IV Estimation

Dr. Alexander Spermann

Summer Term 2012
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 corellated with endogenous x, but uncorrelated with u

Non-testable identifying assumption = exclusion restriction (intuition:
„excludes direct causal effect on outcome“ (van den Berg 2007)
Conditions for an Instrument

1) Cov (z,u) = 0

**Exogeneity condition**
cannot be tested

Endogeneity could be
Cov (z,u) ≠ 0

Implication

Because

Diagram:

```
 y ——— D
    |    |
    v    |
 z ——— u ——— y
```

Instrument: unobserved variables

Outcome: y
Conditions for an Instrument

2) $\text{Cov}(z,x) \neq 0$

relevance condition can be tested

implication

$x$

$z$

correlation

negative

positive
IV Estimator

\[ \hat{\beta}_{IV} = \frac{\text{Cov}(z, y)}{\text{Cov}(z, x)} \]

if z=x i.e. x is exogenous

\[ \hat{\beta}_{IV} = \hat{\beta}_{OLS} = \frac{\text{Cov}(x, y)}{\text{Var}(x)} \]
Example: Return to Education (Mincer equation)

$$\log(wage) = \beta_0 + \beta_1 \text{educ} + \beta_2 \text{abil} + u$$

if no proxy available

$$= \beta_0 + \beta_1 \text{educ} + u$$

educ is obviously endogenous and hides abil in the error term $u$

→ OLS would be biased and inconsistent because OVB i.e. $\text{Cov}(x,u) \neq 0$

→ endogeneity problem

$$\beta_{OLS} = 11\%$$
Instrumental Variables for Education

Plausible arguments for the exogeneity ambition to hold

1) Instrument IQ?
   - Correlated with y
   - Correlated with u (ability)
     \[ \hat{\beta}_{IV} = 12.2\% > \hat{\beta}_{OLS} = 11\% \]
     no instrument for educ

2) Instrument mother’s education?
   - Correlated with x
   - But also correlated with u via child’s ability
     \[ \hat{\beta}_{IV} = 12.2\% > \hat{\beta}_{OLS} = 11\% \]
     no instrument

3) Instrument number of siblings?
   - Negative correlation with x
     (some evidence on that)
   - If no correlation with ability
     OLS underestimates true value
Binary IV


Instrument: quarter of birth of education

<table>
<thead>
<tr>
<th>Q1</th>
<th>Q2</th>
<th>Q3</th>
<th>Q4</th>
</tr>
</thead>
</table>

School starts

Born in Q1 - older than 6 years
Younger than 6 years

→ Later in school
Than born in Q2-Q4
Binary IV

Institutional detail in the U.S.:

- Start school at an older age + leave school with 16 years (birthday) (compulsory schooling laws) with a legal leaving age
- End with less education than others at the legal leaving age
- Born in Q1 means less schooling

Correlation

1. No correlation with ability
2. Correlation with educ

large data set

Weak instrument

\[
\hat{\beta}_{OLS} = 8\%
\]

\[
\hat{\beta}_{IV} = 7.15\%
\]

OLS overestimates Earn less

(cause to positive correlation between education and quarter of birth, i.e. the higher Q the higher educ).
Instrument: College Proximity (Binary Variable)

Card 1995

\[
\log(\text{wage}) = \beta_0 + \beta_1 \text{educ} + \beta_2 \text{exp} + ... + u
\]

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)
   check by regression educ on nearc4 (is it negative?)

\[ \hat{\beta}_{OLS} = 7.5\% \]
\[ \hat{\beta}_{IV} = 13.2\% \]

OLS underestimates true value due to negative correlation
(the closer distance the higher education)

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

95% confidence interval
0.024….0.235

This is the price to pay for a consistent estimator
Instrument binary variable: veteran

\[ \log(earn) = \beta_0 + \beta_1 \text{veteran} + u \]

Angrist 1990, AER

RSN = random sequence numbers randomly assigned to birthdays

Vietnam draft lottery (1970)

Lottery numbers to young men (=instrument for veteran) randomly assigned

Correlated (self selection) H OLS biased and inconsistent

drafted not drafted

lottery numbers 1 100

15.06.2012

IV Estimation
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

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

Wald estimator
Problems of the Wald Estimator

1. Weak instrument
   things could be worse
   
   → inefficency
   → inconsistency

2. Heterogenous treatment framework
   
   → IV not applicable
   → LATE is parameter of interest
Heterogeneous Effects

\[ z \in \{0,1\} \]

\[ \downarrow \]

D=1 or
D=0

\[ \downarrow \]

population

\[ \downarrow \]

subgroups

never takers

always takers

defier

complier

D=0

D=1

D=0

D=1

R=0

R=1

change behaviour due to switch in instrument

change behaviour in line with the instrument before D=0

then \( z=0 \)

then \( z=1 \)

after D=1

In perverse way

monotonicity assumption

no coexistence of defiers and compliers

15.06.2012

IV Estimation
Application JTPA

Control group substitution bias
Treatment group dropout bias

IV could control for that

15.06.2012

IV Estimation
LATE is defined for compliers

\[ \hat{\beta}_{IV,LATE} = \frac{E(J_i \mid X_i, z_i = 1) - E(J_i \mid X_i, z_i = 0)}{P(D_i = 1 \mid X_i, z_i = 1) - P(D_i = 1 \mid X_i, z_i = 0)} \]

more details: Angrist/Pischke 2009
Imbens/ Wooldridge 2009