# Difference-in-Differences

Mixtape Session

MIXTAPE SESSIONS

# Roadmap

Outline

Topical elements of DiD

Workshop

Potential outcomes

# Topics

- Introduction to DiD basics
  - → Potential outcomes review
  - → DiD formula
  - → Covariates
- Differential timing
  - → Heterogeneity and the Bacon decomposition
  - → TWFE bias in estimation of overall and dynamic ATT

# Workshop outline

Three types of solutions

- Aggregated group-time ATT
- Stacked regression
- Explicit Imputation

# Workshop outline

Synthetic control

- Canonical ADH model
- Multiple treated units (matrix completion with nuclear norm regularization)
- Imperfect fit
- Combining synthetic controls

# Roadmap

Outline

Topical elements of DiD

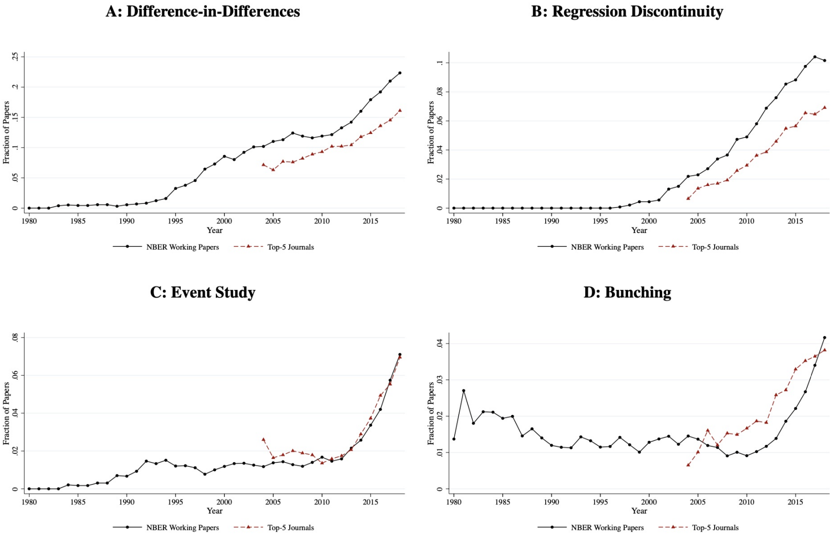
Workshop

Potential outcomes

# What is difference-in-differences (DiD)

- DiD is a very old, relatively straightforward, intuitive research design
- A group of units are assigned some treatment and then compared to a group of units that weren't
- Early usage in several 19th century health policy debates
- Brought into labor economics with Orley Ashenfelter (1978), LaLonde (1986), Card and Krueger (1994)
- Now the most widely used quasi-experimental method

Figure IV: Quasi-Experimental Methods




Notes: This figure shows the fraction of papers referring to each type of quasi-experimental approach. See Table A.I for a list of terms. The series show 5-year moving averages.

# Natural experiments

*“A good way to do econometrics is to look for good natural experiments and use statistical methods that can tidy up the confounding factors that nature has not controlled for us.” – Daniel McFadden (Nobel Laureate recipient with Heckman 1992)*


I ❤ federalism
(for the natural experiments)

# A Little History: Princeton 1970s and 1980s

- What is a theoretical model in economics, particularly labor, and what role will it play in estimation?
- Empirical micro has been divided along two almost philosophical approaches to causal inference over the years
  - → **Model**: Causality is model-based. It only exists within the framework of a theory that says "X causes Y" (e.g., Heckman)
  - → **Design**: Causality is design-based. No causality without *physical* manipulation of a treatment $X$ (e.g., Rubin, Holland)
- This background is an interpretation based on speeches by David Card

# Economist’s models

Economics models typically contain the following

- Preference functions (e.g., quasi concave utility), objective function (e.g., profit)
- Constraints (e.g., time, budgets)
- Endogenous choice variables (e.g., bundles of goods, output)
- Equilibrium (e.g., first order conditions, Nash equilibrium)

# Three economic models within empirical micro

1. **Approximating models**: Consumer demand, labor supply models (e.g., Mincer 1958; 1974)
   - $\rightarrow$ Theory implies $y_i = f_i(x_i)$ with restrictions on $f_i$ (e.g., concavity)
   - $\rightarrow$ Researcher estimates a simpler version

$$y_i = \alpha + x_i\beta + \varepsilon_i$$

2. **Exact models**: Models gives us all causes ("complete DGP")
   - $\rightarrow$ Utility, heterogenous taste, complete demand
   - $\rightarrow$ Estimate model parameters and distribution of heterogeneity
   - $\rightarrow$ Functional form, useful for welfare analysis
3. **Working model**: Design based approaches: No precise model was used or relied for estimation, though it may guide the questions. Princeton industrial relations and labor group, 1970s and 1980s

# Characteristics of design

- Focus is on "physical manipulation of treatments", much like Fisher RCTs and randomization ("No causality without manipulation" – Rubin, Holland)
- Without the explicit role of the model in identification, topics open up
- Part of the dominance of design based approaches is the ability to study almost anything in applied micro so long as physical manipulation of treatments are our focus
- Model-based approaches tend to, on the other hand, be much more closely married to neoclassical topics because you need the neoclassical model for identification

# Causality and labor economics

Keep in mind the sweep of the 20th century econometric approaches to identifying “causal effects”

- **Mid-century**: Macro and linear systems of equations, identification problems
- **1970s**: Micro data shows up, McFadden’s logit, Heckman’s selection model
- **1980s**: Econometric critiques like the Lucas critique, Leamer “specification searching”, LaLonde (1985) critiques program evaluation, Lewis dismisses IV and Heckit models
- **1980s/1990s**: Design emerges within the Princeton labor group, randomized instruments, “clever” natural experiments, “plausibly exogenous”, RDD, difference-in-differences

# Princeton industrial relations group and "design"

- Backdrop is 1970s rising inequality and poverty, returns to education (Katz and Murphy, others)
- Princeton becomes ground zero for design: Albert Rees brings in micro data; advises Orley Ashenfelter
- Chicago and others focus on model driven approaches (Heckman)
- Ashenfelter focuses on job trainings program, likes randomization a lot
- Extensive mentoring: Advising: LaLonde, Angrist, Currie, Levine, Pischke, and on and on
- Other faculty: Alan Krueger, Guido Imbens, Don Rubin
- Adoption of potential outcomes (Krueger notes the NEJM and medical concepts)

# Credibility revolution

- Design approach tends to crowd out the model based approach in the applied micro fields like labor, health, development with exceptions (Wolpin, Rust)
- Nobel Prizes for experiments, RCTs and causal inference (Vernon Smith, Bannerjee, Duflo, Kremer, Card, Angrist, Imbens)
- Structural wins too though (Deaton) so the debate still rages (see Nancy Cartwright and Angus Deaton work)

# 2021 Nobel Prize

Design-based identification goes to Card, Angrist, Imbens, Krueger, Rubin (three of whom win the Nobel Prize)

- David Card for empirical labor (perhaps greatest labor economist of his generation with Krueger, RIP)
- Josh Angrist and Guido Imbens for causal inference (specifically 1990s papers on IV)

# Design contributions

What are the broad contributions of the design approach to causal inference?

- Counterfactuals and causality; research design outlines an “explicit counterfactual”, randomization is best, credible instruments are second best
- Substantive specification tests: randomization tests in RCTs like balance across covariates, pre-treatment comparisons, event studies, falsification
- Here we see the event study in DiD too
- Data quality, replication, data warehouses, journals requiring programs, pre-registration of RCTs

# Design today

Think of identification in design approaches in two ways:

1. **Design-design**: Independence (i.e., randomization). RCT, IV and matching all have an "independence" assumption
2. **Design-model**: Structural (i.e., parallel trends, factor models, smoothness). RDD, DiD, synthetic control

The class largely focuses on the design-model, not the design-design

# Roadmap

Outline

Topical elements of DiD

Workshop

Potential outcomes

# Why an entire workshop on DiD (and synthetic control)?

- **Research advantages**: DiD and synthetic control are often one of the only ways to study large social policies (e.g., Medicaid, minimum wages)
- **Researcher anxiety**: As you may heard, difference-in-differences (and synth) has received considerable scrutiny from econometricians who have shown both problems with old estimators and proposed new ones
- **A New Hope**: These solutions have a lot of similarities and are now widely available code in both R and Stata
- **Bad news**: Without barriers to entry, econometricians will continue writing did and synth so long as there are profits, so expect more DiD and synth for a while

# Pedagogy of the workshop

- My style of teaching DiD and synthetic control is to teach *papers*, not synthesis
- As a result, it can feel like drinking from a firehose to learn so many papers
- Some topics I can't cover, particularly if they are brand new, but I will do my best to give you what I consider to be a selective discussion that I think is important
- Jon Roth will be teaching a June 10th 1-day workshop on "advanced DiD" if you want to follow up this course with even more by one of the economtricians writing in this area

# Roadmap

Outline

Topical elements of DiD

Workshop

Potential outcomes

# Potential outcomes review

- DiD really can't be understood without committing to some common causality language
- Standard language is the potential outcomes model, sometimes called the Rubin-Neyman model
- Don't go over potential outcomes too fast or you'll miss all the fun
- Potential outcomes are thought experiments about worlds that never existed, but which *could have*

# Introduction to Counterfactuals and Causality

- Aliens come and orbit earth, see sick people in hospitals and conclude “these ‘hospitals’ are hurting people”
- Motivated by anger and compassion, they kill the doctors to save the patients
- Sounds stupid, but earthlings do this too - all the time
- Let’s look at the challenges of making causality synonymous with correlations

# #1: Correlation and causality are very different concepts

These are not the same thing:

- Causal question: “If a doctor puts a person with Covid on a ventilator (D), will her health (Y) improve?”
- Correlation question:

$$\frac{Cov(D,Y)}{\sqrt{Var_D}\sqrt{Var_Y}}$$

# #2: Coming first may not mean causality!

- Every morning the rooster crows and then the sun rises
- If the feral cat had killed the rooster the sun would have still risen, so coming first must not be enough
- *Post hoc ergo propter hoc*: “after this, therefore, because of this”

# #3: No correlation does not mean no causality!

- A sailor sails her sailboat across a lake
- Wind blows, and she perfectly counters by turning the rudder
- The same aliens observe from space and say “Look at the way she’s moving that rudder back and forth but going in a straight line. That rudder is broken.” So they send her a new rudder
- They’re wrong but why are they wrong? There is, after all, no correlation
- Question: What if she had been moving the rudder by flipping coins?

# Potential outcomes notation

- Let the treatment be a binary variable:

$$D_{i,t} = \begin{cases} 1 \text{ if hospitalized at time } t \\ 0 \text{ if not hospitalized at time } t \end{cases}$$

  where $i$ indexes an individual observation, such as a person

- Potential outcomes:

$$Y_{i,t}^{j} = \begin{cases} 1 \text{ health if hospitalized at time } t \\ 0 \text{ health if not hospitalized at time } t \end{cases}$$

  where $j$ indexes a counterfactual state of the world

- I'll drop $t$ subscript, but note – these are potential outcomes for the same person at the exact same moment in time

# Moving between worlds

- A potential outcome $Y^1$ and a historical outcome $Y$ are neither conceptually nor notationally the same thing
- Potential outcomes are *hypothetical* possibilities describing states of the world but historical outcomes actually occurred
- We choose among potential outcomes by selecting the treatment

# Important definitions

**Definition 1: Individual treatment effect**

The individual treatment effect, $\delta_i$, equals $Y_i^1 - Y_i^0$

# Important definitions

## Definition 2: Average treatment effect (ATE)

The average treatment effect is the population average of all $i$ individual treatment effects

$$\begin{aligned} E[\delta_i] &= E[Y_i^1 - Y_i^0] \\ &= E[Y_i^1] - E[Y_i^0] \end{aligned}$$

# Important definitions

## Definition 3: Switching equation

An individual's observed health outcomes, $Y$, is determined by treatment assignment, $D_i$, and corresponding potential outcomes:

$$
\begin{aligned}
Y_i &= D_i Y_i^1 + (1 - D_i) Y_i^0 \\
Y_i &= \begin{cases} Y_i^1 \text{ if } D_i = 1 \\ Y_i^0 \text{ if } D_i = 0 \end{cases}
\end{aligned}
$$

# So what's the problem?

**Definition 4: Fundamental problem of causal inference**

If you need both potential outcomes to know causality with certainty, then since it is impossible to observe both $Y_i^1$ and $Y_i^0$ for the same individual, $\delta_i$, is *unknowable*.

# Conditional Average Treatment Effects

## Definition 5: Average Treatment Effect on the Treated (ATT)

The average treatment effect on the treatment group is equal to the average treatment effect conditional on being a treatment group member:

$$\begin{aligned} E[\delta|D=1] &= E[Y^1 - Y^0|D=1] \\ &= E[Y^1|D=1] - E[Y^0|D=1] \end{aligned}$$

As we will see, DiD methods *only* identify the ATT, not the ATE

# SUTVA

- SUTVA stands for “stable unit treatment value assumption” and is a unique representation of the nature of a causal effect in the potential outcomes Rubin-Neyman model
- It concerns the stability of the underlying treatment effects and has a few things relevant to our analysis in this workshop
- These largely concern spillovers, anticipation and hidden variation in treatment – all of which can be threatened in our designs

# SUTVA I: No Spillovers

- “In many situations it may be reasonable to assume that treatments applied to one unit do not affect the outcome for another unit.” (Imbens and Rubin 2015)
- Consider a situation where a state adopts “Right to Work” union policies that affect firms’ ability to unionize
- If the policy results in changes in firm location in the contiguous untreated county, it means the untreated county may be treated
- Isn’t a problem necessarily – after all, it just means the treatment has spillovers next door and that may be *be the treatment*.
- But it does have implications for how you define the treatment and who is used as a comparison (maybe can’t use the neighbor as the control)

# SUTVA II: No Anticipation

- Spillovers can occur in space, but also time – especially challenging with forward looking economic agents (i.e., the ones we assume in economics and other social sciences)
- If a firm knows that a law has passed, but not enforced, that *information about the future costs/benefits* could already lead to entry, exit or marginal adjustments affecting outcomes
- DiD uses as its comparison a baseline, and you want that baseline to be untreated
- Some of this can be checked by rolling back the baseline to check
- Again, be careful about control groups both when the “pre” is chosen as well as the “across”

# SUTVA III: No Hidden Variation in Treatment

- Less widely known element of the SUTVA assumption is homogenous *dosages*
- Note that the potential outcomes are $Y^1$ and $Y^0$ – the ones and zeroes in other words have a presumably homogenous meaning across units (i.e., not $Y^{0.5}$ and $Y^1$)
- Aspirin example – one person gets an aspirin that "works" while another person gets one that is messed up and doesn't work
- Laws are often grouped that may not be exactly the same
- Not necessarily a problem – you may have to consider multiple treatments – but these may pose some challenges according to newer work

# SUTVA IV: Rising Cost and General Equilibrium

- Texas has luck with an early childhood program which raises longrun math test scores for 4th graders
- Biden and US legislators want to copy it
- But the program doesn't easily scale because there aren't enough teachers and we have to hire deeper and deeper into higher cost or lower quality teachers
- It's a kind of hidden variation in treatment associated with rising cost industries
- Similar types of issues if in general equilibrium what is observed in partial equilibrium is not the same

# Let's begin with DiD

- With all this out of the way, let's dig into the DiD material
- We will start with the simplest situation using simple difference in means without covariates
- We will then move into OLS with covariates
- And then move into alternatives to OLS when we have covariates
- Later we go into the more advanced material (e.g., differential timing, continuous treatments)