Measurement Error

Chungsang Tom Lam[1]

[1]Department of Economics
Clemson University

February 25, 2020

Model

- Consider the simple linear regression model with one regressor and no intercept.

$$y = \beta x + \epsilon \tag{1}$$

- If we have data on x and y, we can use the OLS estimator $\hat{\beta}$.
- However, we only have data on $\tilde{x}$ and $\tilde{y}$

"bad things can happen
your data from survey are not
accurate,,太粗糙了。。。"

$$\tilde{x} = x + u \tag{2}$$

$$\tilde{y} = y + v \tag{3}$$

- Therefore, our data $\tilde{x}$ and $\tilde{y}$ are not exactly the same as x and y.
- We are going to look at just Classical measurement error

Measurement error in x

- Consider only measurement error in x (i.e. $\sigma_v^2 = 0$)
- Combining the equations we have

$$y = \beta(\tilde{x} - u) + \epsilon \tag{4}$$

$$y = \beta\tilde{x} + (\epsilon - \beta u) \tag{5}$$

- Then the OLS estimator gives use

$$\hat{\beta} = \frac{cov(\tilde{x}, \tilde{y})}{var(\tilde{x})}$$

$$= \frac{cov(x + u, \boxed{\beta x + \epsilon})}{var(x + u)}$$

equation(5)亦可

OLS : $(X^T X)^{-1} X^T Y$, $\hat{\beta} = \frac{cov(X, Y)}{var(X)}$.

IV : $(Z^T X)^{-1} Z^T Y$, $\hat{\beta} = \frac{cov(Z, Y)}{var(Z, X)}$.

Measurement error in x

- The probability limit of the estimator is

$$plim(\hat{\beta}) = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_u^2}\beta \neq \beta \tag{6}$$

- And the difference is:

$$\frac{\sigma_x^2}{\sigma_x^2 + \sigma_u^2}\beta - \beta = \frac{-\sigma_u^2}{\sigma_x^2 + \sigma_u^2}\beta \tag{7}$$

- Notice that this difference is going to be negative when $\beta > 0$ and positive when $\beta < 0$.
- In other words, the estimator is biased towards zero.

Measurement error in y

- How about the measurement error in y
- Let's say $\sigma_u^2 = 0$ and $\sigma_v^2 > 0$

$$\tilde{y} = \beta x + \epsilon + v \tag{8}$$

- We can still estimate β using $\hat{\beta}$, but the standard error will increase.

which means, it is ok, but less efficient.

Instrumental Variable

- How to estimate β correctly with measurement error in x?
- Again, we need an instrument

$$\hat{\beta}_{IV} = \frac{cov(y, z)}{cov(\tilde{x}, z)} \tag{9}$$

$$= \frac{cov(y, z)}{cov(x + u, z)} \tag{10}$$

- If the instrument z is correlated with x but not correlated with u, then the IV estimate is consistent.