IV Estimation

SS 2011

Alexander Spermann

Evaluation With Non-Experimental Approaches

Selection on Unobservables

Natural Experiment

(exogenous variation in a variable)

DiD

Example:
Card/Krueger (1994)

Minimum wage

IV

Instrument z correlated with endogenous x, but uncorrelated with u

Non-testable identifying assumption

=exclusion restriction

„excludes direct causal effect on outcome“ (van den Berg 2007)

Conditions for an Instrument

1) Cov(z,u) = 0 — **exogeneity condition cannot be tested**

implication

$$y \;\not\leftarrow\; z \;\rightarrow\; D$$

(Instrument)

because

$$z \;\not\rightarrow\; u \;\rightarrow\; y$$

instrument	**unobserv. variables**	**outcome**

Conditions for an Instrument

2) $Cov(z,x) \neq 0$

relevance condition can be tested

implication

$$\begin{array}{c} x \\ \uparrow \\ z \end{array}$$

correlation $\nearrow$ **negative**

correlation $\searrow$ **positive**

IV Estimator

$$\boxed{\hat{\beta}_{IV} = \frac{Cov(z,y)}{Cov(z,x)}}$$

if z = x i.e. x is exogenous

$$\Rightarrow \hat{\beta}_{IV} = \hat{\beta}_{OLS} = \frac{Cov(x,y)}{Cov(x,x)} = \frac{Cov(x,y)}{Var(x)}$$

Example: Return to Education (Mincer equation)

$$\log(wage) = \beta_0 + \beta_1\, educ + \underbrace{\beta_2\, abil + u}_{\text{if no proxy available}}$$

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

$\Longrightarrow$ **OLS would be biased and inconsistant because OVB i.e. Cov(x,u) $\neq$ 0**

$\Longrightarrow$ **endogeneity problem**

$$\boxed{\hat{\beta}_{OLS} = 11\%}$$

Instrumental Variables for Education

1) Instrument IQ?

- Correlated with y
- Correlated with u

$\Rightarrow$ good proxy, but no instrument for ability

2) Instrument mother‘s education?

- Correlated with x
- But also correlated with u via child‘s ability

$\Rightarrow$ no instrument

3) Instrument number of siblings?

- Negative correlated with x

  (some evidence on that)
- If no correlation with ability

$\Rightarrow$ instrument

$$\boxed{\hat{\beta}_{IV} = 12.2\% \;>\; \hat{\beta}_{OLS} = 11\%}$$

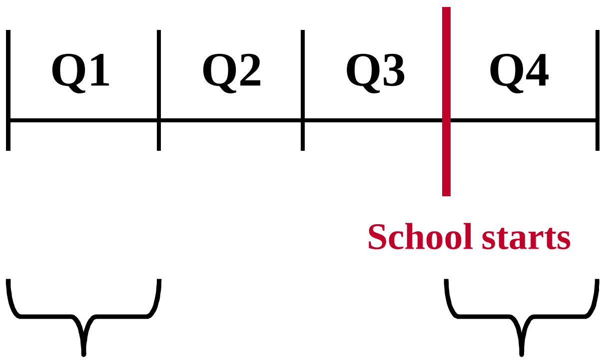
OLS underestimates true value

Binary IV

Angrist/Krueger (1999)

Census data for men 1980, born in the 30s

Instrument: **quarter of birth for education**



Born in Q1 - older than 6 years

→ Later in school than born in Q2 – Q4

younger than 6 years

School start – age policies ?? Bavaria 2009

Binary IV

- Start school at an older age + leave school with 16 years (birthday) (compulsory schooling laws)
- End with less education than others at university
- Born in Q1, earn less

Correlation

1. No correlation with ability

weak instrument

instrument is correlated with other unobserved factors

2. Correlation with educ

$\Longrightarrow$ large data set

$$\boxed{\begin{array}{l} \hat{\beta}_{OLS} = 8\% \\ \hat{\beta}_{IV} = 7.15\% \end{array}}$$

OLS overestimates

Earn less

Instrument: College Proximity (Binary Variable)

Card 1995

$$\log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \ldots + u$$

Instrument

Proximity to college

1 if near college

0 if far from college

Instrument: College Proximity (Binary Variable)

Correlation

1. **No correlation with u**
2. **Correlation with x (educ)**

   ??? by regression educ on nearc4

$$\boxed{\begin{aligned}\hat{\beta}_{OLS} &= 7.5\% \\ \hat{\beta}_{IV} &= 13.2\%\end{aligned}}$$

But large standard errors (18 x OLS s.e.)

→95% confindence interval

0.024 … 0.235

→ This is the price to pay for a consistant estimator

Instrument binary variable: veteran

Angrist 1990, AER

$$\log(earn) = \beta_0 + \beta_1 veteran + u$$

binary
variable

Correlated (self selection)

→ OLS biased and inconsistent

RSN= random sequence numbers

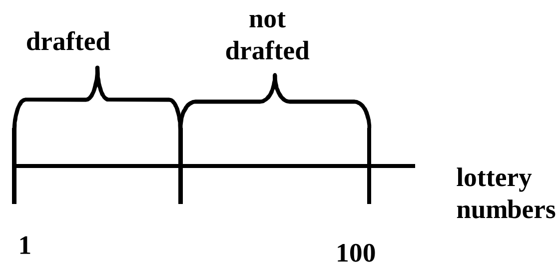
randomly assigned to birthdays

Vietnam draft lottery (1970)

→ Natural experiment

Lottery numbers to young men (=instrument for veteran)

→ randomly assigned



Instrument Binary Variable: Veteran

Correlation

1. Uncorrelated with u due to random assignment
2. Correlated with x (veteran) because low numbers → service in Vietnam

Result

- Veterans earn less ten years later
- Theory: penalty for lack of labor market experience

Dummy Variable Instrument (Caliendo)

Binary instrument z* with {0,1}

Source of exogenous variation to approximate randomised trials

$$\hat{\beta}_{IV} = \frac{E(y \mid x, z^* = 1) - E(y \mid x, z^* = 0)}{Y(D = 1 \mid x, z^* = 1) - Y(D = 1 \mid x, z^* = 0)}$$

Wald estimator

Problems of the Wald Estimator

1. **Weak instrument**

   things could be worse

   → inefficiency

   → inconsistency

2. **Heterogenous treatment framework**

   → IV not applicable

   → LATE is parameter of interest

Heterogenous Effects

$z \in \{0,1\}$

$\downarrow$

$D = 1 \quad or$

$D = 0$

population

subgroups

never takers

D=0

D=1

→?=0

always takers

D=1

D=0

→?=1

defier

change behaviour due to switch in instrument

z=0
z=1

D=0
D=1

in perverse way

complier

change behaviour in line with the instrument

before D=0

then z=0
z=1 **rule change**

after D=1

monotonicity assumption

no coexistence of defiers and compliers

Application JTPA

Control group substitution bias

Treatment group dropout bias

IV could control for that

LATE is defined for compliers

$$\hat{\beta}_{IV,LATE} = \frac{E(J_i \mid X_i, \tilde{z}_i = 1) - E(J_i \mid X_i, \tilde{z}_i = 0)}{P(D_i = 1 \mid X_i, \tilde{z}_i = 1) - P(D_i = 1 \mid X_i, \tilde{z}_i = 0)}$$

more details: Angrist/Pischke 2009

Imbens/Wooldridge 2009