Econometrics 2 – Part 1

Arieda Muço

Central European University

Winter 2023

Case 1: Measurement Error in Y_i

- True relationship

$$Y_i^* \quad = \quad \alpha + X_i'\beta + u_i$$

- But Y_i^* is not observed
- We observe Y_i measured with error

$$Y_i \quad = \quad Y_i^* + \epsilon_i$$

Case 1: Measurement Error in Y_i

- Assumptions:
  - $\epsilon_i \sim iid\left(0, \sigma_\epsilon^2\right)$
  - $Cov\left(X_i, \epsilon_i\right) = 0$
- Estimated Model:

$$Y_i \quad = \quad \alpha + X_i'\beta + \underbrace{(u_i + \epsilon_i)}_{v_i}$$

Note: $Cov\left(v_i, X_i\right) = 0$

- OLS estimator of β is unbiased
- Error variance: $Var\left(u_i + \epsilon_i\right) = \sigma_u^2 + \sigma_\epsilon^2$

Case 2: Measurement Error in X_i

- Consider a bivariate model without constant:

$$\begin{aligned} Y_i &= \rho s_i^* + u_i \\ \text{where} \quad & Cov(s_i^*, u_i) = 0 \end{aligned}$$

- s_i^* true level of schooling
- We observe $s_i = s_i^* + \epsilon_i$
- Classical Measurement Error (CME)
  - $\epsilon_i \sim iid\left(0, \sigma_\epsilon^2\right)$
  - $Cov\left(\epsilon_i, s_i^*\right) = 0$

Estimated model

$$
\begin{aligned}
Y_i &= \rho s_i^* + u_i \\
&= \rho(s_i - \epsilon_i) + u_i = \rho s_i \underbrace{-\rho\epsilon_i + u_i}_{\tilde{u}_i} \\
Y_i &= \rho s_i + \tilde{u}_i \qquad \text{estimated model}
\end{aligned}
$$

Are the error term and the regressor correlated?

$$
\begin{aligned}
Cov\,(s_i, \tilde{u}_i) &= Cov\,(s_i, u_i - \rho\epsilon_i) = Cov\,(s_i, -\rho\epsilon_i) \\
&= -\rho Cov\,(s_i, \epsilon_i) = -\rho Cov\,(s_i^* + \epsilon_i, \epsilon_i) \\
&= -\rho\sigma_\epsilon^2
\end{aligned}
$$

If $\rho > 0$, $Cov\,(s_i, \tilde{u}_i) < 0$

Measurement error bias

$$Y_i = \rho s_i + \tilde{u}_i$$

Estimating the model with OLS gives

$$\begin{aligned}
\tilde{\rho} &= \frac{Cov\left(Y_i, s_i\right)}{Var\left(s_i\right)} \\
&= \frac{Cov\left(\rho s_i + \tilde{u}_i, s_i\right)}{Var\left(s_i\right)} \\
&= \rho + \frac{Cov\left(\tilde{u}_i, s_i\right)}{Var\left(s_i\right)} \\
&= \rho - \rho \frac{\sigma_\epsilon^2}{Var\left(s_i\right)}
\end{aligned}$$

Measurement error bias

Attenuation Bias

$$\tilde{\rho} = \rho - \rho \frac{\sigma_\epsilon^2}{\sigma_s^2} = (1 - \lambda)\, \rho$$

$$\lambda = \frac{\sigma_\epsilon^2}{\sigma_s^2} = \frac{\sigma_\epsilon^2}{\sigma_\epsilon^2 + \sigma_{s^*}^2}$$

- λ noise to signal ratio
- 1-λ is interesting as it tells us that measurement error causes attenuation bias (this fraction is smaller than one)

$$1 - \frac{\sigma_\epsilon^2}{\sigma_\epsilon^2 + \sigma_{s^*}^2} = \frac{\sigma_{s^*}^2}{\sigma_\epsilon^2 + \sigma_{s^*}^2} = \frac{Var(s^*)}{Var(s)}$$

Multivariate Model

$$
\begin{aligned}
Y_i &= \rho s_i^* + X_i \beta + u_i \\
s_i &= s_i^* + \epsilon_i
\end{aligned}
$$

Classical ME: $Cov\left(s_i^*, \epsilon_i\right) = 0$, $Cov\left(X_i, \epsilon_i\right) = 0$

Attenuation Bias

$$
\tilde{\rho} = \rho \frac{\sigma_{r_1^*}^2}{\sigma_{r_1^*}^2 + \sigma_\epsilon^2}
$$

- where r_1^* is the population error from regression of $s_i^* = X_i \beta + r_1^*$
- If s_i is highly correlated with X_i, attenuation bias increases with inclusion of more covariates in the model

Multivariate Model

Consider a situation where

$$\begin{aligned} Cov\left(s_i^*, X_i\right) \neq & \quad 0 \qquad \text{is high} \\ Cov\left(\epsilon_i, X_i\right) = & \quad 0 \end{aligned}$$

Then:

- X_i's that are correlated with s_i^* soak up the signal in s_i, but leave the noise.
- Because X_i is uncorrelated with ϵ_i this exacerbates the ME problem.
- Example:
  X_i parental education, earnings etc.