Consistent estimators

Def

consistency : $\hat{\beta} \xrightarrow[p]{} \beta$ ← true value of unknown parameter

means $\lim_{n \to \infty} \mathbb{P}(|\hat{\beta}_n - \beta| \geq \varepsilon) = 0 \quad \forall \varepsilon > 0$

↗ sample size

So with large sample our estimation of β will be close to the true value more likely (good point to have more funding for research)

Task 0 unbiased | consistent

from SP course we know that

lemma If $E\hat{\beta}_n = \beta$ and $\mathrm{Var}\,\hat{\beta}_n \xrightarrow[n \to \infty]{} 0$

then $\hat{\beta}_n \xrightarrow[p]{} \beta$

(from chebyshev's inequation)

But sometimes unbiased est. can be also consistent

Let's $Y_1 \dots Y_n \overset{iid}{\sim} F(\mu, \sigma^2)$

- is $\overline{Y_n}$ an unbiased and consistent est. of μ ?

1) $E(\overline{Y_n}) = E\left(\frac{1}{n}\sum_{i=1}^{n} Y_n\right) = \frac{1}{n}\sum_{i=1}^{n} EY_i = \frac{n}{n} EY_5 =$
$= \mu \Rightarrow$ unbiased

2) $Var(\overline{Y_n}) = Var\left(\frac{1}{n}\sum Y_i\right) = \frac{1}{n^2} Var \sum Y_i =$
$=$ [independ.] $= \frac{1}{n^2} \cdot \sum Var Y_i = \frac{n}{n^2} Var Y_5 =$
$= \frac{\sigma^2}{n} \xrightarrow[n \to \infty]{} 0 \quad \Rightarrow$ lemma: 1)+2) $\Rightarrow$

$\Rightarrow \overline{Y_n} \xrightarrow[p]{} \mu$

consistent

- is $(\overline{Y_n})^2$ an unbiased and consistent est. of μ^2 ?

1) $E(\overline{Y}^2) = Var(\overline{Y}) + (E\overline{Y})^2 =$

$\Rightarrow E(\bar{Y}_n^2) = Var(\bar{Y}_n) + (E\bar{Y}_n)^2 =$

$$= \underbrace{\frac{\sigma^2}{n}}_{\downarrow 0} + \mu^2 \neq \mu^2 \Rightarrow \text{biased}$$

$bias = \frac{\sigma^2}{n} \xrightarrow[n\to\infty]{} 0$ so $\boxed{\bar{Y}_n^2 \text{ is asymptotically unbiased}}$

2) is $\bar{Y}_n^2 \xrightarrow[p]{} \mu^2$?

we can use continuous mapping theorem

(if $\hat{\beta}_n \xrightarrow[p]{} \beta$ then $g(\hat{\beta}_n) \xrightarrow[p]{} g(\beta)$)

if $\bar{Y}_n \xrightarrow[p]{} \mu \Rightarrow (\bar{Y}_n)^2 \xrightarrow[p]{} \mu^2 \Rightarrow$

$\Rightarrow \boxed{\bar{Y}_n^2 \text{ is consistent est}}$

Task 1 Missed regressor

true DGP (data generating process)

$$y_i = \beta_0 + \beta_1 X_i + \beta_2 Z_i + u_i \quad , \quad E(u_i | X, Z) = 0$$

but we forgot about Z_i and try to est.

$$y_i = \beta_0 + \beta_1 X_i + V_i$$

• so $\hat{\beta}_1 = \frac{\sum (x_i-\bar{x})(y_i-\bar{y})}{\sum (x_i-\bar{x})^2} = \frac{\sum y_i(x_i-\bar{x})}{\sum x_i(x_i-\bar{x})} =$

$$= \frac{\sum (\beta_0+\beta_1 x_i+V_i)(x_i-\bar{x})}{\sum x_i(x_i-\bar{x})} = \beta_0 \frac{\overbrace{\sum(x_i-\bar{x})}^{=0}}{\sum x_i(x_i-\bar{x})} +$$

$$+\beta_1 \frac{\sum x_i(x_i-\bar{x})}{\sum x_i(x_i-\bar{x})} + \frac{\sum V_i(x_i-\bar{x})}{\sum x_i(x_i-\bar{x})} =$$

$$= \beta_1 + \frac{\sum V_i(x_i-\bar{x})}{\sum x_i(x_i-\bar{x})}$$

• $E(\hat{\beta}_1|x) = \beta_1 + E\left[\frac{\sum (V_i(x_i-\bar{x}))}{\sum x_i(x_i-\bar{x})}\Big|x\right]$

$$= \beta_1 + \frac{\sum [E(V_i|x)]\cdot(x_i-\bar{x})}{\sum x_i(x_i-\bar{x})} \neq \beta_1 ?$$

maybe. unbiased

• $\hat{\beta}_1 = \boxed{\frac{scov(x,y)}{svar(x)} \xrightarrow[p]{} \frac{cov(x,y)}{var(x)}} \Rightarrow$

$$\Rightarrow \hat{\beta}_1 = \beta_1 + \frac{\sum v_i(x_i-\bar{x})}{\sum x_i(x_i-\bar{x})} = \beta_1 + \frac{\sum (V_i-\bar{V})(x_i-\bar{x})}{\sum (x_i-\bar{x})^2} =$$

$$= \beta_1 + \frac{scov(V,x)}{svar(x)} \xrightarrow[p]{} \beta_1 + \frac{cov(V,x)}{var(x)}$$

what is it?

- $\frac{cov(V, X)}{var(x)} = \frac{cov(\beta_2 Z + u, x)}{var(x)} =$

$= \frac{cov(\beta_2 Z, x)}{var(x)} + \frac{cov(u, x)}{var(x)} = \beta_2 \frac{cov(Z,x)}{var x} + 0$

$\left(E(u|x)=0 \Rightarrow cov(u,x)=0 \text{ (weaker)} \right)$

- So consistency depends on $cov(Z,X)$

if $cov(Z,X)=0$ then $\hat{\beta}_1 \xrightarrow{p} \beta_1$ consist.

if $cov(Z,x) \neq 0$ then $\hat{\beta}_1 \not\xrightarrow{p} \beta_1$ unconsist

(and Biased)

- If we miss the factor which is correlated with our regressor then the estimator will be unconsistent

---

Task 2 Measuring error

DGP (data gen. process):

$y_i = \beta_0 + \beta_1 X_i + u_i \quad , \quad E(u|x)=0$

But x_i were measured with errors:

we have $\tilde{X}_i = X_i + V_i$ (let's think V_i is indep. with X_i and u_i)

So we have $y_i = \beta_0 + \beta_1 \tilde{X}_i + W_i$

- from previous task

$$\hat{\beta}_1 = \frac{\text{scov}(\tilde{x}, y)}{\text{svar}\,\tilde{x}} \xrightarrow[p]{} \frac{\text{cov}(\tilde{x}, y)}{\text{var}\,\tilde{x}} =$$

$$= \beta_1 + \frac{\text{cov}(W, \tilde{x})}{\text{var}(\tilde{x})} \stackrel{=}{}$$

$$\left[\, y_i = \beta_0 + \beta_1 X_i + u_i = \beta_0 + \beta_1 \tilde{X}_i + \underbrace{\left(\beta_1 X_i - \beta_1 \tilde{X}_i + u_i\right)}_{W_i} \right]$$

$$\stackrel{=}{} \beta_1 + \frac{\text{cov}(\beta_1 X - \beta_1 \tilde{X} + u, \tilde{x})}{\text{var}\,\tilde{x}} =$$

$$= \beta_1 + \frac{\text{cov}(\beta_1 \cdot V + u, \tilde{x})}{\text{var}\,\tilde{x}} = \beta_1 + \frac{\text{cov}(\beta_1 V + u, X + V)}{\text{var}(X+V)} =$$

$$= \beta_1 + \beta_1 \frac{\text{cov}(V, X)}{\text{var}(X+V)} + \beta_1 \frac{\text{cov}(V, V)}{\text{var}(X+V)} +$$

$$+ \underbrace{\frac{\text{cov}(u, X)}{\text{var}(X+V)}}_{=0} + \underbrace{\frac{\text{cov}(u, V)}{\text{var}(X+V)}}_{=0} \stackrel{=}{}$$

because $E(u|X) = 0 \Rightarrow \text{cov}(u, X) = 0$

and u, v indep $\Rightarrow$ $cov(u,v)=0$

$$\textcircled{=} \; \beta_1 + \beta_1 \frac{cov(x,v) + \sigma_v^2}{var(x+v)} \underset{x,v \text{ indep}}{=} \beta_1 + \beta_1 \frac{\sigma_v^2}{\sigma_x^2 + \sigma_v^2}$$

$$= \beta_1 \left(1 - \frac{\sigma_v^2}{\sigma_x^2 + \sigma_v^2}\right) \xrightarrow[\sigma_v^2 \to \infty]{} 0$$

- So if we have measuring error
  then $\hat{\beta}_1$ is unconsistent !
- if the measuring error is very big
  then $\hat{\beta}_1 \approx 0$
- asymptotical biase $-\beta_1 \frac{\sigma_v^2}{\sigma_x^2 + \sigma_v^2}$
  is not corrected with even very large sample