Instrumental Variables

Who Invented Instrumental Variables?

Philip G. Wright

Sewall Wright

Example – The Power of Referrals

Imagine you run a subscription-based sports journalism app

You want to know:

- "Does getting users to refer more friends cause users to stick around and read more articles?"

"Ikea" effect:

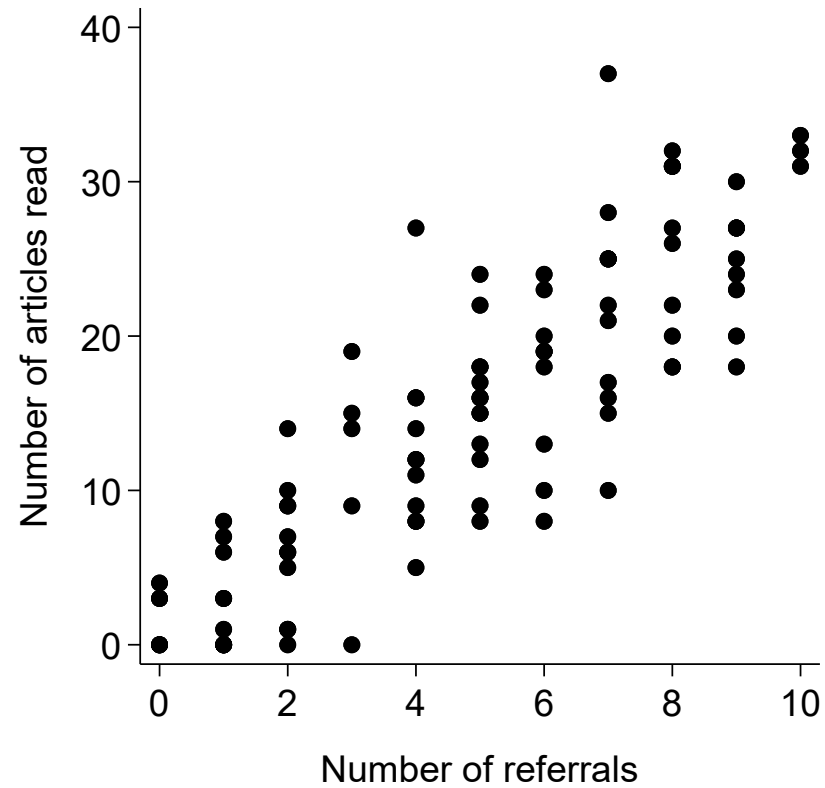- People care more about products that they have invested time contributing to (Norton et al. 2012)

Image Credit: Alain Cohn

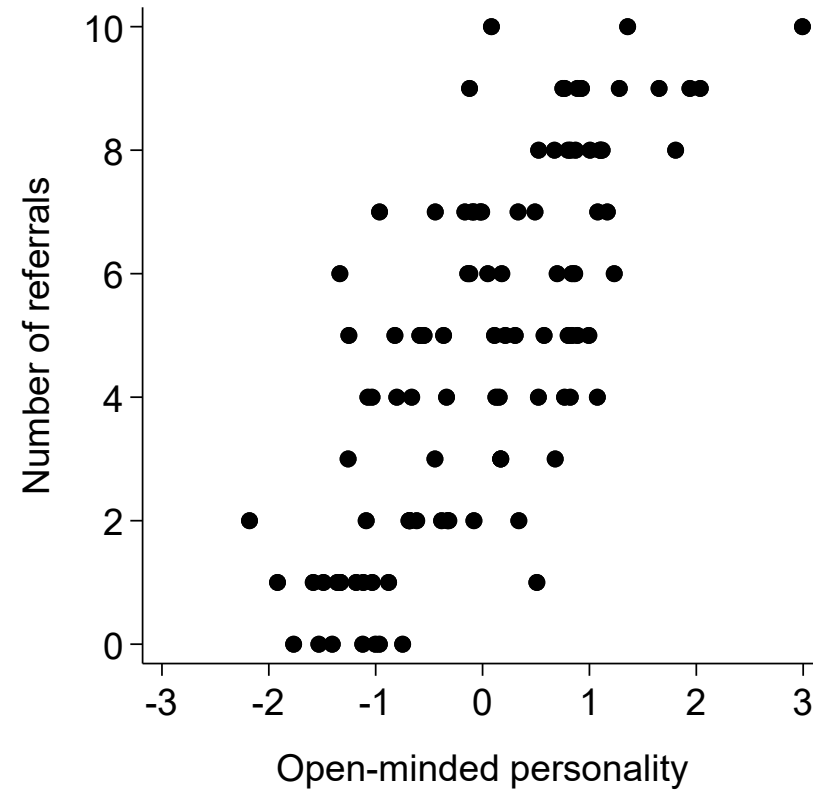# Example – The Power of Referrals

You find that users who refer more friends read more articles:
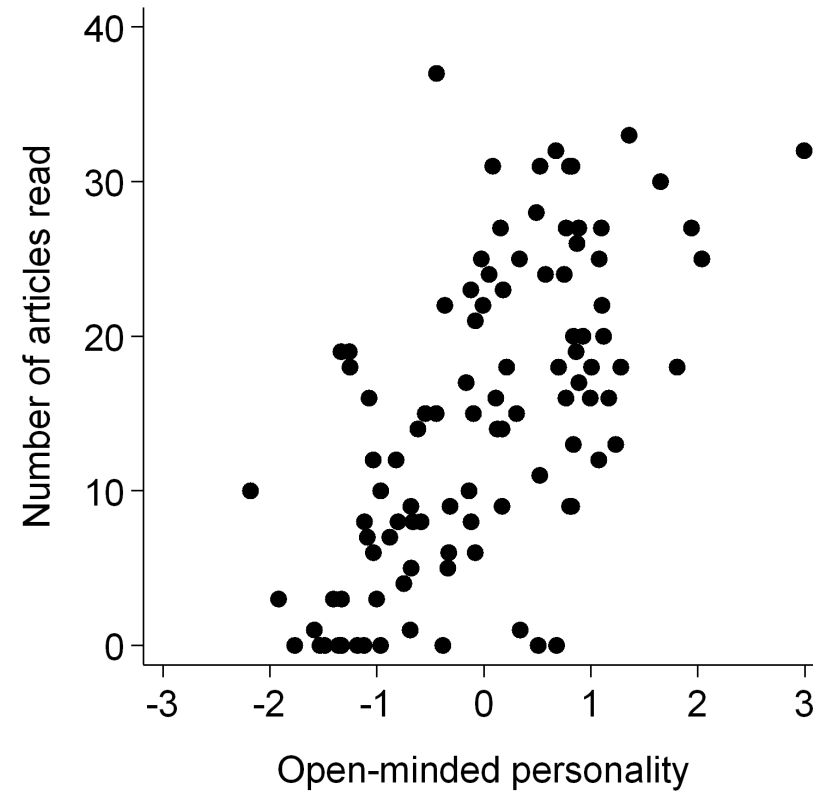
Example – The Power of Referrals

Users who refer more friends are perhaps more open-minded → selection on unobservables

Example – The Power of Referrals

Being open-minded may affect app use independent of referrals:

Basic IV Setup

- Problem: treatment D is affected by unmeasured confounders U

- Identify (or create) an *exogenous* variable Z that affects the treatment D but not the outcome Y

- Z is the called the **instrumental variable** (or "instrument")

- Use the unconfounded variation in the instrument Z to identify treatment effects

Basic IV Setup
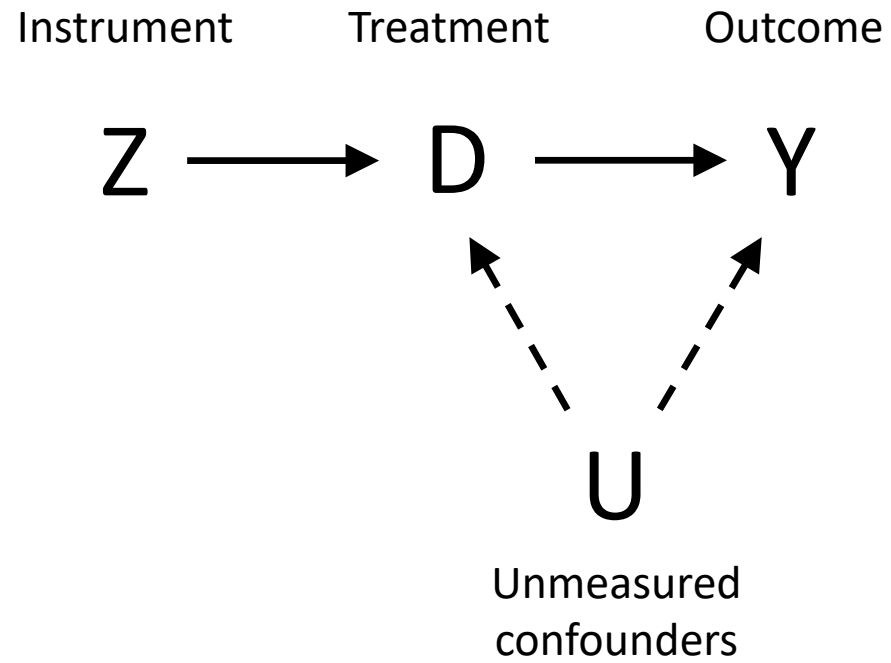
Instrument   Treatment   Outcome

$$Z \longrightarrow D \longrightarrow Y$$

$$U$$

Unmeasured
confounders

IV Assumptions

Requirements for the instrument Z:

1. Needs to cause changes in the treatment

First stage

$$Z \longrightarrow D \longrightarrow Y$$

$$D \longleftarrow U \longrightarrow Y$$
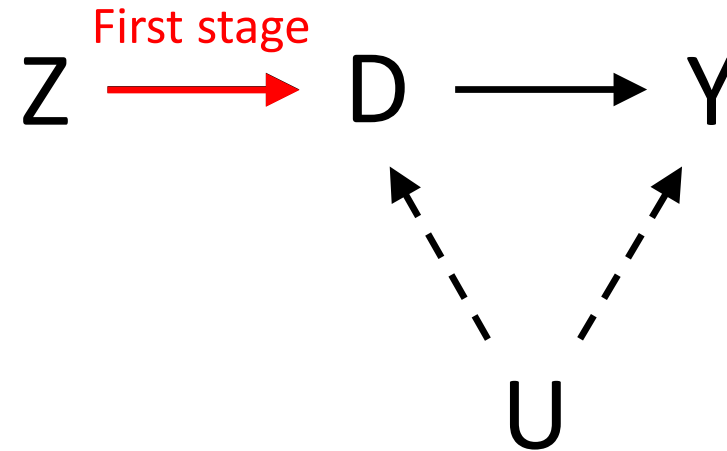
IV Assumptions

Requirements for the instrument Z:

1. Needs to cause changes in the treatment
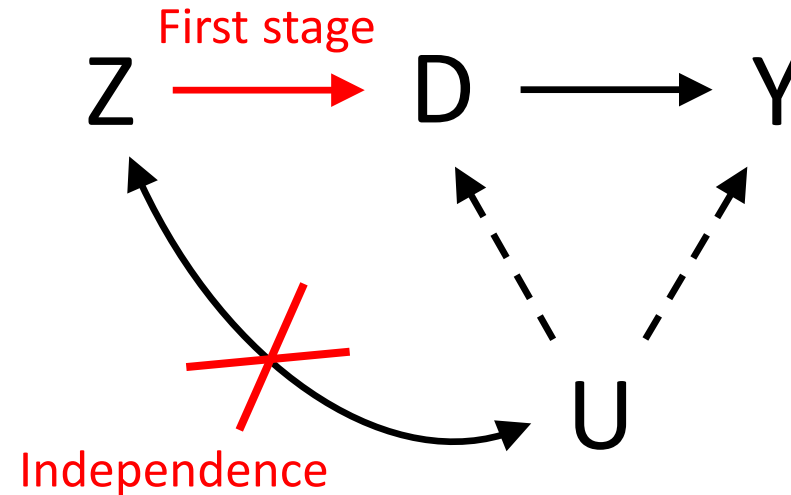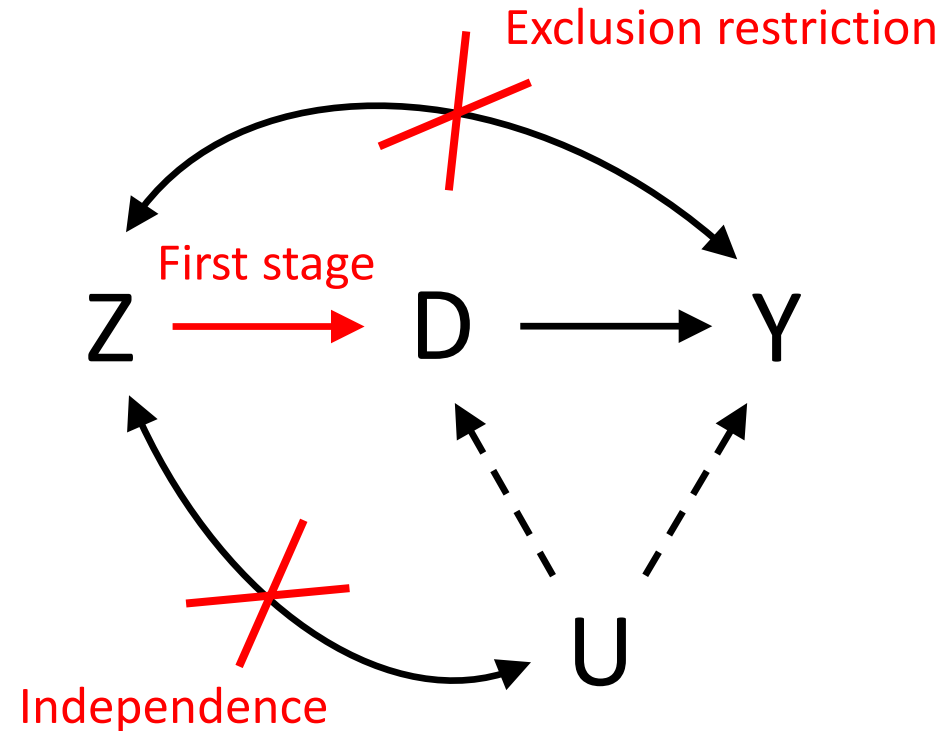
2. "As good as randomly" assigned; cannot influence confounders or vice versa

IV Assumptions

Requirements for the instrument Z:

1. Needs to cause changes in the treatment

2. "As good as randomly"; cannot influence confounders or vice versa

3. Cannot directly influence the outcome

IV Assumptions

1. **First stage**: Z_i has a causal effect on D_i
   - Rule of thumb: *F*-statistic of a joint test whether all instruments are significantly different from zero > 10
   - In case of a single instrument: *t*-statistic $> \sqrt{10}$

2. **Independence**: Z_i is randomly assigned or "as good as randomly assigned"
   - Z is uncorrelated with any possible confounder

3. **Exclusion restriction:** Z_i does *not* have a direct causal effect on Y_i
   - Any effect of Z on Y must go through D
   - Technically <u>not</u> testable

Finding Good Instruments Can Be Difficult

Arbitrary rules created by policies:

- Angrist and Lavy (1999): Maimonides' rule as an IV for class size (test scores as outcome)
- Levitt (1997): Being an election year as an IV for police force size (crime as outcome)

Quasi-random variation in the world:

- Madestam et al (2013): Rainfall as an IV for political rally attendance (voting as outcome)
- Dinkelman (2011): Terrain as an IV for electrification in rural areas (female employment as outcome)

Randomized encouragement designs

Randomized Encouragement Designs

Maybe we can re-use an AB test that encouraged users to refer friends

Suppose you previously ran an experiment where some users were offered a $5 voucher for referring a friend

This experiment serves as an instrument to estimate the causal effect of referrals on app use

- Users refer more friends with $5 referral offer → first stage ✔️
- $5 referral offer was randomly assigned → independence ✔️
- $5 referral offer has no direct effect on app use → exclusion restriction ❓

First Stage

Users who are offered a $5 referral voucher refer more friends ($t = 3.42$):
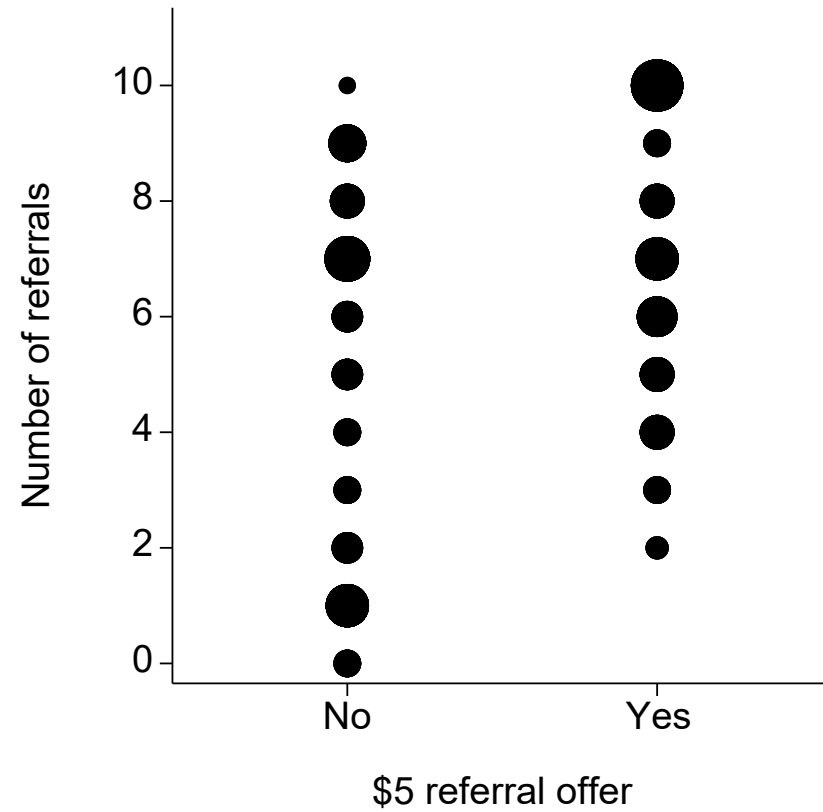
Independence

No correlation between instrument and "unobserved" confounder:
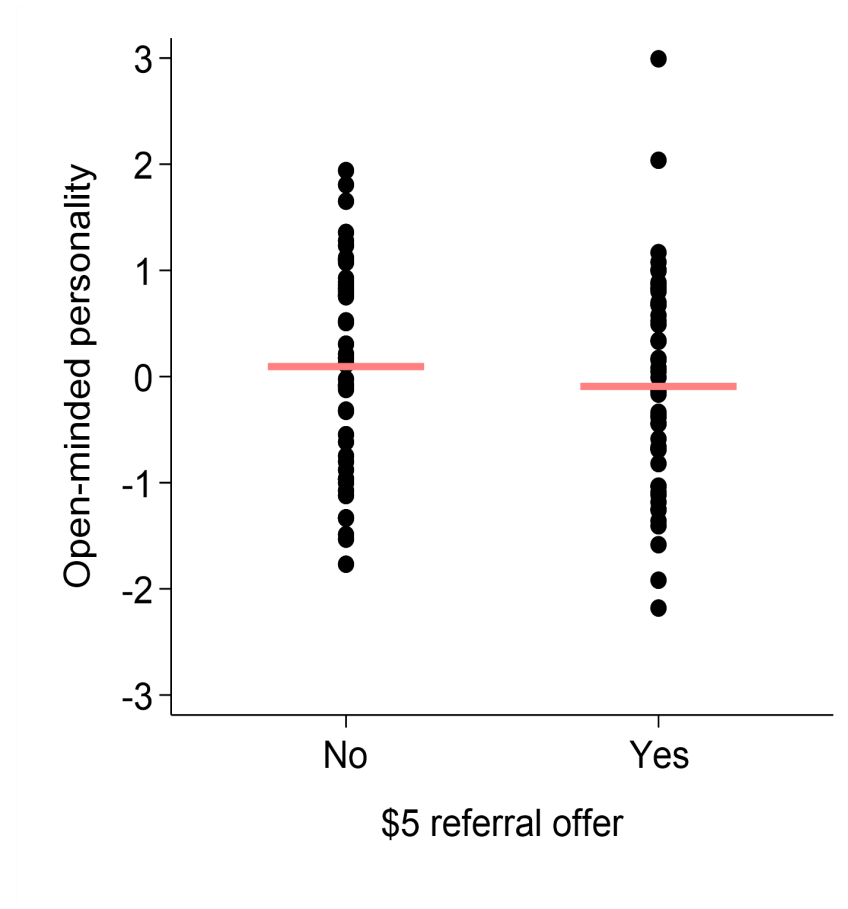
Intention-to-Treat Effect

Being assigned to $5 referral offer increases number of articles read:
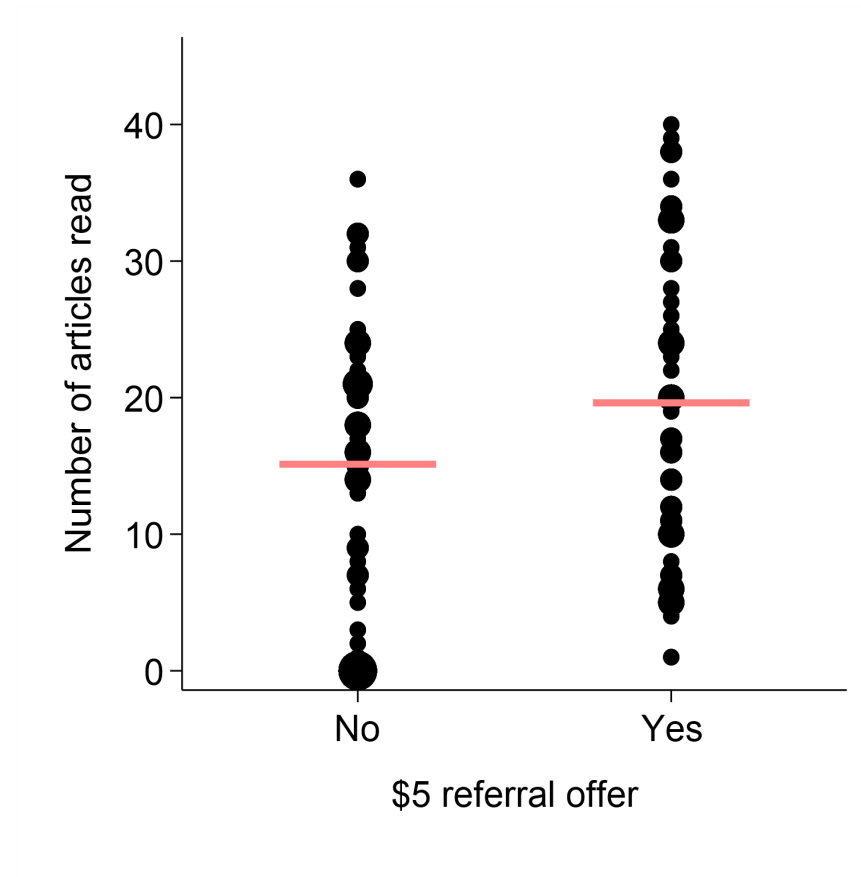
Intention-to-Treat Effect

Intention-to-treat effect (ITT) is the average causal effect of an *offer* of the treatment (i.e. instrument)

- ITT is **not** the causal effect of the treatment on the outcome (which is want we want to know)

- ITT does **not** account for "noncompliance" with treatment assignment (*imperfect compliance*)

How can we estimate the causal effect of referring a friend on app use?

Credits:
Alain Cohn
Assistant Professor of Information

Instrumental Variables with Constant Effects

Basic Idea of IV
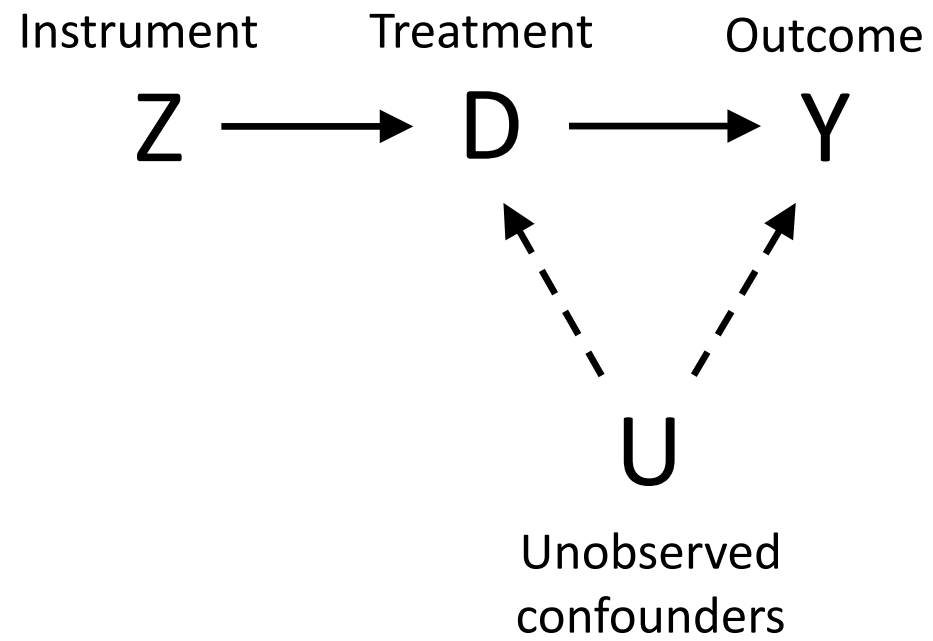
IV Chain Reaction

- **A** is the effect of instrument on treatment (first stage)

- **B** is the causal effect of interest

- **C** is the effect of instrument on outcome via treatment (reduced form)

$$\text{Instrument} \xrightarrow{\ \ A\ \ } \text{Treatment} \xrightarrow{\ \ B\ \ } \text{Outcome}$$

with **C** spanning from Instrument to Outcome.
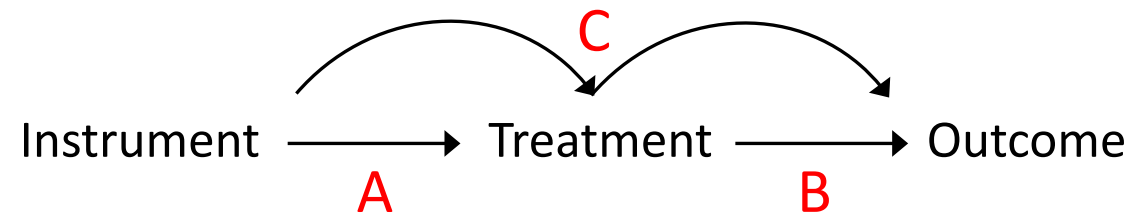
- While we cannot directly estimate **B**, we can recover **B** via estimating **A** and **C** as the instrument provides *exogenous* variation in the treatment:

$$C = A \times B \rightarrow B = \frac{C}{A}$$

IV with Constant Effects

Long regression with constant effects:

$$Y_i = \alpha + \textcolor{red}{\tau} D_i + \gamma U_i + v_i$$

- We assume that $E[D_i v_i] = 0$, so if we meausred U_i, then we would be able to estimate τ

- But $Cov[\underbrace{\gamma U_i + v_i}_{\substack{\text{Error in} \\ \text{short regression}}}, D_i] \neq 0$ because U_i is a common cause of D_i and Y_i

IV with Constant Effects

If we have an instrument Z_i that satisfies *independence* and the *exclusion restriction,* then:

$$Cov(\gamma U_i + v_i, Z_i) = 0$$

- Z_i must be independent of U_i and it has no correlation with v_i (because Z_i only affects Y_i through D_i)

Now we can identify τ:

$$Cov(Y_i, Z_i) = Cov(\alpha + \tau D_i + \gamma U_i + v_i, Z_i)$$
$$= Cov(\alpha, Z_i) + Cov(\tau D_i, Z_i) + Cov(\gamma U_i + v_i, Z_i)$$
$$= 0 + \tau Cov(D_i, Z_i) + 0$$

IV Estimator

The IV estimator is the sample analog of:

$$\tau = \frac{Cov(Y_i, Z_i)}{Cov(D_i, Z_i)} = \underbrace{\frac{Cov(Y_i, Z_i)/V(Z_i)}{Cov(D_i, Z_i)/V(Z_i)}}_{\text{Regression coefficients}}$$

Reduced form coefficient: $Cov(Y_i, Z_i)/V(Z_i)$

First stage coefficient: $Cov(D_i, Z_i)/V(Z_i)$

IV Estimator with Binary Instrument

With a binary instrument, the IV estimator is the sample analog of:

$$\tau = \frac{Cov(Y_i, Z_i)}{Cov(D_i, Z_i)} = \frac{E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]}{E[D_i|Z_i = 1] - E[D_i|Z_i = 0]}$$

Intuitively:

$$\tau = \frac{\text{Effect of instrument on outcome}}{\text{Effect of instrument on treatment}}$$

KIPP

- Knowledge is Power Program (KIPP) is America's largest network of public charter schools

- KIPP schools target low income and minority students

- Can KIPP schools reduce racial achievement gaps?

- Seats at KIPP are scarce → admissions lotteries

- Admissions lotteries as an IV for KIPP school attendance (math scores as outcome)

KIPP

Decision to attend (D_i) is not entirely random:

- Some students are offered a seat but nonetheless choose to go elsewhere
- Others who lost the lottery still find a way in (e.g. because of their siblings)

But being offered a seat (Z_i) is random

Balance Check of Pre-Treatment Variables

	Lynn public fifth graders (1)	KIPP Lynn lottery winners (2)	KIPP applicants		
			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)

Winners and losers have similar characteristics before they apply to KIPP

KIPP

Decision to attend (D_i) is not entirely random:

- Some students are offered a seat but nonetheless choose to go elsewhere
- Others who lost the lottery still find a way in (e.g. because of their siblings)

But being offered a seat (Z_i) is random

- Independence assumption ✅

Winning the lottery affects test scores *only through* a higher probability of enrollment

- Exclusion restriction ✅

First Stage

	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 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

Winners are 74 pp. more likely to attend KIPP

KIPP

Decision to attend (D_i) is not entirely random:
- Some students are offered a seat but nonetheless choose to go elsewhere
- Others who lost the lottery still find a way in (e.g. because of their siblings)

But being offered a seat (Z_i) is random
- Independence assumption ✅

Winning the lottery affects test scores only through a higher probability of enrollment
- Exclusion restriction ✅

Winning the lottery significantly increases the probability of enrollment
- Strong first stage ✅

The Causal Effect of KIPP Attendance
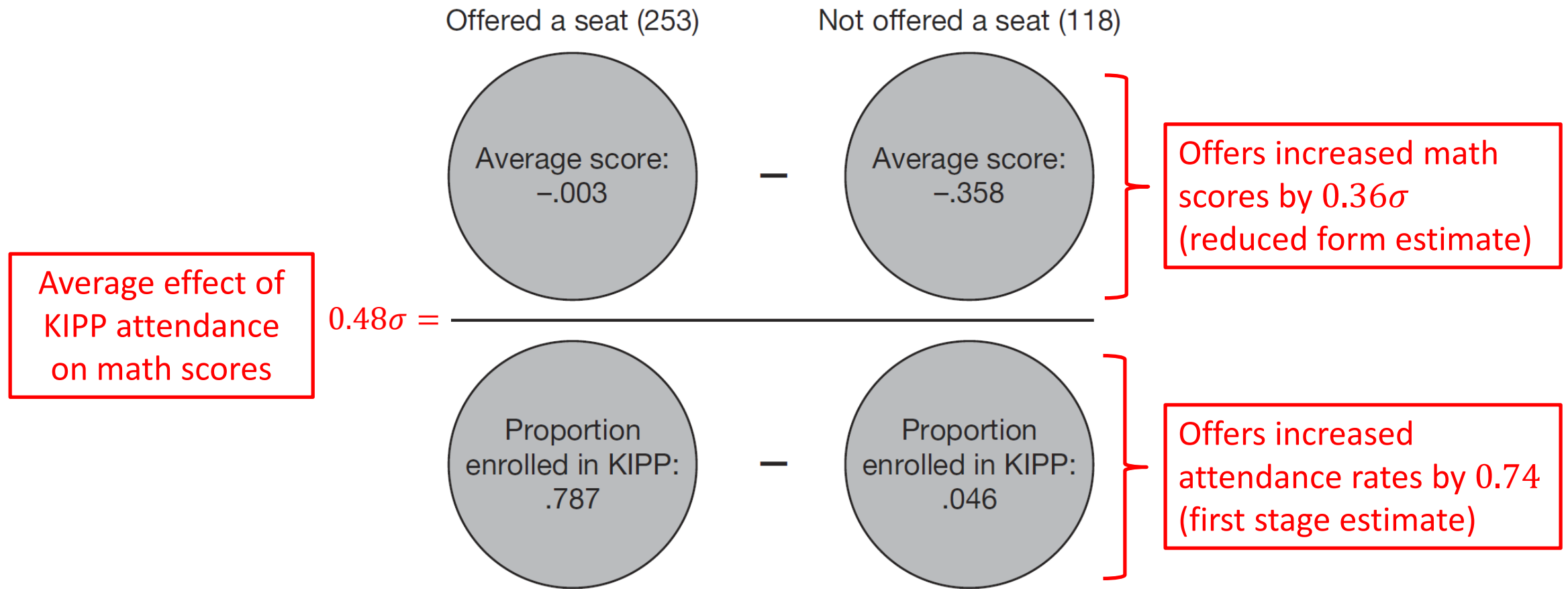
Offered a seat (253)

Not offered a seat (118)

Average score: −.003

Average score: −.358

Offers increased math scores by 0.36σ (reduced form estimate)

$0.48\sigma =$

Average effect of KIPP attendance on math scores

Proportion enrolled in KIPP: .787

Proportion enrolled in KIPP: .046

Offers increased attendance rates by 0.74 (first stage estimate)

Credits:
Alain Cohn
Assistant Professor of Information

IV with Heterogeneous Effects

We allow for each unit to have a unique response to the treatment:

$$Y_i^1 - Y_i^0 = \tau_i$$

With heterogeneous effects, there is a tension between *internal validity* and *external validity*

IV with heterogeneous effects is built on the potential outcomes framework:

$$\underbrace{D_i = Z_i D_i^1 + (1 - Z_i) D_i^0}_{\text{"Switching equation"}} = D_i^0 + \underbrace{(D_i^1 - D_i^0)}_{\text{Causal effect of } Z_i \text{ on } D_i} Z_i$$

IV and Potential Outcomes

With a binary treatment and a binary instrument, there are four different types:

Type	$D_i(Z_i = 1)$	$D_i(Z_i = 0)$
Always Takers	1	1
Never Takers	0	0
Compliers	1	0
Defiers	0	1

IV Assumptions with Heterogeneous Effects

1. **First stage**: Z_i has a causal effect on D_i

2. **Independence**: Z_i is randomly assigned or "as good as randomly assigned"

3. **Exclusion restriction**: Z_i does *not* have a direct causal effect on Y_i

4. **No defiers**: being assigned to the treatment never discourages someone from taking the treatment

$$D_i(Z_i = 1) - D_i(Z_i = 0) \geq 0$$

"No Defiers" Assumption

The no defiers assumption gives us a lot of information:

Type	$D_i(Z_i = 1)$	$D_i(Z_i = 0)$
Always Takers	1	1
Never Takers	0	0
Compliers	1	0
Defiers	0	1

- Anyone with $D_i = 1$ if $Z_i = 0$ must be an always-taker
- Anyone with $D_i = 0$ if $Z_i = 1$ must be a never-taker
- Since Z_i is randomly assigned, we know the proportion of each type in the population

Local Average Treatment Effect

Under the four assumptions, IV estimates the **local average treatment effect (LATE)**:

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

LATE is the ATE among the **compliers**:

- Those that take the treatment when encouraged to do so

How Useful is LATE?

LATE is the causal effect among compliers, who are a subset of the population

With constant treatment effects, compliers have the same treatment effect as non-compliers → LATE = ATE

If we allow for heterogeneous effects → LATE ≠ ATE

We don't know "who" the compliers are because they are defined by two potential outcomes and we only observe one

Better LATE Than Nothing

Compliers are often a subpopulation we want to learn about

Experiments with one-sided non-compliance (e.g. medical trials, job training programs)
→ LATE = ATT

- Only those assigned to treatment can actually take the treatment
- No always-takers (and no defiers by assumption)
- Treated individuals must be compliers

We can calculate the average of any characteristic in the complier group and compare it to the overall population (Abadie 2003)

IV Estimates Tend to be Imprecise

IV splits the variation in treatment into an *exogenous* part and an *endogenous* part

- Some people are treated because of the instrument (compliers)
- Others get treated for reasons unrelated to the instrument (always-takers)

IV estimates are based on the complier group, which is a subset of the data

It's as if we have less data to identify the causal effect → IV estimates tend to produce large standard errors

Minneapolis Domestic Violence Experiment

How should police respond to domestic violence?

MDVE randomly assigned police responses to domestic violence:

- Arrest

- Separation
- Advise

} "Coddled"

Outcome: Binary variable for recividism (i.e. another case of domestic violence within 6 months)

Minneapolis Domestic Violence Experiment

Although participating officers agreed to comply with the randomization protocol …

- Strict adherence to randomization was both unrealistic and inappropriate
- Experiment with imperfect compliance: assigned treatment ≠ delivered treatment

Assigned treatment	Delivered treatment			
		Coddled		
	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)

Coddled = Advise or Separate

Minneapolis Domestic Violence Experiment

- Some batterers assigned to coddling were nonetheless arrested because they were particularly aggressive → selection bias

- Treatment *assignment* is random → intention-to-treat (ITT) effect, instrument for LATE analysis

- ITT effect (reduced form):

$$E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0] = 0.211 - 0.097 = 0.114$$

- Overall recidivism rate: 18%

Local Average Treatment Effect

Average treatment effect among compliers (LATE):

$$\tau_{LATE} = \frac{E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]}{E[D_i|Z_i = 1] - E[D_i|Z_i = 0]} \quad \text{(reduced form estimate)}$$

$$\text{(first stage effect)}$$

$$= \frac{0.114}{0.786} = 0.145 > 0.114 \text{ (ITT)}$$

If we simply compared treated and untreated individuals, we would underestimate the effect of coddling on recidivism:

$$E[Y_i|D_i = 1] - E[Y_i|D_i = 0] = 0.216 - 0.129 = 0.087$$

Average Treatment Effect on the Treated

Always-takers are suspects who were coddled regardless of treatment assignment:

Assigned treatment	Delivered treatment			
		Coddled		
	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)

- One-sided non-compliance: Almost no always-takers → LATE ≈ ATT

Credits:
Alain Cohn
Assistant Professor of Information

Two-Stage Least Squares

Two-Stage Least Squares

Two-stage least squares (2SLS) is a flexible IV method to identify LATE:

- Multiple instruments, continuous instruments
- Control variables

Procedure is basically a sequence of two regressions:

1. Run regression of treatment on covariates and instrument(s)
   - Compute fitted values of treatment ("first-stage fits")
   - Fitted values capture the *exogeneous* part of the treatment explained by the instrument

2. Run regression of outcome on same covariates and fitted values

Two-Stage Least Squares

Step 1: Estimate first-stage effect (with controls) …

$$D_i = \alpha_1 + \phi Z_i + \beta_1 X_i + \epsilon_{1i}$$

… to construct fittes values

$$\widehat{D}_i = \alpha_1 + \phi Z_i + \beta_1 X_i$$

Step 2: Run regression of outcome on first-stage fits (and controls)

$$Y_i = \alpha_2 + \tau_{2SLS}\widehat{D}_i + \beta_2 X_i + \epsilon_{2i}$$

Calculating Standard Errors

Do not "manually" compute IV estimates this way because the standard errors will be wrong

- Standard errors have to be corrected to account for the two-stage design

- There is variability in both stages, not just the second (which makes the standard errors generally larger)

In practice, you should use built-in software packages to compute 2SLS estimates

Family Size and Children's Education

How many children should people have?

- Higher birth rates in developing countries may contribute to poverty trap

- Larger family size may reduce children's education

Causal question: **Does larger family size decrease parental investment in children?**

Parents with many children differ in many ways from those with fewer children → selection bias

Twins "Experiment"

Women sometimes give birth to twins → twin births as an instrument for family size

Idea: compare educational attainment of firstborns with singleton vs. twin siblings (1+1 vs. 1+2)

IV assumptions:
- Independence → twin births are "as good as random"
- First stage → second twin births increase family size
- Exclusion restriction → second twin births affect firstborns' education only through a change in family size

Twins "Experiment"

$$\tau_{LATE} = \frac{\text{Effect of instrument on outcome (reduced form)}}{\text{Effect of instrument on treatment (first stage)}}$$

- First stage: family size increases by 0.3 children with a second twin birth

- Reduced form: firstborns with second twin births are no less educated than those with singleton second births → zero effect

- If reduced form estimate is 0, IV estimate will be 0 too

Are twin births "as good as random"? Maybe not

Sibling Sex Composition

Many parents want to have boys and girls → sex mix of the first two children as an instrument for family size

Idea: compare educational attainment of firstborns with same-sex vs. opposite-sex siblings

IV assumptions:

- Independence → sex is "as good as random"
- First stage → same-sex second births increase family size
- Exclusion restriction → sex mix affects firstborns' education only through a change in family size

Sibling Sex Composition

$$\tau_{LATE} = \frac{\text{Effect of instrument on outcome (reduced form)}}{\text{Effect of instrument on treatment (first stage)}}$$

- First stage: family size increases by 0.08 children with same-sex siblings

- Reduced form: firstborns with same-sex siblings are no less educated than those with mixed-sex siblings → zero effect

- Again: if reduced form estimate is 0, IV estimate will be 0 too

- Is the exclusion restriction satisfied?

Falsification Tests
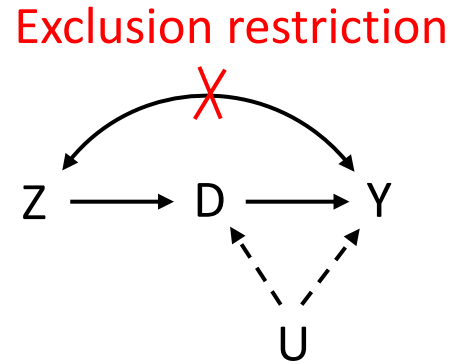
The exclusion restriction cannot be tested directly, but it can be falsified

- Test the reduced form effect of Z_i on Y_i in situations where it is extremely unlikely (or impossible) that Z_i could affect D_i

- Because Z_i cannot affect D_i, the exclusion restriction implies that Z_i has no impact on Y_i

Falsification Tests

Exclusion restriction
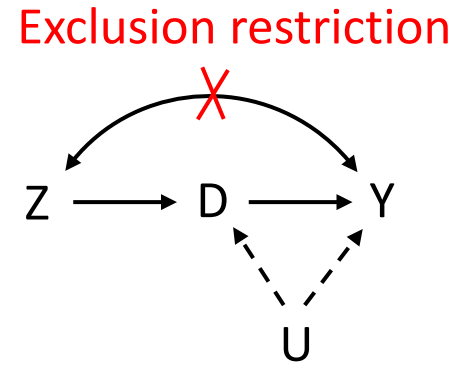
$$\text{LATE} = \frac{\text{reduced form}}{\text{first stage}} \rightarrow \text{reduced form} = \text{first stage} \times \text{LATE}$$

If first stage $= 0$, then reduced form $= 0$

Falsification Tests

Identify families for which the sibling-sex composition is unlikely to affect family size

- Religious women often want 3 or more children (always-takers)
- Highly educated women often prefer to have fewer children (never-takers)

These families have a 0 first-stage effect

Thus, these families should also have a 0 reduced-form effect (which they do)

Family Size and Children's Education

First-stage effects with one and two instruments:

	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

Family Size and Children's Education

OLS and 2SLS estimates:

		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)

IV Tables

	Dependent variable	
	OLS (1)	2SLS (2)
Treatment	…	…
X_1	…	…
X_2	…	…
X_n	…	…
First stage		
Instrument		…
F-statistic for IV in first stage		…
R^2	…	
N	…	…

Recap of IV

- Instrumental variables address selection on unobservables

- Under heterogeneous treatment effects, instrumental variables only identifies LATE (effect among compliers)

- IV has 4 core assumptions

- IV estimates have typically large standard errors

SCHOOL OF INFORMATION

UNIVERSITY OF MICHIGAN

Credits:
Alain Cohn
Assistant Professor of Information