# Five Common IV Designs

## BUSS975: Causal Inference in Financial Research

Professor Ji-Woong Chung

Korea University

# Where we left off

- We built the machinery: the Wald ratio, 2SLS, weak instruments, and the five LATE assumptions. This one is about **where IVs actually come from**.
- IV is a general-purpose strategy: it works wherever a credible shock turns up. But over the years certain strategies have been used so often that they have become **designs** — each with its own literature and its own conventions.
- Knowing the designs tells you where to look for an IV in your own work.

# The five

1. **Lotteries** — randomised offers with imperfect take-up.
2. **Judge / examiner fixed effects** (the "leniency design").
3. **Bartik, or shift-share, IVs.**
4. **Fuzzy regression discontinuity** — which you already know.
5. **Supply shifters for the elasticity of demand** — the original IV, from the 1920s.

- For each one, ask the same four questions:

*What is the shock?* *What is the first stage?*

*Who are the compliers?* *Where does exclusion break?*

- The last question is the one that decides whether the paper survives refereeing.

# 1. Lotteries

# The design: randomisation is not enough

- In many randomised trials, assignment is random but **take-up is voluntary**.
- Those who take it up are precisely those who expect to benefit — **positive selection**.
- So comparing the treated with the untreated is biased **even inside a randomised experiment**.

- **The fix:** instrument actual treatment with the **randomised offer**, and estimate the LATE.

# Oregon Health Insurance Experiment

- In 2008 Oregon expanded Medicaid for low-income adults. Facing excess demand, it allocated places by **lottery**: roughly 30,000 drawn from a list of 85,000.
- Of those selected, about 10,000 enrolled.
- **First stage** (Finkelstein et al. 2012): winning raised the probability of ever being on Medicaid by about **26 percentage points** — large, but nowhere near one. Hence IV rather than a simple comparison.
- **What the design gives you:** the effect of *having Medicaid* for those who enrolled **because** they won — not for those who would have found coverage anyway.

# Oregon, in equations

| | |
|---|---|
| $Y_i$ | utilisation or financial strain — e.g. $\mathbf{1}\{\text{has a usual place of care}\}$ |
| $D_i$ | $\mathbf{1}\{\text{enrolled in Medicaid}\}$ |
| $Z_i$ | $\mathbf{1}\{\text{won the lottery}\}$ |

$$\begin{aligned}
\text{First stage:} \quad & D_i = \pi_0 + \pi_1 Z_i + \nu_i, && \hat{\pi}_1 \approx 0.26 \\
\text{Reduced form:} \quad & Y_i = \gamma_0 + \gamma_1 Z_i + u_i \\
\text{Second stage:} \quad & Y_i = \alpha + \delta \hat{D}_i + e_i, && \hat{\delta} = \hat{\gamma}_1 / \hat{\pi}_1
\end{aligned}$$

- On the next slide, $\hat{\gamma}_1$ is the **ITT column** and $\hat{\delta}$ is the **LATE column**.

# What Oregon found

| Outcome | ITT (reduced form) | LATE (2SLS) |
|---|---|---|
| Has a usual place of care | 9.9% | 33.9% |
| Has a personal doctor | 8.1% | 28.0% |
| Got all needed healthcare | 6.9% | 23.9% |
| Any hospital admission | 0.5% | 2.1% |
| Had a medical collection | −1.6% | −6.4% |
| Owed on medical collection | −$99 | −$390 |

- The **ITT column is the reduced form** — the policy effect of *running the lottery* — what a legislator decides on.
- The **LATE column scales it by the first stage** — the effect of *coverage* on the outcomes.

# 2. Judge and Examiner Fixed Effects

# The recipe

The design needs three ingredients:

1. A **narrow pipeline** every unit must pass through;
2. **Randomly assigned decision-makers** staffing it;
3. **Discretion** among those decision-makers.

- The canonical setting is American courts, where defendants are randomly assigned to judges.
- **Gaudet, Harris and John (1933)** found one judge imprisoned in 34% of his cases and another in nearly 60%, with cases randomly rotated between them.
- Their conclusion: a sentence may indicate *"either the seriousness of his crime or the severity of the judge."*

# One modern example, and how far it travels

- **Stevenson (2018)**, Philadelphia bail magistrates. Randomly assigned judges differ in how affordable they set bail, so some defendants are detained pre-trial **purely by draw**.
- She finds pre-trial detention raises conviction by about **13%**, mostly through guilty pleas, and raises sentence length and court fees.

- The design has spread far beyond courts. Anywhere a queue is staffed by rotating officials with discretion:
  - patent examiners, disability examiners, loan officers, auditors, bankruptcy judges, bank supervisors.

# Leniency, in equations

| | |
|---|---|
| $Y_i$ | convicted $\{0, 1\}$ — also sentence length, court fees |
| $D_i$ | $\mathbf{1}\{\text{detained pre-trial}\}$ |
| $Z_i$ | leave-one-out detention rate of the judge who drew case $i$ |

$$\text{First stage:} \quad D_i = \pi_1 Z_i + \mathbf{c}'_{ct}\gamma_1 + \nu_i$$

$$\text{Second stage:} \quad Y_i = \delta \hat{D}_i + \mathbf{c}'_{ct}\gamma + e_i$$

- $\mathbf{c}_{ct}$ are **court** $\times$ **period fixed effects** — the cell inside which assignment is random. They are not controls; they define the experiment.
- Stevenson's $\hat{\delta}$: detention raises conviction by about 13%.

# Building the instrument

**Not the judge — the judge's *leniency*:** *of all the other cases my judge handled, what fraction did they treat?*

$$Z_i = \frac{1}{n_{j(i)} - 1} \sum_{\substack{k \neq i \\ j(k) = j(i)}} D_k \qquad \text{(leave-one-out, or } \textit{jackknife}\text{, leniency)}$$

| | |
|---|---|
| $i$ | the case you are building the instrument *for* |
| $j(\cdot)$ | the judge who drew case $i$ |
| $n_{j(i)}$ | that judge's caseload; subtract 1 because we dropped case $i$ |
| $D_k$ | whether case $k$ was treated (detained, liquidated, granted) |

- A judge who detains 70% of *everyone else* is a judge your case was unlucky to draw — and that luck is the only thing the instrument uses.

# Two choices that make it work

- **Why leave your own case out?** Keeping it in puts $D_i$ on *both* sides:

$$Z_i^{\text{keep}} = \underbrace{\frac{1}{n_{j(i)}} \sum_{\substack{k \neq i \\ j(k)=j(i)}} D_k}_{\text{the judge}} + \underbrace{\frac{1}{n_{j(i)}} D_i}_{\text{your own case}}$$

  - $Z_i^{\text{keep}}$ then inherits a slice of $D_i$, and so a slice of the structural error $\varepsilon_i$ (how serious the case is). **The instrument is no longer as good as random.**
  - The first stage looks strong for a reason that has nothing to do with the judge.
  - The slice is $1/n_{j(i)}$, so the damage grows when caseloads are short *or* when genuine leniency differences are small.
- **Why residualise first?**
  - Assignment is random *within* a court $\times$ period cell, not nationally.
  - So build leniency inside the group, and carry those fixed effects into both stages.

# Where it breaks

- **Independence** — usually fine, assignment really is random.
- **Exclusion** — the usual worry. A defendant who draws a judge known to be severe expects a worse outcome at trial, so pleads guilty — *whether or not he was detained*. That path runs from $Z$ straight to $Y$, bypassing **detention**, which is the treatment.
- **Monotonicity** — the hardest. Lenient overall, strict on drug charges, and the assumption is gone. Imbens and Angrist flagged this case in 1994: no nonparametric LATE.

# Why not just use judge dummies?

- What is random is **which judge you drew**, so 2SLS uses **one dummy per judge** as instruments:

$$D_i \;=\; \sum_{j=1}^{K} \pi_j \, \mathbf{1}\{j(i) = j\} \;+\; \nu_i \;=\; Z_i'\pi + \nu_i$$

- Here $Z_i$ is the vector of $K$ judge dummies. The first stage sets each $\hat{\pi}_j$ to judge $j$'s detention rate over **all** her cases — **yours included**: $\hat{D}_i = Z_i'\hat{\pi} = Z_i^{\text{keep}}$, the instrument we just ruled out.
- Count your case, and the judge's "strictness" moves with **your own outcome**: a serious case gets detained and makes its own judge look strict.
- Leaving it out is the fix. Applied to the dummies' first stage, it is called **JIVE**.

# How big is the bias? Look at the extremes

- **One case per judge.** The judge's rate *is* your own outcome: $\hat{D}_i = D_i$. 2SLS **is** OLS, with all of OLS's bias.
- **Thousands of cases per judge.** Your case is a vanishing slice of the rate. 2SLS is the clean IV.
- **In between**, 2SLS sits between the two, pulled **toward OLS** — harder the fewer cases each judge has:

$$\text{2SLS bias} \approx \frac{\text{OLS bias}}{F+1}$$

- $F$ asks whether judges' rates differ more than chance alone would make them. With few cases each, chance alone scatters the rates widely, so $F$ is small. Thin caseloads mean a low $F$ *and* a big slice of your own case — the same fact, twice.

# JIVE: leave yourself out

- **The fix:** Estimate the first stage **without case** $i$, then use it to predict case $i$. Same $K$ instruments — different fitted value.

$$\hat{D}_i^{\text{JIVE}} \;=\; Z_i' \hat{\pi}_{(-i)}, \qquad \hat{\pi}_{(-i)} \text{ fitted on everyone but } i.$$

- Regress $Y_i$ on $\hat{D}_i^{\text{JIVE}}$. This is **JIVE** (Angrist, Imbens and Krueger 1999).
- Nothing about your own case is used to predict your own case.

- **You do not run $n$ regressions.** Leave-one-out has a closed form, built from the **leverage** of observation $i$:

$$\hat{D}_i^{\text{JIVE}} \;=\; \frac{Z_i' \hat{\pi} - h_i D_i}{1 - h_i}, \qquad h_i = Z_i' (Z'Z)^{-1} Z_i.$$

- $h_i$ is the $i$th diagonal of the projection ("hat") matrix — how much case $i$ pulls its own fitted value. See Hansen, *Econometrics*, §3.21.

# Your leniency measure *was* JIVE

With judge dummies and nothing else, your fitted value is just your judge's average, $Z_i'\hat{\pi} = \bar{D}_{j(i)}$, and your leverage is your share of that average, $h_i = 1/n_{j(i)}$, where $n_{j(i)}$ is your judge's caseload. So

$$\hat{D}_i^{\text{JIVE}} = \frac{\bar{D}_{j(i)} - D_i/n_{j(i)}}{1 - 1/n_{j(i)}} = \frac{1}{n_{j(i)} - 1} \sum_{\substack{k \neq i \\ j(k) = j(i)}} D_k.$$

- That is **exactly** the leave-one-out leniency we built earlier.
- **Caveat:** Add controls and this breaks — next slide.

# Once you add controls, JIVE breaks too

- Assignment is random only *within* a court $\times$ period group, so we control for group fixed effects. By FWL, the instrument then becomes **your judge's rate − your group's rate**: how strict your judge is *relative to the other judges in your court*.
- JIVE leaves your case out of the first term only. The group's rate is still computed from every case in the group, yours included.
- So your own case walks back in through the second term. The bias returns.

- **UJIVE** (Kolesár 2013) leaves case $i$ out of *both* pieces: your judge's leave-out rate − your group's leave-out rate.
- Without group controls the difference is negligible. With them, UJIVE is the safe default.

# Two more things worth knowing

- **Do not cluster by judge.** It is common, and it is wrong. What matters is how cases were *assigned*.
  - Cases assigned to judges one at a time, independently ⇒ plain **robust** standard errors.
  - Cases assigned in batches — one duty judge takes every case filed that day ⇒ cluster **by day**, the unit that was actually drawn.
- **With UJIVE you need not chase a large $F$.** It stays roughly unbiased even when $F$ is close to 1. That is the point of using it.

- **Read:** Goldsmith-Pinkham, Hull and Kolesár (2026), *Leniency Designs: An Operator's Manual*, *JEP* 40(3). A step-by-step guide — including a weaker, more believable version of monotonicity and a test for it — with an `R` package (`ManyIV`).

# Stevenson's numbers (Remix, Table 7.12)

| | OLS | OLS | 2SLS | 2SLS | JIVE | JIVE |
|---|---|---|---|---|---|---|
| Detention | −0.001 | 0.029 | 0.151 | 0.186 | 0.162 | 0.212 |
| | (0.002) | (0.002) | (0.065) | (0.064) | (0.070) | (0.076) |

$N = 331{,}971$, eight bail judges. Outcome: guilty plea. Odd columns use time controls only; even columns add defendant and offence characteristics.

- **OLS says detention does nothing.** IV says it raises guilty pleas by 15–21 percentage points.
- But Stevenson has only **eight** judges. To see what happens with *many*, we need a bigger pipeline.

# A bigger pipeline: patent examiners

Goldsmith-Pinkham, Hull and Kolesár (2026) re-run Farre-Mensa, Hegde and Ljungqvist (2020): does getting your **first patent approved** help a startup innovate again?

- **Treatment** $D_i$: the application was **approved**.
- **Instrument** $Z_i$: **one dummy per examiner** — examiners differ in how freely they grant.
- **Outcome** $Y_i$: the startup's later innovation
- **Controls** $W_i$: **art unit** $\times$ year.
  - An *art unit* is the small group of examiners who specialise in one technology; an application is assigned quasi-randomly to an examiner **within** it. It is the patent office's court-and-period group — the place where the lottery actually happens.

# The standard errors lie too (GPHK 2026)

| Effect of approval on | UJIVE | 2SLS (examiner dummies) | OLS |
|---|---|---|---|
| Any later application | 0.173 | 0.232 | 0.234 |
| | (0.055) | (0.016) | (0.006) |
| Any later approval | 0.259 | 0.240 | 0.223 |
| | (0.050) | (0.014) | (0.005) |
| $\log(1 + \text{citations})$ | 0.419 | 0.372 | 0.339 |
| | (0.125) | (0.033) | (0.011) |

- **2SLS has landed on OLS** — That is the bias reproduced.
- **And look at the standard errors.** 2SLS reports errors three to four times *smaller* than UJIVE.

# 3. Bartik / Shift-Share IVs

# The question: what does immigration do to local wages?

- **The question.** When immigrants arrive in a city, what happens to the wages of the **native** workers already there?

$$Y_{l,t} = \alpha + \delta D_{l,t} + \rho X_{l,t} + \varepsilon_{l,t}$$

- $Y_{l,t}$: log native wages in city $l$ in year $t$. $D_{l,t}$: immigrant inflow. $X_{l,t}$: controls, including city and year fixed effects.
- **Why OLS cannot answer it.** Immigrants are not sprinkled over cities at random. They go where the work is.
- **What we need:** something that sends immigrants to a city for reasons that have nothing to do with *that city's* labour market.

# The idea: you go where your people already are

- A new arrival from Mexico does not choose a city off a map. They go where there is already a Mexican community.
- So a **national** surge in migration from one country lands disproportionately in the cities that already held people from that country.
- $l$ = a **city**, $k$ = a **source country**, $t$ = the year, $t_0$ = a base year *before* the flows we study.

$$\underbrace{B_{l,t}}_{\text{predicted inflow}} = \sum_{k=1}^{K} \underbrace{z_{l,k,t_0}}_{\substack{\textbf{shares} \\ \text{who was already there}}} \times \underbrace{m_{k,t}}_{\substack{\textbf{shifts} \\ \text{who left home}}}$$

- Los Angeles chose neither the surge out of Mexico nor, in anticipation of it, its 1980 immigrants. The surge was decided in Mexico.

# Shift-share, in equations

| | |
|---|---|
| $Y_{l,t}$ | log wages of *native* workers in city $l$ |
| $D_{l,t}$ | immigrant inflow into city $l$ — the Remix writes this $I_{l,t}$ |
| $Z_{l,t}$ | $B_{l,t} = \sum_k z_{l,k,t_0}\, m_{k,t}$ — inflow predicted by the enclave map |
| $X_{l,t}$ | controls: city and year fixed effects |

$$\text{First stage:} \quad D_{l,t} = \pi_0 + \pi_1 B_{l,t} + \rho_1 X_{l,t} + \nu_{l,t}$$

$$\text{Second stage:} \quad Y_{l,t} = \alpha + \delta \hat{D}_{l,t} + \rho X_{l,t} + e_{l,t}$$

- First stage: do the cities whose enclaves *should* have drawn immigrants actually get them?
- $\delta$ is the effect of immigration on native wages — an average across source countries, weighted towards whichever ones actually move the instrument.

# Identification

**Bartik IV requires exogeneity of either shares OR shifts**.

## Framework A: Shares Exogenous

**Goldsmith-Pinkham, Sorkin, Swift (GPSS, 2020)**

- Identifies via baseline shares $z_{ik}$.
- Equivalent to using shares as individual instruments.
- Shifts $m_{k,t}$ act merely as a weighting matrix.

## Framework B: Shifts Exogenous

**Borusyak, Hull, Jaravel (BHJ, 2022)**

- Identifies via sectoral shocks $m_{kt}$.
- Shares $z_{ik}$ can be completely endogenous.
- Requires many independent shocks ($K \to \infty$).

# How identification fails: Framework A (exogenous shares)

- **Path dependence**
  - Instrument: 1990 manufacturing shares $\rightarrow$ local labor demand.
  - Problem: Rust Belt cities with high manufacturing shares were already declining, so the shares correlate with $\varepsilon$
- **Agglomeration and amenities**
  - Instrument: tech employment share $\rightarrow$ local labor demand.
  - Problem: tech-heavy cities have educated workers, costly housing and VC that move $Y$ directly, so the shares correlate with $\varepsilon$

# How identification fails: Framework B (exogenous shocks)

- **Common macro shocks**
  - National auto growth spikes because interest rates were cut.
  - Problem: a rate cut raises national auto growth (the shift), and auto-heavy cities' spending also responds to the cut directly, so the shift correlates with those cities' $\varepsilon$
- **Few dominant shocks**
  - 2 of 50 industries (e.g., oil & gas, healthcare) drive most of the shock variation.
  - Problem: BHJ needs many independent shocks so unobservables average out. With one or two dominant sectors, any confounder in those sectors biases the estimate.

# Choose your framework first

- **State your framework:** exogenous shares (GPSS) or exogenous shocks (BHJ).
- The framework determines the diagnostics:
  - GPSS: Rotemberg weights; balance and pre-trends for the top-weighted industries.
  - BHJ: shock-level balance; effective number of shocks; exposure-robust SEs.

- **Read:** Breuer (2022), *Bartik Instruments: An Applied Introduction*, *JFR* 7(1). Written for accounting and finance.
- **Read:** Borusyak, Hull and Jaravel (2025), *A Practical Guide to Shift-Share Instruments*, *JEP* 39(1).

# 4. Fuzzy Regression Discontinuity

# Lecture 4, revisited

- In a **sharp** design, $D_i = \mathbf{1}\{X_i \geq c_0\}$ exactly. In a **fuzzy** design, crossing the cutoff raises the probability of treatment without forcing it:

$$\lim_{x \downarrow c_0} \Pr(D_i = 1 \mid X_i = x) \neq \lim_{x \uparrow c_0} \Pr(D_i = 1 \mid X_i = x).$$

- The estimator is a **Wald ratio** of two discontinuities:

$$\delta_{\text{FRD}} = \frac{\lim_{x \downarrow c_0} E[Y|X = x] - \lim_{x \uparrow c_0} E[Y|X = x]}{\lim_{x \downarrow c_0} E[D|X = x] - \lim_{x \uparrow c_0} E[D|X = x]}.$$

- A ratio of two discontinuities is a Wald estimator. So a fuzzy RD **is** an IV design.

# Fuzzy RD is an IV design

- Fuzzy RD **is** IV, with the instrument $Z_i = \mathbf{1}\{X_i \geq c_0\}$:
  - Numerator = reduced-form jump in $Y$; denominator = first-stage jump in $D$.
  - Estimate by 2SLS within a bandwidth, instrumenting $D$ with $Z$ and controlling for a local polynomial in $X$ on each side.
  - Same IV assumptions, holding locally at $c_0$: crossing the cutoff affects $Y$ only through $D$ (exclusion), and no one is pushed *out* of treatment by crossing it (monotonicity). Relevance. SUTVA.
  - So $\delta_{\text{FRD}}$ is a LATE for **compliers at the cutoff**: units whose treatment is switched by crossing $c_0$.

# Fuzzy RD, in equations

| | |
|---|---|
| $Y_i$ | the outcome |
| $D_i$ | treatment actually **received** — not assignment |
| $Z_i$ | $\mathbf{1}\{X_i \geq c_0\}$ — crossing the cutoff |

$$\text{First stage:} \quad D_i = \pi_0 + \pi_1 Z_i + f(X_i - c_0) + \nu_i$$

$$\text{Second stage:} \quad Y_i = \alpha + \delta \hat{D}_i + g(X_i - c_0) + e_i$$

- $f$ and $g$ are the running-variable polynomials, fitted **separately on each side** and only inside the bandwidth, with Lecture 4's kernel weights.
- $\delta$ is a LATE *at the cutoff*, for units whose treatment flips there — doubly local.

# 5. The Elasticity of Demand

# The Identification Problem

- **The Target:** A demand curve is a causal object—the effect of price ($P$) on quantity demanded ($Q$).
- **The Obstacle:** Observed $(P, Q)$ pairs are **equilibrium points** generated by simultaneous shifts in both supply and demand.
- Wright's Solution:
  - **History:** Instrumental variables were invented for this exact problem by **Philip Wright (1928)**.
  - **The Logic:** To trace out a fixed demand curve, you need exogenous variation that shifts **supply only**.

# The question: how price-sensitive are fish buyers?

- **The question.** What is the **price elasticity of demand** for fish? **Kathryn Graddy** set out to measure it at the **Fulton Fish Market**. **She hand-collected the data**, driving in from New Jersey before dawn each morning.
- To hold the good fixed, she narrowed it to one product: **fresh whiting of roughly the same size.**

- **The obstacle is Wright's.** Each trading day hands you a single $(P, Q)$ pair, set by supply and demand *at once*. Regressing $\log Q$ on $\log P$ traces neither curve.
- **The instrument:** the weather at sea in the days *before* the fish reached the market. (Graddy 1995, 2006)

# Fulton Fish, in equations

| | |
|---|---|
| $Y_t$ | $\log Q_t$ — log pounds of whiting sold on day $t$ |
| $D_t$ | $\log P_t$ — log average price that day |
| $Z_t$ | $\texttt{Stormy}_t$ — bad weather at sea over the previous two days |

$$\text{First stage:} \quad \log P_t = \pi_0 + \pi_1 \texttt{Stormy}_t + \mathbf{d}_t'\gamma_1 + \nu_t, \qquad \hat{\pi}_1 \approx 0.35$$

$$\text{Second stage:} \quad \log Q_t = \alpha + \delta \widehat{\log P_t} + \mathbf{d}_t'\gamma + e_t, \qquad \hat{\delta} \approx -1.12$$

- $\mathbf{d}_t$ are day-of-week dummies. The unit is a **day**, not a buyer.
- Both variables are in **logs**, so $\delta$ *is* the elasticity — no conversion required.

# Fulton Fish Market (Remix, Table 7.14)

| Dependent variable: log $Q$ | OLS | 2SLS |
|---|---|---|
| log(Price) | $-0.563^{***}$ | $-1.119^{**}$ |
| | (0.152) | (0.443) |
| | | $[-2.273,\ -0.312]$ |
| *First stage* | | |
| `Stormy` | | $0.346^{***}$ |
| | | (0.072) |
| Effective $F$ | | 22.929 |
| $N$ | 111 | 111 |

**Graddy (2006)**, 111 days at the Fulton Fish Market. $Q$ = log pounds of whiting sold, $P$ = log average daily price. `Stormy` records bad weather at sea over the previous two days. Robust standard errors in parentheses, Anderson–Rubin 95% interval in brackets. All models control for day-of-week dummies. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

# Reading the fish

- Stormy weather raises price by about **35%**. That is a **supply** shock, not a demand shock — storms make fish harder to catch, they do not make New Yorkers hungrier.
- **OLS says the elasticity is** $-0.56$**. IV says** $-1.12$ — twice as elastic.

- **Why does OLS understate it?** Because the observed price–quantity pairs are *equilibria*. Both curves move, so the scatter traces out neither one.

# Even here, it is only a LATE

- The IV elasticity of $-1.12$ says: **for compliers**, a 10% storm-induced price rise cuts quantity by about 11%.
- Now suppose a seller wants to use this number to set a profit-maximising price. Demand looks elastic, so cut prices and sell more — right?
- **No.** That elasticity was identified by *storms*. When the firm cuts its price deliberately, it is not using a storm.
- Buyers read a storm-driven price rise as **temporary** and a posted price cut as **permanent**. Different shock, different response, different elasticity.

# What a LATE is, and is not

- This is the sobering version of LATE:

*It does not tell you what you want to know.*
*It tells you what you can learn, given your instrument.*

- The warning applies to **every** design — the lottery, the judge, the national shock, the threshold, the weather.
- Before reporting a LATE, ask one question: *would the policy I have in mind move the same people, in the same way, as my instrument did?*
- If the answer is no, you have a clean estimate of something other than what you were asked.

# What the five have in common

| Design | The shock | Typical first stage | Hardest assumption |
|---|---|---|---|
| Lotteries | randomised offer | strong | exclusion (usually fine) |
| Judge FE | assigned decision-maker | many weak instruments | **monotonicity** |
| Bartik | national shock $\times$ local share | strong | exclusion (shares or shifts) |
| Fuzzy RD | crossing the cutoff | jump in $\Pr(D = 1)$ | exclusion at the cutoff |
| Elasticity | supply shifter | price response | exclusion (demand channel) |

1. **Instruments are physical events.** A lottery draw, a shift schedule, a national policy, a storm. If you cannot describe the physical event that assigns treatment in the real world, you do not have an IV.
2. **Estimands are Local (LATE).** An IV only identifies the treatment effect for the **compliers**—the specific people moved by that exact shock.

- Always name the compliers in plain English before interpreting your estimate.

# IV in Finance

# Where the shocks are

Finance has fewer clean lotteries than labour economics. The recurring sources are:

- **Regulatory and legal changes** that bind some firms and not others — staggered state laws, index or listing rules, supervisory thresholds.
- **Weather, geography and distance** — historically popular, increasingly contested, because geography correlates with almost everything.
- **Shift-share exposure** — a firm's or region's pre-existing portfolio of industries, suppliers or customers, interacted with aggregate shocks.
- **Randomly assigned decision-makers** — bankruptcy judges, bank examiners, auditors, analysts, patent examiners.
- **Lotteries and quasi-lotteries** — oversubscribed allocations, IPO allocation rules, random audits.

- In finance the agents are **sophisticated, well advised and highly motivated**, so the shock you think is exogenous is often anticipated, lobbied for, or selected into.

# Three questions to ask of any finance IV

1. **Is the shock really a shock to someone?** Name the units that were moved and the units that were not, and say why the difference was not their choice. If you cannot describe the assignment mechanism in a sentence a practitioner would recognise, you do not have one.
2. **Is there exactly one channel?** This is where most finance IVs fail. A legal change that alters takeover exposure also alters managerial incentives, disclosure, and the composition of shareholders. Distance to a city affects financing *and* labour markets, customers and information.
3. **Who are the compliers, and do you care about them?** A regulation that binds only small, constrained firms identifies an effect for small, constrained firms — which may be the opposite of the effect for the firms the policy debate is about.

# The most useful habit

- Write the exclusion restriction down as a **sentence, in words** — not as an equation, and not as a citation.
- Show it to someone who knows the **institution** and not the econometrics.

- If they can name a second channel in under a minute, you have your answer.
- If you cannot write the sentence at all, you did not have an instrument — you had a variable that happened to be correlated with your treatment.

- Read: Jiang, W. (2017). Have Instrumental Variables Brought Us Closer to the Truth? *Review of Corporate Finance Studies* 6, 127–140.

# References


- Angrist, J., K. Graddy and G. Imbens (2000). The Interpretation of IV Estimators in Simultaneous Equations Models. *ReStud* 67, 499–527.
- Angrist, J., G. Imbens and A. Krueger (1999). Jackknife Instrumental Variables Estimation. *Journal of Applied Econometrics* 14, 57–67.
- Baicker, K. et al. (2013). The Oregon Experiment — Effects of Medicaid on Clinical Outcomes. *NEJM* 368, 1713–1722.
- Bartik, T. (1991). *Who Benefits from State and Local Economic Development Policies?* Upjohn Institute.
- Borusyak, K., P. Hull and X. Jaravel (2022). Quasi-Experimental Shift-Share Research Designs. *ReStud* 89, 181–213.
- Borusyak, K., P. Hull and X. Jaravel (2025). A Practical Guide to Shift-Share Instruments. *JEP* 39(1), 181–204.
- Breuer, M. (2022). Bartik Instruments: An Applied Introduction. *Journal of Financial Reporting* 7(1), 49–67. — **read this before building a shift-share instrument.**
- Cunningham, S. *Causal Inference: The Remix*, §7.7 — this deck follows it closely.
- Farre-Mensa, J., D. Hegde and A. Ljungqvist (2020). What Is a Patent Worth? Evidence from the U.S. Patent "Lottery". *Journal of Finance* 75(2), 639–682.
- Finkelstein, A. et al. (2012). The Oregon Health Insurance Experiment: Evidence from the First Year. *QJE* 127, 1057–1106.
- Gaudet, F., G. Harris and C. St. John (1933). Individual Differences in the Sentencing Tendencies of Judges. *Journal of Criminal Law and Criminology* 23, 811–818.
- Goldsmith-Pinkham, P., P. Hull and M. Kolesár (2026). Leniency Designs: An Operator's Manual. *JEP* 40(3), 213–240. — **read this before running a leniency design.**
- Goldsmith-Pinkham, P., I. Sorkin and H. Swift (2020). Bartik Instruments: What, When, Why, and How. *AER* 110, 2586–2624.
- Graddy, K. (1995). Testing for Imperfect Competition at the Fulton Fish Market. *RAND Journal of Economics* 26(1), 75–92.
- Graddy, K. (2006). Markets: The Fulton Fish Market. *JEP* 20, 207–220.
- Hansen, B. (2022). *Econometrics*. Princeton University Press. (§3.21, leave-one-out and leverage.)
- Imbens, G. and J. Angrist (1994). Identification and Estimation of Local Average Treatment Effects. *Econometrica* 62, 467–475.
- Kolesár, M. (2013). Estimation in an Instrumental Variables Model with Treatment Effect Heterogeneity. Working paper. (UJIVE.)
- Jiang, W. (2017). Have Instrumental Variables Brought Us Closer to the Truth? *Review of Corporate Finance Studies* 6, 127–140.
- Stevenson, M. (2018). Distortion of Justice: How the Inability to Pay Bail Affects Case Outcomes. *Journal of Law, Economics, and Organization* 34, 511–542.