Instrumental Variables

Lecture 5

Kristiina Huttunen

Aalto University

Semptember 17, 2021

Mini-course on Causal Inference

Motivation: Selection Problem

- Our aim is to evaluate the impact of S (schooling) on Y (earnings)
- The relationship between outcome Y (earnings) and treatment (S schooling)

$$Y_i = \alpha_0 + \rho S_i + \eta_i$$

$$\eta_i = A_i' \gamma + v_i$$

- Challenge: We do not observe everything (A_i) that affects both selection into treatment S and earnings Y.

Instrumental variable is a solution!

- How to estimate ρ without observing A_i?

$$Y_i = \alpha + \rho S_i + A_i'\gamma + v_i$$

- Instrumental variable (IV) allows us to estimate ρ when A_i is unobserved
- Instrumental variable is a variable (Z_i) that:
  1. Is correlated with causal variable of interest, S_i, $\text{Cov}(Z_i, S_i) \neq 0$
  2. Is uncorrelated with any other determinants of Y_i $\text{Cov}(Z_i, \eta_i) = 0$

- With a valid instrumental variable we can consistently estimate ρ in

$$Y_i = \alpha + \rho S_i + A_i'\gamma + v_i$$

- We can write ρ in terms of the population moments

$$\text{Cov}(Z_i, Y_i) = \rho\text{Cov}(Z_i, S_i) + \text{Cov}(Z_i, \eta_i)$$

- Given the exclusion restriction, $\text{Cov}(Z_i, \eta_i) = 0$, it follows that

$$\rho = \frac{Cov(Z_i,Y_i)}{Cov(Z_i,S_i)} = \frac{\frac{Cov(Z_i,Y_i)}{V(Z_i)}}{\frac{Cov(Z_i,S_i)}{V(Z_i)}}$$

First Stage Regression and Reduced Form Regression

- The coefficient of interest, ρ, is the ratio between regression of Y_i on Z_i (the reduced form) and regression of S_i on Z_i (the first stage).
- First stage

$$S_i = X_i'\pi_{10} + \alpha_1 Z_i + \epsilon_{1i}$$

- Reduced form

$$Y_i = X_i'\rho + \gamma_1 Z_i + \eta_i$$

- Effect of Treatment (S) on Outcome (Y)

$$\rho = \frac{\gamma_1}{\alpha_1} = \frac{\text{reduced form}}{\text{first stage}}$$

Two-stage Least Squares

- First stage

$$S_i = X_i' \pi_{10} + \alpha_1 Z_i + \epsilon_{1i}$$

- Substituting the first-stage fitted values for S_i in equation of interest

$$Y_i = X_i' \theta + \rho \hat{S}_i + u_i$$

Why IV works?

- Intuitive idea behind IV is as follows
- S varies in response to η
- It also varies in response to Z
- Z does not varies as η changes
- We exploit the variation in S that is due to the variation in Z, to identify the effect of S on Y

Good Instruments are hard to find

Some examples of instruments.

- Randomized settings (RCTs): Lottery for selective school offers. Use the lottery (winning/loosing) as instrument for having accepted the offer to the school (E.g. Boston charter schools).
- Assignment of court cases to judges: Law requires randomness! Use the difference in judge's propensity to send people to prison, as instrument for prison sentence.
- Policies/rules: e.g. changes in unemployment benefit levels.
- Note that, in general, choice variables of the agent tend to be bad instruments (e.g. live close to university)!

Example: Effect of Foster Care on Criminal Behavior

Doyle, J. (2007) "Child Protection and Child Outcomes: Measuring the Effects of Foster Care", American Economic Review.

- Children placed in foster care tend to have a higher propensity to commit crime, drop out of school, be on welfare...
- Obviously this tells us nothing about causal effect of foster care (does it help or harm the kids?)

How does foster care (D) affect juvenile delinquency (Y)?

Doyle, J. (2007) "Child Protection and Child Outcomes: Measuring the Effects of Foster Care", American Economic Review.

- A naive estimate: mean comparision

$$Y_i = \alpha + \beta D_i + \epsilon_i$$

- Conditional expectations comparison

$$Y_i = \alpha + \beta D_i + X_i' \gamma + \epsilon_i$$

IV-strategy

Doyle, J. (2007) “Child Protection and Child Outcomes: Measuring the Effects of Foster Care”, American Economic Review.

- Paper exploits the fact there is a rotation system that assigns children to case managers, who decide who will be placed in foster care (removed from home)
- Some case managers have a higher tendency to place children in foster care
- Children assigned to case managers with high tendency to place children to foster care have higher probability to be placed in foster care
- Binary Instrument: High placement propensity (1) /low placement propensity (2).

Assumptions for valid IV

- Is there a first stage? (Do children who are assigned to a case manager with higher previous placement propensity have higher probability for foster care?)

$$D_i = \alpha_0 + \alpha_1 Z_i + \epsilon_i$$

- Does exclusion restriction hold? (Is the case manager placement propensity only affecting future outcomes of these children through the probability to be placed in foster care)

$$Y_i = \beta_0 + \beta_1 D_i + \eta_i$$

$$= \beta_0 + \beta_1[\alpha_0 + \alpha_1 Z_i + \epsilon_i] + \eta_i$$

$$= \gamma_0 + \gamma_1 Z_i + \phi_i \text{ (reduced form)}$$

The Wald estimator

- Since instrument is binary (case manager with high or low previous placement propensity) we can write the IV estimator as Wald estimator

$$\beta_1 = \frac{Cov(Y_i, Z_i)}{Cov(D_i, Z_i)}$$

$$= \frac{E[Y_i|Z_i=1] - E[Y_i|Z_i=0]}{E[D_i|Z_i=1] - E[S_i|Z_i=0]}$$

$$= \frac{\gamma_1}{\alpha_1} = \frac{\text{reduced form}}{\text{first stage}}$$

Table 2
Table of means: instrumental variable estimation.

		Investigator placement propensity			
		High	Low	Difference	p-value
A. First stage	Foster care placement	0.316	0.224	0.092	<0.0001
B. Reduced form	Juvenile delinquency	0.171	0.158	0.013	0.043
C. IV estimate	Change in juvenile delinquency: Change in foster care placement	Difference in B ÷ difference in A:		0.142	
		p-value:		0.035	
	Observations	7792	7889		

Juvenile Delinquency Sample: Children in Cook County who received an abuse/neglect report between July 1, 1990 and December 31, 2000 and were at least 15 in 2000. p-values calculated using standard errors clustered at the investigator level.

Table 3
Introduction of covariates.

	First stage		Reduced form		IV estimates	
Dependent variable:	Foster care placement		Juvenile delinquency		Juvenile delinquency	
Estimator:	OLS (1)	OLS (2)	OLS (3)	OLS (4)	2SLS (5)	2SLS (6)
Investigator placement propensity = high	0.092 (0.013)**	0.091 (0.009)**	0.013 (0.0065)**	0.017 (0.0058)**		
Foster care placement					0.142 (0.067)**	0.183 (0.063)**
F-statistic (Ho: above coefficient = 0)	47.8	102				
Full controls	No	Yes	No	Yes	No	Yes
Mean of Dependent Variable	0.269		0.164		0.164	
Number of Investigators	409					
Observations	15681					

Standard errors are reported, clustered at the investigator level. Full controls include indicators for the type of initial reporter, year of age, sex, race, type of allegation, and ZIP code of residence.

Conclusions from Doyle

- Foster care increases juvenile delinquency
- IV is even higher than OLS
- Doyle explains this by stating that for children on the margin to be placed in FC the impact is more harmful than for others (who benefit more)
- This argument rests on idea that the impact of foster care is heterogenous

Heterogenous potential outcomes

- The discussion of IV up to this point postulates a constant causal effect. In the case of a dummy variable Y_i, this means:

$$Y_{1i} - Y_{0i} = \rho \text{ for all } i$$

- Let us consider (next slides) a more general case where the effect might be heterogeneous...
- Examples: cancer treatment, foster care...

LATE framework

- What does IV estimate if $Y_{1i} - Y_{0i}$ is not the same for everyone?
- LATE = Local Average Treatment Effect
- Let $Y_i(d, z)$ denote the potential outcome for individual i whose treatment status is $D_i = d$ and instrument value $Z_i = z$
- We assume causal chain: instrument (Z_i) affects treatment (D_i) which in turn affects outcome (Y_i).

LATE framework

- D_{1i} is treatment status when $Z_i = 1$
- D_{0i} is treatment status when $Z_i = 0$
- Observed treatment status is

$$D_i = D_{0i} + (D_{1i} - D_{0i})Z_i$$

- For all i we have
- Potential outcomes: $Y_i(0,0)$, $Y_i(1,0)$, $Y_i(0,1)$, $Y_i(1,1)$
- Potential treatments: $D_{0i} = 0$,$D_{0i} = 1$,$D_{1i} = 0$,$D_{1i} = 1$
- Potential assignments: $Z_i = 0$, $Z_i = 1$

Classification of individuals according to treatment and assignment

		$Z_i = 0$	
		D_{0i}=0	D_{0i}=1
Z_i =1	D_{1i}=0	Never-taker	Defier
	D_{1i}=1	Complier	Always taker

LATE assumptions

1. Independence: instrument is as good as randomly designed
2. Exclusion Restriction: affects outcome through single know channel
3. First Stage: $E[D_{1i} - D_{0i}] \neq 0$
4. Monotonicity: $D_{1i} \geq D_{0i}$ for everyone (or vice versa). All those who are affected are affected in the same way.

The last one is a necessary technical assumptions that is needed for IV to have LATE interpretation

LATE

- If the LATE assumptions hold

$$\rho = \frac{E[Y_i|Z_i{=}1]-E[Y_i|Z_i{=}0]}{E[D_i|Z_i{=}1]-E[D_i|Z_i{=}0]} = E[Y_{1i} - Y_{0i}|D_{1i} > D_{0i}]$$

- The IV estimates the impact of treatment for those whose behavior changed because of the instrument
- If treatment effect is heterogenous, different instruments can give different effects

Example: Angrist ad Evans, 1998

Children and Their Parents Labor Supply

- A causal model for the impact of more than two children
$$Y_i = \alpha_0 + \rho_1 D_i + \eta_i$$
- Dependent variable, Y_i: employed, hours worked, weeks worked, earnings
- $D_i = 1[kids > 2]$: More than two children (sample includes only families with at least 2 children)
- Alternative D_i: Number of children
- Two alternative instruments Z_i
  - $Z_i = 1$ Twins at second birth
  - $Z_i = 1$ Same sex sibling at second birth

TABLE 11—COMPARISION OF 2SLS ESTIMATES USING *SAME SEX* AND *TWINS-2* INSTRUMENTS IN 1980 CENSUS DATA

	All women		Married women		Husbands	
Model	(1)	(2)	(1)	(2)	(1)	(2)
Instrument for *More than 2 children*	*Same sex*	*Twins-2*	*Same sex*	*Twins-2*	*Same sex*	*Twins-2*
Dependent variable:						
Worked for pay	−0.125 (0.026)	−0.079 (0.013)	−0.123 (0.028)	−0.087 (0.017)	0.004 (0.009)	−0.001 (0.005)
Weeks worked	−5.82 (1.15)	−3.64 (0.60)	−5.47 (1.23)	−4.21 (0.72)	0.65 (0.61)	−0.35 (0.36)
Hours/week	−4.76 (0.98)	−3.33 (0.51)	−4.91 (1.03)	−3.49 (0.61)	0.57 (0.71)	−0.49 (0.42)
Labor income	−1961.7 (560.5)	−1262.2 (292.8)	−1329.8 (579.1)	−1453.1 (339.8)	−1194.8 (1421.4)	616.8 (836.9)
ln(*Family income*)	−0.021 (0.067)	−0.071 (0.035)	−0.049 (0.057)	−0.025 (0.033)	—	—
ln(*Non-wife income*)	—	—	0.026 (0.068)	0.051 (0.040)	—	—

Notes: The table reports 2SLS estimates of the coefficient on *More than 2 children* in equation (4) in the text using *Same sex* and *Twins-2* as instruments. Other covariates in the models are *Age, Age at first birth,* ages of the first two children, plus indicators for *Boy 1st, Boy 2nd, Black, Hispanic,* and *Other race*. Data are from the 1980 Census. Standard errors are reported in parentheses.

Two Different Instruments, Two Different Estimates...

- Estimates generated by twin instruments lower estimates that are base on same sex instrument. Why?
- Use LATE interpretation: different complier groups
  - Same sex: Parents that had a third child only because they want to have children of different sex
  - Twins: Parents that would not have had more than 2 kids had they not had twins (are there any never takers?)
- Other reasons: Validity

IV in Randomized Trials

- Compliance Problem in Randomized Experiments: Some assigned to treatment group are not treated
- When compliance is voluntary, an as-treated analysis is contaminated by selection bias
- Intention-to-treat analysis preserves independence but are diluted by non compliance
- IV solves this problem: Use random assignment as instrument for actual treatment

IV in Randomized Trials

- Z_i is a dummy variable indicating random assigment to the treatment
- D_i is a dummy indicating whether the treatment was actually received
- There are no always takes (no controld actually treated): $E[D_i|Z_i = 0] = 0$
- Wald Estimator:

$$\rho = \frac{E[Y_i|Z_i=1] - E[Y_i|Z_i=0]}{E[D_i|Z_i=1]} = \frac{ITT}{ComplianceRate} = E[Y_{1i} - Y_{0i}|D_i = 1]$$

- LATE is the Average Effect of Treatment on Treated

Let us wrap it up

- IV estimates are a powerful tool to identify causal links
- But IV power relies on the quality of the instruments
- Two dimensions:
  1. Powerful (can be tested in the first stage!)
  2. Must be exogenous (cannot be tested, but...)
- If treatment effect is heterogenous, we should keep in mind what is the group of compliers (What is LATE)
- Check Josh Angrist's IV lecture on you tube!