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. 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.

The common thread: the assignment mechanism

- 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.

Where finance experiments live

- Finance does have real randomized experiments. Four representative ones:
  - **Blake, Nosko and Tadelis (2015)** switched eBay's paid search off in random markets. Brand-keyword ads had **no measurable benefit** — traffic substituted to the free organic link — and average returns were negative. Attribution had been measuring selection, not causation.
  - **Karlan and Zinman (2009)** mailed 58,000 South African loan offers, randomizing the *offer* rate, the *contract* rate and a repayment incentive. The offer rate identifies **adverse selection**; the contract rate at a fixed offer rate identifies **moral hazard**.
  - **Cole, Giné and Vickery (2017)** randomized rainfall insurance; insured farmers shifted into higher-return, rainfall-sensitive crops.
  - **Duflo and Saez (2003)** paid $20 to attend a benefits fair. Enrollment rose for treated employees *and* untreated colleagues — a clean randomization that violates SUTVA (Part 5).
- **What they share:** the researcher controls **an offer** — a mailer, a price quote, a policy, an invitation. That is why field experiments cluster in **household finance, credit and insurance**.

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. Randomization Inference

Why bother?

- **Randomization-based inference**: the uncertainty in an estimate comes from the **random assignment of treatment**, not from hypothetical sampling out of a larger population.
- The two approaches answer different questions:
  - A conventional standard error asks *"what if I had drawn a different sample?"*
  - Randomization inference asks *"what if the treated had been differently?"*
- In an experiment, only the second thing actually happened. The sample is whoever was there; the **assignment** is what was random.
- That shift buys you *exact* p-values — the probability that chance alone could have produced what you saw.
- The idea is old: **R. A. Fisher**, early twentieth century.

Three situations where it earns its keep

1. **You have the population, not a sample.** With comprehensive or administrative data, a standard error answering "what if I drew another sample?" is not valuable. The real uncertainty is that we never see the counterfactual (Abadie, Diamond and Hainmueller 2010; Abadie et al. 2020).
2. **Few treated units.** Conventional inference leans on large-sample approximations. With a handful of treated units a single observation can swing the estimate, and conventional tests **over-reject** (Young, 2019).
3. **Transparency.** Placebo-based inference is easy to explain and hard to fudge.

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*,

$$H_0 : \delta_i = Y_i^1 - Y_i^0 = 0 \quad \text{for all } i$$

- The sharp null is **stronger**.
- Its payoff is that it is **nonparametric** — no Gaussian assumption, no large-sample approximation, no estimated standard error. It works at $n = 8$.

The sharp null fills in the missing half

- The fundamental problem of causal inference is that we never see both potential outcomes. Under the sharp null, **we do**.
- If $\delta_i = Y_i^1 - Y_i^0 = 0$, then $Y_i^1 = Y_i^0$, and both equal the value we observed.

	What we observe				Under $H_0 : \delta_i = 0$	
Bank	D	Y	Y^0	Y^1	Y^0	Y^1
1	0	2	2	?	2	2
2	1	6	?	6	6	6
3	1	7	?	7	7	7
⋮						
10	0	8	8	?	8	8

It is not the bootstrap

- The two get confused because both produce a distribution by resampling.
- **Bootstrap.** Resample *observations*, with replacement, from your sample. The resulting uncertainty is about **which units ended up in your sample** — sampling uncertainty.
- **Randomization inference.** Hold the units fixed and reshuffle **who was treated**. The uncertainty is about **which units got the treatment** — assignment uncertainty.
- Different questions, different null, different interpretation.

Six steps

Following Cunningham (*The Remix*, §4.7):

1. Choose the **sharp null**.
2. Construct a **test statistic** $t(D, Y)$.
3. Pick a **different** treatment assignment $\tilde{D}$.
4. Compute the test statistic for that pretend assignment.
5. **Cycle** over step 3 for all possible assignments.
6. Divide the **rank** of the true statistic by the number of trials — the exact p-value.

Step 2: what makes a good test statistic

- A test statistic $t(D, Y)$ is just a **scalar** built from the assignment and the outcomes.
- The requirement: it should take **extreme values when the sharp null is false**, and extreme values should be **rare when it is true**.
- The obvious choice is the one we already know — the simple difference in outcomes,

$$t(D, Y) = \bar{Y}_{\text{treated}} - \bar{Y}_{\text{control}} = SDO$$

- For the ten banks, the lottery gave injections to banks $2, 3, 5, 8, 9$:

$$SDO = \frac{6+7+6+7+7}{5} - \frac{2+2+6+7+8}{5} = 6.6 - 5.0 = 1.6$$

Steps 3–5: the permutation loop

- Under the sharp null the outcomes Y **do not change** — only the labels move.
- Suppose banks $1, 3, 5, 8, 9$ had been drawn instead:

$$\tilde{t} = \frac{2+7+6+7+7}{5} - \frac{6+2+6+7+8}{5} = 5.8 - 5.8 = 0.0$$

- Or banks $1, 2, 4, 6, 7$:

$$\tilde{t} = \frac{2+6+2+6+7}{5} - \frac{7+6+7+7+8}{5} = 4.6 - 7.0 = -2.4$$

- Store each one and move on. With five treated out of ten there are $\binom{10}{5} = 252$ such assignments.
- Each $\tilde{t}$ is a number the world *could have produced* if the treatment did nothing. Together they are the **reference distribution**.

Step 6: the exact p-value

$$\Pr\left(|t(\tilde{D}, Y)| \geq |t(D, Y)| \mid \delta_i = 0 \ \forall i\right) = \frac{\sum_{\tilde{D} \in \Omega} I(|t(\tilde{D}, Y)| \geq |t(D, Y)|)}{K}$$

- **what share of the pretend worlds look at least as extreme as the real one?**
- Ω is the set of assignments the randomization could have produced and $K = |\Omega|$. Here $K = 252$.
- "$\delta = 0$" would name the Neyman null (*The Remix*)
- **"Exact"** is meant literally. This is not an approximation to a limiting distribution — it is a finite count.

Randomization inference on ten banks

```
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
```

- $p = 100/252 = 0.397$. With $n = 10$, a difference of 1.6pp is **not** distinguishable from luck of the draw — and we know this *exactly*, not asymptotically.

The assumption hiding in `combn(10, 5)`

- The reference set must match the **actual randomization design**.
- `combn(10, 5)` presumes a completely randomized experiment with a **fixed** five treated — exactly five drawn by lot.
- Had each bank instead been **independently coin-flipped**, the correct reference set would be all $2^{10} = 1024$ assignments, and the p-value would differ.
- Get this wrong and the p-value is not "approximately right." It answers a different question.

The permutation scheme must mirror the physical randomization.

Revisiting step 2: the menu of test statistics

- *SDO* is the default, and a good one **when effects are additive and outliers few**.
- **Log difference** — for skewed outcomes and multiplicative effects:

$$T_{\log} = \frac{1}{N_T} \sum_i D_i \ln Y_i - \frac{1}{N_C} \sum_i (1 - D_i) \ln Y_i$$

- **Quantiles** — robust to outliers in either tail:

$$T_{\text{median}} = \left| \text{median}(Y_T) - \text{median}(Y_C) \right|$$

- **Normalised ranks** — discard magnitudes, keep the ordering. Good with small samples or heavy tails.
- **Kolmogorov–Smirnov** — looks at the whole distribution, not a single summary.
- **None of these changes the null or the validity of the test**. The choice changes only **power**. So choose **before** you look.

Revisiting step 5: when "all assignments" is impossible

- Ten banks give 252 assignments. In *The Remix*, Thornton's (2008) HIV-testing experiment had 2,901 participants, 2,222 of them incentivised. The number of assignments is

$$\binom{2901}{2222} \approx 6.15 \times 10^{680}$$

- So we **sample** the reference distribution: draw a random assignment, recompute the statistic, repeat 1,000 to 10,000 times.
- The p-value is then **approximate** rather than exact — but the error is sampling error you control with more draws, not an asymptotic assumption you must defend.
- This is what you will actually do. Any dataset of a reasonable size makes full enumeration hopeless.

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.
- **Randomization inference** turns the assignment mechanism into a test. Under **Fisher's sharp null** every counterfactual is filled in, so the p-value is an exact count over the assignments the lottery could have produced — no asymptotics, at any sample size.
- The **test statistic is a choice**: means, logs, quantiles, ranks, or the whole distribution (Kolmogorov–Smirnov). Make it before you look, and remember that a treatment can move the *spread* without moving the mean.
- 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.
- Abadie, A., A. Diamond and J. Hainmueller (2010). Synthetic Control Methods for Comparative Case Studies. *JASA* 105, 493–505.
- Abadie, A., S. Athey, G. Imbens and J. Wooldridge (2020). Sampling-Based versus Design-Based Uncertainty in Regression Analysis. *Econometrica* 88, 265–296.
- Fisher, R. A. (1935). *The Design of Experiments*. Oliver and Boyd.
- Thornton, R. (2008). The Demand for, and Impact of, Learning HIV Status. *AER* 98, 1829–1863.
- Young, A. (2019). Channeling Fisher: Randomization Tests and the Statistical Insignificance of Seemingly Significant Experimental Results. *QJE* 134, 557–598.
- 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.
- 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.