Tutorial 1: Measurement Error

Hammad Shaikh

January 12, 2021

Measurement Error Overview

- Problem: Do not perfectly observe a variable

  ↳ Ex. Recall ECO100 grade honest but have imperfect recall (survey)

- Representation: $\underbrace{\tilde{X}}_{\text{observed}} = \underbrace{X}_{\text{truth}} + \underbrace{e}_{\text{error}}$

  ↳ Truth: real grade transcript (administrative data)

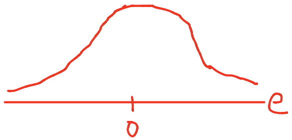
  observed ↓ Self reported ECO100

  truth ↓ transcript

  error ↓ recall

- Classical Measurement Assumpton: $e \sim N(0, \sigma_e^2)$

  ↳ $e \perp X \Rightarrow Cov(e, X) = 0$



Measurement Error and Regression

$Y = \beta_0 + \beta_1 X + \varepsilon$, $\tilde{Y} = Y + e \Rightarrow \begin{cases} E(\tilde{Y}) = E(Y+e) \overset{E(e)=0}{=} E(Y) = \mu_Y \\ V(\tilde{Y}) = V(Y+e) \underset{e \perp Y}{=} \sigma_Y^2 + \underbrace{\sigma_e^2}_{>0} \end{cases}$

$\hookrightarrow \beta_1 = \frac{Cov(Y,X)}{V(X)}$

- Measurement error in the outcome: $\tilde{Y} = Y + e$

$\tilde{Y} = \beta_0 + \beta_1 X + u$, $\hat{\beta}_1 \xrightarrow{p} \frac{Cov(\tilde{Y},X)}{V(X)} = \frac{Cov(Y+e,X)}{V(X)} \underset{e \perp X}{=} \frac{Cov(Y,X)}{V(X)} = \beta_1$

$\hat{\beta}_1 \xrightarrow{p} \beta_1$ (consistent)

- Measurement error in the treatment: $\tilde{X} = X + e$

$Y = \beta_0 + \beta_1 \tilde{X} + u \rightarrow \hat{\beta}_1 \xrightarrow{p} \frac{Cov(Y,\tilde{X})}{V(\tilde{X})} = \frac{Cov(\beta_0 + \beta_1 X + \varepsilon, X+e)}{V(X+e)} \underset{\substack{e \perp X \\ e \perp \varepsilon}}{=} \frac{Cov(\beta_1 X, X)}{\sigma_X^2 + \sigma_e^2} = \beta_1 \cdot \frac{\sigma_X^2}{\sigma_X^2 + \sigma_e^2}$

$\hat{\beta}_1 \xrightarrow{p} \beta_1 \cdot \underbrace{\frac{\sigma_X^2}{\sigma_X^2 + \sigma_e^2}}_{\in (0,1)} \Rightarrow \hat{\beta}_1 \xrightarrow{p} \beta_1^*$

$\sigma_e^2 \rightarrow \infty \Rightarrow \hat{\beta}_1 \rightarrow 0 \Rightarrow \hat{\beta}_1$ biased towards 0 (attenuation bias)

Measurement Error Practice

$$E\left[\frac{\sum(\tilde{Y}_i-\bar{\tilde{Y}})(X_i-\bar{X})}{\sum(X_i-\bar{X})^2}\right] = E\left[\frac{\sum(X_i-\bar{X})\tilde{Y}_i}{\sum(X_i-\bar{X})^2}\right] \underset{\tilde{Y}_i=Y_i+e}{=} \underbrace{E\left[\frac{\sum(X_i-\bar{X})Y_i}{\sum(X_i-\bar{X})^2}\right]}_{\beta_1} + \underbrace{E\left[\frac{\sum(X_i-\bar{X})}{\sum(X_i-\bar{X})^2}\right]E(e)}_{0}$$

True or False?

- If measurement error in a dependent variable has zero mean and is uncorrelated with the independent variables, the OLS estimator will be unbiased. T

  $\hookrightarrow \tilde{Y} = Y + e,\ E(e)=0,\ e \perp X,\ E(\hat{\beta}_1)=\beta_1?$

  $\hookrightarrow \tilde{Y} = \beta_0 + \beta_1 X + u,\ E(\hat{\beta}_1) = E\left[\frac{\widehat{Cov}(\tilde{Y},X)}{\hat{V}(X)}\right] = E\left[\frac{\sum(\tilde{Y}_i-\bar{\tilde{Y}})(X_i-\bar{X})}{\sum(X_i-\bar{X})^2}\right]$

- If measurement error in an observed independent variable is uncorrelated with the corresponding (unobserved) true variable, the OLS estimator is unbiased. F

  $\hookrightarrow \tilde{X} = X + e,\ e \perp X,$ True: $Y = \beta_0 + \beta_1 X + \varepsilon,\ X \perp \varepsilon,\ \tilde{X} \perp \varepsilon$

  $\hookrightarrow Y = \beta_0 + \beta_1 \tilde{X} + u,$ Hint: $Cov(\tilde{X}, u) = ?$

Start from truth $\Rightarrow$ Estimable $\Rightarrow$ sign $Cov(\tilde{X}, \eta)$

$$Y = \beta_0 + \beta_1 X + \varepsilon \Rightarrow Y = \beta_0 + \beta_1 \tilde{X} + \underbrace{\varepsilon - \beta_1 e}_{\eta} \Rightarrow Cov(\tilde{X}, \varepsilon - \beta_1 e)$$

$$X = \tilde{X} - e$$

$$= -\beta_1 \underbrace{Cov(\tilde{X}, e)}_{\neq 0}$$

$Cov(\tilde{X}, \eta) \neq 0 \Rightarrow$ endogeneity bias