Instrumental Variables

Econometrics II

R. Mora

Department of Economics
Universidad Carlos III de Madrid
Master in Industrial Organization and Markets

Outline

Motivation

OLS

$$y = \beta_0 + \beta_1 x + u, \qquad cov(x,u) = 0$$

$$\begin{tikzcd}
 & x \arrow[rr] & & \beta_0 + \beta_1 x \arrow[r] & y \\
\text{unobservable factors} \arrow[rr] & & u \arrow[urr] & &
\end{tikzcd}$$

- For the population, the effect on y of unobserved factors is unrelated to the controls
- OLS obtains the estimates of β by imposing this property in the sample

Why Use Instrumental Variables?

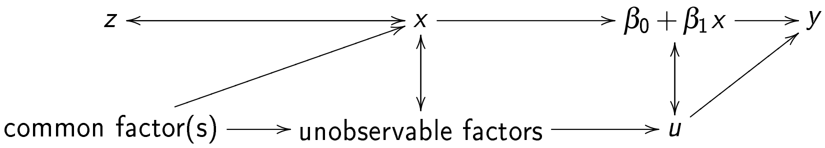
Suppose that both x and the unobservable factors are affected by one common factor:

$$\begin{tikzcd}
 & x \arrow[r] \arrow[d, leftrightarrow] & \beta_0 + \beta_1 x \arrow[r] \arrow[d, leftrightarrow] & y \\
\text{common factor(s)} \arrow[ur] \arrow[r] & \text{unobservable factors} \arrow[r] & u \arrow[ur] &
\end{tikzcd}$$

$$cov(x, u) \neq 0$$

- OLS exploits in the sample a property which is false for the population

Using an Instrument



$$y = \beta_0 + \beta_1 x + u, \; cov(z, u) = 0$$

- z is related to x
- z is unrelated to u (z only affects y through x)
  - we can exploit this property in the sample

Instruments

$$y = \beta_0 + \beta_1 x + u, \qquad cov(x,u) \neq 0$$

An instrument z must be relevant and exogenous

- z is **relevant** if it correlates with controls:

$$cov(x,z) \neq 0$$

- z is **exogenous** if controls capture all its effects on the dependent variable:

$$cov(u,z) = 0$$

- each exogenous control is an instrument of itself

Some Examples of Instruments

Returns to College Education

- $wages = \beta_0 + \beta_1 college + u$

- people freely choose to go to college:
  - people with larger ability are more likely to go to college: $cov(college, ability) \neq 0$
  - ability affects wages independently from college: $cov(ability, u) \neq 0$

$$\begin{tikzcd}
 & college \arrow[r] \arrow[d, leftrightarrow] & \beta_0 + \beta_1 x \arrow[r] \arrow[d, leftrightarrow] & y \\
ability \arrow[ur] \arrow[r] & \text{unobservable factors} \arrow[r] & u \arrow[ur] &
\end{tikzcd}$$

$$cov(college, u) \neq 0$$

Two examples in College Education

- A good instrument:
  - makes going to college more likely (**relevant**)
  - does not affect wages directly (**exogenous**)

Distance between pre-college residence and college

- proximity makes college more likely (relevance)
- residence is usually the parents' decision (exogeneity)

Father's education

- an educated father will tend to inform the child better about the profits of education (relevance)
- father's education is father's decision (exogeneity)

The returns to Compulsory Attendance Laws in the US

$$
\begin{tikzcd}
& education \arrow[r] \arrow[d, leftrightarrow] & \beta_0 + \beta_1 x \arrow[r] \arrow[d, leftrightarrow] & y \\
ability \arrow[ur] \arrow[r] & \text{unobservable factors} \arrow[r] & u \arrow[ur] &
\end{tikzcd}
$$

$$cov(education, u) \neq 0$$

Unobservable ability is likely related to years of education, but...

- children start schooling in the academic year they are 6 BY JANUARY 1ST
- children must remain in school until they are 16 BY THE SCHOOL ENTRY DATE

Month of Birth as Instrument

- Suppose you want to study the returns of those who do not want to continue studying after primary education:
- Month of birth correlates with years of education (relevance):
  - Children born in December enter schooling younger than children born in January
  - Children born in January require less schooling than those born in December to attain the legal dropout age
- Month of birth (presumably) does not correlate with ability (exogeneity)
- Month of birth is a good instrument for years of education

Lifetime Earnings and War Veterans in the US

What is the effect of going to war on future earnings?

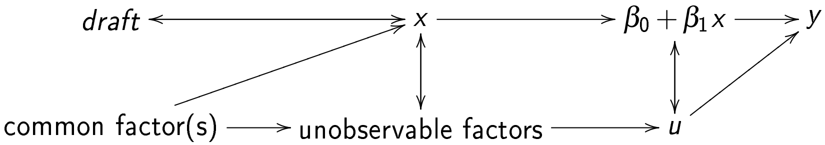
- Individuals with fewer alternatives in life are more likely to join the army and go to war and, independently, have lower future earnings

- Veteran status is likely to be correlated with unobservables:

$$
\begin{tikzcd}
 & \text{veteran} \arrow[r] \arrow[d, leftrightarrow] & \beta_0 + \beta_1 x \arrow[r] \arrow[d, leftrightarrow] & y \\
\text{fewer alternatives in life} \arrow[r] \arrow[ur] & \text{unobservable factors} \arrow[r] & u \arrow[ur] &
\end{tikzcd}
$$

$$cov(veteran, u) \neq 0$$

The draft



The Draft

- being drafted affects the probability of going to the war (relevance)
- being drafted is purely random (exogeneity)

Checking the Validity of the Instruments

Exogeneity

- use common sense and economic theory to decide if it makes sense to assume

$$cov(z, u) = 0$$

Relevance

- regress $x = \alpha_0 + \alpha_1 z + \varepsilon$
- test $H_0 : \alpha_1 = 0$

Now suppose we have a valid instrument z, what do we do with it?

IV Estimation & Inference

IV Estimation in the SRM

$y = \beta_0 + \beta_1 x + u$

- $cov(x,u) \neq 0$ (x is endogenous and OLS is inconsistent)
- $cov(x,z) \neq 0$ (z is relevant)
- $cov(z,u) = 0$ (z is exogenous)

- given these assumptions, $cov(y,z) = \beta_1 cov(x,z)$
- thus $\beta_1 = \frac{cov(y,z)}{cov(x,z)}$

$$\hat{\beta}_1^{IV} = \frac{\hat{cov}(y_i,z_i)}{\hat{cov}(x_i,z_i)}$$

Inference with IV Estimation

IV estimates are asymptotically normal

under homoskedasticity

$$AsyVar(\beta_1) = \frac{\sigma^2}{N\sigma_x^2\rho^2}$$

- where ρ^2 is the square of the correlation between x and z

- the estimate of the variance would be the sample analog

IV versus OLS estimation

- standard errors in IV case differs from OLS only in R^2 from regressing x on z
- since $R^2 < 1$, IV standard errors are larger
- however, IV is consistent, while OLS is inconsistent
- the stronger the correlation between z and x, the smaller the IV standard errors

The Effect of Poor Instruments

- what if $cov(z, u) \neq 0$?
- the IV estimator will be inconsistent
- however, it can still be better than OLS
- asymptotic bias in IV will be smaller than asymptotic bias in OLS if

$$\frac{corr(z,u)}{corr(x,u)} < corr(x, u)$$

IV estimation in the general case

$$y_1 = \beta_0 + \beta_1 z_1 + \beta_2 y_2 + u$$

$$cov(z_1, u) = 0$$

$$cov(y_2, u) \neq 0$$

- y_2 is endogenous, but there is an instrument for each endogenous variable in y_2, $cov(z_2, u) = 0$

IV estimation in the general case

- the IV estimator exploits in the sample the population conditions

$$cov(z_1, u) = 0$$

$$cov(z_2, u) = 0$$

$$\sum \left(y_{1i} - \hat{\beta}_0 - \hat{\beta}_1 z_1 - \hat{\beta}_2 y_2 \right) = 0$$
$$\sum z_{1ij} \left(y_{1i} - \hat{\beta}_0 - \hat{\beta}_1 z_1 - \hat{\beta}_2 y_2 \right) = 0$$
$$\sum z_{2ij} \left(y_{1i} - \hat{\beta}_0 - \hat{\beta}_1 z_1 - \hat{\beta}_2 y_2 \right) = 0$$

A simple example: estimating a demand function

a supply and demand system of equations

- supply function: $q = \gamma_0 + \beta^s p + \gamma x^s + u^s$
- demand function: $q = \alpha_0 + \beta^d p + \alpha x^d + u^d$

At equilibrium, $q = q(x^s, x^d, u^s, u^d)$, $p = p(x^s, x^d, u^s, u^d)$

$$cov(p, u^d) \neq 0$$

"identification" of β^d using a "supply shifter"

- $cov(x^s, p) \neq 0$ (relevance) (because p is a function of x^s)
- $cov(x^s, u^d) = 0$ (exogeneity) (otherwise, x^s is not really a "supply shifter")

Summary

- When a control is correlated with the error term, then OLS is inconsistent
- To implement IV we need an instrument: a variable which affects the dependent variable only via the control
- For example, if we want to estimate the price elasticity in a demand equation, we need a "supply shifter"