E(X\varepsilon) = E(E(X\varepsilon | X)) = E(X \cdot E(\varepsilon | X)) = 0$$

Deterministic Regressors:

True: $y_i = \beta_1 + \beta_2 x_i + \overset{0}{\beta_3} W + \varepsilon_i$

Est: $y_i = \hat{\beta}_1 + \hat{\beta}_2 x_i + \hat{\beta}_3 w_i + \hat{\varepsilon}_i$

$E(\varepsilon_i) = 0$

$E(\varepsilon_i^2) = \sigma^2$

$E(\varepsilon_i \varepsilon_j) = 0$

① $$Var(\hat{\beta}_k) = \frac{\sigma_\varepsilon^2}{TSS_k \cdot (1 - R_k^2)}$$

R_k^2 : R^2 in $X_k | X_{-k}$

$$Var(\hat{\beta}_2) = \frac{\sigma_\varepsilon^2}{TSS_x \cdot (1 - \hat{\rho}_{x,w}^2)}$$

② $df \downarrow$ t-test $t \sim t_{n-k}$

③ $$\hat{\beta}_2 = \frac{\widehat{Cov}(X,Y) \cdot \widehat{Var}(W) - \widehat{Cov}(W,Y) \cdot \widehat{Cov}(X,W)}{\widehat{Var}(X) \cdot \widehat{Var}(W) - (Cov(X,W))^2}$$

RESET (Ramsey) Test

$$y_i = \gamma_1 + \gamma_2 X_{1i} + \gamma_3 X_{2i} + u_i \quad \Rightarrow \hat{y}$$

$$\hat{y}^2 = (\hat{\gamma}_1 + \hat{\gamma}_2 X_1 + \hat{\gamma}_3 X_2)^2$$

$$y_i = \gamma_1 + \gamma_2 X_{1i} + \gamma_3 X_{2i} + \gamma_4 \hat{y}^2 + \left[\gamma_5 \hat{y}^3\right] + \varepsilon$$

t-test $H_0: \gamma_4 = 0$ F-test $H_0: \gamma_4 = \gamma_5 = 0$

$t \sim t_{N-4}$

Proxy Variables

$$y_i = \beta_1 + \beta_2 k_i + \beta_3 \, tech_i + u_i$$

exp_i $\quad cov(exp_i, tech_i) \neq 0$

– proxy $\quad cov(exp_i, u_i) = 0$

$$tech_i = \lambda + \gamma \, exp_i + \varepsilon_i$$

$$y_i = \beta_1 + \beta_2 k_i + \beta_3 (\lambda + \gamma \, exp_i + \varepsilon_i) + u_i$$

$$y_i = (\beta_1 + \beta_3 \cdot \lambda) + \beta_2 k_i + \underbrace{\beta_3 \gamma}\, exp_i + \underbrace{\beta_3 \varepsilon_i + u_i}$$