Erasmus
School of
Economics

MOOC Econometrics

Lecture 4.3 on Endogeneity: Estimation under endogeneity

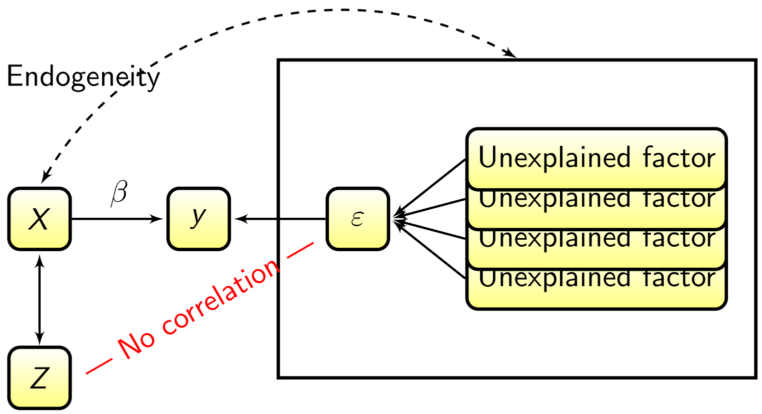
Dennis Fok

Erasmus University Rotterdam

What have we so far?

- Endogeneity is a common problem
- Endogeneity causes OLS to be inconsistent
- Estimation requires another estimation technique

"Solving endogeneity": Graphical representation



Instrumental variable estimation

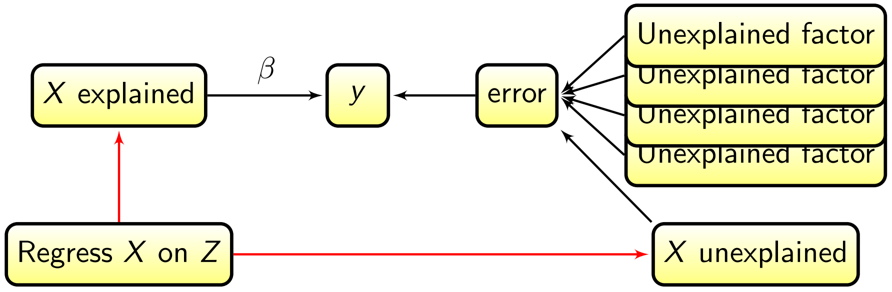
- Z variables are *instruments* if
  - Z and X are correlated
  - Z does not correlate with ε
- Correlation between instruments and y is only due to X

$$\text{Cov}(Z, y) = \text{Cov}(Z, X\beta + \varepsilon) = \text{Cov}(Z, X\beta) + \underbrace{\text{Cov}(Z, \varepsilon)}_{=0}$$
$$= \text{Cov}(Z, X)\beta$$

- Use instruments to estimate β

"Solving endogeneity": Graphical representation

1. Use Z to decompose X in explained and unexplained part
2. Effect size of explained part on y equals β
3. Unexplained part is added to error term



Endogeneity is solved as

- X unexplained not correlated with X explained
- X explained is exogenous

2SLS in matrix notation

Given model

$$y = X\beta + \varepsilon, \quad \mathrm{Var}[\varepsilon] = \sigma^2 I$$

and instruments Z

1. Regress X on Z to get explained part:
   - Model: $X = Z\gamma + \eta$
   - OLS estimate: $(Z'Z)^{-1}Z'X$
   - Fitted value: $\hat{X} = \underbrace{Z(Z'Z)^{-1}Z'}_{H_Z} X = H_Z X$
2. Regress y on $\hat{X}$:

$$\begin{aligned} b_{2SLS} &= (\hat{X}'\hat{X})^{-1}\hat{X}'y \\ &= (X'H_Z'H_ZX)^{-1}X'H_Z'y \\ &= (X'H_ZX)^{-1}X'H_Zy \end{aligned}$$

Use: $H_Z = H_Z' = H_Z'H_Z$

Properties 2SLS

- Variance of b_{2SLS}: $\mathrm{Var}[b_{2SLS}] = \sigma^2(X'H_ZX)^{-1}$
- Estimating σ^2:
  - $\hat{\sigma}^2 = \frac{1}{n-k}(y - Xb_{2SLS})'(y - Xb_{2SLS})$
  - Do **not** use residuals (or reported standard errors) of second stage regression!

Derivation of variance (use $\mathrm{Var}[\varepsilon] = \sigma^2 I$):

$$\begin{aligned} b_{2SLS} &= (X'H_ZX)^{-1}X'H_Zy = (X'H_ZX)^{-1}X'H_Z(X\beta + \varepsilon) \\ &= \beta + (X'H_ZX)^{-1}X'H_Z\varepsilon \\ \mathrm{Var}[b_{2SLS}] &= \mathrm{Var}[(X'H_ZX)^{-1}X'H_Z\varepsilon] \\ &= (X'H_ZX)^{-1}X'H_Z\mathrm{Var}[\varepsilon]\left((X'H_ZX)^{-1}X'H_Z\right)' \\ &= (X'H_ZX)^{-1}X'H_Z(\sigma^2 I)H_Z'X(X'H_ZX)^{-1} \\ &= \sigma^2(X'H_ZX)^{-1}\underbrace{X'\underbrace{H_ZH_Z'}_{H_Z}X(X'H_ZX)^{-1}}_{I} = \sigma^2(X'H_ZX)^{-1} \end{aligned}$$

Properties of 2SLS

- 2SLS is consistent if (when $n \to \infty$)
  - Z and ε not correlated: $\frac{1}{n}Z'\varepsilon \to 0$
  - Z not multicollinear: $\frac{1}{n}Z'Z \to Q_{ZZ}$, and Q_{ZZ} invertible
  - X and Z sufficiently correlated: $\frac{1}{n}X'Z \to Q_{XZ}$, and Q_{XZ} rank k

Sketch of proof:

$$\begin{aligned} b_{2SLS} &= \beta + (X'H_ZX)^{-1}X'H_Z\varepsilon = \beta + (X'Z(Z'Z)^{-1}Z'X)^{-1}X'Z(Z'Z)^{-1}Z'\varepsilon \\ &= \underbrace{\beta + \underbrace{\left(\frac{1}{n}X'Z\left(\frac{1}{n}Z'Z\right)^{-1}\frac{1}{n}Z'X\right)^{-1}}_{\downarrow \; (Q_{XZ}Q_{ZZ}^{-1}Q_{XZ}')^{-1}} \underbrace{\frac{1}{n}X'Z\left(\frac{1}{n}Z'Z\right)^{-1}}_{\downarrow \; Q_{XZ}Q_{ZZ}^{-1}} \underbrace{\frac{1}{n}Z'\varepsilon}_{\downarrow \; 0}}_{=\beta+0} \end{aligned}$$

Finding instruments

What are good instruments?

- All exogenous variables in X (incl. constant)
- Other instruments are always needed:
  - At least one for every endogenous variable
  - Want: strong correlation between Z and X
  - Need: no correlation between Z and ε

Examples of instruments

Explain obtained grade using attendance:

Potential instruments:

- Travel time home to university
- Policy change to obligatory attendance

Test

What variable would be an instrument for price when modeling consumer sales of ice cream using sales $= \alpha + \beta$price $+ \varepsilon$?

Potential instruments?

1. Prices of raw materials (valid)
2. ~~Competitor prices~~ (direct influence on sales, so part of ε)
3. ~~Outside temperature~~ (direct influence on sales, so part of ε)

Summary

If X is in fact exogenous

- OLS and 2SLS both consistent
- Variance OLS smaller than variance 2SLS!

$\rightarrow$ Use OLS

If X is endogenous

- 2SLS is consistent
- OLS inconsistent

$\rightarrow$ Use 2SLS

TRAINING EXERCISE 4.3

- 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).