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."

GMT

- $y = X\beta + u$
- r.p.1 no perf. multicollinearity
- $E(u \mid X) = 0$

$\rightarrow$ $Var(u \mid X) = \sigma^2 I$ $\quad \begin{bmatrix} \sigma^2 & & 0 \\ & \ddots & \\ 0 & & \sigma^2 \end{bmatrix}$

$\Rightarrow$ $E(\hat{\beta} \mid X) = \beta$

$\hat{\beta}$ - BLUE (in class of linear by y conditionally unbiased estimators)

Question 2. (UoL Exam) A variable Y is determined by the model

$$Y = \beta_1 + \beta_2 Z + v,$$

where Z is a stochastic variable and v is a disturbance term that satisfies the Gauss–Markov conditions. The explanatory variable is subject to measurement error and is measured as X where $X = Z + w$ and w is the measurement error, distributed independently of v.
Describe analytically the consequences of using OLS to fit this model. It is assumed that expected value of w is 0, and w is distributed independently of Z.

$$Y = \beta_1 + \beta_2 X \underbrace{- \beta_2 w + v}_{u}$$

Endogeneity
- Omitted variable
- Measurement error
- Simultaneity

$$\hat{\beta}_2 = \beta_2 + \frac{\widehat{Cov}(X, u)}{\widehat{Var}(X)} \xrightarrow{p}$$

$$\beta_2 + \frac{Cov(X, u)}{Var(X)} = \frac{Cov(Z + w, -\beta_2 w + v)}{Var(X)} =$$

$$= \beta_2 + \frac{\overbrace{Cov(Z, -\beta_2 w)}^{=0} + Cov(w, -\beta_2 w)}{Var(X)} =$$

$$= \beta_2 - \beta_2 \frac{Var(w)}{Var(Z + w)} = \beta_2 - \beta_2 \frac{\sigma_w^2}{\sigma_z^2 + \sigma_w^2}$$

$$= \beta_2 \cdot \frac{\sigma_z^2}{\sigma_z^2 + \sigma_w^2}$$

Instrumental variables (2 step OLS)

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

$\uparrow$ endogenous

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

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

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

$\searrow$ exogenous $cov(z_i, \varepsilon_i) = 0$

z – instrumental variable

1 STEP: $\hat{x}_i = \hat{\theta}_1 + \hat{\theta}_2 z_i$ $\qquad \hat{\theta}_2 = \dfrac{\widehat{cov}(z,x)}{\widehat{var}(z)}$

2 STEP: $y_i = \hat{\beta}_1 + \hat{\beta}_2 \cdot \hat{x}_i + \varepsilon_i$

$$\hat{\beta}_2^{TSLS} = \frac{\widehat{cov}(y,z)}{\widehat{cov}(x,z)} = \frac{\Sigma (y_i - \bar{y})(z_i - \bar{z})}{\Sigma (x_i - \bar{x})(z_i - \bar{z})}$$

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

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

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

$$\hat{\beta}_2 \xrightarrow{p} \frac{\operatorname{cov}(y_i, z_i)}{\operatorname{cov}(x_i, z_i)} =$$

$$= \frac{\operatorname{cov}(\beta_1 + \beta_2 \cdot x_i + \varepsilon_i,\ z_i)}{\operatorname{cov}(x_i, z_i)} =$$

exogenous

$$= \frac{\beta_2 \operatorname{cov}(x_i, z_i) + \underbrace{\operatorname{cov}(\varepsilon_i, z_i)}_{=0}}{\operatorname{cov}(x_i, z_i)}$$

$$= \frac{\beta_2 \cdot \operatorname{cov}(x_i, z_i)}{\underbrace{\operatorname{cov}(x_i, z_i)}_{\neq 0}} = \beta_2$$

relevance

Q1. Demand: $\ln Q_i = \beta_1 + \overset{\beta_2 < 0}{\beta_2} \ln P_i + \varepsilon_i \longrightarrow OLS$

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

$\gamma_2 > 0$

a) $\hat{\beta}_2^{OLS}$ - inconsistent?

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

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

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

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

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

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

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

$$\beta_2 + \frac{cov(\ln P, \varepsilon)}{var(\ln P)} = \beta_2 + \underbrace{\frac{\sigma_\varepsilon^2}{(\gamma_2 - \beta_2)\,\sigma_{\ln P}^2}}_{>0} \neq \beta_2$$

$$\gamma_2 - \beta_2 > 0$$

$\beta_2 < 0 \implies \beta_2$ will be closer to 0

b) $cov(\ln T_i, \varepsilon_i) = 0$

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

$\implies \ln T_i$ – valid instrument

1 step $\ln P_i \mid \ln T_i \implies \widehat{\ln P_i}$

2 step $\ln Q_i \mid \widehat{\ln P_i}$

$$\Downarrow$$

$$\ln Q_i = \beta_1 + \beta_2 \widehat{\ln P_i} + \varepsilon_i$$

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