Applied Microeconometrics I

Lecture 6: Instrumental variables

Tuomas Pekkarinen

Aalto University

September 30 2021

`Lecture Slides`

What did we do last time?

- Problems with the CIA
- Bad controls:X that are themselves caused by D
- Example: Effect of college ($C_i = 1$) among white collar workers ($W_i = 1$)
- Assume that C_i is randomly assigned

$$\{Y_{0i}, Y_{1i}, W_{0i}, W_{1i}\} \perp\!\!\!\perp C_i$$

What did we do last time?

- Can we estimate: $E[Y_{1i} - Y_{0i}|W_{1i} = 1]$?
- We can only observe:

$$E[Y_i|W_i = 1, C_i = 1] - E[Y_i|W_i = 1, C_i = 0]$$

What did we do last time?

$$
\begin{aligned}
& E[Y_i|W_i = 1, C_i = 1] - E[Y_i|W_i = 1, C_i = 0] \\
= \quad & E[Y_{1i}|W_{1i} = 1, C_i = 1] - E[Y_{0i}|W_{0i} = 1, C_i = 0] \\
= \quad & E[Y_{1i}|W_{1i} = 1] - E[Y_{0i}|W_{0i} = 1] \\
= \quad & E[Y_{1i}|W_{1i} = 1] - E[Y_{0i}|W_{1i} = 1] \\
& + E[Y_{0i}|W_{1i} = 1] - E[Y_{0i}|W_{0i} = 1] \\
= \quad & \underbrace{E[Y_{1i} - Y_{0i}|W_{1i} = 1]}_{\text{Causal effect}} + \underbrace{E[Y_{0i}|W_{1i} = 1] - E[Y_{0i}|W_{0i} = 1]}_{\text{Selection bias}}
\end{aligned}
$$

What did we do last time?

- Measurement error in the independent variable

$$
\begin{aligned}
Y_i &= \mu + \tau X_i + v_i \\
\tilde{X}_i &= X_i + e_i
\end{aligned}
$$

- We assume that $Cov(X, e) = 0$ which implies that $Cov(\tilde{X}, e) = Var(e)$
- Running

$$Y_i = \mu + \tau \tilde{X} - \tau e_i = v_i$$

yields

$$
\begin{aligned}
\hat{\tau}_{OLS} &= \frac{Cov(\tilde{X}, Y)}{Var(\tilde{X})} \\
&= \tau \left(\frac{Var(X)}{Var(X) + Var(e)} \right)
\end{aligned}
$$

What did we do last time?

- How do controls potentially reduce bias?

$$\begin{aligned} Y_i &= \beta^s S_i + e_i^s \\ Y_i &= \beta S_i + \gamma X_i + e_i \\ X_i &= \rho S_i + u_i \end{aligned}$$

- Then $\beta^s - \beta = \gamma\rho$. So if $\beta^s \simeq \beta$, we can conclude that $\gamma = 0$ or that $\rho = 0$

What did we do last time?

- What if we have to use proxy control $\tilde{X}_i = X_i + u_i$ with $Cov(X, u) = 0$ and $Cov(X, S) = Cov(\tilde{X}, S)$?
- Running

$$Y = \beta^m S_i + \gamma \tilde{X}_i + e_i^m$$

yields

$$\begin{aligned} \hat{\gamma}^m &= \Lambda\gamma \\ \hat{\beta}^m &= \beta + \gamma\rho(1 - \Lambda) \end{aligned}$$

where $\Lambda = \frac{Var(S)Var(X) - Cov(S,X)^2}{[Var(X)+Var(u)]Var(S) - Cov(S,X)^2} < 1$

What did we do last time?

- So we have that $\beta^m - \beta^s = \Lambda\gamma\rho < \gamma\rho = \beta - \beta^s$
- It may seem that $\beta^m \simeq \beta^s$ even though $\beta - \beta^s \neq 0$
- But we can always run:

$$\tilde{X}_i = \rho S_i + u_i + v_i$$

which yields an unbiased estimate of ρ

Motivation

- Arriving at causal estimates by simply controlling for observables is challenging if not impossible
- Our data are unlikely to be rich enough to allow for credible *ceteris paribus* claims
- **Instrumental variables** are an often used strategy to arrive at causal inference when controlling for observables is not enough
- Idea: Look for variables (instruments) that generate partial or incomplete random assignment to our treatment of interest

Instrumental variables: Basics

- Suppose we are interested in the effect of D on Y
- We are unlucky because:
  - There is no RCT where D would be assigned randomly
  - We are unable to control for all the factors that are correlated with D and have an effect on Y
- Instrument Z is a variable that (incompletely) plays the part of the RCT in this kind of situation

Instrumental variables: Basics

- For Z to work as an instrument it has to fulfill the following conditions
  1. "First stage": Z has a causal effect on D
  2. "Independence": Z is as good as randomly assigned
  3. "Exclusion restriction": Z has an effect on Y only through D
- "Chain reaction" from Z to Y
- Z is assigned as good as random and has an effect on D without having a direct effect on Y

Instrumental variables: Basics

- Think of the following examples. Do they work as an instrument?
  - Draft lottery as an instrument for military service: **Yes**
  - Twin births as an instrument for family size: **Probably**
  - Parental education as an instrument for child's education: **No**
- Best instruments are like randomized trials that allocate the variable of interest across units

Instrumental variables: Interpretation

Reference: Angrist et al, 1996

- Consider a case where both Z and D take values 0 and 1
- **First stage**: Z assigns some individuals to treatment D

$$\phi = E[D_i|Z_i = 1] - E[D_i|Z_i = 0]$$

- **Reduced form**: We can estimate the causal effect of Z on Y:

$$\rho = E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]$$

- With the first stage and the reduced form we can calculate the **Local Average Treatment Effect (LATE)**

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

- LATE is the causal effect that instrumental variables identify

Instrumental variables:Interpretation

- How to interpret LATE?
- Think about instrumental variables as RCT's with **incomplete compliance**
- $Z_i = 1$ is a randomly allocated assignment to treatment $D_i = 1$
- However, we cannot fully control that:
  - That all those with $Z_i = 1$ actually get the treatment $D_i = 1$
  - That none of those with $Z_i = 0$ gets the treatment $D_i = 1$
- That is, we cannot rule out the existence of **never-takers** and **always-takers**
- Use the notation of counterfactual treatment outcomes
- We have

$$D_{1i} = D_{0i} = 0 \text{ for never-takers, and}$$

$$D_{1i} = D_{0i} = 1 \text{ for always-takers}$$

Potential treatments

	Not assigned to treatment: $Z_i=0$	Assigned to treatment: $Z_i=1$
Does not take the treatment: $D_i=0$	$D_{0i}=0$	$D_{1i}=0$
Does take the treatment: $D_i=1$	$D_{0i}=1$	$D_{1i}=1$

Instrumental variables: Interpretation

- However, suppose we can rule out the existence of those who do not take the treatment when assigned and take the treatment when not assigned:

$$D_{1i} = 0 \text{ and } D_{0i} = 1$$

- These people are called **defiers** and the assumption that rules them out is referred to as **monotonicity**
- Those who only take the treatment when assigned are called **compliers**

$$D_{1i} = 1 \text{ and } D_{0i} = 0$$

Instrumental variables: Interpretation

- The effects of treatment may vary across these groups: **compliers, always-takers, never-takers, defiers**
- In this setting there are two potential assignments $[Z_i = 1, Z_i = 0]$
- Four treatment-assignment combinations $[D_i(0) = 0, D_i(0) = 1, D_i(1) = 0, D_i(1) = 1]$
- Four potential outcome-treatment combinations $[Y_i(0,0), Y_i(1,0), Y_i(0,1), Y_i(1,1)]$

The Four Groups

		Not assigned to treatment: $Z_i=0$	
		Does not take the treatment: $D_i=0$	Does take the treatment: $D_i=1$
Assigned to treatment: $Z_i=1$	Does not take the treatment: $D_i=0$	*Never-takers* $Y_i(0,1)-Y_i(0,0)=0$	*Defiers* $Y_i(0,1)-Y_i(1,0)=$ $-[Y_i(1)-Y_i(0)]$
	Does take the treatment: $D_i=1$	*Compliers* $Y_i(1,1)-Y_i(0,0)=$ $[Y_i(1)-Y_i(0)]$	*Always-takers* $Y_i(1,1)-Y_i(1,0)=0$

Instrumental variables: Interpretation

- Combining all the assumptions more formally:
  1. Independence: $\{Y_i(D_{1i}, 1), Y_i(D_{0i}, 0), D_{1i}, D_{i0}\} \perp\!\!\!\perp Z_i$
  2. Exclusion: $Y_i(d, 0) = Y_i(d, 1) \equiv Y_{di}$ for $d = 0, 1$
  3. First stage: $E[D_{1i} - D_{0i}] \neq 0$
  4. Monotonicity: $D_{1i} \geq D_{0i} \forall i$ or vice versa
- Which of these assumptions can we test? (3) Yes, (1) sort of, (2) and (4) No
- Under these assumptions IV estimates:

$$E[Y_i(1) - Y_i(0) | D_{1i} > D_{i0}] = \frac{E[Y_i|Z_i=1]-E[Y_i|Z_i=0]}{E[D_i|Z_i=1]-E[D_i|Z_i=0]}$$

- This is the effect of the treatment on those who only take the treatment because they were assigned to it by Z_i (Compliers)

Instrumental variables: Interpretation

- Why might LATE differ from the average treatment effect on the treated?
- Never-takers, always-takers, and compliers probably have a reason to behave as they do
  - Never-takers don't want the treatment under any circumstances
  - Always-takers want it no matter what
  - The compliers only take it if our instrument tells them to
- Potential outcomes, and hence the treatment effects, may differ across these groups
- Instrumental variables only give us the treatment effect for compliers
- Whether that is interesting or not depends on the application

Example

Who benefits from KIPP?

Angrist et al (2012)

- Controversial (in the US) topic: Charter schools
- Is attending charter schools good for student achievement?
- Kids attending charter schools are a selected group
- Student achievement is affected by a myriad of factors that may also affect the probability of attending a charter
- In Massachusetts entry to over-subscribed charter schools is decided by lottery

Example

Who benefits from KIPP?

Angrist et al (2012)

- Angrist at al obtain data on the applicants who participated in a lottery to gain entry to a KIPP Charter School in Lynn, MA
- They use winning the entry lottery (=gaining entry) as an instrument for attending KIPP
- Think about the following questions
  1. Is winning the entry lottery as good as randomly assigned?
  2. Does winning the entry lottery have a direct effect on student achievement?
  3. Does winning the entry lottery have an effect on attending KIPP?
  4. Are there likely to be defiers?

Example

Who benefits from KIPP? Observable characteristics

Table 1: Descriptive Statistics and Covariate Balance

	Means			Balance regressions	
	Lynn Public 5th graders	KIPP Lynn 5th graders	KIPP Lynn lottery applicants	No controls	Demographic controls
	(1)	(2)	(3)	(4)	(5)
Hispanic	0.418	0.565	0.538	-0.052	-
				(0.053)	
Black	0.173	0.235	0.256	0.033	-
				(0.044)	
White	0.296	0.168	0.179	-0.017	-
				(0.040)	
Asian	0.108	0.021	0.022	0.028*	-
				(0.015)	
Female	0.480	0.474	0.489	-0.002	-
				(0.054)	
Free/reduced price lunch	0.770	0.842	0.825	-0.031	-
				(0.041)	
Special Education	0.185	0.189	0.200	-0.009	-
				(0.043)	
Limited English Proficiency	0.221	0.172	0.206	-0.074	
				(0.047)	
Baseline Math Score	-0.307	-0.336	-0.389	0.097	0.034
				(0.114)	(0.107)
Baseline Verbal Score	-0.356	-0.399	-0.443	0.039	-0.036
				(0.119)	(0.105)
F-value from joint test				0.814	0.184
p-value from F-test				0.615	0.832
N for demographics	3964	285	446	446	446
N for baseline Math	3808	284	435	435	435
N for baseline ELA	3805	284	436	436	436

Example

Who benefits from KIPP? Lotteries

Table 2: KIPP Academy Lynn Lotteries

Lottery Cohort (1)	Calendar years observed (2)	Grades observed (3)	Number of applicants (4)	Number of applicants in lottery sample (5)	Percent offered (6)	Percent attended (7)	Average years at KAL (winners) (8)
2005-2006	2006-2009	5-8	138	106	0.925	0.670	2.56
2006-2007	2007-2009	5-7	117	86	0.674	0.535	2.29
2007-2008	2008-2009	5-6	167	118	0.627	0.534	1.68
2008-2009	2009	5	207	136	0.537	0.397	0.70
All cohorts	2006-2009	5-8	629	446	0.679	0.525	1.85

Example

Who benefits from KIPP?

- Unsurprisingly, things work beautifully
- Observable characteristics are balanced across lottery winners and losers
- We wouldn't expect winning a lottery to have any direct effects on achievement
- 303 children (=0.679*446) were offered a slot as a result of the lottery
- However only 221 (73%) of winners actually attend and somehow 5 (3.5%) losers also attend

Example

Who benefits from KIPP?

- So things are nearly perfect but not quite because not all winners actually attend
- Lottery is not a controlled RCT but an RCT with imperfect compliance, ie an instrument
- First stage: We see a clear effect of winning a lottery on attending
- Reduced form: Compare outcomes of losers and winners
- LATE: Divide the reduced form with the first stage
- Angrist et al are interested in the effects on test scores in standard deviation units

Example

Who benefits from KIPP? Effects on test cscores

Table 4: Lottery Results

		all applicants				Lynn public schools at baseline	
		First Stage	Reduced Form	2SLS	OLS	2SLS	OLS
Subject	Controls	(1)	(2)	(3)	(4)	(5)	(6)
Math	Basic	1.218***	0.437***	0.359***	0.301***	0.352***	0.304***
		(0.065)	(0.117)	(0.096)	(0.048)	(0.110)	(0.054)
		842	842	842	842	683	683
	Demographics	1.225***	0.399***	0.325***	0.312***	0.324***	0.332***
		(0.067)	(0.106)	(0.084)	(0.041)	(0.099)	(0.046)
		842	842	842	842	683	683
	Demographics &	1.221***	0.430***	0.352***	0.314***	0.352***	0.344***
	Baseline Scores	(0.068)	(0.067)	(0.053)	(0.032)	(0.064)	(0.038)
		833	833	833	833	675	675
ELA	Basic	1.218***	0.189	0.155	0.169***	0.224*	0.166***
		(0.065)	(0.118)	(0.096)	(0.049)	(0.115)	(0.057)
		843	843	843	843	684	684
	Demographics	1.228***	0.124	0.101	0.170***	0.159*	0.179***
		(0.068)	(0.098)	(0.078)	(0.041)	(0.092)	(0.046)
		843	843	843	843	684	684
	Demographics &	1.228***	0.164**	0.133**	0.174***	0.150**	0.185***
	Baseline Scores	(0.068)	(0.073)	(0.059)	(0.031)	(0.069)	(0.036)
		833	833	833	833	677	677

Example

Who benefits from KIPP?

- The first stage implies that attendance increase by 1.2 years (perfect compliance would imply 1.75)
- Winners score about 0.4 sd's higher than losers in math
- The LATE is 0.35 sd's
- This result is robust to including controls

When is LATE same as the effect on the treated?

- There is an important special case when instrumental variables actually give the treatment on the treated
- If there are no always takers so that $E[D_i|Z_i = 0] = 0$
- Then:

$$E[Y_i(1) - Y_i(0)|D_{1i} > D_{oi}] = \frac{E[Y_i|Z_i=1]-E[Y_i|Z_i=0]}{E[D_i|Z_i=1]}$$

- In these cases IV estimates the effect on the treated population

Example

Instruments and criminology

Angrist (2006)

- Angrist (2006) revisits a famous RCT on the treatment of domestic disturbance by the police force in Minneapolis
- RCT tried to address the question whether the officer should arrest the offender or "coddle" (=advise/separate)
- Upon arriving at the scene the officers were supposed to randomize by drawing a card with a coded color for each treatment
- The goal of the RCT was to find out how coddling affects recidivism

Example

Instruments and criminology: Assigned and delivered treatments

Table 1. Assigned and delivered treatments in spousal assault cases.

	Delivered treatment			
		Coddled		
Assigned treatment	*Arrest*	*Advise*	*Separate*	*Total*
Arrest	98.9 (91)	0.0 (0)	1.1 (1)	29.3 (92)
Advise	17.6 (19)	77.8 (84)	4.6 (5)	34.4 (108)
Separate	22.8 (26)	4.4 (5)	72.8 (83)	36.3 (114)
Total	43.4 (136)	28.3 (89)	28.3 (89)	100.0 (314)

The table shows statistics from Sherman and Berk (1984), Table 1.

Example

Instruments and criminology

- We see that when told to coddle 80% ($\frac{(84+5)+(5+83)}{108+114}$) actually coddled
- However, when not told to arrest only 1% ($\frac{1}{92}$) coddled
- Hence, there practically are no always-takers in this experiment

Example

Instruments and criminology: First stage and reduced form

Table 2. First stage and reduced forms for Model 1.

	Endogenous variable is coddled			
	First stage		*Reduced form (ITT)*	
	(1)	*(2)**	*(3)*	*(4)**
Coddled-assigned	0.786 (0.043)	0.773 (0.043)	0.114 (0.047)	0.108 (0.041)
Weapon		−0.064 (0.045)		−0.004 (0.042)
Chem. influence		−0.088 (0.040)		0.052 (0.038)
Dep. var. mean	0.567 (Coddled–delivered)		0.178 (Re-arrested)	

The table reports OLS estimates of the first-stage and reduced form for Model 1 in the text.
*Other covariates include year and quarter dummies, and dummies for non-white and mixed race.

- We see that being told to coddle lead to 78.6 percentage point increase in coddling
- We are interested in the effect of coddling on re-arrest rates
- The reduced form effect is 11.4 percentage points

Example

Instruments and criminology: OLS and IV

Table 3. OLS and 2SLS estimates for Model 1.

	Endogenous variable is coddled			
	OLS		*IV/2SLS*	
	(1)	*(2)**	*(3)*	*(4)**
Coddled–delivered	0.087 (0.044)	0.070 (0.038)	0.145 (0.060)	0.140 (0.053)
Weapon		0.010 (0.043)		0.005 (0.043)
Chem. influence		0.057 (0.039)		0.064 (0.039)

The Table reports OLS and 2SLS estimates of the structural equation in Model 1.
*Other covariates include year and quarter dummies, and dummies for non-white and mixed race.

Example

Instruments and criminology

- We see that if we would only compare coddles and arrests the effect would be 8.7 percentage points
- The reduced form effect is 11.4 percentage points
- LATE estimate is 14.5 percentage points (which is what we get if divide the reduced form with the first stage)
- Why do these estimates differ even though this was an RCT?
- Police officers didn't comply with the coddle assignment if they thought that an arrest was necessary
- Because this non-compliance is in practice one sided there are no always-takers
- Therefore all the treated are compliers

What did we do last time?

- Instrumental variables: Randmly assigned Z affects D which in turn may affect Y without Z directly affecting Y
- Needs to fulfill:
  1. First stage: Z affects D
  2. Independence: Z is as good as randomly assigned
  3. Exclusion restriction: Z has an effect on Y only through D
- RCT with incomplete compliance
- Assignment to treatment: $Z_i = \{0, 1\}$
- Counterfactual treatments $\{D_{0i}, D_{1i}\}$
- Incomplete compliance: $D_{0i} = \{0, 1\}$; $D_{1i} = \{0, 1\}$

What did we do last time?

- Four groups:
  1. Never takers: $D_{0i} = D_{1i} = 0$
  2. Always takers: $D_{0i} = D_{1i} = 1$
  3. Defiers: $D_{0i} = 1, D_{1i} = 0$
  4. Compliers: $D_{0i} = 0, D_{1i} = 1$
- Instrumental variable estimates the effect on compliers

What did we do last time?

- More formally
  1. Independence: $\{Y_i(D_{1i},1), Y_i(D_{0i},0), D_{1i}, D_{i0}\} \perp\!\!\!\perp Z_i$
  2. Exclusion: $Y_i(d,0) = Y_i(d,1) \equiv Y_{di}$ for $d = 0, 1$
  3. First stage: $E[D_{1i} - D_{0i}] \neq 0$
  4. Monotonicity: $D_{1i} \geq D_{0i} \forall i$ or vice versa
- Under these assumptions IV estimates:

$$E[Y_i(1) - Y_i(0)|D_{1i} > D_{i0}] = \frac{E[Y_i|Z_i=1]-E[Y_i|Z_i=0]}{E[D_i|Z_i=1]-E[D_i|Z_i=0]}$$