Stochastic Regressors

Endogeneous regr. $\text{corr}(x_i, \varepsilon_i) \neq 0$

$$\begin{tikzcd}
L^s \arrow[r] & L^2 \arrow[d] & \\
a.s. \arrow[r] & \textcircled{p} \arrow[r] & \textcircled{d}
\end{tikzcd}$$

in prob $\lim_{n \to \infty} \Pr(|X_n - X| < \varepsilon) = 1$

in dist $\lim_{n \to \infty} F_n(x) = F(x)$

LLN: $y_1, \dots, y_n$ i.i.d. $E(y_i) = \mu$ $\text{Var}(y_i) < \infty$ $= \sigma^2$

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

CLT: $y_1, \dots, y_n$ i.i.d. $E(y_i) = \mu$ $\text{Var}(y_i) < \infty$ $= \sigma^2$

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

Slutsky thm: $X_n \xrightarrow{p} a$

$$g(X_n) \xrightarrow{p} g(a)$$

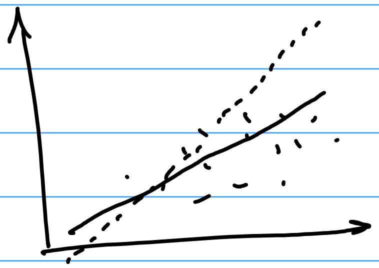
TGM

1) $y_i = \beta_0 + \beta_1 x_{1i} + \dots + \beta_k x_{ki} + \varepsilon_i$

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

3) $E(x_{ji}^4) < \infty, \qquad E(y_i^4) < \infty$

$j = \overline{1,k}$



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

5) no perf. m.c.

$\hat{\beta}_{OLS}$ - consistent and as. normal

Col. 1. $\quad \hat{Cov}(x,y) \xrightarrow{P} Cov(x_i, y_i)$

$$Cov(x_i, y_i) = E(X_i\, y_i) - E(X_i)\cdot E(y_i)$$

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

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

$$\bar{x}\cdot\bar{y} \xrightarrow{P} E(X_i)\cdot E(y_i) \quad \leftarrow \text{by Slutsky thm}$$

$$\overline{x\cdot y} - \bar{x}\cdot\bar{y} \xrightarrow{P} E(x_i\cdot y_i) - E(x_i)E(y_i)$$

Ex. 1.

$$y_i = \alpha + \beta x_i + \varepsilon_i$$

① $x_i \perp \varepsilon_i$

$\Rightarrow \hat{\beta}$ – unbiased and consistent

② $cov(x_i, \varepsilon_i) = 0$

$\Rightarrow \hat{\beta}$ – consistent

(cond. unbiased)

③ $cov(x_i, \varepsilon_i) \neq 0$

$\Rightarrow \hat{\beta}$ – inconsistent and biased

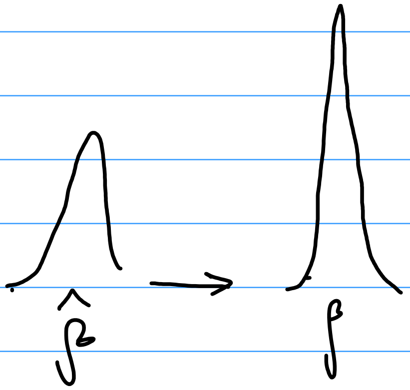
$$\hat{\beta} = \frac{\hat{cov}(x,y)}{\hat{var}(x)} \xrightarrow{p} \frac{cov(x_i, y_i)}{var(x_i)}$$

$$plim(\hat{\beta}) = plim\left(\frac{\hat{cov}(x,y)}{\hat{var}(x)}\right) =$$

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

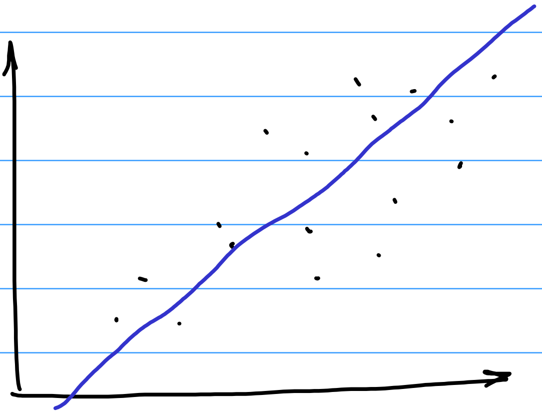
$$\textcircled{\star} \left[plim \frac{\Sigma X}{\Sigma y} = plim \frac{\Sigma X/N}{\Sigma y/N} = \frac{E(x)}{E(y)} \right]$$

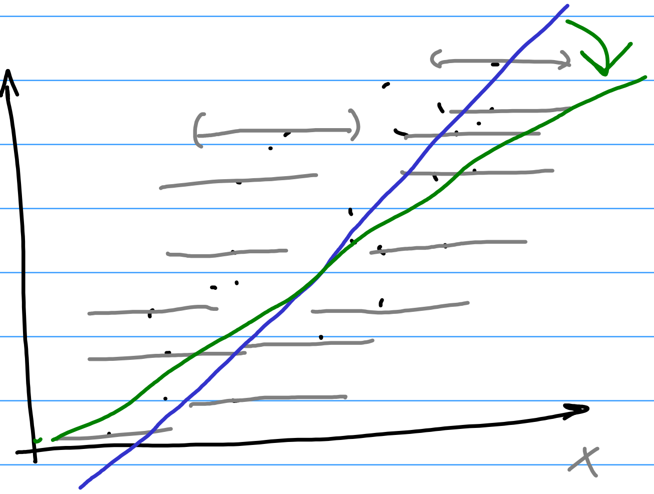
$$\textcircled{=} \frac{cov(x_i, \alpha + \beta x_i + \varepsilon_i)}{var(x_i)} = \beta \frac{\cancel{cov(x_i, x_i)}}{\cancel{var(x_i)}} + \underbrace{\frac{\overset{0}{cov(\varepsilon_i, x_i)}}{var(x_i)}}_{\substack{bias \\ =0}} = \beta$$



Endogeneity
- → Omitted variable
- → measurement error (of X)
- → simultaneity ($y \leftarrow x$ & $x \leftarrow y$)

Measurement error





True regr.: $y_i = \beta_1 + \beta_2 \cdot x_i^*$ ← true x

$x_i = x_i^* + \varepsilon_i$

data

Est. regr: $y_i = \beta_1 + \beta_2 x_i + u_i$

- $y_i = \beta_1 + \beta_2 (x_i - \varepsilon_i) = \beta_1 + \beta_2 x_i \overbrace{- \beta_2 \varepsilon_i}^{u_i}$

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

$$= \beta_2 - \beta_2 \frac{Var(\varepsilon_i)}{Var(x_i^* + \varepsilon_i)} = \beta_2 - \beta_2 \frac{\sigma_\varepsilon^2}{\sigma_{x^*}^2 + \sigma_\varepsilon^2} = \frac{\sigma_{x^*}^2}{\sigma_{x^*}^2 + \sigma_\varepsilon^2} \cdot \beta_2$$