Causality and Potential Outcomes

BUSS975: Causal Inference in Financial Research

Professor Ji-Woong Chung

Korea University

Outline

1. What causality is — and isn't
2. The potential outcomes framework
3. Selection bias: the Perfect Regulator
4. Randomization and independence
5. SUTVA
6. The regression bridge
7. Randomization inference

1. What Causality Is — and Isn't

Motivation

- As researchers, we want to make **causal** statements:
  - What is the effect of a corporate tax change on firms' leverage?
  - What is the effect of CEO equity ownership on risk-taking?
  - What is the effect of a disclosure mandate on the cost of capital?
- We are rarely satisfied with saying variables are merely "associated."
- But nearly all of our data is **observational**: nobody randomized leverage, ownership, or regulation for us.
- This course is about when — and how — observational data can answer causal questions.

Three goals of empirical analysis

1. **Description** — how things are or were.
   - Has industry concentration increased over time?
2. **Prediction** — guessing an unknown value, without intervening.
   - What will this firm's earnings be next quarter?
3. **Causality** — how *changing* one variable would change another, all else equal.
   - How would a disclosure mandate change firms' cost of capital?

- All three are legitimate. But they require different tools, and confusing them is the original sin of empirical work.

Fallacy 1: prediction is not causation

- Momentum: past returns *predict* future returns.
  - Does buying past winners *cause* them to keep winning? No — the predictive relationship says nothing about intervention.
- A model can predict superbly while being causally empty.
  - This is the promise (and the limit) of much of machine learning.
- **Causal inference asks a different question**: what would happen if we *changed* something?
  - Prediction asks what Y **is**, among the units where $X = x$.
  - Causation asks what Y **would be**, if we set $X = x$.

Fallacy 2: temporal ordering is not causation

- *Post hoc ergo propter hoc*: "after this, therefore because of this."
- The rooster crows, then the sun rises.
- Finance version: a struggling firm fires its CEO, then performance recovers.
  - Did the new CEO cause the recovery?
  - Or is this **mean reversion** — firms replace CEOs at the trough?
- Timing evidence is useful (we will use it in event studies and DiD), but timing alone proves nothing.

Fallacy 3: no correlation does not mean no causation

- A sailor moves the rudder constantly, yet the boat sails straight.
  - Correlation between rudder and course: ≈ 0. Causal effect: enormous.
  - Why? The rudder *responds optimally* to the wind.
- Finance is full of optimizing agents doing exactly this:
  - Firms hedge **because** they are exposed: hedging may show zero correlation with distress even if it strongly prevents distress.
  - The Fed adjusts rates to offset shocks: policy can look uncorrelated with output.
- **Lesson**: when treatment is chosen optimally, causal effects can be *invisible* in correlations.

The common thread: the assignment mechanism

- In all three fallacies, what breaks the link between correlation and causation is **how treatment got assigned**:
  - Who chose to be treated? Why? Based on what?
- The central message of this lecture (and this course):

Correlations are properties of the data.
*Causal effects are properties of the data **plus** the assignment mechanism.*

- To formalize this, we need a language for "what would have happened otherwise."

Defining causality: counterfactuals

- Our (narrow) definition: *ceteris paribus*, how does D affect Y?
- To know the effect of D on unit i, we need to compare:
  - Y_i **with** the treatment, and
  - Y_i **without** the treatment,
  - with everything else the same.
- One of these is always hypothetical — a **counterfactual**.
- **The Fundamental Problem of Causal Inference** (Holland 1986): we can never observe both states for the same unit at the same time.

2. The Potential Outcomes Framework

States of the world

- Imagine each unit exists in two potential states of the world, identical in every respect except one: treatment.
- Binary treatment indicator:

$$D_i = \begin{cases} 1 & \text{if unit } i \text{ is treated} \\ 0 & \text{otherwise} \end{cases}$$

- Two **potential outcomes** for every unit:
  - Y_i^1: the outcome unit i *would* have if treated,
  - Y_i^0: the outcome unit i *would* have if untreated.
- Both are defined for every unit — treated or not. That is the trick.

The switching equation

- What we *observe* is only one of the two:

$$Y_i = D_i \, Y_i^1 + (1 - D_i) \, Y_i^0$$

- Treatment "switches on" which potential outcome is realized.
- Notice what this equation does **not** contain: no error term, no functional form, no linearity. It is a definition, not a model.

Individual treatment effects

- The causal effect of the treatment on unit i:

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

- δ_i carries an i subscript.
- Effects are allowed to **differ across units** from the start.
- The Fundamental Problem, restated: δ_i is never observed for any unit, because one of its two ingredients is always missing.
  - Causal inference is a **missing data problem** — but the missing value never existed to begin with.

A running example: capital injections

- Setting: a banking crisis. The government can inject capital into banks (think TARP, 2008).
- Outcome Y: bank lending growth (%) over the following year.
- Y_i^1: bank i's lending growth *with* an injection.
- Y_i^0: bank i's lending growth *without* one.
- Question: what is the causal effect of capital injections on lending?

Ten banks, two potential outcomes

Bank	Y^1	Y^0	δ
1	9	2	7
2	6	1	5
3	7	3	4
4	5	2	3
5	6	4	2
6	7	6	1
7	7	7	0
8	7	8	-1
9	7	9	-2
10	9	8	1

- Weak banks (low Y^0) gain a lot; healthy banks gain little — or are even hurt.

Average treatment effects

- **Average Treatment Effect (ATE)**:

$$ATE = E[\delta_i] = E[Y_i^1] - E[Y_i^0]$$

- In our table: $ATE = \frac{7+5+4+3+2+1+0-1-2+1}{10} = 2.0$
- On average, injections raise lending growth by 2 percentage points.
- But note the heterogeneity: effects range from $+7$ to -2.
  - An average can hide winners and losers.

ATT and ATU: different questions, different answers

- **Average Treatment effect on the Treated**:

$$ATT = E[\delta_i \mid D_i = 1]$$

- **Average Treatment effect on the Untreated**:

$$ATU = E[\delta_i \mid D_i = 0]$$

- These are **different policy questions**:
  - ATT: "Did TARP help the banks that received it?"
  - ATE: "Would injections help a randomly chosen bank?"
  - ATU: "Would extending the program to the remaining banks help them?"
- When effects are heterogeneous and assignment is not random, $ATT \neq ATE \neq ATU$ — and each estimator you will meet in this course targets a *particular* one of these.

What we actually observe

Bank	D	Y^1	Y^0	δ
1	1	9	?	?
2	1	6	?	?
3	1	7	?	?
4	1	5	?	?
5	1	6	?	?
6	0	?	6	?
7	0	?	7	?
8	0	?	8	?
9	0	?	9	?
10	0	?	8	?

- Half of the table is missing — forever. Everything we do from here on is a strategy for filling in the question marks *on average*.

3. Selection Bias: The Perfect Regulator

The naive comparison

- With observational data, the natural move is to compare treated and untreated:

$$SDO = E[Y_i^1 \mid D_i = 1] - E[Y_i^0 \mid D_i = 0]$$

- **Simple Difference in Outcomes** (SDO): mean outcome of the treated minus mean outcome of the untreated.
- Every cross-sectional regression of Y on D is, at heart, an SDO.
- Question: when does SDO recover the ATE?

The Perfect Regulator

- Suppose the regulator has perfect foresight: it knows every bank's δ_i and injects capital into the five banks that **gain the most** (banks 1–5).
- This is not a villain — it is an *optimizing* regulator, allocating scarce capital where it helps most.
  - Exactly what a good policymaker (or a good CEO, or a good doctor) should do.
- Treated: banks 1–5. Untreated: banks 6–10.
- Now compute the SDO from the observable data.

The Perfect Regulator's data

- Mean outcome among treated: $E[Y^1 \mid D = 1] = \frac{9+6+7+5+6}{5} = 6.6$
- Mean outcome among untreated: $E[Y^0 \mid D = 0] = \frac{6+7+8+9+8}{5} = 7.6$

$$SDO = 6.6 - 7.6 = -1.0$$

- The truth: $ATE = +2.0$, and $ATT = \frac{7+5+4+3+2}{5} = +4.2$.
- The naive comparison gets the **sign wrong**: injections look *harmful* precisely because they were allocated *well*.
- This is the TARP debate in one table: "bailed-out banks lent less than healthy banks" is an SDO, not a causal effect.

The decomposition

- The SDO can always be decomposed — this is an identity, not an assumption (π = share of units treated):

$$\underbrace{E[Y^1 \mid D{=}1] - E[Y^0 \mid D{=}0]}_{\text{SDO}} = ATE + \underbrace{E[Y^0 \mid D{=}1] - E[Y^0 \mid D{=}0]}_{\text{selection bias}} + \underbrace{(1-\pi)\,(ATT - ATU)}_{\text{heterogeneous TE bias}}$$

- **Selection bias**: treated and untreated differ in their *untreated* outcomes.
- **Heterogeneous treatment effect bias**: the treated gain differently than the untreated would.
- Memorize this equation. Every research design in this course is an attempt to kill these two terms.

Proving the decomposition

Let $\pi = P(D{=}1)$.

Step 1. Add and subtract $E[Y^0 \mid D{=}1]$ — the *counterfactual* mean for the treated:

$$\begin{aligned} SDO &= E[Y^1 \mid D{=}1] - E[Y^0 \mid D{=}0] \\ &= \underbrace{E[Y^1 \mid D{=}1] - E[Y^0 \mid D{=}1]}_{ATT} + \underbrace{E[Y^0 \mid D{=}1] - E[Y^0 \mid D{=}0]}_{\text{selection bias}} \end{aligned}$$

Step 2. By the law of total expectation, $ATE = \pi\, ATT + (1-\pi)\, ATU$. Subtract ATE from ATT:

$$ATT - ATE \;=\; (1-\pi)\, ATT - (1-\pi)\, ATU \;=\; (1-\pi)(ATT - ATU)$$

Substituting $ATT = ATE + (1-\pi)(ATT - ATU)$ into Step 1 gives the three-term decomposition. ■

The decomposition, in our example

- Selection bias:

$$E[Y^0 \mid D{=}1] - E[Y^0 \mid D{=}0] = \frac{2+1+3+2+4}{5} - \frac{6+7+8+9+8}{5} = 2.4 - 7.6 = -5.2$$

  - Treated banks were *much weaker* to begin with.
- Heterogeneous TE bias, with $\pi = 0.5$:

$$(1 - 0.5)(ATT - ATU) = 0.5 \times (4.2 - (-0.2)) = +2.2$$

- Check: $2.0 - 5.2 + 2.2 = -1.0 = SDO$ ✓
- Both bias terms are large; they partially offset, and the net result is a sign flip.

Selection on gains is endemic in finance

- The Perfect Doctor (Cunningham's version) treats patients who benefit most. Finance is full of Perfect Doctors:
  - Regulators bail out the banks where intervention matters most.
  - Boards grant equity to the CEOs they expect to respond.
  - Firms go public when going public helps them most.
  - Banks lend to the borrowers they screen as creditworthy.
- Agents **optimize**. Optimization means treatment is assigned on expected gains — the worst case for naive comparisons.
- This is why finance cannot simply borrow "big data" logic: more observations do not fix a biased assignment mechanism.

Simulation: triage vs. coin flip

```r
set.seed(975)
n    <- 100000
y0   <- rnorm(n, 7, 2)      # lending growth, no injection
gain <- rnorm(n, 2, 1)      # true individual gain (so ATE = 2)
y1   <- y0 + gain

D1 <- as.integer(y0 < quantile(y0, 0.30))  # triage: weakest 30%
D2 <- rbinom(n, 1, 0.30)                   # coin flip: random 30%

sdo <- function(D) mean(y1[D == 1]) - mean(y0[D == 0])
round(c(ATE = mean(gain), triage = sdo(D1), coin = sdo(D2)), 2)
```

```
   ATE triage   coin
  2.00  -1.31   1.98
```

What the simulation tells us

- The true effect is $+2$ in both worlds — same banks, same gains, only the **assignment rule** differs.
- Triage (treat the weakest): SDO is strongly *negative*. An observer concludes bailouts destroy lending.
- Coin flip: SDO $\approx$ ATE. Same data-generating process, correct answer.
- With $n = 100{,}000$: the bias does not shrink with sample size. **Selection bias is not a small-sample problem.**

4. Randomization and Independence

The independence assumption

- What made the coin flip work? Treatment assignment was **independent of potential outcomes**:

$$(Y^1, Y^0) \perp\!\!\!\perp D$$

- In words: knowing a bank's treatment status tells you *nothing* about its potential outcomes — neither its baseline Y^0 nor its gain δ.
- Under independence:
  - $E[Y^0 \mid D{=}1] = E[Y^0 \mid D{=}0] \Rightarrow$ **selection bias** $= 0$
  - $E[\delta \mid D{=}1] = E[\delta \mid D{=}0] \Rightarrow ATT = ATU = ATE \Rightarrow$ **heterogeneity bias** $= 0$
- Both terms die, and $SDO = ATE$. Randomization is the assignment mechanism that *guarantees* independence.

What randomization balances

- Physical randomization balances treated and control groups on:
  - observable characteristics (size, leverage, ratings, . . .),
  - **unobservable** characteristics (management quality, risk appetite, . . .),
  - potential outcomes themselves, and
  - treatment gains δ_i.
- No control variables needed. No model needed. The balance is created by the *mechanism*, not by the econometrician.
- Practice: in any experiment (or natural experiment), the first table is a **balance table** — compare pre-treatment covariates across groups. Balance on observables is the testable *symptom* of a sound mechanism.

Case: eBay and the value of paid search

- **The setting.** In paid search advertising, a firm bids so that its ad appears when a user searches a keyword. eBay spent heavily on this.
- Two kinds of keyword:
  - **Brand** keywords contain the firm's own name — "eBay typewriter."
  - **Non-brand** keywords do not — "used typewriter."
- **How the industry measured returns.** *Attribution*: if a user clicks the ad and then buys, credit the sale to the ad. By this arithmetic, paid search looked extraordinarily profitable.
- **The problem.** Somebody typing "eBay typewriter" was already going to eBay. Take the ad away and they click the organic eBay link one line below.
- Attribution **conditions on clicking**. It never intervenes. It is an SDO, not an *ATE*.

eBay: what the experiments found

- Blake, Nosko, and Tadelis (2015, *Econometrica*) ran large-scale field experiments at eBay, switching paid search off in randomly selected markets.
- **Brand keywords**: **no measurable short-run benefit.** Traffic substituted almost entirely to the free organic link. eBay was paying for clicks it would have received anyway.
- **Non-brand keywords**: ads *did* work — but only for **new and infrequent** users.
  - Frequent users were unaffected, yet accounted for most of the advertising spend.
  - Net of costs, **average returns were negative**.
- Experimental returns were "a fraction of conventional non-experimental estimates."
- **Two lessons.** Attribution measures selection, not causation. And the average effect concealed enormous heterogeneity — precisely the variation in δ_i we defined earlier.

A genuine finance RCT: Karlan and Zinman (2009)

- "*Observing Unobservables: Identifying Information Asymmetries with a Consumer Credit Field Experiment,*" **Econometrica** 77.
- **Setting.** A major South African consumer lender mailed loan offers to **58,000** former clients.
- **Three independent randomizations:**
  1. The **offer rate** printed on the direct-mail solicitation.
  2. A **contract rate**, revealed only *after* the client accepted the offer — sometimes a surprise discount.
  3. A **dynamic repayment incentive**, also a surprise: preferential pricing on future loans for borrowers who stayed in good standing.
- The surprises carry the design. Anything revealed *after* acceptance cannot have influenced **who** accepted.

Karlan and Zinman: separating two unobservables

- Credit markets have two classic private-information problems, and observational data cannot tell them apart:
  - **Adverse selection** (hidden *type*): high rates attract intrinsically worse borrowers.
  - **Moral hazard** (hidden *action*): high rates weaken the incentive to repay.
- Both predict the same correlation — higher rate, more default. No amount of loan data separates them.
- The experiment does:
  - Vary the **offer** rate $\Rightarrow$ changes *who selects in* $\Rightarrow$ identifies **selection**.
  - Hold the offer rate fixed and vary the **contract** rate $\Rightarrow$ same borrowers, different incentives $\Rightarrow$ identifies **moral hazard**.
- **Findings**: strong evidence of moral hazard, weaker evidence of hidden information; roughly **13–21%** of default attributable to moral hazard.
- Randomization used not to estimate one number, but to **decompose a mechanism**.

Other field experiments in finance

- **Cole, Giné, and Vickery (2017, *RFS*)** — rainfall insurance in India. Randomly insured farmers shifted into **higher-return, rainfall-sensitive cash crops**, *before* rainfall was realized.
- **Bertrand et al. (2010, *QJE*)** — direct-mail loan offers in South Africa, randomizing price *and* content. A **photo of an attractive woman** raised demand as much as a **25% rate cut**.
- **Duflo and Saez (2003, *QJE*)** — $20 to attend a benefits fair raised enrollment, and raised it too for **untreated colleagues in treated departments**.
- That last one is a clean randomization that still violates SUTVA (Part 5).

Where finance experiments live

- Notice what these studies have in common: the researcher controls **an offer**.
  - A mailer, a price quote, an insurance policy, an invitation.
- That is why field experiments cluster in **household finance, credit, and insurance** — markets with many small units and a lender or insurer willing to randomize.

Why finance rarely gets to randomize

- Nobody will randomize leverage, M&A, CEO pay, bankruptcy, or bank regulation:
  - stakes too high, units too big, ethics, general-equilibrium effects.
- So corporate finance and asset pricing live almost entirely in observational data.
- **The rest of this course is a toolkit for *earning* independence without a coin flip:**
  - Unconfoundedness / matching: independence *conditional on observables*.
  - RDD: independence *near a threshold*.
  - IV: independence *of an instrument*.
  - DiD / panel: independence *of timing, in changes*.
- Every method = an argument that some comparison mimics the coin flip. Your job as a referee is to ask: *whose coin, and who flipped it?*

5. SUTVA

The fine print under the potential outcomes

- Writing Y_i^1, Y_i^0 at all presumes the **Stable Unit Treatment Value Assumption**:

1. **No interference**: unit i's potential outcomes do not depend on *other units'* treatment status.
2. **No hidden versions of treatment**: "treated" means the same thing for every unit — one treatment, one dose.

- If SUTVA fails, Y_i^1 is not even well-defined: bank i's outcome under treatment depends on *who else* was treated.

Interference is the norm in finance

- Units in finance **compete and interact**. Examples:
  - **Peer effects in capital structure** (Leary and Roberts 2014): firm i's leverage responds to peers' leverage. Treating one firm treats its rivals.
  - **Bailouts and competition**: a capital injection into bank i lets it bid for deposits and borrowers — *taking them from control banks*.
  - **Banking DiDs**: treated banks' customers migrate to control banks; the control group's outcome moves *because of* the treatment.
- If treatment hurts controls, the treated-control gap **overstates** the true effect; if it helps them (spillovers), it understates.

"The control group is treated too"

- In a competitive industry, there may be no untreated state of the world to observe:
  - Treatment effects estimated off treated-vs-control gaps are **relative** (partial-equilibrium) effects.
  - The **aggregate** (general-equilibrium) effect can be smaller, zero, or of opposite sign.
- Example: a subsidy to treated firms raises their market share. Industry output may be unchanged — pure reallocation. The DiD shows a large "effect."
- Always ask: *what happens to the market, not just the treated units?*
- We will return to this in DiD (choosing clean controls) and in the reflection problem (peer effects).

Hidden versions of treatment

- SUTVA also requires the treatment to be *one thing*.
- Finance treatments are rarely uniform in dose:
  - "Having a credit rating" spans AAA to CCC.
  - "Being bailed out" ranged from $1bn to $45bn in TARP.
  - "Adopting an anti-takeover provision" bundles very different provisions.
- If doses differ and units select their dose, the single δ_i hides a dose-response function — and the estimand becomes an uninterpretable mixture.

What to do about varying dose

- **Define** the treatment precisely. What exactly is the "1" in $D_i = 1$?
- **Report** the *distribution* of dose, not just a count of treated units.
- If doses vary widely, do **not** collapse them into a binary. Instead, one of:
  - use the dose itself as the treatment and estimate a **dose-response function**, $E[Y \mid \text{dose}]$;
  - restrict the sample to a narrow dose band, so that "treated" means one thing;
  - interact the treatment dummy with dose, and report the effect at several dose levels.
- Whichever you choose, state which estimand it delivers:

"the effect of a \$1bn injection" is not "the effect of being bailed out."

6. The Regression Bridge

Same problem, older dialect

- You already know a language for causality — the regression language:

$$y = \beta_0 + \beta_1 x + u$$

- Causal interpretation of β_1 requires **conditional mean independence**:

$$E[u \mid x] = E[u] = 0$$

- The potential outcomes framework and the CMI framework describe the **same problem** in two dialects. Let's build the dictionary.

The dictionary

Potential outcomes dialect	Regression dialect
$(Y^1, Y^0) \perp\!\!\!\perp D$	$E[u \mid x] = E[u]$ (CMI)
Selection bias	$\mathrm{cov}(x, u) \neq 0$ (OVB, simultaneity)
Heterogeneous δ_i	Random coefficients b_i
ATE	Average partial effect (APE)
SDO	OLS coefficient of Y on D

- The PO dialect is sharper about *estimands* (ATE vs. ATT) and *assignment*.
- The regression dialect is sharper about *direction of bias* — and it is the language referees speak. You need both.

Omitted variable bias

- You estimate $y = \beta_0 + \beta_1 x + u$, but the true model is

$$y = \beta_0 + \beta_1 x + \beta_2 z + v$$

- Then

$$\hat{\beta}_1 \xrightarrow{p} \beta_1 + \underbrace{\frac{\mathrm{cov}(x, z)}{\mathrm{var}(x)}}_{\delta_{xz}} \beta_2$$

- The bias is the *effect of the omitted variable* (β_2) times *its regression on the included one* (δ_{xz}).
- Consistent only if $\mathrm{cov}(x, z) = 0$ or $\beta_2 = 0$: the omitted variable must matter **and** be correlated with x.

Signing the bias

$$\text{bias} = \delta_{xz}\,\beta_2$$

	$\text{cov}(x,z) > 0$	$\text{cov}(x,z) < 0$
$\beta_2 > 0$	$+$	$-$
$\beta_2 < 0$	$-$	$+$

- Classic example: ln(*wage*) on education, omitting ability.
  - Ability raises wages ($\beta_2 > 0$) and correlates positively with education $\Rightarrow$ upward bias.
- Finance example: firm value on governance, omitting CEO talent.
  - Talented CEOs may sort into well-governed firms and raise value $\Rightarrow$ governance coefficient too big.
- With **multiple** omitted variables (or multiple included regressors), signs interact and the algebra stops being your friend. Sign one bias at a time, or don't sign at all.

The omitted variable is Y^0

- The switching equation factors as $Y_i = Y_i^0 + D_i\delta_i$. Split each term into its mean plus a deviation, writing $\delta = E[\delta_i] = ATE$:

$$Y_i \;=\; \underbrace{E[Y^0]}_{\beta_0} \;+\; \underbrace{\delta}_{\beta_1} D_i \;+\; \underbrace{\left(Y_i^0 - E[Y^0]\right)}_{\text{baseline}} \;+\; \underbrace{D_i(\delta_i - \delta)}_{\text{gain}}$$

- That **is** a regression of Y on a constant and D. The last two terms are the error u_i.
- So the "omitted variable" is the baseline potential outcome Y_i^0, together with the individual gain δ_i.
- OVB — in causal settings, it *is* selection bias, written in the older dialect.

Two ways treatment can be selected

- D can be correlated with **either** piece of the error:

$$u_i = \underbrace{\left(Y_i^0 - E[Y^0]\right)}_{\text{baseline}} + \underbrace{D_i(\delta_i - \delta)}_{\text{gain}}$$

- **Selection on levels** — treatment goes to units with unusual Y_i^0: how they would have done *anyway*, untreated.
  - Bailing out the weakest banks. Training the least employable workers.
- **Selection on gains** — treatment goes to units with unusual δ_i: how much they would *benefit*.
  - The Perfect Regulator. Any optimizing agent allocating a scarce treatment.
- The two are distinct, and **either one alone** breaks the naive comparison.
- Note our own running example sorts on **gains**, not levels. Defining selection bias as "$\text{cov}(D, Y^0) \neq 0$" would misclassify it.

Both biases, from one lemma

- **Lemma.** For *binary* D with $\pi = P(D{=}1)$, and any random variable W:

$$\operatorname{cov}(D, W) = \pi(1-\pi)\Big(E[W \mid D{=}1] - E[W \mid D{=}0]\Big)$$

  - *Proof.* $E[DW] = \pi E[W|D{=}1]$ since D is binary, and $E[W] = \pi E[W|D{=}1] + (1-\pi)E[W|D{=}0]$. Substitute into $\operatorname{cov}(D, W) = E[DW] - E[D]E[W]$ and collect. ■

- **Take** $W = Y^0$**:**

$$\operatorname{cov}(D, Y^0) = \pi(1-\pi)\Big(E[Y^0|D{=}1] - E[Y^0|D{=}0]\Big)$$

- **Take** $W = D_i(\delta_i - \delta)$**.** It is 0 whenever $D = 0$, so $E[W|D{=}0] = 0$, while $E[W|D{=}1] = ATT - ATE$:

$$\operatorname{cov}\big(D,\, D_i(\delta_i - \delta)\big) = \pi(1-\pi)\big(ATT - ATE\big)$$

- Since $\operatorname{var}(D) = \pi(1-\pi)$, dividing each by $\operatorname{var}(D)$ **cancels that factor** — leaving the selection bias term and $ATT - ATE$.

The two dialects give the same equation

- Since $\hat{\beta}_1 \xrightarrow{p} \delta + \mathrm{cov}(D, u)/\mathrm{var}(D)$, and using $ATT - ATE = (1-\pi)(ATT - ATU)$ from earlier:

$$\hat{\beta}_1 \xrightarrow{p} \underbrace{ATE}_{\delta} + \underbrace{\Big(E[Y^0|D{=}1] - E[Y^0|D{=}0]\Big)}_{\text{selection bias}} + \underbrace{(1-\pi)(ATT - ATU)}_{\text{heterogeneity bias}}$$

- The right-hand side is exactly the *SDO* decomposition we proved in Part 3.
- The regression route and the potential-outcomes route are not analogies. **They are the same three terms.**
- So $\hat{\beta}_1$ is inconsistent for the *ATE* — but perfectly *consistent* for the *SDO*. Once again: OLS delivers the estimand you asked for, not the one you wanted.

Simultaneity: when y fires back

- Suppose the regressor responds to the outcome:

$$y = \beta_0 + \beta_1 x + u, \qquad x = \alpha_0 + \alpha_1 y + \varepsilon$$

- Solve for x:

$$x = \frac{\alpha_0 + \alpha_1 \beta_0}{1 - \alpha_1 \beta_1} + \frac{\alpha_1 u + \varepsilon}{1 - \alpha_1 \beta_1}$$

- x is mechanically correlated with u: CMI fails, OLS is biased — **reverse causality**.
- Finance examples: leverage and performance; board independence and performance; analyst coverage and returns. In equilibrium, almost every firm characteristic responds to almost every outcome.
- Note: lagging x does **not** fix this if x is persistent and anticipates y. It hides the problem; it does not solve it.

Random coefficients: the formal bridge

- Let every firm have its own line:

$$y_i = a_i + b_i x_i, \qquad a_i = \alpha + c_i, \quad b_i = \beta + d_i$$

- With binary x, this **is** the potential outcomes model: $b_i = \delta_i$, $a_i = Y_i^0$.
- Rewrite: $y_i = \alpha + \beta x_i + \underbrace{(c_i + d_i x_i)}_{u_i}$
- OLS identifies the **average partial effect** $\beta = E[b_i]$ (the ATE!) only if

$$E[c_i \mid x_i] = 0 \ \text{(no selection on levels)}, \qquad E[d_i \mid x_i] = 0 \ \text{(no selection on gains)}$$

- These are precisely "selection bias $= 0$" and "ATT $=$ ATU." The two dialects have converged.

Bad controls: a preview

- Adding controls is not free. If a control x_2 is itself **affected by the treatment**:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u, \qquad x_2 = \gamma_0 + \gamma_1 x_1 + \varepsilon$$

- Then conditioning on x_2 blocks part of the effect you want — and can *introduce* selection bias where none existed.
- Rule of thumb: controls must be **pre-treatment** (determined before D).
- Why conditioning on a post-treatment variable creates bias is subtle — it is the centerpiece of the next lecture (DAGs and colliders).

7. Randomization Inference

Two null hypotheses

- Standard (Neyman) null: the **average** effect is zero, $E[\delta_i] = 0$.
  - Inference relies on asymptotics: large n, estimated standard errors.
- Fisher's **sharp null**: the effect is zero *for every unit*,

$$H_0 : \delta_i = 0 \text{ for all } i$$

- Under the sharp null, both potential outcomes are known for every unit: $Y_i^1 = Y_i^0 = Y_i$ (the observed value).
- That makes *exact* inference possible — no asymptotics, no standard errors, any sample size.

The permutation logic

1. Under the sharp null, the observed outcomes would have been the same under **any** treatment assignment.
2. So: re-randomize D on paper, in every way the actual randomization could have come out.
3. Recompute the test statistic (e.g., the SDO) for each hypothetical assignment.
4. The exact p-value is the share of assignments giving a statistic at least as extreme as the observed one.

- The randomness comes from the **assignment mechanism itself** — the same mechanism that justified the causal comparison justifies the inference.

Randomization inference on ten banks

- Suppose the ten banks were randomized: five drawn by lot got injections. Observed SDO $= 1.6$.

```
y <- c(2, 6, 7, 2, 6, 6, 7, 7, 7, 8)      # observed lending growth
treated <- c(2, 3, 5, 8, 9)               # banks drawn by lot
sdo_obs <- mean(y[treated]) - mean(y[-treated])

perms <- combn(10, 5)                      # all 252 possible assignments
sdo_all <- apply(perms, 2, function(tr) mean(y[tr]) - mean(y[-tr]))
p_exact <- mean(abs(sdo_all) >= abs(sdo_obs))
round(c(SDO = sdo_obs, exact_p = p_exact), 3)
```

```
   SDO exact_p
 1.600   0.397
```

- With $n = 10$, a difference of 1.6pp is *not* statistically distinguishable from luck-of-the-draw — and we know this **exactly**, not asymptotically.

Where you will see this again

- Randomization inference is not a curiosity:
  - **Synthetic control**: placebo tests permute the treatment across donor units — that *is* randomization inference.
  - **Small-N settings**: one treated state, one regulatory event, a handful of treated firms — common in finance, hopeless for asymptotics.
  - Cluster-robust inference with few clusters: permutation tests as the honest fallback.
- The deeper point: *inference should mirror the assignment mechanism.* Where the mechanism is uncertain, so is the p-value.

Wrapping Up

Summary [Part 1]

- Causality is defined by **counterfactuals**; the fundamental problem is that one potential outcome is always missing.
- Estimands differ: **ATE, ATT, ATU** answer different policy questions. Know which one your design identifies.
- The naive comparison decomposes as

$$SDO = ATE + \text{selection bias} + (1 - \pi)(ATT - ATU)$$

- Optimizing agents generate selection **on gains** — the Perfect Regulator gets the sign wrong *because* it allocates well.
- More data does not fix a biased assignment mechanism.

Summary [Part 2]

- **Randomization** kills both bias terms by making $(Y^1, Y^0) \perp\!\!\!\perp D$. Balance tables are its testable symptom.
- **SUTVA** fails easily in finance: competition, peer effects, and heterogeneous doses. Ask what happens to the market, not just the treated.
- The **regression dialect** translates one-for-one: CMI $\Leftrightarrow$ independence; OVB $\Leftrightarrow$ selection bias; random coefficients $\Leftrightarrow$ heterogeneous δ_i.
- **Randomization inference** gives exact p-values from the assignment mechanism — remember it when you meet synthetic control.
- Next lecture: DAGs — a graphical language for deciding *what to control for, and what not to*.

References

- Cunningham, S. *Causal Inference: The Remix*, Ch. 4.
- Holland, P. (1986). Statistics and Causal Inference. *JASA* 81, 945–960.
- Blake, T., C. Nosko, and S. Tadelis (2015). Consumer Heterogeneity and Paid Search Effectiveness: A Large-Scale Field Experiment. *Econometrica* 83, 155–174.
- Karlan, D. and J. Zinman (2009). Observing Unobservables: Identifying Information Asymmetries with a Consumer Credit Field Experiment. *Econometrica* 77, 1993–2008.
- Bertrand, M., D. Karlan, S. Mullainathan, E. Shafir, and J. Zinman (2010). What's Advertising Content Worth? *Quarterly Journal of Economics* 125, 263–306.
- Cole, S., X. Giné, and J. Vickery (2017). How Does Risk Management Influence Production Decisions? *Review of Financial Studies* 30, 1935–1970.
- Duflo, E. and E. Saez (2003). The Role of Information and Social Interactions in Retirement Plan Decisions. *Quarterly Journal of Economics* 118, 815–842.
- Leary, M. and M. Roberts (2014). Do Peer Firms Affect Corporate Financial Policy? *Journal of Finance* 69, 139–178.
- Angrist, J. and J.-S. Pischke (2009). *Mostly Harmless Econometrics*. Princeton UP.