Erasmus School of Economics

MOOC Econometrics

Lecture 4.2 on Endogeneity: Consequences

Dennis Fok

Erasmus University Rotterdam

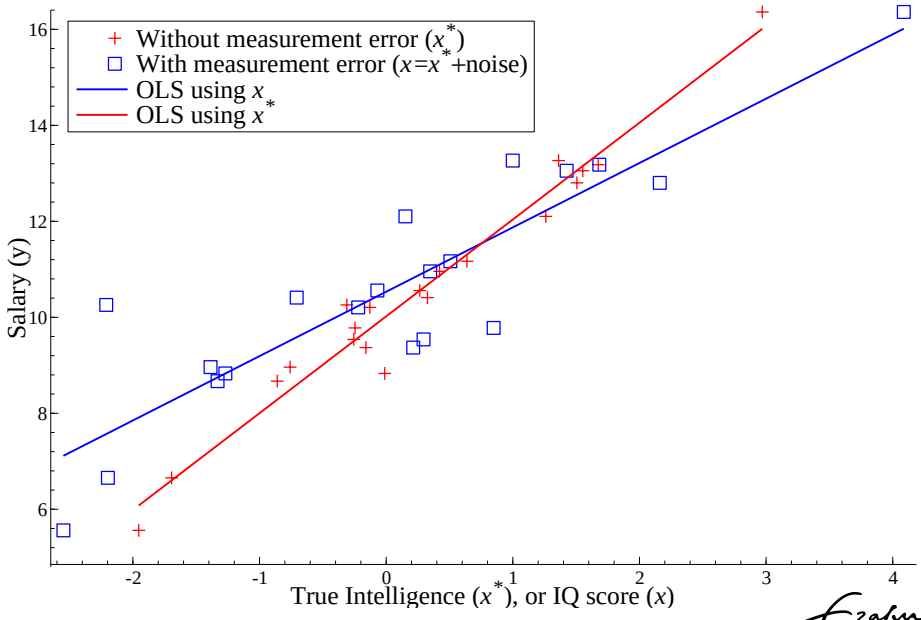
Endogeneity

- Common problem in economics
  1. Omitted variables
  2. Strategic behavior
  3. Measurement errors

  $\rightarrow X$ is correlated with ε
- Endogeneity violates the basic assumptions

$\rightarrow$ How bad is this?

Simulated example, $y = 1 + 2x^* + u$



Measurement error example

Under measurement error (and endogeneity in general):

- we obtain the wrong coefficients!

Test

Can we say anything about the direction of the bias?

Direction of bias in the measurement error case

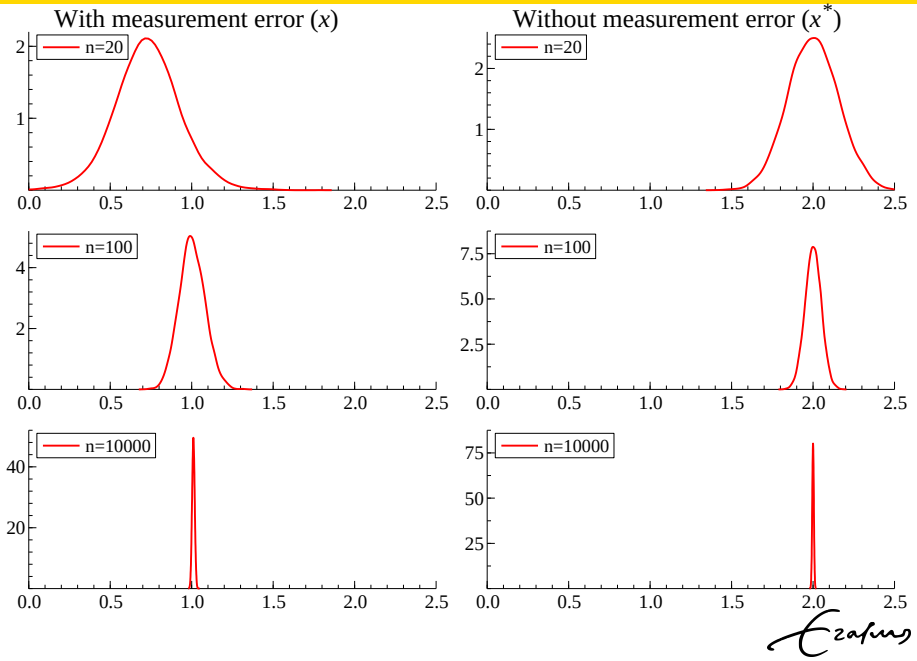
OLS is "biased towards zero"
→ OLS underestimates true effect

Intuitively:

- x-values on the *left* likely have negative measurement errors
- x-values on the *right* likely have positive measurement errors

Measurement errors "stretch" the scatter in the horizontal direction
→ a flatter regression line

Distribution of estimator for different n, true value= 2



Consistency: formal argumentation

If X is endogenous:

- If n grows the OLS estimator converges to the wrong value.
  → OLS is inconsistent

Consider the standard model $y = X\beta + \varepsilon$ and the OLS estimator

$$\begin{aligned} b &= (X'X)^{-1}X'y = (X'X)^{-1}X'(X\beta + \varepsilon) \\ &= (X'X)^{-1}X'X\beta + (X'X)^{-1}X'\varepsilon \\ &= \beta + (X'X)^{-1}X'\varepsilon \end{aligned}$$

So, b can be split into

1. True parameter value β
2. Random deviation $(X'X)^{-1}X'\varepsilon$

Asymptotic properties

What happens to b as $n \to \infty$?

Recall: $b = \beta + (X'X)^{-1}X'\varepsilon$

- β is constant
- Elements of $(X'X)$ and $X'\varepsilon$ are sums over observations:

$$X'X = \begin{pmatrix} \sum_{i=1}^n x_{1i}^2 & \sum_{i=1}^n x_{1i}x_{2i} & \cdots & \sum_{i=1}^n x_{1i}x_{ki} \\ \sum_{i=1}^n x_{1i}x_{2i} & \sum_{i=1}^n x_{2i}^2 & \cdots & \sum_{i=1}^n x_{2i}x_{ki} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^n x_{ki}x_{1i} & \sum_{i=1}^n x_{ki}x_{2i} & \cdots & \sum_{i=1}^n x_{ki}^2 \end{pmatrix}, X'\varepsilon = \begin{pmatrix} \sum_{i=1}^n x_{1i}\varepsilon_i \\ \sum_{i=1}^n x_{2i}\varepsilon_i \\ \vdots \\ \sum_{i=1}^n x_{ki}\varepsilon_i \end{pmatrix}$$

→ these diverge as $n \to \infty$

Asymptotic properties

Rewrite $b = \beta + (\frac{1}{n}X'X)^{-1}(\frac{1}{n}X'\varepsilon)$

- $(\frac{1}{n}X'X)$ is an average
  $\rightarrow$ in general converges to, say, Q
- $(\frac{1}{n}X'\varepsilon)$ also converges in general

Consistency result:
b converges to β as $n \rightarrow \infty$ if

1. $\frac{1}{n}X'X$ converges to Q, *and*
2. Q^{-1} exists, *and*
3. $\frac{1}{n}X'\varepsilon$ converges to 0
   - No correlation between X and ε (for large n)
   - X is exogenous

X endogenous: b does not converge to β!

Small sample properties

So far we discussed what happens for $n \rightarrow \infty$

Test

Why can't we derive the bias?

To obtain the bias

- need to evaluate

$$\begin{aligned} \mathrm{E}[b] &= \mathrm{E}[(X'X)^{-1}X'y] = \mathrm{E}[(X'X)^{-1}X'(X\beta+\varepsilon] \\ &= \mathrm{E}[\beta + (X'X)^{-1}X'\varepsilon] = \beta + \underbrace{\mathrm{E}[(X'X)^{-1}X'\varepsilon]}_{=?}. \end{aligned}$$

- X is stochastic
- cannot simplify final expectation (without further assumptions)

OLS in presence of endogeneity

If X endogenous

- X correlated with ε
- OLS estimator for β is not consistent
- Even with in infinite amount of data: OLS does not give useful estimates

TRAINING EXERCISE 4.2

- Train yourself by making the training exercise (see the website).
- After making this exercise, check your answers by studying the webcast solution (also available on the website).