ECON61001 Econometric Methods

Lecture 6: Endogeneity & Instrumental Variables

Nicky Grant

University of Manchester

November 2016

Today's Lecture

OLS is biased and inconsistent when $E[u_i x_i] \neq 0$ [**Endogeneity**]

Problem encountered frequently in practise

Require alternative to OLS

Study properties of **Instrumental Variables Estimator**

IV is consistent under alternative assumptions to OLS

OLS & Large Sample Assumptions

LS1 $y_i = x_i'\beta_0 + u_t$ **Linearity**

LS2 $E[u_i|X] = 0$ **Exogeneity**

LS3 $E[u_i^2|X] = \sigma_0^2$ **Homoskedasticity**

LS4 $E[u_j u_k|X] = 0 \ \forall j \neq k$ **Serially Uncorrelated Errors**

LS5 $(x_i, u_i) \ \ i = 1, .., N$ forms i.i.d sequence

LS6 $\mathrm{Rank}(X) = k$ & $E[x_i x_i']$ finite **No multicollinearity & finite variance of x_i**

$X = (x_1, ..x_N)'$ x_i is $k \times 1$.

This lecture considers violations of LS2

Endogeneity & OLS

$$y_i = x_i'\beta_0 + u_i$$

$$\hat{\beta}_{OLS} \xrightarrow{p} \beta_0 + E[x_i x_i']^{-1} E[x_i u_i] := \beta_*$$

Exogeneity $(E[x_i u_i] = 0)$ $\quad \beta_* = \beta_0$ (Consistency)

Endogeneity $(E[x_i u_i] \neq 0)$ $\quad \beta_* \neq \beta_0$ (Biased)

OLS estimates β_* where $E[x_i x_i']^{-1} E[x_i u_i]$ is the **bias**.

Common Causes of Endogeneity

- **Omitted Variable Bias**

$$y_i = x_i'\beta_0 + w_i'\delta_0 + v_i$$

If we omit w_i, i.e we run the regression (assuming v_i is exogenous)

$$y_i = x_i'\beta_0 + u_i \qquad u_i = w_i'\delta_0 + v_i$$

$$E[x_i u_i] = E[x_i w_i']\delta_0 \neq 0 \text{ when } \delta_0 \neq 0 \ \& \ E[x_i w_i'] \neq 0$$

- **Measurement Error Bias** in the explanatory variables x_i (when measurement error correlated with u_i).

Classic Example- Returns to Schooling

$$w_i = \alpha_0 + \beta_0 s_i + u_i$$

Natural ability a_i determines wages but is not measurable

Example of omitted variable bias $u_i = \delta_0 a_i + v_i$

$$\beta_{0*} = \beta_0 + \delta_0 \frac{cov(s_i, a_i)}{var(s_i)}$$

$$\delta_0 > 0,\ cov(s_i, a_i) > 0 \text{ hence } \beta_* > \beta_0$$

OLS overestimates returns to education

Instrumental Variables

Instrumental Variables alternative estimator to OLS.

Linearity $y_i = x_i\beta_0 + u_i$ and maintain i.i.d assumption

A $k \times 1$ Instrumental Variable z_i is assumed to satisfy

IV1 Exogeneity: $E[z_i u_i] = 0$

IV2 Identified: $E[x_i z_i']$ is full rank

$$\text{IV1 \& IV2} \Rightarrow \Rrightarrow \beta_0 = E[z_i x_i']^{-1} E[z_i y_i]$$

IV Estimator satisfies sample version $\frac{1}{N}\sum_{i=1}^{N} z_i(y_i - x_i'\beta) = 0$

$$\hat{\beta}_{IV} = (\sum_{i=1}^{N} z_i x_i')^{-1} \sum_{i=1}^{N} z_i y_i$$

Special Case: 1 Endogenous Regressor & 1 Instrument

$$y = \beta_{00} + \beta_{01} x_{1i} + u_i$$

$$x_i = (1, x_{i1})', \; z_i = (1, z_{i1})'$$

IV1 $E[z_i u_i] = (E[u_i], E[z_{i1} u_i])' = 0$, both $\Rightarrow$ $Cov(z_{i1}, u_i) = 0$.

IV2 $E[x_i z_i'] = \begin{pmatrix} 1 & E[z_{i1}] \\ E[x_{i1}] & E[z_{i1} x_{i1}] \end{pmatrix}$ full rank $\equiv$ $Cov(z_{i1}, x_{i1}) \neq 0$

Asymptotic Properties of IV Estimator

$$\hat{\beta}_{IV} = (\sum_{i=1}^{N} z_i x_i')^{-1} \sum_{i=1}^{N} x_i y_i$$
$$= \beta_0 + \left(\frac{1}{N} \sum_{i=1}^{N} z_i x_i' \right)^{-1} \frac{1}{N} \sum_{i=1}^{N} x_i u_i$$

$\left(\frac{1}{N} \sum_{i=1}^{N} z_i x_i' \right)^{-1} \xrightarrow{p} E[z_i x_i']^{-1}$ by i.i.d Slutsky if IV2 holds.

$\frac{1}{N} \sum_{i=1}^{N} x_i u_i \xrightarrow{p} E[z_i u_i] = 0$ by i.i.d and IV1 (exogeneity)

$\sqrt{N}(\hat{\beta}_{IV} - \beta_0) \xrightarrow{d} N(0, V)$ for some matrix V under set of assumptions similar to LS replacing LS2 with $IV1, IV2$ (Lecture 7).

Example of Instruments: Wage-Schooling

$$w_i = \alpha_0 + \beta_0 s_i + u_i$$

Example Instruments

- Angrist & Kruger (1991) - **Quarter of Birth**
- Angrist & Kruger (1992)- **Vietnam war lottery**
- Card (1995)- **Distance to School.**

May be uncorrelated with ability (IV1)

IV2 may fails as correlation between these instruments and s_i small

Other Examples of Instruments

Child development- Alcohol consumption during pregnancy

OLS likely underestimates the effect (alc. + correlated with other unmeasurable health/behavioural factors with − effect on child development).

Instruments - random assignment to counselling, genetic markers (genes which make people ill/unable to process alcohol.)

GDP-Institutional quality

OLS likely overestimates the effect (Inst. − correlated with unmeasurable factors (e.g political instability, corruption) that − effect GDP)

Instruments Settler mortality in 17-19c.

The Colonial Origins of Comparative Development: An Empirical Investigation, Acemoglu & Robinson (2001).

Empirical Example: Acemoglu & Robinson (2001)

Rich countries more likely to afford better institutions (property rights, less distortionary policies, patents etc),
OLS overestimate effect of institutions on GDP

AR(2001) use **settler mortality rates** between 17-19c as an instrument for institutions(protection against expropriation risk)

(potential) settler mortality $\Rightarrow$ settlements

$\Rightarrow$ early institutions $\Rightarrow$ current institutions

$\Rightarrow$ current performance.

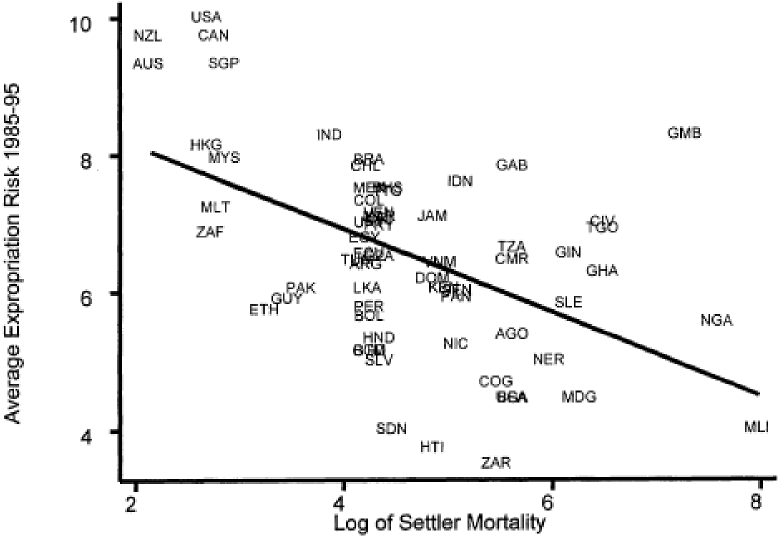

FIGURE 3. FIRST-STAGE RELATIONSHIP BETWEEN SETTLER MORTALITY AND EXPROPRIATION RISK

Source: AR(2001) pg1384

TABLE 2—OLS REGRESSIONS

	Whole world (1)	Base sample (2)	Whole world (3)	Whole world (4)	Base sample (5)	Base sample (6)	Whole world (7)	Base sample (8)
	Dependent variable is log GDP per capita in 1995						Dependent variable is log output per worker in 1988	
Average protection against expropriation risk, 1985–1995	0.54 (0.04)	0.52 (0.06)	0.47 (0.06)	0.43 (0.05)	0.47 (0.06)	0.41 (0.06)	0.45 (0.04)	0.46 (0.06)
Latitude			0.89 (0.49)	0.37 (0.51)	1.60 (0.70)	0.92 (0.63)		
Asia dummy				−0.62 (0.19)		−0.60 (0.23)		
Africa dummy				−1.00 (0.15)		−0.90 (0.17)		
"Other" continent dummy				−0.25 (0.20)		−0.04 (0.32)		
R^2	0.62	0.54	0.63	0.73	0.56	0.69	0.55	0.49
Number of observations	110	64	110	110	64	64	108	61

Notes: Dependent variable: columns (1)–(6), log GDP per capita (PPP basis) in 1995, current prices (from the World Bank's World Development Indicators 1999); columns (7)–(8), log output per worker in 1988 from Hall and Jones (1999). Average protection against expropriation risk is measured on a scale from 0 to 10, where a higher score means more protection against expropriation, averaged over 1985 to 1995, from Political Risk Services. Standard errors are in parentheses. In regressions with continent dummies, the dummy for America is omitted. See Appendix Table A1 for more detailed variable definitions and sources. Of the countries in our base sample, Hall and Jones do not report output per worker in the Bahamas, Ethiopia, and Vietnam.

Source: AR(2001) pg.1379

Latitude: Measure of distance from the equator (scaled to lie between 0,1 where 0 =equator)
Other Continent Dummy Dummy if any other continent (not Africa, Asia or America).

TABLE 4—IV REGRESSIONS OF LOG GDP PER CAPITA

	Base sample (1)	Base sample (2)	Base sample without Neo-Europes (3)	Base sample without Neo-Europes (4)	Base sample without Africa (5)	Base sample without Africa (6)	Base sample with continent dummies (7)	Base sample with continent dummies (8)	Base sample, dependent variable is log output per worker (9)
Panel A: Two-Stage Least Squares									
Average protection against expropriation risk 1985–1995	0.94 (0.16)	1.00 (0.22)	1.28 (0.36)	1.21 (0.35)	0.58 (0.10)	0.58 (0.12)	0.98 (0.30)	1.10 (0.46)	0.98 (0.17)
Latitude		−0.65 (1.34)		0.94 (1.46)		0.04 (0.84)		−1.20 (1.8)	
Asia dummy							−0.92 (0.40)	−1.10 (0.52)	
Africa dummy							−0.46 (0.36)	−0.44 (0.42)	
"Other" continent dummy							−0.94 (0.85)	−0.99 (1.0)	
Panel B: First Stage for Average Protection Against Expropriation Risk in 1985–1995									
Log European settler mortality	−0.61 (0.13)	−0.51 (0.14)	−0.39 (0.13)	−0.39 (0.14)	−1.20 (0.22)	−1.10 (0.24)	−0.43 (0.17)	−0.34 (0.18)	−0.63 (0.13)
Latitude		2.00 (1.34)		−0.11 (1.50)		0.99 (1.43)		2.00 (1.40)	
Asia dummy							0.33 (0.49)	0.47 (0.50)	
Africa dummy							−0.27 (0.41)	−0.26 (0.41)	
"Other" continent dummy							1.24 (0.84)	1.1 (0.84)	
R^2	0.27	0.30	0.13	0.13	0.47	0.47	0.30	0.33	0.28
Panel C: Ordinary Least Squares									
Average protection against expropriation risk 1985–1995	0.52 (0.06)	0.47 (0.06)	0.49 (0.08)	0.47 (0.07)	0.48 (0.07)	0.47 (0.07)	0.42 (0.06)	0.40 (0.06)	0.46 (0.06)
Number of observations	64	64	60	60	37	37	64	64	61

Source: AR(2001) pg. 1386

Concluding Remarks

Endogenous ($E[x_i u_i] \neq 0$) OLS inconsistent

IV consistent when z_i exogenous and sufficiently correlated with x_i

Empirical examples: GDP/institutions

Next Week

Introduce 2 Stage Least Squares (more instruments than regressors)

Asymptotic Properties of IV/2SLS

Discuss Weak Instrument Problem

Reading

Greene: Chapter 8.1, 8.2, 8.3, 8.8.

Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments Joshua D. Angrist and Alan B. Krueger, Journal of Political Economy (2001)
http://economics.mit.edu/files/18

The Colonial Origins of Comparative Development: An Empirical Investigation, Acemoglu & Robinson, The American Economic Review (2001)
http://economics.mit.edu/files/4123