* PS 5 assigned Today
  Due in 2 weeks (December 1st)
* Midterm scores
  posted to my.ucla.edu
* Midterm discussed in TA section. Sol'ns posted in two weeks

Case 2:

$y_i = \beta x_i + \varepsilon_i$ assume $E[x_i] = E[\varepsilon_i] = 0$

assume $E[x_i \varepsilon_i] = 0$ and $E[v_i \varepsilon_i] = 0 \leftarrow$ not a bad assumption
and $E[x_i v_i] = 0 \leftarrow$ similarly

But we observe $x_i^* = x_i + v_i$ instead. v_i

Regress y_i on x_i^*

$$\Rightarrow y_i = \beta(x_i^* - v_i) + \varepsilon_i$$
$$= \beta x_i^* + \underbrace{(\varepsilon_i - \beta v_i)}_{\varepsilon_i^*}$$
$$= \beta x_i^* + \varepsilon_i^*$$

OLS will be good if $\text{cov}(x_i^*, \varepsilon_i^*) = 0$

$$E[x_i^* \varepsilon_i^*] = E[(x_i + v_i)(\varepsilon_i - \beta v_i)]$$
$$= \underbrace{E[x_i \varepsilon_i]}_{\substack{=0 \\ \text{by assumpt}}} + \underbrace{E[v_i \varepsilon_i]}_{\substack{=0 \\ \text{by assumption}}} - \beta \underbrace{E[x_i v_i]}_{\substack{=0 \\ \text{by assumption}}} - \beta \underbrace{E[v_i^2]}_{=\sigma_v^2}$$
$$= -\beta \sigma_v^2$$

When independent variable is measured with error, there is a problem.

$$b = \frac{\sum_{i=1}^N x_i^* y_i}{\sum_{i=1}^N (x_i^*)^2} = \frac{\sum_{i=1}^N (x_i^*)(\beta x_i^* + \varepsilon_i^*)}{\sum_{i=1}^N (x_i^*)^2} = \beta + \frac{\sum_{i=1}^N x_i^* \varepsilon_i^*}{\sum_{i=1}^N (x_i^*)^2}$$
$$= \beta + \frac{\frac{1}{N}\sum_{i=1}^N x_i^* \varepsilon_i^*}{\frac{1}{N}\sum_{i=1}^N (x_i^*)^2}$$

$$\Rightarrow b \approx \beta + \frac{E[x_i^* \varepsilon_i^*]}{E[(x_i^*)^2]} = \beta - \beta\sigma_v^2$$

But $E[(x_i^*)^2] = E[(x_i+v_i)^2]$

$= E[x_i^2 + 2x_i v_i + v_i^2]$

$= E[x_i^2] + 2\underbrace{E[x_i v_i]}_{\substack{=0 \\ \text{by assump}}} + E[v_i^2]$

$= \sigma_x^2 + \sigma_v^2$

$\Rightarrow b \approx \beta - \dfrac{\beta\sigma_v^2}{\sigma_x^2+\sigma_v^2} = \beta\left(1 - \dfrac{\sigma_v^2}{\sigma_x^2+\sigma_v^2}\right)$

$= \beta\left(\dfrac{\sigma_x^2}{\sigma_x^2+\sigma_v^2}\right)$ signal, noise } signal to noise ratio

↳ b/t 0 and 1

Thus, b is <u>biased towards zero.</u>

How do we fix this? Instrumental variable technique.

$y_i = x_i^* \beta + \varepsilon_i^*$

Suppose $\exists z_i$ s.t. (i) $E[z_i \varepsilon_i^*] = 0$

(ii) $E[z_i x_i^*] \neq 0$

$\varepsilon_i^* = y_i - x_i^* \beta$

$\Rightarrow z_i \varepsilon_i^* = z_i y_i - z_i x_i^* \beta$

$\Rightarrow \underbrace{E[z_i \varepsilon_i^*]}_{=0} = E[z_i y_i] - \beta E[z_i x_i^*]$

$\Rightarrow \beta = \dfrac{E[z_i y_i]}{E[z_i x_i^*]} \approx \dfrac{\frac{1}{N}\sum_{i=1}^{N} z_i y_i}{\frac{1}{N}\sum_{i=1}^{N} z_i x_i^*}$

$= \dfrac{\sum_{i=1}^{N} z_i y_i}{\sum_{i=1}^{N} z_i x_i^*} \equiv b_{IV}$

Cf OLS: $b_{OLS} = \dfrac{\sum_{i=1}^{N} x_i^* y_i}{\sum_{i=1}^{N} x_i^* x_i^*}$. Replace x_i^* with z_i

How do we find such a z_i?

Assume that

$$z_i = x_i + w_i$$

[Recall that $x_i^* = x_i + v_i$

$$z_i \varepsilon_i^* = (x_i + w_i)(\varepsilon_i - \beta v_i)$$
$$= x_i \varepsilon_i + w_i \varepsilon_i - \beta x_i v_i - \beta w_i v_i$$

measurement error
economics error

$$\Rightarrow E[z_i \varepsilon_i^*] = \underbrace{E[x_i \varepsilon_i]}_{\substack{=0 \text{ by} \\ \text{assumption}}} + \underbrace{E[w_i \varepsilon_i]}_{\substack{=0 \text{ by} \\ \text{assumption}}} - \beta \underbrace{E[x_i v_i]}_{\substack{=0 \text{ by} \\ \text{assumption}}} - \beta E[w_i v_i]$$

Thus $E[z_i \varepsilon_i^*] = -\beta E[w_i v_i] \Rightarrow E[z_i \varepsilon_i^*] = 0$ iff $E[w_i v_i] = 0$

$E[w_i v_i] = 0$ makes sense. Errors b/t research assistants should not be correlated.

Omitted Variable Bias

$$y_i = \beta_2 x_{2i} + \beta_3 x_{3i} + \varepsilon_i \quad \leftarrow \text{true mean}$$

$$E[x_{2i} \varepsilon_i] = 0$$
$$E[x_{3i} \varepsilon_i] = 0$$

Suppose we do not observe x_{3i}. Then regress y_i on x_{2i}

$$y_i = \beta_2^* x_{2i} + \varepsilon_i$$

$$\Rightarrow b_2^* = \frac{\sum_{i=1}^N x_{2i} y_i}{\sum_{i=1}^N x_{2i}^2} = \frac{\sum_{i=1}^N x_{2i}(\beta_2 x_{2i} + \beta_3 x_{3i} + \varepsilon_i)}{\sum_{i=1}^N x_{2i}^2}$$

$$= \beta_2 + \beta_3 \frac{\sum_{i=1}^N x_{2i} x_{3i}}{\sum_{i=1}^N x_{2i}^2} + \frac{\sum_{i=1}^N x_{2i} \varepsilon_i}{\sum_{i=1}^N x_{2i}^2}$$

$$= \beta_2 + \beta_3 \frac{\sum_{i=1}^N x_{2i} x_{3i}}{\sum_{i=1}^N x_{2i}^2} + \underbrace{\frac{\frac{1}{N}\sum_{i=1}^N x_{2i} \varepsilon_i}{\frac{1}{N}\sum_{i=1}^N x_{2i}^2}}_{\approx \frac{E[x_{2i}\varepsilon_i]}{E[x_{2i}^2]} = 0}$$

$$\Rightarrow b_2^* \approx \beta_2 + \beta_3 \frac{\sum_{i=1}^{N} x_{2i} x_{3i}}{\sum_{i=1}^{N} x_{2i}^2}$$
$$\approx \frac{\partial x_{3i}}{\partial x_{2i}}$$

Thus, there will be a bias unless $\beta_3 = 0$ or $E[x_{2i} x_{3i}] = 0$

$$E[x_{2i}(\beta_3 x_{3i} + \varepsilon_i)] = \beta E[x_{2i} x_{3i}] + \underbrace{E[x_{2i}\varepsilon_i]}_{=0}$$
$$\neq 0$$

Thus, we can't just replace
$$y_i = \beta_2 x_{2i} + \beta_3 x_{3i} + \varepsilon_i$$
with $y_i = \beta_2 x_{2i} + \varepsilon_i^*$ where $\varepsilon_i^* = \beta_3 x_{3i} + \varepsilon_i$

Then what do we do?
Find z_i s.t. $E[z_i x_{3i}] = 0$ $E[z_i x_{2i}] \neq 0$ } This cannot be tested
$E[z_i \varepsilon_i] = 0$
$\Rightarrow E[z_i(\beta_3 x_{3i} + \varepsilon_i)] = 0$

The game in applied micro. is to find such instrumental random variables.

* Can we test the validity of astrology?