Econ 103: Introduction to Econometrics

Lecture 20 — Treatment Effects & Difference-in-Differences

Ryan Longmuir

UCLA

Summer Session C, 2026

Reading: Hill, Griffiths & Lim (5th ed.), §7.5–7.6; Stock & Watson (4th ed.), §13.1, 13.3–13.4.

Where we are — and the question we've circled all term

From Lecture 1's scatterplots to Lecture 18's omitted-variable bias, one warning has recurred: correlation is not causation. A regression slope measures association — it is causal only if the regressor is *exogenous* (SR2/MR2).

Today we close the loop. When *is* a coefficient a causal effect? The cleanest answer comes from the **potential-outcomes** framework:

- potential outcomes, treatment effects, and the **ATE**;
- selection bias — and why it's just omitted-variable bias;
- **randomized experiments** (Project STAR) as the gold standard;
- **difference-in-differences** when you *can't* randomize;
- and a final word on correlation vs. causation.

Today's plan

1. The potential-outcomes framework
2. Selection bias
3. Randomized experiments
4. Difference-in-differences
5. Correlation vs. causation, revisited

Part 1

The potential-outcomes framework

Two outcomes, one observed

Let $d_i = 1$ if individual i is **treated**, 0 if not. Imagine *both* futures:

$$y_{1i} = \text{outcome if treated}, \qquad y_{0i} = \text{outcome if not}.$$

The causal effect for individual i is the difference $y_{1i} - y_{0i}$.

The fundamental problem of causal inference

We only ever see *one* outcome per person:

$$y_i = y_{0i} + (y_{1i} - y_{0i})\, d_i.$$

The other potential outcome — the **counterfactual** — is forever missing. So the individual effect $y_{1i} - y_{0i}$ is unknowable.

The framework's gift is honesty: effects *differ* across people, and we must settle for an average, not an individual, effect.

The average treatment effect

Since individual effects are hidden, we target the **average treatment effect**:

$$\tau_{\mathrm{ATE}} = \mathbb{E}(y_{1i} - y_{0i}).$$

The natural estimator compares average outcomes across groups — which is just a regression on a treatment dummy:

$$y_i = \alpha + \tau\, d_i + e_i \qquad \Longrightarrow \qquad \hat{\tau} = \bar{y}_{\text{treated}} - \bar{y}_{\text{control}}.$$

This is the **difference estimator** — the dummy-variable regression of Lecture 19, now carrying a causal name.

The question is whether $\hat{\tau}$ actually estimates τ_{ATE}. Sometimes yes, often no — and the gap is selection bias.

Part 2

Selection bias

The difference estimator, decomposed

Split the group gap into two pieces:

$$\underbrace{\mathbb{E}(y \mid d{=}1) - \mathbb{E}(y \mid d{=}0)}_{\text{what } \hat{\tau} \text{ estimates}} = \underbrace{\mathbb{E}(y_1 - y_0 \mid d{=}1)}_{\text{effect on the treated (ATT)}} + \underbrace{\left[\mathbb{E}(y_0 \mid d{=}1) - \mathbb{E}(y_0 \mid d{=}0)\right]}_{\text{selection bias}}.$$

Selection bias is the difference in the groups' *untreated* outcomes: the treated and control groups weren't comparable to begin with.

Example (Do hospitals make you sicker? (HGL))

Survey: people who'd been hospitalized rated their health worse (3.21) than those who hadn't (3.93). Hospitals don't cause illness — sick people select into hospitals. The naive difference $3.21 - 3.93 = -0.72$ *is almost all selection bias.*

Selection bias is omitted-variable bias

In $y_i = \alpha + \tau d_i + e_i$, the error e holds everything else — including pre-existing health. If sicker people choose treatment, then d_i is correlated with e_i:

$$\mathbb{E}(e_i \mid d_i) \neq 0 \quad \Longrightarrow \quad \hat{\tau} \text{ is biased.}$$

This is *exactly* the omitted-variable / endogeneity problem from Lectures 7, 13, and 18 — the confounder (health) is correlated with the "regressor" (treatment) and drives the outcome.

The whole game

A treatment-dummy coefficient is causal *only if* treatment is uncorrelated with the omitted factors — i.e. d_i is (as-if) randomly assigned. How do we get that? Sometimes we can engineer it.

Part 3

Randomized experiments

Randomization kills selection bias

Randomly assign treatment, and d_i becomes statistically independent of the potential outcomes. Then the untreated groups are comparable, selection bias vanishes, and

$$\tau_{\text{ATE}} = \mathbb{E}(y \mid d{=}1) - \mathbb{E}(y \mid d{=}0) = \hat{\tau} \text{ (unbiased).}$$

- The **randomized controlled experiment** (RCT) is the gold standard — the benchmark every observational study is judged against.
- Common in medicine (drug vs. placebo); rarer in economics (cost, ethics, feasibility) — but not unheard of.
- With randomization, the *simple* difference estimator — a dummy regression — recovers the causal effect. No fancy econometrics required.

Project STAR: class size, randomized

Tennessee, 1985–89: kindergartners randomly assigned within schools to small (13–17) vs. regular (22–25) classes. A genuine RCT.

$$\text{TOTALSCORE} = \beta_1 + \beta_2\,\text{SMALL} + e \quad \Longrightarrow \quad \hat{\beta}_2 = 13.9 \text{ points (significant).}$$

- Because assignment was random, $\hat{\beta}_2$ *is* the causal effect of small classes — not contaminated by which families chose them.
- Add a control (teacher experience): the estimate barely moves ($13.9 \to 14.0$) — the signature of good randomization (controls are *uncorrelated* with treatment, so they only shrink the standard error).
- Check the randomization: regress SMALL on student traits (a linear probability model, Lecture 19) — nothing is significant, $\hat{p} \approx 0.47 \approx$ a coin flip. Assignment really was random.

Part 4

Difference-in-differences

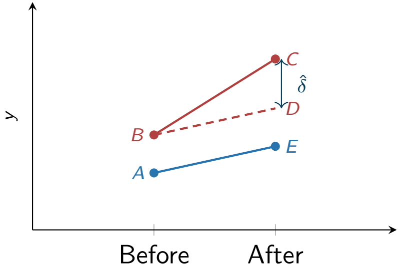
When you can't randomize: natural experiments

Most economic data are observational. But sometimes a policy change hits one group and not another — a **natural (quasi-)experiment** where treatment is "as if" random.

Difference-in-differences uses before/after data on a treatment and a control group:

$$\hat{\delta} = (\bar{y}^{\text{treat}}_{\text{after}} - \bar{y}^{\text{treat}}_{\text{before}}) - (\bar{y}^{\text{ctrl}}_{\text{after}} - \bar{y}^{\text{ctrl}}_{\text{before}}).$$

The control's change nets out the common trend; what's left is the treatment's effect.



Dashed BD = the counterfactual (parallel trend).

DiD as one regression

With a treatment dummy and a time dummy, the difference-in-differences estimator is the coefficient on their interaction:

$$y_{it} = \beta_1 + \beta_2\,\mathrm{TREAT}_i + \beta_3\,\mathrm{AFTER}_t + \delta\,(\mathrm{TREAT}_i \times \mathrm{AFTER}_t) + e_{it}.$$

- β_2: fixed gap between groups; β_3: common time trend; δ: the treatment effect (exactly the slope-dummy idea of Lecture 19).
- Key assumption: parallel trends — absent treatment, the two groups would have moved together. This is the counterfactual; it is an *assumption*, not something the data prove.

With panel data (the *same* units before and after), first-differencing removes every fixed unit characteristic c_i — another route to the same $\hat{\delta}$.

Card & Krueger: the minimum wage

April 1992: New Jersey raised its minimum wage \$4.25 → \$5.05; neighboring Pennsylvania held at \$4.25. Card & Krueger surveyed fast-food employment (FTE) before and after.

$$\hat{\delta} = (21.03 - 21.17)_{\text{NJ}-\text{PA, after}} - (20.44 - 23.33)_{\text{NJ}-\text{PA, before}} = +2.75.$$

- Employment in NJ *did not fall* (if anything, rose by ≈ 2.75 FTE) — contrary to the textbook competitive prediction.
- Robust to adding chain/ownership/region controls, and to a panel first-difference specification ($\hat{\delta} \approx 2.75$ throughout).
- A landmark result that reshaped the minimum-wage debate — and a model of how a clean comparison group turns observational data into credible causal evidence.

Part 5

Correlation vs. causation, revisited

The through-line of the whole course

When is a regression coefficient causal?

Only when the regressor is (as-if) randomly assigned — exogenous, so it's uncorrelated with everything in the error. Everything else has been a way to get there or to check it.

- RCT — engineer randomization (gold standard; Project STAR).
- DiD / natural experiments — borrow "as-if" randomness from a policy change (Card–Krueger).
- Controls & proxies (Lecture 18) — make treatment as-good-as-random *conditional* on observables.
- Beware spurious correlation: Maine's divorce rate and U.S. margarine consumption correlate at 0.99 — and mean nothing.

Most economic data are observational, so a causal claim demands a credible source of variation — not just a small p-value.

Recap — and the whole course in one arc

Today

- potential outcomes: y_{1i}, y_{0i}; never both
- $\tau_{\text{ATE}} = \mathbb{E}(y_1 - y_0)$; difference estimator = dummy regression
- selection bias = OVB; randomization removes it
- STAR: small class $+13.9$ pts (causal)
- DiD δ on TREAT$\times$AFTER; Card–Krueger $+2.75$
- causal $\Leftrightarrow$ exogenous regressor

The journey (L1–L20)

- probability & the CLT (L2–4)
- simple regression: estimate, BLUE, infer (L5–10)
- fit, functional form (L11–12)
- multiple regression & inference (L13–17)
- specification, dummies, causality (L18–20)

Beyond 103: instrumental variables, panel data, time series, probit/logit.

You can now...

estimate an economic relationship, quantify its uncertainty, test hypotheses about it, *and* reason about whether it is causal. That is what it means to do econometrics. Good luck on the final!

Questions?

Thank you — it's been a great term.