Stochastic Regressors

Question 1. (UoL Exam). Explain what is correct, mistaken, confused or incomplete in the following statement:

"When an explanatory variable in a regression model has a random component, it is described as a stochastic regressor. When a stochastic regressor is used in a regression model, the Gauss-Markov condition that the explanatory variables should be independent of the disturbance term is violated. Consequently OLS regression estimates will be biased. However, they will be consistent because the bias will disappear in large samples."

↳ if bias disappears ⇒ consistent

$\text{cov}(X, \varepsilon) = 0$ $\quad E(\hat{\beta}) \neq \beta$

Endogeneous regressor $\quad \text{cov}(X, \varepsilon) \neq 0$

1. omitted variable
2. measurement errors
3. simultaneity

GMT : 1. $y = X\beta + u$

2. w.p. 1 no perfect multicollinearity

3. $E(u|X) = 0$

$$\text{Var}(u|X) = \sigma^2 I \qquad \begin{bmatrix} \sigma^2 & & 0 \\ & \ddots & \\ 0 & & \sigma^2 \end{bmatrix}$$

$\hat{\beta}_{OLS}$ – BLUE (linear by y, conditionally unbiased)

$$E(\hat{\beta} | X) = \beta$$

Q1. $$Y = \beta_2 Z + u$$

$$X = Z + w \qquad E(w) = 0$$

w and Z are ind.

$$Y = \beta_2 X + \underbrace{u - \beta_2 w}_{v}$$

$$\hat{\beta}_2 = \frac{\sum X_i Y_i}{\sum X_i^2} = \frac{\sum X_i \cdot (\beta_2 \cdot X_i + u_i - \beta_2 w_i)}{\sum X_i^2} =$$

$$= \beta_2 + \frac{\sum X_i u_i}{\sum X_i^2} - \beta_2 \frac{\sum X_i w_i}{\sum X_i^2} =$$

$$= \beta_2 + \frac{\sum (X_i - \bar{X})(u_i - \bar{u}) + n\bar{X}\bar{u}}{\sum (X_i - \bar{X})^2 + n\bar{X}^2} -$$

$$- \beta_2 \frac{\sum (X_i - \bar{X}) \cdot (w_i - \bar{w}) + n \cdot \bar{X} \cdot \bar{w}}{\sum (X_i - \bar{X})^2 + n\bar{X}^2}$$

$$\xrightarrow{p} \beta_2 + \frac{\overset{=0}{Cov(X, u)} + \overset{=0}{\mu_X \mu_u}}{Var(X) + \mu_X^2} -$$

$$- \beta_2 \cdot \frac{Cov(X, w) + \overset{=0}{\mu_X \mu_w}}{Var(X) + \mu_X^2} =$$

$$= \beta_2 - \beta_2 \cdot \frac{Cov(z + w, w)}{var(x) + \mu_x^2} =$$

$$= \beta_2 - \beta_2 \frac{\sigma_w^2}{\sigma_x^2 + \mu_x^2}$$

Instrumental variables

(2 step least squares)

$y_i = \beta_1 + \beta_2 \cdot x_i + \varepsilon_i$

↑ endogenous

$cov(x_i, \varepsilon_i) \neq 0$

$$\begin{tikzcd} X \arrow[r, leftrightarrow] \arrow[d, leftrightarrow] & \varepsilon \\ z \arrow[ur, dotted, no head] & \end{tikzcd}$$

$Z \nearrow$ relevant $cov(z_i, x_i) \neq 0$.

$Z \searrow$ exogenous $Cov(z_i, \varepsilon_i) = 0$.

↑ z – instrumental variable
(valid)

1 step. $\hat{x}_i = \hat{\theta}_1 + \hat{\theta}_2 \cdot z_i$ $\quad \hat{\theta}_2 = \frac{\widehat{Cov}(x, z)}{\widehat{Var}(x)}$

2 step. $y_i = \beta_1 + \beta_2 \cdot \hat{x}_i + \varepsilon_i$

$\hat{\beta}_2^{TSLS}$ (IV)

$$\hat{\beta}_2^{TSLS} = \frac{\widehat{cov}(y,z)}{\widehat{cov}(x,z)}$$

$$\hat{\beta}_2^{TSLS} = \frac{\widehat{cov}(\hat{x},y)}{\widehat{var}(\hat{x})} = \frac{\widehat{cov}(\hat{\theta}_1 + \hat{\theta}_2 \cdot z, y)}{\widehat{var}(\hat{\theta}_1 + \hat{\theta}_2 z)} =$$

$$= \frac{\hat{\theta}_2 \, \widehat{cov}(z,y)}{\hat{\theta}_2^2 \, \widehat{var}(z)} = \frac{\widehat{cov}(z,y)}{\hat{\theta}_2 \, \widehat{var}(z)} =$$

$$= \frac{\widehat{cov}(z,y)}{\frac{\widehat{cov}(x,z)}{\widehat{var}(x)} \widehat{var}(z)} = \frac{\widehat{cov}(z,y)}{\widehat{cov}(x,z)}$$

$$\hat{\beta}_2^{TSLS} \xrightarrow{p} \frac{cov(z_i,y_i)}{cov(X_i,z_i)} =$$

$$= \frac{cov(z_i, \beta_1 + \beta_2 X_i + \varepsilon_i)}{cov(X_i, z_i)} =$$

$$= \beta_2 \cdot \overset{\text{relevance}}{\frac{cov(z_i,X_i)}{\underset{\neq 0}{cov(z_i,X_i)}}} + \frac{\overset{\text{exogenous}}{\overset{0}{\overset{\|}{cov(z_i,\varepsilon_i)}}}}{cov(z_i,X_i)} = \beta_2$$

Q_2. $\begin{cases} \text{Demand: } \ln Q_i = \beta_1 + \overset{\beta_2 < 0}{\beta_2} \ln P_i + \varepsilon_i \\ \text{Supply: } \ln Q_i = \gamma_1 + \underset{\gamma_2 > 0}{\gamma_2} \cdot \ln P_i + \gamma_3 \cdot \ln T_i + u_i \end{cases}$

a) $\ln Q_i = \hat{\beta}_1 + \hat{\beta}_2 \ln P_i + \varepsilon_i$

1. omitted var.
2. meas. err.
3. simult.

$\hat{\beta}_2^{OLS}$ consistent?

$$\beta_1 + \beta_2 \ln P_i + \varepsilon_i = \gamma_1 + \gamma_2 \cdot \ln P_i + \gamma_3 \ln T_i + u_i$$

$$\ln P_i = \frac{\beta_1 - \gamma_1 - \gamma_3 \cdot \ln T_i - u_i + \varepsilon_i}{\gamma_2 - \beta_2}$$

$$\operatorname{cov}(\ln P_i, \varepsilon_i) = \operatorname{cov}\left(\frac{\beta_1 - \gamma_1 - \gamma_3 \cdot \ln T_i - u_i + \varepsilon_i}{\gamma_2 - \beta_2}, \varepsilon_i \right) =$$

$$= \frac{1}{\gamma_2 - \beta_2} \operatorname{cov}(- \gamma_3 \cdot \ln T_i - u_i + \varepsilon_i, \varepsilon_i) =$$

$$= \frac{1}{\gamma_2 - \beta_2} \left(- \gamma_3 \cdot \underbrace{\operatorname{cov}(\ln T_i, \varepsilon_i)}_{=0} - \underbrace{\operatorname{cov}(u_i, \varepsilon_i)}_{=0} + \sigma_\varepsilon^2 \right)$$

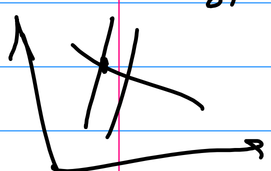
$$= \frac{\sigma_\varepsilon^2}{\gamma_2 - \beta_2} \neq 0$$

$$\hat{\beta}_2^{OLS} = \frac{\widehat{cov}(\ln P, \ln Q)}{\widehat{var}(\ln p)} = \beta_2 + \frac{\widehat{cov}(\ln p, \varepsilon)}{var(\ln P)} \xrightarrow{p}$$

$$\beta_2 + \frac{\sigma_\varepsilon^2 / \gamma_2 - \beta_2}{\underbrace{\sigma^2_{\ln p}}_{> 0}} \neq \beta_2$$

$$(\gamma_2 - \beta_2) > 0$$

since $\beta_2 < 0$ it will be closer to 0.



b) Demand: $\ln Q_i = \beta_1 + \beta_2 \ln P_i + \varepsilon_i$

Supply: $\ln Q_i = \gamma_1 + \gamma_2 \cdot \ln P_i + \gamma_3 \cdot \ln T_i + u_i$

$cov(\ln T_i, \varepsilon_i) = 0 \Rightarrow$ exogenous

$cov(\ln T_i, \ln P_i) \neq 0 \Rightarrow$ relevant

$\ln T_i$ - valid instrument

1) $\ln P_i \mid \ln T_i$

2) $\ln Q \mid \widehat{\ln P_i}$ $\quad \hat{\beta}_2^{TSLS}$ - consistent