Instrumental Variables

Econometrics II

Raquel Carrasco & Ricardo Mora

Department of Economics
Universidad Carlos III de Madrid
Máster Universitario en Desarrollo y Crecimiento Económico

Outline

Why Use Instrumental Variables?

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

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

- OLS exploits in the sample a property which is false for the population
- we want to exploit in the sample a property which is true for the population

Instruments

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

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

An instrument z is a variable whose influence on the dependent variable is only via a control

- z is relevant in the sense that it correlates with controls:

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

- z is exogenous in the sense that controls capture all its effects on the dependent variable:

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

- each exogenous control is an instrument of itself

College Education (1/2)

Returns to college education among young workers

- $wages = \beta_0 + \beta_1 college + u$
- people freely choose to go to college: $cov(college, u) \neq 0$
- a good instrument is a variable in the sample that:
    - makes going to college more likely (relevance)
    - does not affect wages directly (exogeneity)

College Education (2/2)

Distance between pre-college residence and college

- individuals who live in the proximity of college will be more likely to go to college (relevance)
- pre-college 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 (1/2)

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

- children start schooling in the year when they are 6 BY JANUARY 1ST
- thus, children born in the same year enter school in the same year
- children must remain in school until they are 16 BY THE SCHOOL ENTRY DATE
- those born in January may leave one year before those born in December

The returns to Compulsory Attendance Laws in the US (2/2)

Think of those students restricted by the attendance laws

- month of birth correlates with months of education (relevance)
- month of birth (presumably) does not correlate with ability (exogeneity)

Lifetime Earnings and War Veterans in the US

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

- individuals with fewer alternatives are more likely to join the army and go to war
- thus, a dummy for veteran status is likely to be correlated with unobservables

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

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

- $cov(x,u) \neq 0$ (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 $\beta_1 = \frac{cov(y,z)}{cov(x,z)}$

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

IV versus OLS estimation

- IV only exploits the variance in the control which is correlated with the instrument
- IV standard errors are larger than OLS standard errors
- however, IV is consistent, while OLS is inconsistent
- the stronger the correlation between z and x, the smaller the IV standard errors

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, $cov(z_2, u) = 0$

- the IV estimator exploits in the sample these population conditions

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")

One IV Estimator per Instrument

- it is possible to have more than one instrument for each variable

$wages = \beta_0 + \beta_1 educ + u$

- $cov(educ, u) \neq 0$

Two instruments:

- father's education: *fed*
- mother's education: *med*

which instrument should we use?

$$\hat{\beta}_1^{fed} = \frac{\hat{cov}(wages,fed)}{\hat{cov}(educ,fed)} \neq \hat{\beta}_1^{med} = \frac{\hat{cov}(wages,med)}{\hat{cov}(educ,med)}$$

Which Instrument Should We Use?

using only one instrument is inefficient

- $\hat{\beta}_1^{fed}$ only exploits $cov(fed, u) = 0$
- $\hat{\beta}_1^{med}$ only exploits $cov(med, u) = 0$

the most efficient estimator uses a combination of both

$$\alpha * cov(fed, u) + (1 - \alpha) * cov(med, u) = 0$$

- this is known as "two-stages least squares"

2SLS in the general case

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

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

$$cov(z_2^1, u) = 0$$

$$cov(z_2^2, u) = 0$$

- the 2SLS estimator exploits in the sample these three sets of population conditions
- the weights for the second and third sets of conditions depend on how good instruments z_2^1 and z_2^2 are.

Two ways of obtaining the 2SLS estimates

First step

OLS y_2 on z_1 and z_2, compute $\hat{y}_2$

Second step: two versions

- Version A:

  IV y_1 on z_1 and z_2 using $(z_1, \hat{y}_2)$ as instruments of (z_1, y_2)

- Version B:

  OLS y_2 on z_1 and $\hat{y}_2$

STATA and Two-Stages Least Squares (3/3)

- versions A and B of step 2 give exactly the same output
- but let STATA do the estimation for you to get the correct (robust) standard errors
  - help ivregress
- also use STATA test comand to test for linear restrictions
  - help ivregress postestimation
- you need at least as many instruments as there are endogenous variables

Testing for endogeneity: a Hausman test

- since OLS is preferred to IV, we'd like to be able to test for endogeneity to avoid IV
- if we do not have endogeneity, both OLS and IV are consistent, although OLS is more efficient
- if we have endogeneity, only IV is consistent
- A Hausman test for endogeneity: H_0 : OLS and IV are consistent

$$\text{under } H_0,\ H \rightarrow \chi^2_q$$

Testing for endogeneity: a t test

- First step: regress potentially endogenous variable y_2 on all exogenous variables and compute residual $\hat{v}$
- (under endogeneity, $\hat{v}$ should be correlated with the error u)
- OLS equation of interest including endogenous variable and residual $\hat{v}$
- (this is like adding the missing variable which captures the correlation between y_2 and u)
- under exogeneity, the slope for $\hat{v}$ should not be significant

Testing overidentifying restrictions

- if there is just one instrument for each endogenous variable, we can't test whether the instrument is uncorrelated with the error
- we say the model is just identified
- if we have multiple instruments for each endogenous variable, it is possible to test if the "overidentifying" instruments are good instruments
- this is called testing for overidentifying restrictions

The OverID test

- estimate the structural model using IV and obtain the residuals
- regress the residuals on all the exogenous variables and obtain the R^2

under the null that all instruments are uncorrelated with the error

$$LM = nR^2 \to \chi^2_q$$

- where q is the number of extra instruments

Summary

- when a control is likely 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 dependent variable
- if we want to estimate the price elasticity in a demand equation, we need a “supply shifter”
- we can use more than one instrument efficiently using 2SLS
- we can test for endogeneity and also for the validity of the extra instruments