corinna.birner@stud-mail.uni-wuerzburg.de
max.mueller@stud-mail.uni-wuerzburg.de

Understanding Metrics based on Mastering Metrics

Corinna Birner & Max Müller

University of Würzburg

July 2, 2020

Chapter 3: Instrumental Variables

Overview

- today we will continue our journey on the path from cause to effect
- we just learned that regression is a powerful tool for estimation
- sometimes the forces of nature, especially human nature, can manipulate our treatment in a way that can help us find causal effects using the instrumental variable (IV) method
- we will therefore look at three case studies today:

Charter Schools in the USA

- public schools in the USA often have a bad reputation
- charter schools are public schools but operated by a NGO
- charter schools have more autonomy and can structure their curricula more independently
- a NGO gets a charter for a limited period and the renewal depends on good performance

do charter schools provide better public education?

- one approach to education is the No Excuses approach which uses long school days and selective teacher hiring and focuses on discipline and math and reading skills
- one emplemantic example of this approach is the Knowledge Is Power Program (KIPP)
- this program recruits teachers from the best colleges to teach in low-performing school districts
- KIPP's students: 95 % black and hispanic, 80 % poor
- nonwhite KIPP students have higher test scores than nonwhite students from nearby schools

- Is there a causal effect of the KIPP System?

Case Study: KIPP Lynn

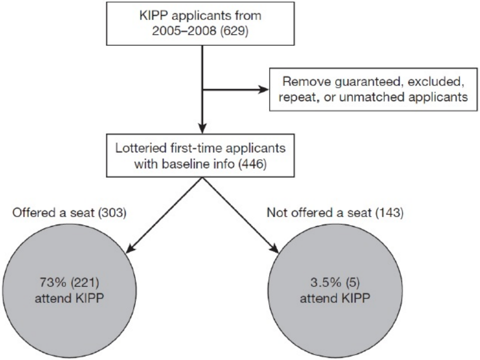
- in order to answer this question we will look at KIPP at the city of Lynn in Massachusetts
- KIPP Lynn opened in 2004 and after one year over 200 students applied for only 90 seats
- Massachusetts law requires scarce charter seats to be allocated by lottery
- the decission to go to the charter school is never entirely random
- however, we can still compare applicants who are and who are not offered a seat as a result of the random admission lottery
- we can use the IV tool on the lottery in order to frame a naturally occuring randomized trial

Case Study: KIPP Lynn

- IV turns randomized offer effects into causal estimates of the efffect of charter attendance
- IV captures effects on the sort of child who enrolls in KIPP when offered a seat but wouldn't get in otherwise
- this group is called compliers
- KIPP Study details:
  - lotteries from 2005 to 2008
  - 446 applicants
  - some applicnts were excluded (too old) or bypass the lottery (enrolled sibling)
  - outcome: standardized math/verbal score

Application and enrollment process for KIPP Lynn

Application and enrollment data from KIPP Lynn lotteries



Note: Numbers of Knowledge Is Power Program (KIPP) applicants are shown in parentheses.

- removed because of age, enrolled siblings, etc.
- only 73% attended KIPP, the others moved away or decided to attend a different school
- 3,5 % were offered a seat afterwards although they had lost in the lottery

KIPP Lynn Baseline Characteristics

		KIPP applicants			
	Lynn public fifth graders (1)	KIPP Lynn lottery winners (2)	Winners vs. losers (3)	Attended KIPP (4)	Attended KIPP vs. others (5)
Panel A. Baseline characteristics					
Hispanic	.418	.510	−.058 (.058)	.539	.012 (.054)
Black	.173	.257	.026 (.047)	.240	−.001 (.043)
Female	.480	.494	−.008 (.059)	.495	−.009 (.055)
Free/Reduced price lunch	.770	.814	−.032 (.046)	.828	.011 (.042)
Baseline (4th grade) math score	−.307	−.290	.102 (.120)	−.289	.069 (.109)
Baseline (4th grade) verbal score	−.356	−.386	.063 (.125)	−.368	.088 (.114)
Panel B. Outcomes					
Attended KIPP	.000	.787	.741 (.037)	1.000	1.000 —
Math score	−.363	−.003	.355 (.115)	.095	.467 (.103)
Verbal score	−.417	−.262	.113 (.122)	−.211	.211 (.109)
Sample size	3,964	253	371	204	371

Notes: This table describes baseline characteristics of Lynn fifth graders and reports estimated offer effects for Knowledge Is Power Program (KIPP) Lynn applicants. Means appear in columns (1), (2), and (4). Column (3) shows differences between lottery winners and losers. These are coefficients from regressions that control for risk sets, namely, dummies for year and grade of application and the presence of a sibling applicant. Column (5) shows differences between KIPP students and applicants who did not attend KIPP. Standard errors are reported in parentheses.

- only small differences between winners and losers in gender, ethnicity, economic status
- baseline scores $-.3\sigma$ below state mean
- insignificant pre-treatment differences between winners and losers (column 3)

KIPP Lynn Baseline Characteristics

	Lynn public fifth graders (1)	KIPP applicants: KIPP Lynn lottery winners (2)	Winners vs. losers (3)	Attended KIPP (4)	Attended KIPP vs. others (5)
Panel A. Baseline characteristics					
Hispanic	.418	.510	−.058 (.058)	.539	.012 (.054)
Black	.173	.257	.026 (.047)	.240	−.001 (.043)
Female	.480	.494	−.008 (.059)	.495	−.009 (.055)
Free/Reduced price lunch	.770	.814	−.032 (.046)	.828	.011 (.042)
Baseline (4th grade) math score	−.307	−.290	.102 (.120)	−.289	.069 (.109)
Baseline (4th grade) verbal score	−.356	−.386	.063 (.125)	−.368	.088 (.114)
Panel B. Outcomes					
Attended KIPP	.000	.787	.741 (.037)	1.000	1.000 —
Math score	−.363	−.003	.355 (.115)	.095	.467 (.103)
Verbal score	−.417	−.262	.113 (.122)	−.211	.211 (.109)
Sample size	3,964	253	371	204	371

Notes: This table describes baseline characteristics of Lynn fifth graders and reports estimated offer effects for Knowledge Is Power Program (KIPP) Lynn applicants. Means appear in columns (1), (2), and (4). Column (3) shows differences between lottery winners and losers. These are coefficients from regressions that control for risk sets, namely, dummies for year and grade of application and the presence of a sibling applicant. Column (5) shows differences between KIPP students and applicants who did not attend KIPP. Standard errors are reported in parentheses.

- there might be a selection bias for students enrolled and nonenrolled
- lottery winners who chose to go elsewhere might care less about school
- no significant differences in column 5
- selection bias might not be important in this context

KIPP Lynn - Outcomes

		KIPP applicants			
	Lynn public fifth graders (1)	KIPP Lynn lottery winners (2)	Winners vs. losers (3)	Attended KIPP (4)	Attended KIPP vs. others (5)
Panel A. Baseline characteristics					
Hispanic	.418	.510	−.058 (.058)	.539	.012 (.054)
Black	.173	.257	.026 (.047)	.240	−.001 (.043)
Female	.480	.494	−.008 (.059)	.495	−.009 (.055)
Free/Reduced price lunch	.770	.814	−.032 (.046)	.828	.011 (.042)
Baseline (4th grade) math score	−.307	−.290	.102 (.120)	−.289	.069 (.109)
Baseline (4th grade) verbal score	−.356	−.386	.063 (.125)	−.368	.088 (.114)
Panel B. Outcomes					
Attended KIPP	.000	.787	.741 (.037)	1.000	1.000 —
Math score	−.363	−.003	.355 (.115)	.095	.467 (.103)
Verbal score	−.417	−.262	.113 (.122)	−.211	.211 (.109)
Sample size	3,964	253	371	204	371

Notes: This table describes baseline characteristics of Lynn fifth graders and reports estimated offer effects for Knowledge Is Power Program (KIPP) Lynn applicants. Means appear in columns (1), (2), and (4). Column (3) shows differences between lottery winners and losers. These are coefficients from regressions that control for risk sets, namely, dummies for year and grade of application and the presence of a sibling applicant. Column (5) shows differences between KIPP students and applicants who did not attend KIPP. Standard errors are reported in parentheses.

- math score of applicants who where offered a seat (post-treatment) is 0
- math score of average 5th grader is -.36
- average causal effect: the offer of a seat at KIPP boosts math score by .36σ

Technical note on results in differences

- results in column 3 are not differences in average
- they come from a regression of scores on a dummy for KIPP offer, year, grade of application and presence of a sibbling
- this is needed because the probability of winning varies from year to year, between applicants from 5th and 6th grade and children with siblings

KIPP Lynn Outcomes

- what does the offer effect of $.36\sigma$ tell us about the effect of KIPP attendance?
- we need to turn the offer effect into an attendance effect by using the IV
- the instrumental variable here is a dummy indicating KIPP applicants who receive an offer
- our instrument must meet three requirements:

Instrumental Variables

- first stage: the instrument has a causal effect on the variable whose effects we're trying to capture
- independence assumption: the instrument is unrelated to the omitted variables
- exclusion restriction: the instrument affects outcomes through a single channel

The IV estimator in the KIPP Lynn

Using these three assumptions, the IV method gets us a chain reaction leading from the instrument to student achievement

- first link connects randomily assigned offers with KIPP attendance
- second link connects KIPP attendance with achievement
- with the independence assumption and the exclusion restriction the product of these two links generates the effect of offer on test scores:

$$Effect\ of\ offers\ on\ scores =$$

$$(Effect\ of\ scores\ on\ attendance\ x\ Effect\ of\ attendance\ on\ scores)$$

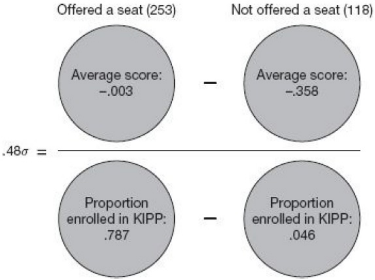
The effect of KIPP attendance is:

$$Effect\ of\ attendance\ on\ scores = \frac{Effect\ of\ offers\ on\ scores}{Effect\ of\ offers\ on\ attendance}$$

The IV estimator in the KIPP Lynn

- KIPP offers are assumed to affect test scores via KIPP attendance alone
- Offers increase attendance rates by .74%
- to get the attendance effect we multiply the effect of the ofers in scores by $\frac{1}{.74}$
- raw difference is $.355\sigma$, so the effect turns out to be $.48\sigma$

KIPP Lynn - Outcomes



- difference in math scores is $.355\sigma$
- offers in attendance is .741
- the effect of the KIPP enrollment is $.48\sigma$

The IV estimator

Our IV estimation needs three ingredients:

- the instrument Z_i: randomizes the treatment, here Dummy that indicates if KIPP seat was offered
- the treatment variable D_i (sometimes called endogenous variable), here Dummy that indicates KIPP attendance
- the outcome variable Y_i: here the 5th grade math score

The IV estimator

Our IV chain reaction has multiple components:

- the randomizer that is called instrument Z_i (here: KIPP offer)
- first stage: effect of Z_i on D_i: link from the instrument to the causal variable of interest (here: effect of offer on KIPP attendance)
- the direct effect of the instrument on outcomes Z_i on Y_i is called reduced form (here: effect of offers on scores)
- the causal effect of interest D_i on Y_i is determined by the ratio of reduced form to first-stage estimates and is called local average treatment effect (LATE)

Estimating the LATE

the first stage:

$$E[D_i|Z_i = 1] - E[D_i|Z_i = 0] \ , \ we \ call \ this \ \phi$$

difference in KIPP attendance between those offered and not offered a seat

the reduced form:

$$E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0] \ , \ we \ call \ this \ \rho$$

difference in average scores between those offered and not offered a seat

the LATE:

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

difference in scores between winners and losers divided by difference in attendance between winners and losers

LATE in the KIPP Lynn

- the LATE is the average causal effect or children whose enrollment status (KIPP vs. public) is determined solely by the KIPP lottery
- we can think of four types of children in this scenario
  - those who don't go to KIPP, even if they win a seat: never-takers
  - those who go to KIPP, even if they lose in the lottery: always-taker
  - those who go to KIPP if they win and don't go if they lose: compliers
  - those who to KIPP only if they lose: defiers
- the compliers are important for the IV because their treatment status is determined by the instrument
- we indicate this group with the dummy C_i

What does LATE tell us?

The four types of children

		Lottery losers $Z_i = 0$	
		Doesn't attend KIPP $D_i = 0$	Attends KIPP $D_i = 1$
Lottery winners $Z_i = 1$	Doesn't attend KIPP $D_i = 0$	Never-takers (*Normando*)	Defiers
	Attends KIPP $D_i = 1$	Compliers (*Camila*)	Always-takers (*Alvaro*)

- four types of children: never-takers, always-takers, defiers, compliers
- LATE theorem says that the ratio of reduced form to first stage os LATE, the average causal effect of treatment on compliers

The LATE

- LATE doesn't necesarrily describe causal effect for never-takers and always-takers
- if you want to look at average causal effects for the entire treated population, you need to look at the treatment effect on the treated (TOT): $E[Y_{1i} - Y_{0i} | D_i = 1]$
- TOT includes always-takers and compliers
- LATE and TOT are usually not the same
- you should always think about the external validity of a particulat LATE

The Minneapolis Domestic Violence Experiment (MDVE)

- this experiment tried to answer the question how police should respond to a case of domestic violence
- arresting might aggravate the problem when the agressors return home but not arresting might signal tolerance to violence
- the MDVE was conducted in the 1980s in Minneapolis
- participating police officers who where called to a scene of domestic violence had to react randomly in one of the three possible treatment ways:
  - arrest the agressor
  - order suspect off premises for 8 hours (separation
  - counselling/mediation by the officers (advice)
- Outcome examined: reoccurence of a domestic assault within 6 months
- Randomization device: randomly color-coded report form
- participation was voluntarily

MDVE

	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)

Notes: This table shows percentages and counts for the distribution of assigned and delivered treatments in the Minneapolis Domestic Violence Experiment (MDVE). The first three columns show row percentages. The last column reports column percentages. The number of cases appears in parentheses.

- in practice the officer's reaction often deviated
- treatment delivered was not random
- a case assigned to be coddled was coddled with a probability of .797
- a case assigned to be arrested was coddled with a probability of .011

The Minneapolis Domestic Violence Experiment (MDVE)

- naive analysis of the MDVE based on treatment delivered is misleading
- comparison of those who were and were not coddled is contaminated by selection bias
- we need an IV analysis in order to capture the effect of treatment on the treated!

IV in the MDVE

- First Stage: difference between the probability of being coddled when assigned to be coddled and the probability of being coddled when assigned to be arrested
- We use Z_i indicating assignment to coddling and D_i indicating incidents where coddling was delivered:

$$E[D_i|Z_i = 1] - E[D_i|Z_i = 0] = .797 - .011 = .786$$

- we now calculate the effect of random assignment to coddling on the outcome variable Y_i:

$$E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0] = .211 - .097 = .114$$

- overall recidivism rate was 18 %, so the difference of 11 % is substantial

IV in the MDVE

- in experiments where treatment assigned differs from treatment delivered, we need to calculate the effects of random assignment
- these effects are called intention-to-treat (ITT)
- an ITT analysis captures the causal effect of being assigned to treatment
- ITT effect does not take the noncompliance of those assigned to be coddled and turning out to be arrested into account
- ITT is too small!
- therefore, we have to divide the ITT effects by the difference in compliance rates between treatment and control groups in order to capture the causal effect of coddling on compliers who were coddled as a result of the experiment

IV in the MDVE

- ITT is a reduced form for a randomly assigned instrument (here assignment to being coddled)
- the IV causal chain begins with random assignment to treatment, then goes through treatment delivered and in the end affects outcomes
- the LATE from the MDVE is $\frac{.114}{.786} = .145$, showing a large coddling effect even in comparison with the corresponding ITT estimates
- exclusion restriction: requires that randomly assigned form color affect recidivism only through the decision to arrest or coddle

IV in the MDVE

- If we had ignored the nonrandom nature, we would have calculated a smaller effect:

$$E[Y_i|D_i = 1] - E[Y_i|D_i = 0] = .216 - .129 = .087$$

- as seen in Chapter 1, comparison of treated and untreated without random assignment equals the causal effect plus selection bias
- in the MDVE, batteres who were arrested although being assigned to coddled were especially violent
- batteres who were coddled were less likely to abuse again in any case
- selection bias does matter here!
- IV analysis eliminates selection bias, capturing the average causal effect on compliers

MDVE Conclusions

- always-takers would be suspected batteres who were coddled without being assigned to coddled
- In the MDVE we have practilly no always-takers
- therefore, all of the treated are compliers and LATE is TOT
- when applying this property of the MDVE, we see that LATE is the average causal effect of coddling on the coddled
- when some of those randomly assigned to treatment go untreated but no one assigned to control group gets treated, IV methods using random intention to treat as an instrument for treatment delivered captures TOT

The Population Bomb

- We want to take a look at the relationship between family size and living standards
- We might indeed expect increases in family size to be associated with increased poverty and reduced education
  - $\rightarrow$ more mouths to feed means less for each
- Gary Becker: introduced the notion of a quantity-quality tradeoff, the idea that reductions in family size increase parental investment in children, e.g. invest more in their schooling
- Is there a causal connection between family size and children's education?
- Identification Problem: inability in principle to identify a best estimate of the value(s) of one or more parameters in a regression.

Random Assignment

- Marked differences in observable characteristics across families of different sizes raise the red flag of selection bias.
- women with different numbers of children are observably different
- But: we must acknowledge the possibility of important unobserved differences associated with family size as well
- Solution: random assignment:
  1. Draw a sample of families with one child.
  2. In some of these households, randomly distribute an additional child.
  3. Wait 20 years and collect data on the educational attainment of firstborns who did and did not get an extra sibling.
- Unlikely experiment to happen, we need another way.

Differences in Family Size

- Random variation in family size by sometimes having more than one child in the form of twins
- Large sample from Israel exploits this.
- Good case study: diverse population, including many people who were born in developing countries and into large families.
- Consider a family in which the second birth is a singleton. On average, such families include 3.6 children. A second twin birth increases average family size by .32,
- The .32 twins differential in family size reflects a difference in probabilities: the likelihood of having a third child increases from about .7 with a singleton second birth to a certainty

Using of IV Estimates

- We use twin births as an instrument for family size.
- The comparison of schooling between firstborns with twin and singleton siblings constitutes the reduced form for an IV estimate.
- IV estimates are constructed from the ratio of reduced-form to first-stage estimates
- Here: The fact that the twins reduced form and associated IV estimates are close to zero weighs against the view that a larger family of origin reduces children's schooling.
- But: potentially omitted variables due to non random assignment.

Another Option

- Many parents hope for a diversified sibling-sex portfolio: Families whose first two children are both boys or both girls are more likely to have a third.
- Sex of a newborn is randomly assigned.
- Parental preferences for mixed sibling-sex composition generate an instrument.
- First-born Israeli adults who have a second-born sibling of the opposite sex grew up in households with about 3.60 children.
- But firstborns whose second-born sibling is of the same sex were raised in families with 3.68 children.
- The same-sex first stage for Israeli firstborns is about .08.

The IV Variable

Instrumental variable is a dummy variable that equals 1 for families whose first two children are both male or both female and equals 0 for families with one boy and one girl.

But: the educational attainment of first-born Israeli adults is unaffected by their siblings' sex composition.

IV has an effect on the treatment (family size) and should only effect the outcome (Education) through the treatment (family size) $\rightarrow$ *Exclusion restriction*.

But: Might the sex-mix of the first two children affect childrens educational outcomes for other reasons?

Maybe same sex childs have the benefit of using the same clothing etc. so the family is a bit richer which increases parental investment in their children's schooling.

Test exclusion restriction

- Test exclusion restriction:
- Because the fertility of always-takers and never-takers is unchanged by sibling sex composition, any relationship between sex-mix instruments and outcomes in samples with few compliers may signal violations of the underlying exclusion restriction.
- Help with using LATE: The ratio of reduced-form to first-stage parameters, that is:

$$\lambda = \frac{\rho}{\phi}$$

Recap: First Stage and Reduced Form

- When the first stage, ϕ, is zero, the reduced form should be zero as well.

$\rightarrow$ Where the effect of the instrument on the treatment variable is zero, the reduced form should be zero.

- Reduced-form estimate with no evidence of a corresponding first stage is cause for worry, because this suggests some channel other than the treatment variable (in this case, family size) links instruments with outcomes.

$\rightarrow$ Would mean that the instrument does not effect the treatment variables, but the outcome, so it would violate the exclusion restriction.

- Also the instrument should not correlate with the error term.

2SLS Regression

- Assume twins and sex-mix instruments both satisfy the required assumptions and capture similar average causal effects.
  - → Combining the two IV estimates, they increase statistical precision.
- Potential bias: We must control for confounding factors
- Solution: Two-stage least squares (2SLS) generalizes IV in two ways:
  i. 2SLS estimates use multiple instruments efficiently.
  ii. 2SLS estimates control for covariates, thereby mitigating OVB from imperfect instruments.

Using regular Regression

- Reduced Form: dummy variable for multiple second births denoted by Z_i, the reduced-form effect can be written as the coefficient ρ in the regression equation:

$$Y_i = \alpha_0 + \rho Z_i + \epsilon_{0i}$$

- First Stage: Likewise, the first-stage effect of Z_i is the coefficient ϕ in the first-stage equation:

$$D_i = \alpha_1 + \phi Z_i + \epsilon_{1i}$$

- LATE is the ratio of the slope coefficients:

$$\lambda = \frac{\rho}{\phi}$$

Using 2SLS Regression

- The 2SLS procedure offers an alternative way of computing $\lambda = \frac{\rho}{\phi}$
- In the 2SLS first stage, we estimate $D_i = \alpha_1 + \phi Z_i + \epsilon_{1i}$ and save the fitted values.
- These "first-stage fits" are defined as:

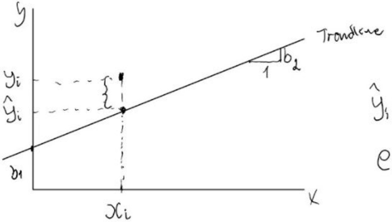
$$\hat{D}_i = \alpha_1 + \phi Z_i$$

- The 2SLS second stage regresses Yi on Di, as in:

$$Y_i = \alpha_2 + \lambda_{2SLS} \hat{D}_i + \epsilon_{2i}$$

- The value of λ_{2SLS} generated by this second step is identical to the ratio of reduced form to first-stage regression coefficients, $\frac{\rho}{\phi}$.

Remember: Fitted Values



$$\hat{y}_i = b_1 + b_2 x_i$$

$$e_i = y_i - \hat{y}_i$$

- Fitted value $\hat{y}_i$ is the prediction of the mean response value when you input the values of the predictors into the model.
- Also called predicted values.
- The difference between fitted values and real values is my residuum.

Using 2SLS Regression

- Adding maternal age, denoted A_i, the reduced form and first stage look like (influence of twins on Education and family size):

$$\begin{aligned} Reduced\ Form &: Y_i = \alpha_0 + \rho Z_i + \gamma_0 A_i + \epsilon_{0i} \\ First\ Stage &: D_i = \alpha_1 + \phi Z_i + \gamma_1 A_i + \epsilon_{1i} \end{aligned}$$

- Again, we need the first Stage fitted Values: $\hat{D}_i = \alpha_1 + \phi Z_i + \gamma_1 A_i$
- 2SLS estimates are again constructed by regressing Y_i on both $\hat{D}_i$ and A_i. Hence, the 2SLS second-stage equation is (Influence of Family Size on Education):

$$Y_i = \alpha_2 + \lambda_{2SLS} \hat{D}_i + \gamma_2 A_i + \epsilon_{2i}$$

Benefits of 2SLS

- Benefit of this model: The 2SLS setup allows as many control variables as you like, provided they appear in both the first and second stages.
- Econometrics software packages compute 2SLS estimates directly, reducing the scope for mistakes
- Another benefit: Multiple instruments useable: e.g. dummy for same-sex siblings (W_i), need to add a dummy, B_i, indicating first-born boys.
- But: 2SLS is nothing else then 2 OLS Regressions: A first and a second stage.
- Remember: OLS: statistical algorithm to minimize the sum of least squares.

Using 2SLS Regression

- With two instruments, W_i and Z_i, and the extra control variable, B_i, the 2SLS first stage becomes:

$$D_i = \alpha_1 + \phi_t Z_i + \phi_s W_i + \gamma_1 A_i + \delta_1 B_i + \epsilon_{1i}$$

- reduced form:

$$Y_i = \alpha_0 + \rho_t Z_i + \rho_s W_i + \gamma_0 A_i + \delta_0 B_i + \epsilon_{0i}$$

- Second-stage estimates with two instruments and two covariates are generated by the regression equation:

$$Y_i = \alpha_2 + \lambda_{2SLS} \hat{D}_i + \gamma_2 A_i + \delta_2 B_i + \epsilon_{2i}$$

- This produces a weighted average of the estimates we get using the instruments Z_i and W_i one at a time, while controlling for covariates A_i and B_i.

Quality-Quantity Trade Off

TABLE 3.4
Quantity-quality first stages

	Twins instruments		Same-sex instruments		Twins and same-sex instruments
	(1)	(2)	(3)	(4)	(5)
Second-born twins	.320 (.052)	.437 (.050)			.449 (.050)
Same-sex sibships			.079 (.012)	.073 (.010)	.076 (.010)
Male		−.018 (.010)		−.020 (.010)	−.020 (.010)
Controls	No	Yes	No	Yes	Yes

Notes: This table reports coefficients from a regression of the number of children on instruments and covariates. The sample size is 89,445. Standard errors are reported in parentheses.

- Second born twins increase the number of children to 0.32 more
- First-born Israeli adults whose second-born siblings were twin were raised in families with about .44 more children than those raised in families where the second birth was a singleton.

OLS and 2SLS Estimates

TABLE 3.5
OLS and 2SLS estimates of the quantity-quality trade-off

		2SLS estimates		
Dependent variable	OLS estimates (1)	Twins instruments (2)	Same-sex instruments (3)	Twins and same-sex instruments (4)
Years of schooling	−.145 (.005)	.174 (.166)	.318 (.210)	.237 (.128)
High school graduate	−.029 (.001)	.030 (.028)	.001 (.033)	.017 (.021)
Some college (for age ≥ 24)	−.023 (.001)	.017 (.052)	.078 (.054)	.048 (.037)
College graduate (for age ≥ 24)	−.015 (.001)	−.021 (.045)	.125 (.053)	.052 (.032)

Notes: This table reports OLS and 2SLS estimates of the effect of family size on schooling. OLS estimates appear in column (1). Columns (2), (3), and (4) show 2SLS estimates constructed using the instruments indicated in column headings. Sample sizes are 89,445 for rows (1) and (2); 50,561 for row (3); and 50,535 for row (4). Standard errors are reported in parentheses.

- OLS shows a strong negative relation between family size and education outcomes.
- 2SLS estimates in column (3) show uniformly positive effects of family size on education.

Summary

- The pooled second-stage estimates are not very different from those generated using the instruments one at a time, but the standard errors are appreciably smaller.
  - $\rightarrow$ Can increase precision by pooling multiple instruments
- Findings here suggest that the strong negative association between family size and schooling is driven in large part and perhaps entirely by selection bias.

In a nutshell

- Foundation has three layers:
  i the first-stage requires instruments that affect the causal channel of interest;
  ii the independence assumption requires instruments to be as good as randomly assigned;
  iii the exclusion restriction asserts that a single causal channel connects instruments with outcomes.
- Check those assumptions:
  i First stage: By looking for a strong relationship between instruments and the proposed causal channel;
  ii Independence assumption: By checking covariate balance with the instrument switched off and on, as in a randomized trial.

Appendix

- IV setup with one instrument and no covariates.
- The first stage links instrument and treatment:

$$D_i = \alpha_1 + \phi Z_i + \epsilon_{1i}$$

- The reduced form links instrument and outcomes:

$$Y_i = \alpha_0 + \rho Z_i + \gamma_0 A_i + \epsilon_{0i}$$

- The 2SLS second stage is the regression of outcomes on first-stage fitted values:

$$Y_i = \alpha_2 + \lambda_{2SLS} \hat{D}_i + \gamma_2 A_i + \epsilon_{2i}$$

- With ratio of covariances ρ *and* ϕ, λ is called the IV formula:

$$\lambda = \frac{\rho}{\phi}$$

- If you have Covariates, e.g. A_i, include them in the model, no change to the IV formula (or 2sls Formula)

2SLS Standard Errors

- First the 2SLS residual is constructed using:

$$\eta_i = Y_i - \alpha_2 - \lambda_{2SLS} D_i - \gamma_2 A_i.$$

- $\sigma\eta$ is the standard deviation of η_i
- The standard error for lambda is then given by:

$$SE(\hat{\lambda}_{2SLS}) = \frac{\sigma_\eta}{\sqrt{n}} \times \frac{1}{\sigma_{\hat{D}}}$$

- 2SLS combines multiple instruments in an effort to generate precise estimates of a single causal effect.
- Problem: 2SLS estimates with many weak instruments can be misleading.
- How do I know that an instrument is weak: one that isn't highly correlated with the regressor being instrumented, so the first-stage coefficient associated with this instrument is small or imprecisely estimated. Then 2SLS is close to OLS.

2SLS Bias

- Finite sample bias: 2SLS estimates in a many weak IV scenario tell you little about the causal relationship of interest.
- Solution: An alternative to 2SLS, called the limited information maximum likelihood estimator (LIML for short) is less affected by finite sample bias. You'd like LIML estimates and 2SLS estimates to be close to one another, since the former are unlikely to be biased even with many weak instruments.
- But: finite sample bias does not occur, when you only use one instrument.
- Reduced-form estimates that are small and not significantly different from zero provide a strong and unbiased hint that the causal relationship of interest is weak or nonexistent.

Thank you very much for listening