Stochastic Regressors

Endogenous regressor $\quad Corr(x_i, \varepsilon_i) \neq 0$

$$\begin{array}{ccccc} L^s & \xrightarrow{s \geq 2} & L^2 & & \\ & & \downarrow & & \\ a.s. & \longrightarrow & p & \longrightarrow & d \end{array}$$

in Prob $\quad \lim\limits_{n \to \infty} P_r(|X_n - X| > \varepsilon) = 0 \qquad \begin{cases} 1) & E(X_n) = X \\ 2) & Var(X_n) \to 0 \end{cases}$

in distr. $\quad \lim\limits_{n \to \infty} F_n(x) = F(x)$

LLN: $\quad y_1, \ldots, y_n \quad$ i.i.d. $\quad E(y_1) = \mu \quad Var(y_1) = \sigma^2 < \infty$

$$\bar{y} \xrightarrow{p} \mu$$

CLT: $\quad y_1, \ldots, y_n \quad$ i.i.d. $\quad E(y_1) = \mu \quad Var(y_1) = \sigma^2 < \infty$

$$\frac{\sqrt{n}(\bar{y} - \mu)}{\sigma} \xrightarrow{d} N(0,1)$$

Slutsky: $\quad X_n \xrightarrow{p} a \quad \Rightarrow \quad g(X_n) \xrightarrow{p} g(a)$

$$\begin{array}{l} X_n \xrightarrow{p} a \\ y_n \xrightarrow{p} b \end{array} \quad \Rightarrow \quad X_n + y_n \xrightarrow{p} a + b$$

GMT

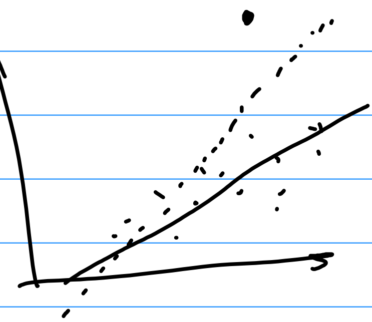
1. $y_i = \beta_0 + \beta_1 x_{1i} + \ldots + \beta_k x_{ki} + \varepsilon_i$

2. $\{ (x_{1i}, \ldots, x_{ki}, y_i), \ i = \overline{1,n} \}$ i.i.d.

3. $E(x_{ji}^4) < \infty \quad , j = \overline{1,k}$

   $E(y_i^4) < \infty$

   <=> no outliers



→ 4. $E(\varepsilon_i \mid x_{1i}, \ldots, x_{ki}) = 0$

   ↳ $E(\varepsilon_i) = 0$

   ↳ $cov(\varepsilon_i, x_i) = 0$

5. $rank(X) = k + 1 \quad$ w.p. 1

1-5 => $\hat{\beta}_{OLS}$ - consistent & as. normal

1) $x_i \perp \varepsilon_i$

=> $\hat{\beta}$ - unbiased
consistent

2) $cov(x_i, \varepsilon_i) = 0$

=> $\hat{\beta}$ - consistent
(conditionally
unbiased)

3) $cov(x_i, \varepsilon_i) \neq 0$ => inconst.

Col. 1 : $\hat{Var}(x) \longrightarrow Var(x_i)$

$\hat{cov}(x, y) \longrightarrow cov(x_i, y_i)$

$$Cov(x_i, y_i) = E(x_i y_i) - E(x_i)E(y_i)$$

$$\hat{cov}(x, y) = \overline{xy} - \bar{x} \cdot \bar{y}$$

$$\left.\begin{array}{l} \bar{x} \xrightarrow{p} E(x_i) \\ \bar{y} \xrightarrow{p} E(y_i) \\ \overline{xy} \xrightarrow{p} E(x_i y_i) \end{array}\right\} \text{by LLN}$$

$$\left.\begin{array}{l} \bar{x} \cdot \bar{y} \xrightarrow{p} E(x_i) \cdot E(y_i) \\ \overline{xy} - \bar{x} \cdot \bar{y} \xrightarrow{p} E(x_i y_i) - E(x_i) \cdot E(y_i) \end{array}\right\} \text{by Slutsky}$$

Prob. 1. $y_i = \alpha + \beta x_i + \varepsilon_i$

$$\hat{\beta} = \frac{\Sigma (x_i - \bar{x})(y_i - \bar{y})}{\Sigma (x_i - \bar{x})^2}$$

$$\text{plim}\ \hat{\beta} = \text{plim} \frac{\Sigma (x_i - \bar{x})(y_i - \bar{y}) \cdot 1/N}{\Sigma (x_i - \bar{x})^2 \cdot 1/N}$$

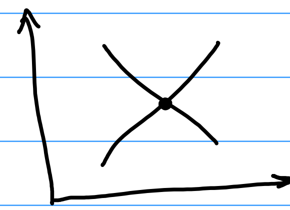
$$= \frac{\text{plim}\ \hat{cov}(x, y)}{\text{plim}\ \hat{var}(x)} = \frac{cov(x_i, y_i)}{Var(x_i)} =$$

$$= \frac{Cov(X_i,\ \alpha + \beta \cdot X_i + \varepsilon_i)}{Var(X_i)} = \beta \frac{Cov(\cancel{X_i}, X_i)}{\cancel{Var(X_i)}} + \frac{Cov(X_i, \varepsilon_i)}{Var(X_i)}$$

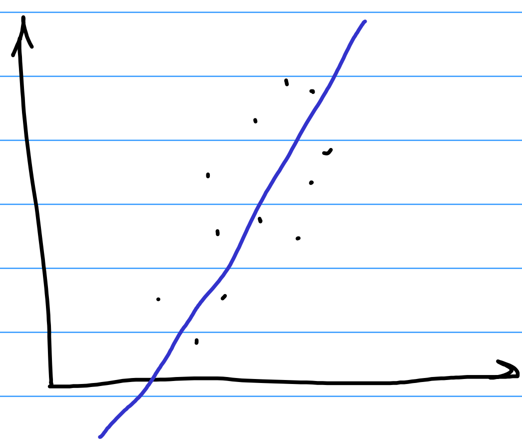
$$= \beta + \underbrace{\frac{\overset{\to 0}{Cov(X_i, \varepsilon_i)}}{Var(X_i)}}_{bias = 0} = \beta \quad \Rightarrow \text{consistent}$$

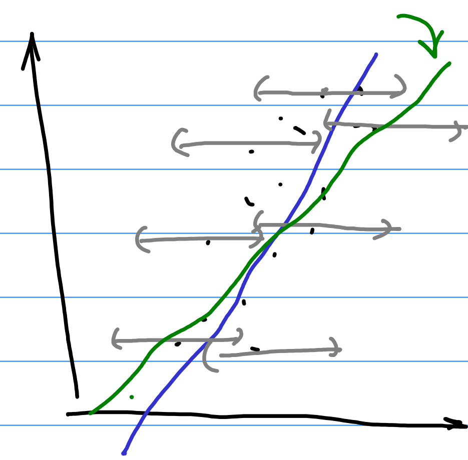
Endogeneity
- → omitted variable
- → measurement error (for X)
- → simultaneity ($y \leftarrow X$ & $y \rightarrow X$)

$\left\{ \begin{array}{l} P \\ Q \end{array} \right.$



Prob. 2 :





True model $\quad y_i = \beta_1 + \beta_2 X_i^*$ ← true X $\quad \Rightarrow y_i = \beta_1 + \beta_2 (X_i - \varepsilon_i)$

$X_i = X_i^* + \varepsilon_i$ ← meas. error $\quad = \beta_1 + \beta_2 \cdot X_i - \beta_2 \cdot \varepsilon_i$

↑ est X (data)

$\Rightarrow u_i = -\beta_2 \cdot \varepsilon_i$

Est model $y_i = \beta_1 + \beta_2 \cdot X_i + \underbrace{u_i}$

$$\hat{\beta}_2 \xrightarrow{p} \beta_2 + \frac{cov(x_i, u_i)}{Var(x_i)} = \beta_2 + \frac{cov(X_i^* + \varepsilon_i, -\beta_2 \cdot \varepsilon_i)}{Var(X_i^* + \varepsilon_i)} =$$

$$= \beta_2 - \beta_2 \frac{\sigma_\varepsilon^2}{\sigma_{x^*}^2 + \sigma_\varepsilon^2} = \beta_2 \cdot \underbrace{\frac{\sigma_{x^*}^2}{\sigma_{x^*}^2 + \sigma_\varepsilon^2}}_{<1\ !} \neq \beta_2$$

=> inconsistent

β_2 is biased towards 0