2. Instrumental Variables

PhD Applied Methods

Duncan Webb
NovaSBE

Spring 2026

Housekeeping

Exercise B (Power calculations)

- Will only be due in the **last week** — I'll cover power in the last lecture
- Feel free to submit earlier if you want

Grading

- I aim to send you a personal report + grade by Friday, along with solutions
- Please don't share the solutions with future students. I trust you!

PSET 1 — Tips

Overall you did **very well**! A few tips going forward:

- **Interpretation matters**: Don't just do the derivation — add a sentence explaining what the result *means*
- **Independence assumptions**: If $Z \perp Y$, this means:
  - $\mathbb{E}[Y|Z] = \mathbb{E}[Y]$
  - $\mathbb{E}[Y|Z > c] = \mathbb{E}[Y]$
  - Use this to *simplify* your expressions, not just state them
- **Collider bias**: Some confusion here. Remember: conditioning on a common *effect* creates a spurious association between its causes
- **Two OVBs at once**: When two omitted variables create biases in opposite directions, cancellation is a **knife-edge** result — you can't verify it empirically because you don't observe the omitted variables

Why add control variables?

Two main reasons to add control variables in a regression:

1. For identification of the causal effect

- Control for confounders
- Make the conditional independence assumption more plausible

2. To reduce variance of the error term

- Increase precision of estimates
- Reduce standard errors
- Increase statistical power

With a large set of potential controls, **which should we choose?**

Identification through DAGs

- **Directed Acyclic Graphs** (DAGs): graphical representation of causal relationships

- We can encode the relationship between D and Y using an arrow

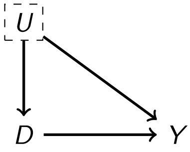
- Direction emphasizes that D causes Y, not vice versa

- More intuitive than equations for complex settings

$$D \longrightarrow Y$$

Confounders

- U affects both treatment D and outcome Y
- U is a **confounder**
- Creates association between D and Y that is *not* the causal effect
- Why? D and Y correlated through their common cause U



Example: Education (D) and wages (Y)
Ability (U) affects both choices
High ability $\implies$ more education *and* higher wages
$\implies$ Observe D and Y correlated even without causal effect

Controlling for confounders

- If X is observable confounder, can control for it
- Compare $D = 1$ vs. $D = 0$ *within* values of X
- Removes spurious association through X
- Recovers causal effect of D on Y

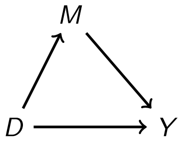
$$\begin{tikzcd} X \arrow[d] \arrow[dr] & \\ D \arrow[r] & Y \end{tikzcd}$$

Conditioning on X makes D and Y independent (conditional on X)

Example continued:
Control for ability (or test scores as proxy)
Compare high-ability people with/without education
Compare low-ability people with/without education
$\implies$ Removes confounding by ability

Bad controls: Post-treatment variables

- M is caused by treatment D
- Also called “mediator”
- Controlling for M blocks part of causal effect
- Shuts down mechanism: $D \rightarrow M \rightarrow Y$



Example: Job training $\rightarrow$ skills $\rightarrow$ earnings
Controlling for skills misses indirect effect

Bad controls: Colliders

- C is caused by *both* D and Y
- Called a "collider"
- D and Y are independent
- But if you condition on C, creates spurious correlation

$$\begin{tikzcd} & C & \\ D \arrow[ur] \arrow[rr] & & Y \arrow[ul] \end{tikzcd}$$

Intuition:
Knowing C gives information about D and Y
If D high but C low, can infer Y must be low
$\implies$ Conditioning on C makes D and Y correlated
Even though no causal relationship!

Collider example: Beauty and talent

Setting: Study beauty (D) and talent (Y) among actors

- Beauty increases probability of becoming actor
- Talent increases probability of becoming actor
- Sample: only people who became actors

Problem: "Becoming an actor" is a collider

- Both beauty and talent cause it
- By conditioning on actors, we create spurious correlation

Result: Among actors, beauty and talent negatively correlated!

- Very beautiful actors succeed with low talent
- Less beautiful actors need high talent to succeed
- This correlation is *not* causal - it's an artifact of sample selection

Summary: Which controls?

Type	Control?	Why?
Confounder	Yes	Removes confounding
Post-treatment	No	Blocks causal mechanism
Collider	No	Creates spurious correlation

Key takeaway: Not all controls are good! Need to think carefully about the causal structure

Controls for precision

Even in RCTs (where identification is clean), controls increase precision

Recall: $V(\hat{\beta}) = \sigma^2(X'X)^{-1}$ where σ^2 is residual variance

If X_i predicts Y_i, then including controls reduces σ^2

$\implies$ Lower standard errors, higher power

TABLE V

OLS AND REDUCED-FORM ESTIMATES OF EFFECT OF CLASS-SIZE ASSIGNMENT ON AVERAGE PERCENTILE OF STANFORD ACHIEVEMENT TEST

Explanatory variable	Reduced form: initial class size (5)	(6)	(7)	(8)
Small class	4.82	5.37	5.36	5.37
	(2.19)	(1.25)	(1.21)	(1.19)
Regular/aide class	.12	.29	.53	.31
	(2.23)	(1.13)	(1.09)	(1.07)
White/Asian (1 = yes	—	—	8.35	8.44
			(1.35)	(1.36)
Girl (1 = yes)	—	—	4.48	4.39
			(.63)	(.63)
Free lunch (1 = yes)	—	—	−13.15	−13.07
			(.77)	(.77)
White teacher	—	—	—	−.57
				(2.10)
Teacher experience	—	—	—	.26
				(.10)
Master's degree	—	—	—	−.51
				(1.06)
School fixed effects	No	Yes	Yes	Yes
R^2	.01	.25	.31	.31

STAR experiment: Controls reduce SEs, point estimates similar

Which controls for precision?

Ideal controls:

- Strongly correlated with outcome Y_i
- Pre-treatment (measured before randomization)
- Not affected by treatment

Common choices:

- Baseline outcome (e.g., pre-treatment earnings)
- Demographics (age, gender, education)
- Stratification variables

With many potential controls, how to choose?

Control selection with LASSO (for RCTs)

Post-Double-Selection LASSO (Belloni et al., 2014)

Procedure:

1. Run LASSO of Y_i on all controls (select predictors of outcome)
2. Run LASSO of D_i on all controls (select predictors of treatment)
3. Final regression: union of variables from steps 1 and 2

LASSO: penalized regression that automatically sets some coefficients to zero

$$\min_{\beta,\gamma} \sum_{i=1}^{n} (Y_i - \beta D_i - X_i' \gamma)^2 + \lambda \sum_{j=1}^{p} |\gamma_j| \tag{1}$$

Advantages: Data-driven, reduces researcher discretion, helps avoid p-hacking

Outline

Why instrumental variables?

The problem: Treatment is often endogenous — people who choose treatment differ in unobservable ways

Example: Does college education increase wages?

- Selection bias: $\mathbb{E}[Y_i(1) - Y_i(0)|D_i = 1] \neq \mathbb{E}[Y_i(1) - Y_i(0)]$

Solution: Find an **instrument** — a source of exogenous variation in treatment

- Affects treatment but not the outcome directly
- Examples: Draft lottery, proximity to college, scholarship eligibility

This week's goal: Master IV theory and practice, including three key challenges:

1. Justifying the **exclusion restriction**
2. Understanding the **Local Average Treatment Effect** (LATE)
3. Dealing with **weak instruments**

Roadmap for today

1. **Choosing controls**: Good and bad controls, DAGs, and LASSO for control selection
2. **Basics of instrumental variables**: What is an instrument? 2SLS and the Wald estimator
3. **The exclusion restriction challenge**: Why ”as-good-as-random” is not enough
4. **Heterogeneous treatment effects and LATE**: Compliers, always-takers, never-takers — why IV estimates effects only for those who respond to the instrument
5. **Weak instruments**: Finite-sample bias and the first-stage F-statistic

Outline

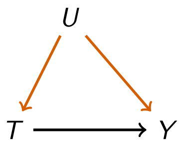
The identification problem

Recall from lecture 1: we want the **causal effect** of treatment T on outcome Y

But there are unobservable factors U (ability, motivation, etc.) that affect **both** T and Y

This creates two paths from T to Y:

- The **causal path**: $T \rightarrow Y$ (what we want)
- The **confounding path**: $T \leftarrow U \rightarrow Y$ (the problem)



Red arrows = confounding path

Why is the effect not identifiable?

When we compare outcomes by treatment status, we get:

$$\mathbb{E}[Y_i | T_i = 1] - \mathbb{E}[Y_i | T_i = 0] = \underbrace{\text{Causal effect}}_{\text{what we want}} + \underbrace{\text{Selection bias}}_{\text{because } U \text{ differs}}$$

Question: Why does U differ across treatment groups?

Answer: Because $U \to T$ — people with different unobservables *select* into different treatment levels

Example: Does college (T) increase wages (Y)?

- Ability (U) affects both college attendance *and* wages directly
- People who attend college have higher ability on average
- So the wage gap overstates the causal effect of college

The IV idea

We need variation in T that is **unrelated to the confounder** U

Question for you: Can you think of something that affects whether people go to college, but is unrelated to their ability?

What we need is a variable Z that:

- Pushes people into or out of treatment (Z affects T)
- Has no connection to the unobservable U
- Has no direct effect on Y except through T

Key insight: If Z has these properties, then any correlation between Z and Y *must be caused by* T

What is an instrumental variable?

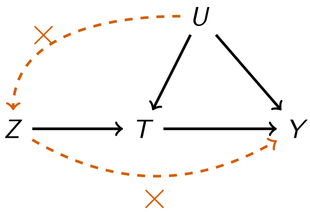
An **instrumental variable** Z satisfies:

1. Relevance: Z affects T

- $\text{Cov}(Z_i, T_i) \neq 0$

2. Exclusion / Exogeneity: Z affects Y **only through** T

- No direct $Z \to Y$ arrow
- No $U \to Z$ arrow (independence)
- $\mathbb{E}[\varepsilon_i | Z_i] = 0$



Red dashed = assumed absent

Why does an instrument work?

Follow the logic step by step:

1. Z has no direct effect on Y (exclusion restriction)
2. Z is uncorrelated with U (exogeneity)
3. Therefore, any correlation between Z and Y can **only** arise because $Z \to T \to Y$
4. So if we observe that Z correlates with Y, this reveals that T **causally affects** Y

To **quantify** the causal effect, note that:

Effect of Z on Y = (Effect of Z on T) $\times$ (Effect of T on Y)

Reduced form = **First stage** $\times$ β

To recover β: **divide the reduced form by the first stage**

The Wald estimator

Suppose Z_i is binary ($= 0$ or 1). Start from:

$$Y_i = \alpha + \beta T_i + \varepsilon_i \quad \text{where } \mathbb{E}[\varepsilon_i | Z_i] = 0$$

Taking conditional expectations:

$$\mathbb{E}[Y_i | Z_i = 1] = \alpha + \beta\, \mathbb{E}[T_i | Z_i = 1] \tag{2}$$
$$\mathbb{E}[Y_i | Z_i = 0] = \alpha + \beta\, \mathbb{E}[T_i | Z_i = 0] \tag{3}$$

Subtracting and solving for β:

$$\beta_{\text{Wald}} = \frac{\mathbb{E}[Y_i | Z_i = 1] - \mathbb{E}[Y_i | Z_i = 0]}{\mathbb{E}[T_i | Z_i = 1] - \mathbb{E}[T_i | Z_i = 0]} = \frac{\text{Reduced form}}{\text{First stage}}$$

The Wald estimator: connection to lecture 1

You already met this in lecture 1 for imperfect compliance! There, random assignment played the role of Z.

$$\beta_{\text{Wald}} = \frac{\mathbb{E}[Y_i|Z_i = 1] - \mathbb{E}[Y_i|Z_i = 0]}{\mathbb{E}[T_i|Z_i = 1] - \mathbb{E}[T_i|Z_i = 0]} = \frac{\text{Reduced form}}{\text{First stage}}$$

Key idea: Divide the effect of Z on Y by the effect of Z on T to recover the causal effect of T on Y

Two-Stage Least Squares (2SLS)

When Z_i is not binary, we generalize using **2SLS**:

First stage: Regress T on Z to get predicted treatment

$$T_i = \pi_0 + \pi_1 Z_i + v_i \quad \Rightarrow \quad \hat{T}_i = \hat{\pi}_0 + \hat{\pi}_1 Z_i$$

Second stage: Regress Y on predicted $\hat{T}$

$$Y_i = \alpha + \beta \hat{T}_i + u_i$$

Question: Why use $\hat{T}_i$ instead of T_i?

Answer: $\hat{T}_i$ contains **only the variation in T that comes from Z** — the exogenous part. The endogenous variation (driven by U) stays in the residual v_i.

The reduced form perspective

We can also see this through the **reduced form**. Substituting the first stage into the outcome equation:

$$
\begin{aligned}
Y_i &= \alpha + \beta(\pi_0 + \pi_1 Z_i + v_i) + \varepsilon_i && (4) \\
&= \underbrace{(\alpha + \beta\pi_0)}_{\text{constant}} + \underbrace{(\beta\pi_1)}_{\text{reduced form coeff}} Z_i + \underbrace{(\beta v_i + \varepsilon_i)}_{\text{composite error}} && (5)
\end{aligned}
$$

The coefficient on Z in the reduced form is $\beta \times \pi_1$. So:

$$\boxed{\beta = \frac{\text{Reduced form coeff on } Z}{\text{First stage coeff on } Z} = \frac{\beta\pi_1}{\pi_1}}$$

Same logic: all of Z's effect on Y flows through T, so divide out the first stage

The IV model: notation

Let's consolidate the notation. The IV model consists of:

Outcome equation (structural):

$$Y_i = \alpha + \beta T_i + \varepsilon_i$$

- β = causal effect of interest
- ε_i = structural error, captures everything affecting Y besides T
- Key problem: $\text{Cov}(\varepsilon_i, T_i) \neq 0$ — this is the endogeneity

First stage:

$$T_i = \pi_0 + \pi_1 Z_i + v_i$$

- π_1 = effect of instrument on treatment
- v_i = first-stage residual (contains the endogenous variation in T)

The IV model: reduced form

Reduced form (substitute first stage into outcome equation):

$$Y_i = (\alpha + \beta\pi_0) + (\beta\pi_1)Z_i + (\beta v_i + \varepsilon_i)$$

This shows how the three equations relate:

- The reduced form coefficient on Z equals $\beta \times \pi_1$
- Dividing by the first stage π_1 recovers β

Formal statement of IV assumptions

Assumption 1 (Relevance): $\pi_1 \neq 0$, i.e., $\text{Cov}(Z_i, T_i) \neq 0$
"The instrument moves treatment." **Testable** (first-stage F-stat)

Assumption 2 (Exogeneity / Exclusion): $\mathbb{E}[\varepsilon_i | Z_i] = 0$ (implies $\text{Cov}(Z_i, \varepsilon_i) = 0$)
"The instrument is uncorrelated with the structural error." **Not testable**

Under these two assumptions, the IV estimator is **consistent**:

$$\hat{\beta}_{IV} = \frac{\widehat{\text{Cov}}(Y_i, Z_i)}{\widehat{\text{Cov}}(T_i, Z_i)} \xrightarrow{p} \frac{\text{Cov}(Y_i, Z_i)}{\text{Cov}(T_i, Z_i)} = \frac{\beta \cdot \text{Cov}(T_i, Z_i)}{\text{Cov}(T_i, Z_i)} = \beta$$

Examples: Where do instruments come from?

Random events / natural experiments:

- Vietnam draft lottery $\rightarrow$ military service (Angrist, 1990)
- Quarter of birth $\rightarrow$ years of schooling (Angrist & Krueger, 1991)
- Same-sex siblings $\rightarrow$ family size (Angrist & Evans, 1998)

Institutional rules / policy discontinuities:

- Election cycles $\rightarrow$ police hiring (Levitt, 1997)
- School construction program $\rightarrow$ years of schooling (Duflo, 2001)

Exercise: For each example, identify Z, T, Y, and U. What does the exclusion restriction require?

Summary and what comes next

So far: IV gives us a way to estimate causal effects when treatment is endogenous, by exploiting exogenous variation from an instrument

We need two assumptions:

1. **Relevance:** $\text{Cov}(Z_i, T_i) \neq 0$ (testable)
2. **Exclusion:** $\mathbb{E}[\varepsilon_i | Z_i] = 0$ (not testable – must be argued)

The tricky part starts now. Two major issues with this setup:

1. **Exclusion restriction**: Even with a "random" instrument, exclusion is hard to justify. Why?
2. **Homogeneous effects**: We assumed β is the same for everyone. What happens when treatment effects vary across individuals?
   - This leads to the **Local Average Treatment Effect** (LATE)

Outline

Why is the exclusion restriction challenging?

- Recall the key (untestable) feature for IV: exclusion restriction
- In the context of the DAG, the intuition is that Z only affects Y through T
- Intuitively, it feels like something randomly assigned or nearly random should satisfy this, so long as it affects T
- <u>This is not sufficient</u>
  - You need to think critically about the IV

$$
\begin{tikzcd}
& U \arrow[d] \arrow[dr] & \\
Z \arrow[r] & T \arrow[r] & Y
\end{tikzcd}
$$

Why is the exclusion restriction challenging?

- Consider two examples. First, using Vietnam war lottery numbers as an IV for military service, studying the impact on mortality.
  - Y: death, T: vietnam vet, Z: lottery number
- Lottery number was randomly assigned as a function of birthdate
  - Well-defined design-based view of Z allocation!
- Does that necessarily satisfy exclusion restriction? Seems like a pretty slam dunk IV
  - Clearly affects veteran status
  - Clearly random!

Why is the exclusion restriction challenging?

- Does that necessarily satisfy exclusion restriction?
  - Not necessarily!
- Why? Consider one simple example: being drafted induces you to change your behavior to avoid the draft
  - Stay in school
  - Flee to Canada
- This would violate the exclusion restriction!

Why is the exclusion restriction challenging?

- Second, consider rainfall as an instrument for income in agriculture environments (many crops are heavily dependent on it)
  - This is not uncommon in development papers, as Sarsons (2015) points out
  - Y: conflict, T: income, Z: rainfall
- Exclusion restriction is that rainfall has no effect on conflict <u>beyond</u> income
  - While the logic seems reasonable, Sarsons (2015) shows that places with dams (which protect against the income shocks due to rain) have similar conflict to those without dams
- Plausible that while rain is "<u>random</u>", it might have many channels

Exclusion Restrictions

- Even with a variable that is near-random in its allocation, the exclusion restriction is not always satisfied
  - Worse yet, it's a fundamentally untestable restriction
- Using an IV requires thinking carefully about justifying the exclusion restriction
  - It can also be useful to think about what violations in the restriction implies

Outline

We are now going to discuss a fundamental concern about IV methods.

That IVs do not in general the **Average Treatment Effect**, but the “**Local Average Treatment Effect**” on a subpopulation of individuals called compliers.

Hypothetical model

- Estimate the wage impact of college rather than high school education
- Education is endogenous
- Instrument for education: whether the individuals are eligible to a college scholarship (to fix ideas, assume this has been randomized)

Notations:

$$W : \text{wage} \tag{6}$$
$$S = 1 \text{ if college} \tag{7}$$
$$Z = 1 \text{ if scholarship} \tag{8}$$

Z=0 (no scholarship)	**Z=1** (scholarship)
100 High school	0 High school
0 College	100 College
Average wage: 100	Average wage: 130

Causal impact = ITT / [E(X—Z=1) - E(X—Z=0)] = (130-100) / (1 - 0) = 30

Causal impact = ITT / [E(X—Z=1) - E(X—Z=0)] = (130-110) / (1 - 0.2) = 25

Z=0 (no scholarship)	**Z=1** (scholarship)
80 High school	0 High school
20 College	100 College
Average wage: 110	Average wage: 130

ATE = (122 - 110) / (0.8 - 0.2) = 20

Z=0 (no scholarship)	**Z=1** (scholarship)
80 High school 20 College	20 High school 80 College
Average wage: 110	Average wage: 122

Z=0 (no scholarship)	**Z=1** (scholarship)
HIGH SCHOOL PARENTS	
80 High school	0 High school
0 College	80 College
Average wage: 100	Average wage: 120
COLLEGE PARENTS	
0 High school	0 High school
20 College	20 College
Average wage: 125	Average wage: 125

Z=0 (no scholarship)	Z=1 (scholarship)
HIGH SCHOOL PARENTS	
80 High school	0 High school
0 College	80 College
COLLEGE PARENTS	
0 High school	0 High school
20 College	20 College
Average wage: 105	Average wage: 121

Sum up

- When we have separate information for High school parents and College parents: High school parents: +20 effect College parents: not identified
- When we have global estimation stacking HS and C parents: +20 effect

College parents population does not seem to contribute

Sum up

There is no way we can learn something on the impact among college parents population because there is no experiment actually going on in that population

- In this example, all the reduced form effect comes from HS parents population: 121-105=16
- And they represent a change in college participation in 80% of the sample
- Thus, the effect $16/0.8 = 20$ results only from HS parents population

Let's call them **compliers** because they comply with the treatment assignment

Z=0 (no scholarship)	Z=1 (scholarship)
High school 80 (HS parents)	College 80 (HS parents)
College 20 (College parents)	College 20 (College parents)
Average wage: 105	Average wage: 121

Impact is identified on the share of population who moves from HS to College

Now add Never takers:

Z=0 (no scholarship)	Z=1 (scholarship)
High school 10 (Never takers)	High school 10 (Never takers)
High school 80 (Compliers)	College 80 (Compliers)
College 10 (Always takers)	College 10 (Always takers)
Average wage: 105	Average wage: 121

All the change in the reduced form: 121-105 is due to compliers What is the share of compliers in the sample? 80% Thus impact: $16/0.8 = 20$

- **20 is the effect on the compliers** (should it be different for the other populations)
- Information: 90 HS, 10 College for $Z = 0$ and 10 HS, 90 College for $Z = 1$
- How do we know there are 80% compliers? Can we name them?

Now add Defiers:

Z=0 (no scholarship)	Z=1 (scholarship)
High school 5 (Never takers)	High school 5 (Never takers)
High school 80 (Compliers)	College 80 (Compliers)
College 10 (Always takers)	College 10 (Always takers)
College 5 (Defiers)	High school 5 (Defiers)
Average wage: 105	Average wage: 121

Formalizing

$T(Z)$ is a random variable that assigns an individual response T to the value of the instrument Z

Every person may respond differently to the instrument

$$\begin{aligned} \text{Compliers} \quad & T_i(0) = 0 \quad T_i(1) = 1 && (9) \\ \text{Never-takers} \quad & T_i(0) = 0 \quad T_i(1) = 0 && (10) \\ \text{Always-takers} \quad & T_i(0) = 1 \quad T_i(1) = 1 && (11) \\ \text{Defiers} \quad & T_i(0) = 1 \quad T_i(1) = 0 && (12) \end{aligned}$$

Note: can generalize to more values of the instrument than just $(0, 1)$

Hypothesis 1 (Independence)

Z is independent from $(Y_0, Y_1, T(0), T(1))$

In particular implies that people with some sensitivity to the instrument (described by the set $\{T(0), T(1)\}$) are not more or less likely to draw a specific value of z

Hypothesis 2 (Monotonicity)

either $T_i(0) \geq T_i(1) \quad \forall i$ or $T_i(0) \leq T_i(1) \quad \forall i$

i.e.: all agents' response to the instrument is (weakly) in the same direction

For instance: a mother with one boy-one girl who has a third child would also have a third child if she had two boys (the effect of same-sex is never to reduce fertility)

Monotonicity is equivalent to the absence of defiers

Reduced form

$$E(Y|Z=1) = E(Y_0 + T(Y_1 - Y_0)|Z=1) \tag{13}$$
$$= E(Y_0 + T(1)(Y_1 - Y_0)) \tag{14}$$

Thus

$$E(Y|Z=1) - E(Y|Z=0) = E(Y_0 + T(1)(Y_1 - Y_0)) - E(Y_0 + T(0)(Y_1 - Y_0)) \tag{15}$$
$$= E[(T(1) - T(0))(Y_1 - Y_0)] \tag{16}$$

Reduced form

$$\begin{aligned}
E[(T(1) - T(0))(Y_1 - Y_0)] = & \qquad (17) \\
& E[(Y_1 - Y_0)|T(1) - T(0) = 1]P(T(1) - T(0) = 1) \qquad (18) \\
& + E[0 \times (Y_1 - Y_0)|T(1) - T(0) = 0]P(T(1) - T(0) = 0) \qquad (19) \\
& + E[-1 \times (Y_1 - Y_0)|T(1) - T(0) = -1]P(T(1) - T(0) = -1) \qquad (20)
\end{aligned}$$

$$\begin{aligned}
& = E[(Y_1 - Y_0)|C]P(C) \qquad (21) \\
& + E[0 \times (Y_1 - Y_0)|A \text{ or } N]P(A \text{ or } N) \qquad (22) \\
& + E[-1 \times (Y_1 - Y_0)|D]P(D) \qquad (23)
\end{aligned}$$

Role of monotonicity

Assume $T(1) \geq T(0)$; then $T(1) - T(0) = -1$ is impossible; there are no defiers

Thus:

$$E(Y|Z=1) - E(Y|Z=0) = E[(Y_1 - Y_0)|T(1) - T(0) = 1]P(T(1) - T(0) = 1) \tag{24}$$

with

$$\begin{aligned} P(T(1) - T(0) = 1) &= E(T(1) - T(0)) && (25) \\ &= E(T|Z=1) - E(T|Z=0) && (26) \\ &= P(T=1|Z=1) - P(T=1|Z=0) && (27) \end{aligned}$$

LATE

Under hypothesis 1 (*Independence*) and 2 (*Monotonicity*), the Wald estimator is:

$$W = \frac{E(Y|Z=1) - E(Y|Z=0)}{P(T=1|Z=1) - P(T=1|Z=0)} \tag{28}$$

$$= E[(Y_1 - Y_0)|T(1) - T(0) = 1] = LATE \tag{29}$$

Local Average Treatment Effect: treatment effect on those that change their behavior (T) under the instrument (compliers)

LATE with more than 2 values

When instrument takes more than 2 values, $LATE_{Z_1,Z_2}$ can be defined for each pair of values of the instrument (Z_1, Z_2).

The IV estimator uses all values of Z at a time: can be interpreted as a weighted sum of the LATEs, where the weights depend on the local impact of the instrument

What about the ATE?

So we cannot use IVs to estimate the ATE if:

1. There is treatment heterogeneity ($E(Y_1 - Y_0)$ is not constant), and
2. **This heterogeneity is related to treatment behavior**:

$$E(Y_1 - Y_0) \neq E(Y_1 - Y_0 | \text{Compliers}) \tag{30}$$

This is called “essential heterogeneity”.

In this case, $LATE \neq ATE$.

Implications

- IV has no clear interpretation if there is essential heterogeneity or if there are defiers
- Different instruments can identify different parameters because they estimate the impacts on different populations
- The gap between OLS and IV mix the result of bias reduction and change in the populations that contribute to the estimation

This is a **major reason** why IV estimations have fallen out of favor among economists, along with the difficulty of justifying the exclusion restriction

Thinking about the LATE: examples

1. Scholarship $\rightarrow$ secondary education $\rightarrow$ wage at 25
2. Vietnam draft $\rightarrow$ military service $\rightarrow$ death
3. Rainfall shocks $\rightarrow$ household agricultural income $\rightarrow$ civil conflict

Outline

- So far we have been thinking about **identification**, but less about **estimation**
- Now let's discuss the main issue regarding estimation of IVs - **weak instruments**

An instrument is said to be **weak** if it explains little of the endogenous variable

Weak instruments

$$Y_i = T_i\beta + \epsilon_i$$
$$T_i = Z_i\pi_1 + u_i$$

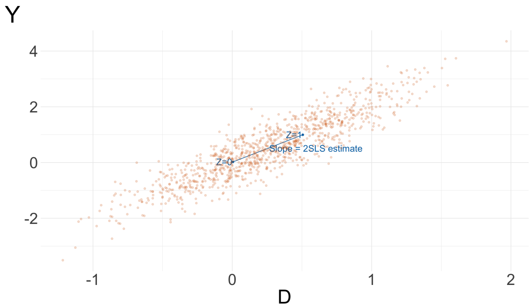
- Recall that one of the key assumptions for our estimation procedure was relevance
  - $\pi_1 \neq 0$, or $Cov(Z_i, T_i) \neq 0$
- Why is this necessary? Consider the 2SLS estimator for β_{IV} in the simplest case:

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

- If $Cov(D_i, Z_i) = 0$, this estimate is obviously undefined! But what about if it's very small?
  - Small variations in it will move around $\hat{\beta}$ in a big way. That's what statistical uncertainty will do
  - One easy way to see this: graphically

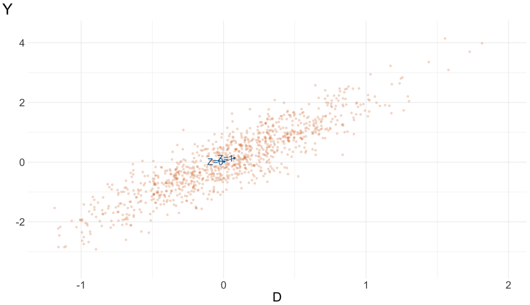
Weak instruments

- Simple 2SLS simulation, with binary instrument
  - First stage coef = 0.5, true beta = 2
- Note that the estimation on the x-axis comes from variation in the first stage
- The larger this is, the stronger the first stage
- However, if the first stage is weak, this interval is quite short, even if the variation in D stays the same



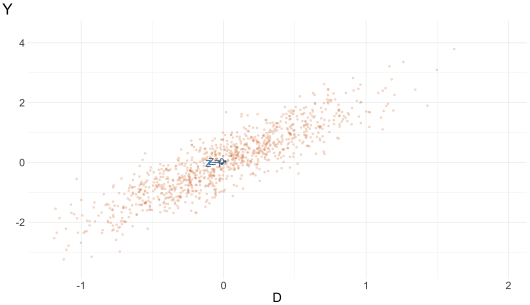
Weak instruments

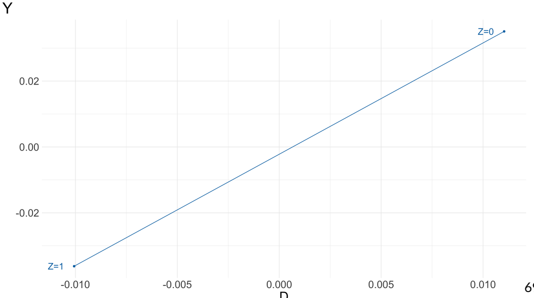
- With a first stage coefficient of 0.1, it becomes hard to distinguish the points
  - Note: I hold fixed the overall variance of D here to keep the correct comparison!
- Given that the model is correctly specified, with enough data it should converge to the right β
- But small shifts in the x-axis will massively swing the estimate!



Weak instruments

- With a first stage coefficient of 0.01, the problem is even worse
- We see that the relevant variation being exploited is tiny
- A small change in the x-axis points would even flip the sign!
- What does that do to our estimation procedure?





Examples of weak instruments?

1. Birth month (Z) $\rightarrow$ years of schooling (T) $\rightarrow$ adult wages (Y)
   [Angrist & Krueger 1991, later critiqued by Bound, Jaeger & Baker 1995]
2. Colonial settler mortality (Z) $\rightarrow$ institutional quality (T) $\rightarrow$ modern-day GDP (Y)
   [Acemoglu, Johnson, Robinson 2001]

What concretely happens if we have a weak instrument?

1. **Loss of precision**
2. **Bias in finite samples**

1. Loss of precision

Recall expression for variance of an IV estimator in the simplest case:

$$Var(\hat{\beta}_{IV}) = \frac{V(u)}{N} \cdot \frac{1}{Var(T)} \cdot \frac{1}{\pi_1^2}$$

when $Y = \alpha + \beta T + u$ and $T = \pi_1 Z + v$

So as π_1 gets smaller, $Var(\hat{\beta}_{IV})$ increases more than linearly

2. Bias in finite samples

Even though an IV estimator is consistent, it is still biased in finite samples.

- **Consistency**: $\hat{\beta}_{2SLS} \xrightarrow{p} \beta$
- **Biased**: $\mathbb{E}[\hat{\beta}_{2SLS}] \neq \beta$

The bias is towards β_{OLS}

$$T = Z\pi + v \tag{31}$$

We want to replace T with what is in the 2nd stage. We need to estimate π using $\hat{\pi}$

- We would require $\hat{T} = Z\pi$
- But in finite sample $\hat{\pi} \neq \pi$ so $\hat{T} \neq Z\pi$
- The least square criteria to estimate $\hat{\pi}$ ”get $\hat{T}$ close to T”
- The mistake is towards ”$\hat{T}$ looks like T too much”: ”overfit”
- So $\hat{\beta}_{2SLS}$ looks too much like $\hat{\beta}_{OLS}$

Determinants of bias

Expression for bias of IV estimator:

$$E(\hat{\beta}_{2SLS}) - \beta \approx \frac{\text{cov}(\varepsilon, v)}{\sigma_v^2}\left[\frac{1}{1+F}\right] \tag{32}$$

where F is an F-test statistic of the regression of T on Z, i.e.,

$$F = \frac{R^2_{T,Z}/K}{(1-R^2_{T,Z})/(N-K)} \tag{33}$$

where K is the number of instruments (usually $K = 1$), $R^2_{T,Z}$ is the R^2 in the regression of T and Z

Determinants of bias

$$E(\hat{\beta}_{2SLS}) - \beta \approx \frac{\text{cov}(\varepsilon, v)}{\sigma_v^2}\left[\frac{1}{1+F}\right] \tag{34}$$

$$F = \frac{R^2_{T,Z}/K}{(1-R^2_{T,Z})/(N-K)} \tag{35}$$

- Correlation between ε and v (source of bias)
- F (measure of weak instruments), mostly driven by how much the instruments explain T ($R^2_{T,Z}$) (weak instrument when R^2 is small)

Determinants of bias

$$E(\hat{\beta}_{2SLS}) - \beta \approx \frac{\text{cov}(\varepsilon, v)}{\sigma_v^2} \left[\frac{1}{1+F} \right] \tag{36}$$

$$F = \frac{R^2_{T,Z}/K}{(1 - R^2_{T,Z})/(N-K)} \tag{37}$$

Implications:

- If $R^2_{T,z}$ is small enough, even large n cannot impede strong bias
- Adding instruments is a bad idea if instruments are weak (increase K but hardly increases $R^2_{T,Z}$)

Testing for weak instruments

The F-stat for $\pi = 0$ (significance of excluded instruments in first step) is proportional to the bias

But also depends on other parameters K, N and cov(ε, v)

Stock & Yogo (2005) derive formal tests: Roughly, if $F > 10$, reject that the 2SLS bias will be more than 10% of the OLS bias

Summing up

We covered the basics of IVs. They are a way of estimating causal effects that don't rely on an experimenter randomly allocating treatments.

But they come with a number of very important challenges:

1. Justifying the **exclusion restriction**
2. Understanding what the **LATE** is really measuring
3. Dealing with **weak instruments**

Feedback

How is the course going?

Please take 1 minute to give anonymous feedback:

https://forms.gle/x2S7Ju5EsbPTzn5H9