# Instrumental Variables

## BUSS975: Causal Inference in Financial Research

Professor Ji-Woong Chung

Korea University

# Where we left off

- **Lecture 4 (RD).** Identification came from a *rule*: a known threshold in a running variable. One assumption — continuity — and an estimand at one point.
- **Today (IV).** Identification comes from a *shock*: something in the world that pushes units into and out of treatment for reasons unrelated to their outcomes.
- The problem IV addresses is the one matching and RD could not touch: **selection on unobservables**. You believe $D$ is correlated with something you cannot measure, and no amount of conditioning will fix it.
- The price is steep. RD has one main identifying assumption. IV, under heterogeneous effects, has **five** — and the most important one is untestable.
- "Finding a valid IV is like finding a four-leaf clover."

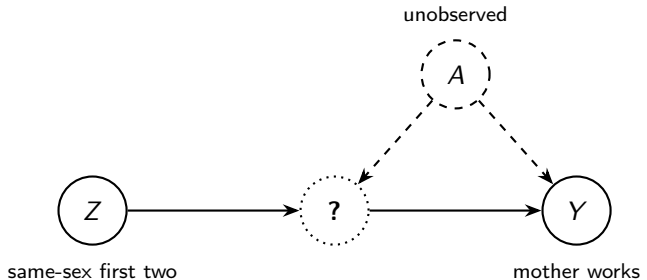
# 7.1 The Strangeness Principle

# A strange thing happened at Thanksgiving

- Your brother, four years into a PhD, produces a chart at dinner. Mothers whose **first two children are the same sex** — two boys or two girls — are about **one percentage point less likely** to work for pay than mothers whose first two are a boy and a girl.
- Why on earth would the *sexes* of two children have anything to do with whether their mother holds a job? Two boys are no more burdensome than one of each.
- Nobody can explain it. That baffled reaction is the point.
- This is a real result: **Angrist and Evans (1998)**.

- The DAG you are implicitly drawing has a **hole** in it.

# The hole in the DAG



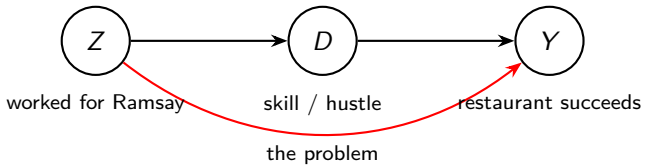

- You know the IV, the outcome, and maybe the confounders. **You do not know the treatment** — and that missing link is the source of the strangeness.
- Fill the hole with **family size**: parents who want a mix of sexes and get two of a kind try for a third child, and a third child changes labour supply.
- Sex ratios are valid IVs for family size *because* they move family size and are otherwise unrelated to the mother's labour supply.

# The strangeness principle

*A valid IV neither directly causes the outcome, nor is correlated with anything else that does —* ***except through the treatment.***

- In a valid DAG the only route from $Z$ to $Y$ runs *through* $D$.
- Now look at the raw $Z$–$Y$ correlation, in which $D$ never appears. You see a real association and **no visible reason for one**, because the only thing that could explain it has been left out of the picture. *That* is the strangeness — and it is a symptom of validity, not a defect.
- If the $Z$–$Y$ association looks perfectly natural on its face, that usually means $Z$ reaches $Y$ by some route *other* than $D$ — which is the next slide.

# When nothing feels strange, be suspicious



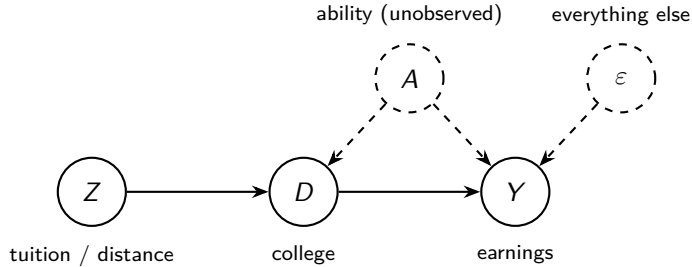

- Cousin Jacob became a chef. He was a line cook on one of Gordon Ramsay's shows, and credits his own hustle for everything since.
- The table laughs. There is **nothing strange** about someone who worked with Ramsay going on to succeed — his name alone brings investors, press, partnerships and early customers.
- That red arrow is a direct path from IV to outcome. It is a violation of the **exclusion restriction**, and it cannot be fixed by any estimator.

# IVs are treatment assignment mechanisms

- **A valid IV corresponds to something in the real world that literally moves units into and out of treatment.** Coin flips. Draft numbers. Lotteries.
- The best IVs come from **institutional knowledge** of how the treatment is actually assigned.
- Talk to people. Ask them, in their own words, how some ended up treated and others did not.
- You will notice this is the same discipline that §6.4 demanded of RD: *know the rule*. Here: *know the shock*.

# 7.2 The Wald Estimator

# When OLS is biased




Suppose the true model of earnings is

$$Y_i = \alpha + \delta D_i + \gamma A_i + \varepsilon_i,$$

but $A$ is not in your data, so you estimate $Y_i = \alpha + \delta D_i + \eta_i$ with $\eta_i = \gamma A_i + \varepsilon_i$.

- $A$ **and** $\varepsilon$ **are both unobserved**, but they differ in one decisive way: $A$ shifts who goes to college, while everything else, $\varepsilon$, does not. Hence $C(D, A) \neq 0$ but $C(D, \varepsilon) = 0$.

# The omitted variable bias formula

Write $C(\cdot, \cdot)$ for covariance and $V(\cdot)$ for variance. OLS is a scaled covariance:

$$\hat{\delta} = \frac{C(Y, D)}{V(D)}.$$

Substitute the *long* model for $Y$ and expand:

$$\begin{aligned} \hat{\delta} &= \frac{\delta V(D) + \gamma\, C(D, A) + C(D, \varepsilon)}{V(D)} \\ &= \delta \;+\; \gamma\, \frac{C(A, D)}{V(D)} \qquad \text{since } C(D, \varepsilon) = 0. \end{aligned}$$

# The trick: take the covariance with $Z$ instead

Keep the long model and compute $C(Y, Z)$ rather than $C(Y, D)$:

$$C(Y, Z) \;=\; \delta\, C(D, Z) \;+\; \gamma\, C(A, Z) \;+\; C(\varepsilon, Z).$$

- If the DAG is true, $C(A, Z) = 0$ and $C(\varepsilon, Z) = 0$.

$$C(Y, Z) = \delta\, C(D, Z).$$

- Divide both sides by $C(D, Z)$:

$$\boxed{\hat{\delta}_{\text{Wald}} \;=\; \frac{C(Y, Z)}{C(D, Z)} \;=\; \delta}$$

- This ratio is the Wald estimator, named after Abraham Wald (Wald 1940).

# What makes it work, and what cannot be checked

The Wald estimator recovers $\delta$ when:

1. **Exclusion.** $Z$ does not cause $Y$ directly — any effect runs through $D$. Equivalently, $Z$ is *excluded from* the $Y$ equation: write it in and its coefficient is zero. Hence the standard phrase **"excluded instrument"**.
2. **Independence.** $Z$ is independent of the confounder $A$ and the error $\varepsilon$.
3. **A nonzero first stage.** $C(D, Z) \neq 0$.

- Under these, $\operatorname{plim} \hat{\delta} = \delta$: **consistency**. In finite samples it differs, and by how much depends largely on the strength of that denominator — §7.4.
- Exclusion cannot be tested. Arguments for it are *narrative*.
- Good narratives rest on a realistic understanding of how treatment is actually assigned.
- Defeating an IV paper requires only creating doubt about exclusion.

# Reduced form and first stage

$$\underbrace{\frac{C(Y,Z)}{C(D,Z)}}_{\hat{\delta}_{\text{Wald}}} = \frac{\textbf{Reduced form}}{\textbf{First stage}}$$

- The **reduced form**, $C(Y,Z)$, is the strange association itself — the $Z$–$Y$ correlation that should not be there.
- The **first stage**, $C(D,Z)$, is the association between IV and treatment. It is what *resolves* the strangeness.

# 7.3 Two Stage Least Squares

# Wald *is* 2SLS

Start from the Wald ratio and multiply through by $V(Z)/V(Z)$:

$$\hat{\delta}_{\text{Wald}} = \frac{C(Y,Z)}{C(D,Z)} = \frac{C(Y,Z)/V(Z)}{C(D,Z)/V(Z)} = \frac{\hat{\pi}_1\, C(Y,Z)}{\hat{\pi}_1^2\, V(Z)} = \frac{C(Y,\hat{\pi}_1 Z)}{V(\hat{\pi}_1 Z)},$$

where $\hat{\pi}_1 = C(D,Z)/V(Z)$ is the first-stage slope from $D = \pi_0 + \pi_1 Z + \nu$.

- But $\hat{\pi}_1 Z$ is, up to the constant, the **fitted value** $\hat{D}$ from that first stage. So

$$\hat{\delta}_{2SLS} = \frac{C(\hat{D},Y)}{V(\hat{D})}.$$

- We are no longer using $D$. We are using the part of $D$ that the IV explains — the part that is, by assumption, unrelated to $A$.
- **Without covariates, Wald and 2SLS are identical.**

# Four ways, one number

```
suppressMessages({library(haven); library(AER)})
card <- as.data.frame(read_dta("data/card.dta"))
d <- card[!is.na(card$lwage) & !is.na(card$educ) & !is.na(card$nearc4), ]

w1 <- cov(d$lwage, d$nearc4) / cov(d$educ, d$nearc4)        # ratio of covariances
w2 <- coef(lm(lwage ~ nearc4, d))[2] / coef(lm(educ ~ nearc4, d))[2]  # ratio of OLS
fs <- lm(educ ~ nearc4, d)
w3 <- coef(lm(d$lwage ~ fitted(fs)))[2]                     # manual two stages
w4 <- coef(ivreg(lwage ~ educ | nearc4, data = d))[2]       # one command
setNames(unname(c(w1, w2, w3, w4)),
         c("covariances", "OLS ratio", "manual 2SLS", "ivreg"))
```

```
covariances   OLS ratio manual 2SLS       ivreg
  0.1880626   0.1880626   0.1880626   0.1880626
```

Card (1995), $N = 3{,}010$. Compare OLS on the same data: 0.0521.

# One warning about doing it by hand

- Manual two-step 2SLS is fine to **see** the mechanics, never to **report**: second-stage SEs treat $\hat{D}$ as data, so they are too small.
- Use `ivreg` (R) or `ivregress 2sls` (Stata) for the correct variance.
- More IVs than endogenous regressors means the model is **over-identified**, and 2SLS uses a weighted combination of the IVs. That is where §7.4's trouble begins.
- Over-identification test (Sargan, or Hansen's $J$ with robust SEs): do the IV-specific Wald estimates agree? A rejection is usually read as "an IV is invalid"; §7.5 gives a second reason.

# 7.4 Weak Instruments

# Quarter of birth and compulsory schooling

- **Angrist and Krueger (1991).** Two institutional facts, combined:
  1. Children start first grade based on a **31 December** cutoff. Two children born a day apart, 31 December and 1 January, start a year apart — purely by chance.
  2. Compulsory schooling laws let students leave at **16**.
- A child born in the fourth quarter therefore reaches 16 having completed *more* schooling than one born in the first quarter.
- So quarter of birth shifts years of schooling — a little, for some people — and is otherwise, one hopes, unrelated to earnings.
- **The strangeness test:** tell someone that people born in January earn less than people born in October. They will frown. Good.

# Present the Wald estimator as two pictures

- Angrist and Krueger show the **first stage** — years of schooling by quarter of birth, cohort by cohort — as a figure. Fourth- and third-quarter births sit visibly higher.
- They show the **reduced form** — log weekly earnings by quarter of birth — the same way. Jagged, but the 3s and 4s sit above the 1s and 2s.
- Those two figures *are* the numerator and the denominator of the Wald estimator.
- **The rule for your own work:** visualise the reduced form and the first stage separately, even if you will ultimately report 2SLS. A plot of $Y$ against $\hat{D}$ is opaque and invites suspicion, because the data have already been processed.
- A second thing the figure reveals: the relationship **fades for later cohorts**. By the 1947 births it has nearly vanished.

# A falsification test built into the first stage (Table 7.2)

- Quarter of birth predicts **total schooling** and **high school graduation**, both negative and significant: born earlier in the year, less schooling. That is the first stage.
- It does **not** predict **college graduation**. And that is the point.
- Compulsory schooling binds only through high school. If quarter of birth predicted college completion, postgraduate degrees and total years alike, the IV would clearly be picking up something *other* than the compulsory-schooling channel — and exclusion would be in doubt.
- **The general lesson.** Where your proposed mechanism implies the IV should have *no* effect, show that it has no effect. Untestable assumptions have testable implications; go find them.

# More IVs, more precision — at what price?

- Angrist and Krueger wanted more variation in predicted schooling, which should mean more precision. So they interacted the three quarter dummies with state and year of birth: **180 IVs**. It looked like a free lunch. It was not. Many extra IVs were **nearly useless**.
- **Bound, Jaeger and Baker (1995)** showed what this does. Write both equations with their own errors, $Y_i = \delta D_i + \varepsilon_i$ and $D_i = Z_i'\pi + \nu_i$. Endogeneity *is* $\sigma_{\varepsilon\nu} \equiv C(\varepsilon_i, \nu_i) \neq 0$: the part of $D$ the IVs cannot explain moves with the structural error. Then

$$E[\hat{\delta}_{2SLS} - \delta] \approx \underbrace{\frac{\sigma_{\varepsilon\nu}}{\sigma_\nu^2}}_{\text{the OLS bias}} \cdot \frac{1}{F+1},$$

  where $F$ is the first-stage $F$-statistic.
- As $F \to \infty$ the bias vanishes; as $F \to 0$ **2SLS converges on the OLS bias**. Adding weak instruments lowers $F$ and so moves you back toward the thing you were trying to escape.

# How should we measure weakness?

- **The old rule.** Stock and Yogo (2005): first-stage $F > 10$. Based on Monte Carlo under homoskedasticity. For twenty years this was the standard.
- **The number 10 is now considered too low**, especially with multiple endogenous regressors, heteroskedasticity, or non-normal errors. Lee et al. (2022) argue the relevant threshold can be above **100**.
- Weak instruments do not merely bias point estimates. They **distort inference**: standard errors come out too small and $t$-tests mislead.
- **What to report now**, following Andrews, Stock and Sun (2019):
  1. The **effective $F$-statistic** of Montiel Olea and Pflueger (2013), which accounts for heteroskedasticity. In the just-identified case it reduces to the robust first-stage $F$.
  2. **Anderson–Rubin confidence intervals** — *regardless* of the value of $F$.

# Anderson–Rubin intervals

- A conventional 2SLS interval is built **around a point estimate**: $\hat{\delta} \pm 1.96\,\widehat{SE}(\hat{\delta})$. It presumes $\hat{\delta}$ is well behaved, which under weak instruments it is not.
- An AR interval instead **inverts a hypothesis test**. For each candidate value $\delta_0$, it asks: could this value plausibly have generated the data? Values that could not are excluded.
- AR intervals are valid **whether or not the IV is strong**.
  - Strong IV $\Rightarrow$ the interval is tight.
  - Weak instrument $\Rightarrow$ the interval expands, sometimes to include both large positive and large negative values.

# So what do you do with a weak instrument?

- **Use a just-identified model with your single strongest IV.** Fewer, stronger beats more, weaker — the opposite of the 1991 intuition.
- **Report AR intervals** and an effective $F$. Let the reader see the weakness rather than discover it.
- If you have a weak instrument problem, the only real solutions are to get a stronger IV or to stop doing IV. You are probably better off attacking the question another way.

# 7.5 Local Average Treatment Effects

# Heckman's objection

- Everything so far assumed **homogeneous** treatment effects: one $\delta$, the same for everyone. Drop that and the picture changes completely.
- **Heckman (1990)** showed that IV identifies an average causal effect only under very strong conditions, later called **identification at infinity**: the IV would have to shift essentially the *entire* population into or out of treatment.
- Real IVs do nothing of the sort. The draft lottery did not compel service for everyone with a low number, nor exempt everyone with a high one.
- So Heckman's critique appeared to undercut the most natural reading of Angrist's results.
- And yet: the US government was literally randomising young men into the military. How could that reveal *nothing* causal?
- **Imbens and Angrist (1994)** worked backwards from that tension. If IVs cannot shift everyone, what exactly do they identify?

# Potential treatment status

The move is to apply potential outcomes to the **first stage** as well as the second:

$D_i^1 = i$'s treatment status when $Z_i = 1$, $\quad D_i^0 = i$'s treatment status when $Z_i = 0$.

- Person $i$ has **two hypothetical treatment statuses**, one for each value of the IV — only one of which is ever realised.
- The switching equation now applies to treatment:

$$D_i = D_i^0 + \underbrace{(D_i^1 - D_i^0)}_{\pi_{1i}} Z_i,$$

where $\pi_{1i}$ is the **heterogeneous causal effect of the IV on treatment**.

# The five assumptions

1. **SUTVA** — no spillovers, in the treatment *or* in the IV.
2. **Independence** — the IV is as good as randomly assigned.
3. **Exclusion** — the IV reaches $Y$ only through $D$.
4. **Nonzero first stage** — the IV actually moves $D$.
5. **Monotonicity** — it moves everyone the same way, or not at all.

- Compare RD, which rests on one assumption. This is why IV papers spend most of their length on assumptions rather than estimates.

# Assumption 1: SUTVA

**Person $i$'s potential treatment status and potential outcomes depend only on *their own* IV value and *their own* treatment — not on anyone else's.**

- Potential outcomes always required this. What is new is that it now applies to the **first stage** too: $D_i^1$ and $D_i^0$ must be well defined for $i$ whatever numbers everyone else draws.
- **Violation through the IV.** I had no intention of enlisting and was not drafted. My best friend was drafted, so I went with him. My $D_i$ depends on *his* $Z_i$.
- **Violation through the treatment.** Mass conscription thins the civilian labour force, so wages change even for men who never served. Then $Y_i^0$ is not a fixed quantity — it depends on how many others were treated.
- Peer effects, market-clearing prices and congestion are the usual suspects.

# Assumption 2: Independence

$$\{\, Y_i(D_i^1, 1),\ Y_i(D_i^0, 0),\ D_i^1,\ D_i^0 \,\} \perp\!\!\!\perp Z_i$$

- "As good as randomly assigned": the IV is independent of potential outcomes **and** of potential treatment status. A draft number was drawn like a bingo ball — it knew nothing about who you were.
- **What it buys, first:** the reduced form becomes causal,

$$E[Y_i \mid Z_i = 1] - E[Y_i \mid Z_i = 0] = E[Y_i(D_i^1, 1)] - E[Y_i(D_i^0, 0)].$$

- **Second:** the first stage becomes causal,

$$E[D_i \mid Z_i = 1] - E[D_i \mid Z_i = 0] = E[D_i^1 - D_i^0].$$

# Assumption 3: Exclusion

$$Y_i(D_i, 0) = Y_i(D_i, 1) \qquad \text{for } D_i = 0, 1$$

- Any effect of $Z$ on $Y$ must travel **through** $D$.
- **The Vietnam violation.** A low lottery number pushed some men into service — and pushed others into graduate school for a deferment. Schooling affects earnings directly, so the lottery reaches earnings by two routes, not one.
- **Independence does not imply exclusion.** The lottery really was random, which secures independence. Exclusion is a separate claim about the *channel* for causal effects of the IV.

# Assumption 4: Nonzero first stage

$$E[D_i^1 - D_i^0] \neq 0$$

- The IV must actually move the average probability of treatment. Did a low lottery number raise the chance of serving? If so, the first stage holds.
- It is the **denominator** of the Wald ratio. At zero the estimand does not exist; near zero it is unstable — which was the whole of §7.4.
- **The one testable assumption**: it involves $D$ and $Z$ only, and you have data on both. Report the first-stage coefficient, its robust SE, and the effective $F$.
- Note what is *not* required: the IV need not move everyone, only a nonzero fraction. That concession is exactly what will make the estimand **local**.

# Assumption 5: Monotonicity

Recall the switching equation for the first stage, $D_i = D_i^0 + \pi_{1i} Z_i$, so that

$$\underbrace{\pi_{1i} = D_i^1 - D_i^0}_{i\text{'s own causal effect of } Z \text{ on } D} \qquad \Longrightarrow \qquad \text{either } \pi_{1i} \geq 0 \text{ for all } i, \quad \text{or } \pi_{1i} \leq 0 \text{ for all } i.$$

- The IV may leave some people untouched, but everyone it does move must move in the **same direction**. Draft eligibility pushes men into service or leaves them alone — never out.
- Equivalently: **no defiers.** Nobody serves *only because* their number was *not* called.
- Without it, compliers' and defiers' effects enter the Wald ratio with opposite signs and can cancel. The estimand need not be a weighted average of the underlying individual causal effects.

# The LATE theorem

If all five assumptions hold, IV identifies

$$\delta_{\text{LATE}} = \frac{\text{effect of } Z \text{ on } Y}{\text{effect of } Z \text{ on } D} = \frac{E[Y_i(D_i^1, 1) - Y_i(D_i^0, 0)]}{E[D_i^1 - D_i^0]} = E\left[\, Y_i^1 - Y_i^0 \mid D_i^1 - D_i^0 = 1 \,\right].$$

- The average is taken **only over compliers** — the people whose treatment status the IV actually changed. That conditioning is what the word *local* means.
- Heckman was right that IV cannot deliver the ATE. Imbens and Angrist's answer was not to argue with him, but to ask what IV *does* deliver, and to name it.
- **Angrist (1990) therefore estimated** the effect of military service on earnings *for men who served because their number was called* — which is arguably exactly the right policy parameter for a question about conscription.
- Whether the complier group is the group you care about varies entirely by application. You must argue it, not assume it.

# Proof, and where each assumption enters

A potential outcome has *two* arguments, $Y_i(d, z)$ — four per person. **Exclusion** says $Y_i(d,1) = Y_i(d,0) \equiv Y_i^d$, collapsing them to two: only now is $\Delta_i \equiv Y_i^1 - Y_i^0$ well defined, and only now does $Z$ reach $Y$ through $D$ alone. (**SUTVA** makes them person-$i$ quantities.)

$$Y_i = Y_i^0 + \Delta_i D_i, \quad D_i = D_i^0 + (D_i^1 - D_i^0)Z_i \implies Y_i = Y_i^0 + \Delta_i\big[D_i^0 + (D_i^1 - D_i^0)Z_i\big].$$

**Independence** Evaluate that last line at $Z_i = 1$ and at $Z_i = 0$:

$$Y_i\big|_{Z_i=1} = Y_i^0 + \Delta_i D_i^1, \qquad Y_i\big|_{Z_i=0} = Y_i^0 + \Delta_i D_i^0.$$

Independence leaves the distribution of $(Y_i^0, \Delta_i, D_i^0, D_i^1)$ unchanged by $Z_i$, so take expectations and difference:

$$\underbrace{E[Y_i \mid Z_i{=}1] - E[Y_i \mid Z_i{=}0]}_{\text{reduced form}} = E\big[\Delta_i(D_i^1 - D_i^0)\big], \qquad \underbrace{E[D_i \mid Z_i{=}1] - E[D_i \mid Z_i{=}0]}_{\text{first stage}} = E\big[D_i^1 - D_i^0\big].$$

**Monotonicity** Now $D_i^1 - D_i^0 \in \{0, 1\}$: it is the **complier indicator**, and an indicator inside an expectation is a conditional mean times a probability:

$$E\big[\Delta_i(D_i^1 - D_i^0)\big] = E[\Delta_i \mid C]\,P(C), \qquad E\big[D_i^1 - D_i^0\big] = P(C), \qquad C \equiv \{D_i^1 - D_i^0 = 1\}.$$

**Nonzero first stage** $P(C) > 0$, so divide. The $P(C)$ cancels: $\delta_{\text{Wald}} = E[\Delta_i \mid C]$. ■

# The four types

| Type | In the draft example | $D_i^0$ | $D_i^1$ |
|---|---|---|---|
| **Compliers** | serve if drafted, not otherwise | 0 | 1 |
| Always-takers | patriots: serve either way | 1 | 1 |
| Never-takers | conscientious objectors; medically exempt | 0 | 0 |
| *Defiers* | serve only if *not* drafted | 1 | 0 |

- **Only compliers contribute.** Always-takers and never-takers have $D_i^1 = D_i^0$, so they drop out of the numerator *and* the denominator of the Wald ratio on their own — no assumption required. Defiers are the only group that has to be assumed away.
- Under homogeneous effects the partition is irrelevant, because compliers have the same effect as everyone else.

# If monotonicity fails

Recall $\pi_{1i} = D_i^1 - D_i^0$. Reduced form $= E[\Delta_i \pi_{1i}]$ and first stage $= E[\pi_{1i}]$ as before. But now $\pi_{1i} \in \{-1, 0, 1\}$:

$$E[\Delta_i \pi_{1i}] = \underbrace{P(C)\, E[\Delta_i \mid C]}_{\pi_{1i}=1} \underbrace{- P(F)\, E[\Delta_i \mid F]}_{\pi_{1i}=-1} + \underbrace{0}_{\pi_{1i}=0}, \qquad E[\pi_{1i}] = P(C) - P(F),$$

$$\delta_{\text{Wald}} = \frac{P(C)\, E[\Delta_i \mid C] \;-\; P(F)\, E[\Delta_i \mid F]}{P(C) - P(F)}.$$

- The weights still sum to one, but **one of them is negative.** That is not an average of anything: it can land outside the range of every individual effect.
- **Example.** 50% compliers with $\Delta_i = 2$, 30% defiers with $\Delta_i = 6$. Then $\delta_{\text{Wald}} = (1.0 - 1.8)/(0.5 - 0.3) = -\mathbf{4}$ — a negative estimate where treatment helps *everybody*.
- **Homogeneous effects.** If $\Delta_i = \delta$ for all $i$ the numerator is $\delta\,[P(C) - P(F)]$, so the ratio is $\delta$ either way. Monotonicity is a price we pay for heterogeneity.
- Defiers are unobservable, so monotonicity is argued, not tested.

# Compliers are defined by the IV

- **Change the IV and you change the estimand.** Both IVs in this lecture target the *same* treatment (years of schooling) and the *same* outcome (log wages):

| | Angrist–Krueger (1991) | Card (1995) |
|---|---|---|
| IV | quarter of birth | 4-year college in the county |
| Compliers | teens who would quit at 16 | kids for whom living at home decided it |
| Margin priced | a year of *high school* | the *college* years |
| 2SLS | $\approx 0.09$ | 0.188 |

- The two IVs do not price the same *year* of schooling, let alone for the same people. Neither is "the" return to education.
- So "one of you is wrong" does not follow — and the **over-identification test of §7.3 can reject for this reason alone.** Two valid IVs with different compliers *should* disagree; rejection means "not the same estimand", not "invalid IV".
- **Practical advice.** Report the complier share and their observable characteristics.

# 7.6 Card (1995): College in the County

# The design

$$Y_i = \alpha + \delta S_i + \gamma X_i + \varepsilon_i,$$

$Y$ log earnings, $S$ years of schooling, and $\varepsilon$ containing **ability**. If ability raises both schooling and wages, OLS is biased up.

- **The IV:** whether the respondent grew up in a county with a **four-year college**.
- **Why it moves schooling:** proximity lowers the cost of attendance.
- **Who are the compliers?** Not those who would always attend, nor those who never would. The compliers are people for whom the cost of leaving home is the binding constraint.
- **The strangeness test:** why should the presence of a college in your childhood county affect your wages at 30? Only through schooling, one hopes. Is that believable?

# The estimates

| | Schooling |
|---|---|
| Estimated effect of schooling | 0.188 |
| | (0.026) |
| | [0.144, 0.251] |
| *First stage* | |
| College in the county instrument | 0.829 |
| | (0.107) |
| Olea and Pflueger $F$-statistic | 60.374 |
| Obs | 3,010 |

- Robust SEs in parentheses, **Anderson–Rubin** 95% interval in brackets. Computed from `card.dta`, not copied — this reproduces Remix Table 7.6.
- Living in a college county buys **0.83 of a year** of schooling, and $F \approx 60$ is comfortably strong.
- **2SLS says 18.8% per year of schooling. OLS on the same data says 5.2%.** More than three times larger — which is backwards, and is the next slide.

# Why is 2SLS *larger* than OLS?

- If the problem were **ability bias**, OLS would be biased *upward*, so 2SLS should come in **lower**.
- **Measurement error?** Classical measurement error in schooling attenuates OLS toward zero, and IV repairs it. But people know their own years of schooling fairly well!
- **Compliers genuinely have higher returns?** The compliers are people for whom the marginal cost of attending was the binding constraint. If they were **underinvesting** in education because of that constraint, the return to the schooling they were induced to get is high.
- This is a case where LATE $\neq$ ATE is the *interesting* finding: it says something about liquidity constraints in education.
- It is also unfalsifiable from the data alone. Treat it as an interpretation, and say so.

# Aside: classical measurement error attenuates OLS

We want $Y_i = \alpha + \delta D_i^* + u_i$, but we observe $D_i = D_i^* + e_i$. **Classical** error means $e_i$ is pure noise: uncorrelated with $D_i^*$ and with $u_i$. Substitute $D_i^* = D_i - e_i$:

$$Y_i = \alpha + \delta D_i + \underbrace{(u_i - \delta e_i)}_{\text{new error}}, \qquad C(D_i,\ u_i - \delta e_i) = -\,\delta\,\sigma_e^2 \neq 0.$$

$$\operatorname{plim} \hat{\delta}_{\text{OLS}} = \delta \cdot \lambda, \text{ where } \lambda = \frac{\sigma_{D^*}^2}{(\sigma_{D^*}^2 + \sigma_e^2)}$$

- **Attenuation:** the bias is toward zero.
- Error in a *control* variable is worse: it biases the coefficients on everything else, in no predictable direction.

# Aside: why IV repairs it

$Z$ must be correlated with the **true** $D^*$, and uncorrelated with $u$ *and* with the measurement error $e$. Then the noise simply cancels:

$$\frac{C(Y,Z)}{C(D,Z)} = \frac{\delta\, C(D^*,Z)}{C(D^*+e,\ Z)} = \frac{\delta\, C(D^*,Z)}{C(D^*,Z)} = \delta.$$

- **So where do you find such a $Z$? Independent measurement of the same thing.** Each measure's error is its own, so each can instrument the other.
- **Ashenfelter and Krueger (1994)**: they instrument a twin's *self-reported* schooling with the *sibling's* report of it, and the attenuation-corrected return comes out higher.
- In finance: two data vendors' versions of the same variable, a restated figure against the original, a survey measure against an administrative one.

# 7.8 Presenting Your IV Results

# Know your assignment mechanism

- Spend the time. Call people. Ask them how units ended up treated. Be your own worst critic about why you think the IV is random and why you think exclusion holds.
- **And be willing to lose.** If your IV is bad, finding that out early is a *good* outcome — a bad IV crowds out your time and creativity while taking you somewhere false.
- The empirical crisis in 1970s labour economics came largely from assuming model conditions held rather than establishing that they did. That is the mistake to avoid.

# What an IV table must contain

1. **The OLS estimate**, so the reader can see what you are correcting and in which direction.
2. **The 2SLS estimate** with standard errors clustered appropriately.
3. **The first-stage coefficient on the excluded IV**, with its standard error — usually in a bottom panel.
4. **The effective $F$-statistic** (Olea–Pflueger), not merely the classical one.
5. **Anderson–Rubin confidence intervals**, whatever the $F$.
6. **The number of observations** and, where the design implies it, the complier share.

# A checklist for an IV paper

1. **Name the shock.** What physically moved units into treatment? Who was moved, who was not, and why was that not their choice?
2. **State the exclusion restriction in words**, as a sentence about the world, and then argue it. Expect to spend a page.
3. **Run the falsification tests your mechanism implies.** Where the IV should have no effect, show it has none — Angrist and Krueger's college-graduation result is the template.
4. **Plot all three**: the IV, the first stage, the reduced form.
5. **Report the effective $F$ and Anderson–Rubin intervals**. Prefer one strong IV to many weak ones.
6. **Describe the compliers** — their share, their characteristics, and whether they are the population the question is about.
7. **Show the reduced form on its own.** If it is not there, no amount of scaling by a first stage will create it.
8. **Address monotonicity explicitly** if your IV is a decision-maker, a score, or anything with discretion behind it.
9. **State the estimand as a LATE**, and resist calling it "the" effect.

# 7.9 Concluding Remarks

# What IV buys

- It is the design in this course that addresses **selection on unobservables** head-on. Matching conditions on what you measured; RD needs a rule; IV needs only a shock.
- When the shock is real — a lottery, a draft number, a randomly assigned judge — the argument is as close to an experiment as observational data gets.
- It reframes what you are looking for. The question stops being *"what should I control for?"* and becomes *"what, in the world, moved these units around?"* That is almost always the better question.

# What it costs

- **Five assumptions, four of them untestable.** RD has one. This is the honest reason many applied researchers now avoid IV — not that it fails to identify something, but that it is hard to imagine an IV satisfying all five.
- **The estimand is a LATE** for a group you cannot name and did not choose. Its usefulness depends entirely on how close the compliers' effect is to everybody else's, which you cannot check.
- **Weak instruments are worse than they look.**
- **Exclusion is an argument, not a test.** That makes IV papers contestable in a way RD papers are not — defeating a LATE requires only creating doubt.

# References


- Andrews, I., J. Stock and L. Sun (2019). Weak Instruments in IV Regression: Theory and Practice. *Annual Review of Economics* 11, 727–753.
- Angrist, J. (1990). Lifetime Earnings and the Vietnam Era Draft Lottery. *AER* 80, 313–336.
- Angrist, J. and W. Evans (1998). Children and Their Parents' Labor Supply. *AER* 88, 450–477.
- Angrist, J. and A. Krueger (1991). Does Compulsory School Attendance Affect Schooling and Earnings? *QJE* 106, 979–1014.
- Angrist, J. and A. Krueger (2001). Instrumental Variables and the Search for Identification. *JEP* 15, 69–85.
- Ashenfelter, O. and A. Krueger (1994). Estimates of the Economic Return to Schooling from a New Sample of Twins. *AER* 84, 1157–1173.
- Bound, J., D. Jaeger and R. Baker (1995). Problems with Instrumental Variables Estimation. *JASA* 90, 443–450.
- Card, D. (1995). Using Geographic Variation in College Proximity to Estimate the Return to Schooling. In *Aspects of Labour Market Behaviour*.
- Cunningham, S. *Causal Inference: The Remix*, Ch. 7 — this lecture follows it closely.
- Heckman, J. (1990). Varieties of Selection Bias. *AER P&P* 80, 313–318.
- Imbens, G. and J. Angrist (1994). Identification and Estimation of Local Average Treatment Effects. *Econometrica* 62, 467–475.
- Lee, D., J. McCrary, M. Moreira and J. Porter (2022). Valid $t$-ratio Inference for IV. *AER* 112, 3260–3290.
- Montiel Olea, J. and C. Pflueger (2013). A Robust Test for Weak Instruments. *JBES* 31, 358–369.
- Stock, J. and M. Yogo (2005). Testing for Weak Instruments in Linear IV Regression. In *Identification and Inference for Econometric Models*.
- Wald, A. (1940). The Fitting of Straight Lines if Both Variables Are Subject to Error. *Annals of Mathematical Statistics* 11, 284–300.