

# Causal Inference II

MIXTAPE SESSION

---



# Roadmap

## Introduction

Managing expectations

Origins of diff-in-diff in public health

Potential outcomes

Identification, Estimation and Inference

Design Stage

## Parallel Trends Violations

Results versus Evidence

Compositional Changes and Cross Sections

Event Studies

Triple differences

Falsifications

## Conclusion

# Introduction

- Thank you coming to Mixtape Sessions' Causal Inference 2 workshop on difference-in-differences!
- Scott Cunningham, Ben H. Williams Professor of Economics at Baylor University in Texas USA
- 9:00am to 5:00pm CST, 15 min breaks every hour, 1 hour lunch at 12:00
- Lecture, discussion, exercises, application with "live coding" exercises

# What my pedagogy is like

- High energy, eclectic approach to teaching
- Move between the econometrics, history of thought, videos, applications, code, spreadsheets, exercises
- Workshop is intended to take someone from knowing nothing about difference-in-differences to a broad level of competency on advanced topics
- Ask questions at any point; I'll do my best to answer them but will sometimes defer it to the break

# Class goals

Pedagogical goal is to break down the procedures into plain English, rebuilding it into something you can and want to use, but also:

1. **Confidence:** You will feel like you have a good enough understanding of diff-in-diff, both in its basics and some more contemporary issues, so that it seems like an intuitive and useful tool you could imagine using
2. **Comprehension:** You will have learned a lot both conceptually and in the specifics, particularly with regards to issues around identification and estimation in the diff-in-diff
3. **Competency:** You will have more knowledge of programming syntax in Stata and R so that later you can apply this in your own work

# Day 1 outline

Basics of difference-in-differences (Beginners but also my own personal take on this)

- Potential outcomes review and the ATT parameter
- DiD equation (“four averages and three differences”), parallel trends and estimation with OLS specification and inference
- Simple exercises illustrating potential outcomes, parameters, diff-in-diff and parallel trends
- Five step evidence of difference-in-differences
- Event study
- Triple differences (design) vs Falsifications (not a design)
- **Homework!**

# Day 2 outline

## Violations of parallel trends

- Repeated cross sections and compositional change
- Why do we include covariates?
  - Heckman, Ichimura and Todd (1997)
  - Abadie (2005)
  - Sant'Anna and Zhao (2020)
  - Twoway fixed effects with time varying controls
- Lalonde coding exercise

# Day 3 outline

## Pathologies of and Solutions to Two-way Fixed Effects with Differential Timing

- Panel fixed effects and strict exogeneity (Cunningham 2021)
- “Old school” example using Cheng and Hoekstra (2013) “Castle Doctrine” paper
- Bacon decomposition (Goodman-Bacon 2021)
- Aggregate ATT( $g,t$ ) method (Callaway and Sant’Anna 2021)
- Simple illustrations in simulations
- Walk through R shiny app

# Day 4 outline

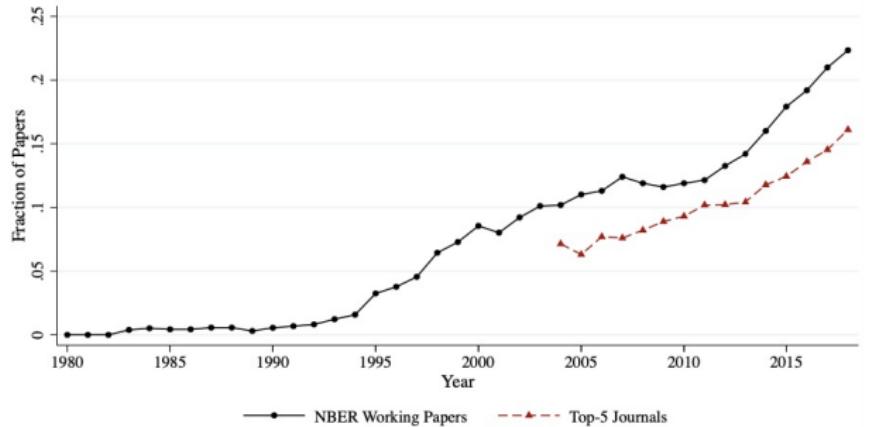
## Event studies and staggered, Imputation estimators

- Decomposition event study leads and lags under staggered adoption  
(Sun and Abraham 2021)
- Imputation estimators
  - Borusyak, et al. (2023) “robust efficient imputation estimator”
  - Gardner, et al. (2023) two stage difference in differences
- Bringing it together with an application from my research

# Growing popularity in economics

Figure: Currie, et al. (2020)

## A: Difference-in-Differences



With some exception (e.g., Heckman, Ichimura and Todd 1997; Abadie 2005; Bertrand, Duflo and Mullainathan 2004), econometricians had not given it much notice

# What is difference-in-differences (DiD)

- DiD is when a group of units are assigned some treatment and then compared to a group of units that weren't before and after
- One of the most widely used quasi-experimental methods in economics and increasingly in industry
- Predates the randomized experiment by 80 years, but uses basic experimental ideas about treatment and control groups (just not randomized)
- Uses panel or repeated cross section datasets, binary treatments usually, and often covariates
- We'll do a quick run through the social history of diff-in-diff to set the stage for our workshop this week

# Ignaz Semmelweis and washing hands

- Early 1820s, Vienna passed legislation requiring that if a pregnant women giving birth went to a public hospital (free care), then depending on the day of week and time of day, she would be routed to either the midwife wing or the physician wing (most likely resulting in random assignment)
- But by the 1840s, Ignaz Semmelweis noticed that pregnant women died after delivery in the (male) wing at a rate of 13-18%, but only 3% in the (female) midwife wing – cause was puerperal or “childbed” fever
- Somehow this was also well known – women would give birth in the street rather than go to the physician if they were unlucky enough to have their water break on the wrong day and time

# Ignaz Semmelweis and washing hands

- Ignaz Semmelweis conjectures after a lot of observation that the cause is the teaching faculty teaching anatomy using cadavers and then delivering babies *without washing hands*
- New training happens to one but not the other and Semmelweis thinks the mortality is caused by working with cadavers
- Convinced the hospital to have physicians wash their hands in chlorine but not the midwives, creating a type of difference-in-differences design

# Semmelweis diff-in-diff evidence



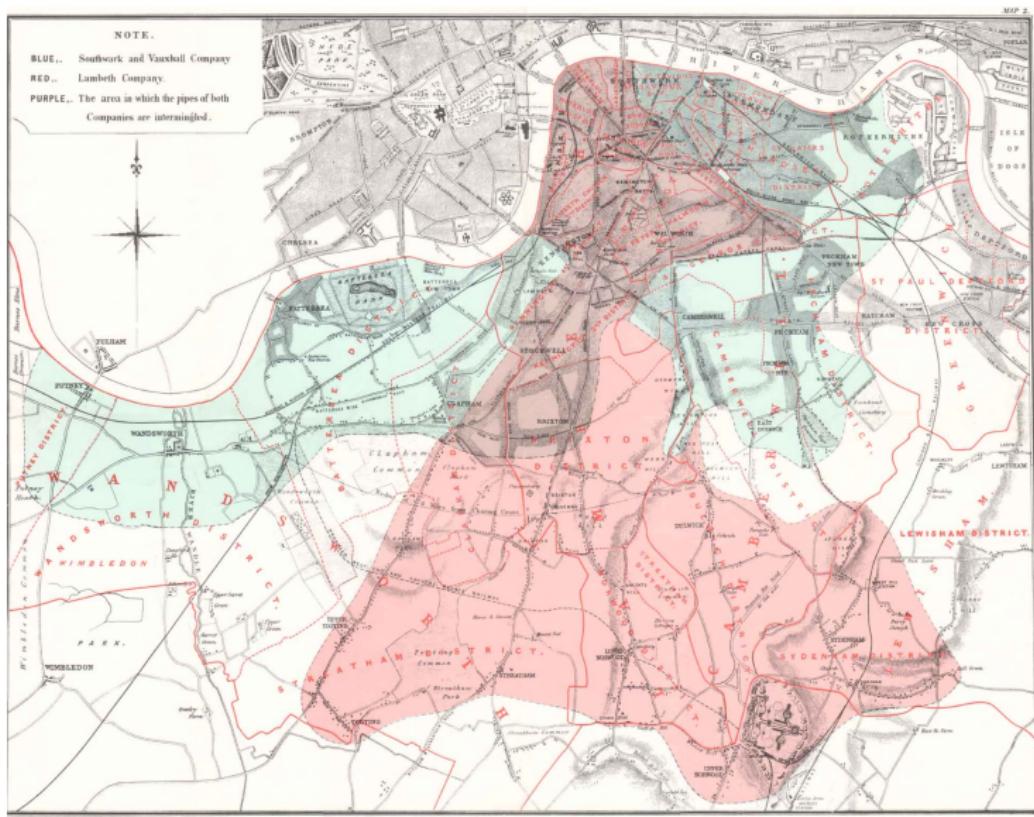
# Evidence Rejected

- Diff-in-diff evidence was rejected by Semmelweis' superiors claiming it was the hospital's new ventilation system
- Dominant theory of disease spread was caused by "odors" or miasma or "humors"
- Semmelweis began showing signs of irritability, perhaps onset of dementia, became publicly abusive, was committed to a mental hospital and within two weeks died from wounds he received while in residence
- Despite the strength of evidence, difference-in-differences was rejected – a theme we will see continue

# John Snow and cholera

- Three major waves of cholera in the early to mid 1800s in London, largely thought to be spread by miasma ("dirty air")
- John Snow believed cholera was spread through the Thames water supply through an invisible creature that entered the body through food and drink, caused the body to expel water, placing the creature back in the Thames and causing epidemic waves
- London passes ordinance requiring water utility companies to move inlet pipe further up the Thames, above the city center, but not everyone complies
- Natural experiment: Lambeth water company moves its pipe between 1849 and 1854; Southwark and Vauxhall water company delayed

*Figure: Two water utility companies in London 1854*



# Difference-in-differences

Table: Lambeth and Southwark and Vauxhall, 1849 and 1854

Companies	Time	Outcome	$D_1$	$D_2$
Lambeth	Before	$Y = L$		
	After	$Y = L + L_t + D$	$L_t + D$	
Southwark and Vauxhall	Before	$Y = SV$		$D + (L_t - SV_t)$
	After	$Y = SV + SV_t$	$SV_t$	

$$\hat{\delta}_{did} = D + (L_t - SV_t)$$

This method yields an unbiased estimate of  $D$  if  $L_t = SV_t$ , but note that  $L_t$  is a counterfactual trend and therefore not known

## Two rivers into causal inference

**Orley Ashenfelter**

↓  
Princeton Industrial Relations Section

↓  
Quasi-Experimental Design

↓  
David Card

↓  
Alan Krueger

↓

**Don Rubin**

↓  
Harvard Statistics

↓  
Experimental Design

↓  
Potential Outcomes

↓  
Treatment Effects

↓



**Harvard Economics**

## Background I: Harvard Stats and Potential Outcomes

- Don Rubin, former chair of Harvard stats, is the main source of potential outcomes, building on Jerzy Neyman's 1923 work.
- Rubin's influential 1970s papers advocated for causal inference using contrasts of  $Y(1)$  and  $Y(0)$ .
- Neyman's notation, initially in Polish, was translated into English in 1990, likely due to Rubin.
- Rubin expanded Neyman's ideas from experiments to observational studies, leading to developments like propensity score methods.
- Economics was slow to adopt these methods initially.

## Background II: Princeton Industrial Relations Section

- Late 1970s and early 1980s: little “credibility” in empirical labor studies.
- Princeton Industrial Relations Section: older than the economics dept, rigorous, non-partisan focus on US “manpower”, highly empirical.
- Key faculty are Orley Ashenfelter, David Card, Alan Krueger.
- Key students include Bob Lalonde, Josh Angrist, Steve Pischke, John Dinardo, Janet Currie, Anne Case, and many more.
- Listen to David Card:  
[https://youtu.be/1soLdywFb\\_Q?si=BCVqYeRz6jYiwHTQ&t=1580](https://youtu.be/1soLdywFb_Q?si=BCVqYeRz6jYiwHTQ&t=1580)

## Background II: Princeton Industrial Relations Section

### Example of Princeton Paradigm Emerging

- Lalonde (1986) was a groundbreaking study, recently reviewed by Guido Imbens and Yiqing Xu (2024)
- Lalonde, a student of Card and Orley, analyzed an RCT on a job training program, finding an average treatment effect of +\$800.
- He then replaced the experimental control group with survey data, reran econometric methods, and couldn't replicate the results.
- Orley and Card emphasized randomization in their 1985 Restat article, advocating for its exploitation in studies.

## Orley Ashenfelter and diff-in-diff

- Diff-in-diff gets rediscovered by Orley Ashenfelter from Princeton
- Leaves academia to work in Washington DC to study job training programs for low skill workers
- Coins the phrase "difference-in-differences" so as to avoid having to explain regressions to bureaucrats (3:53)  
<https://youtu.be/WnB3EJ8K7lg?si=uE4clqUIPzvbxm0r&t=2>
- More associated with David Card (Mariel boatlift, minimum wage), but it was earlier that he and Orley worked with the method, and ironically largely, rejected its usefulness for the questions they were working on

# Identification vs Estimation

- We must start by making a distinction between the parameter we are attempting to identify and the manner in which we will estimate it
- Identification requires first stating explicitly our goal expressed using potential outcomes
- But often people skip this step and go directly to the numerical calculation like Orley was doing so let's do that

# Four Averages and Three Subtractions

$$Y_{ist} = \alpha_0 + \alpha_1 Treat_{is} + \alpha_2 Post_t + \delta(Treat_{is} \times Post_t) + \varepsilon_{ist}$$

$$\widehat{\delta} = \left( \bar{y}_k^{post(k)} - \bar{y}_k^{pre(k)} \right) - \left( \bar{y}_U^{post(k)} - \bar{y}_U^{pre(k)} \right)$$

- Orley claims that the OLS estimator of  $\delta$  and the “four averages and three subtractions” calculation are numerically identical  
<https://youtu.be/WnB3EJ8K7lg?t=126>
- They are and I want you to see that with a numerical example in equivalence.do and equivalence.R

# Introducing Potential Outcomes to DiD

- Research question versus causal question – not the same thing
- Research question would be you are wanting to know effect of job training programs on earnings
- Causal question is expressed using potential outcomes
- Causal questions are usually averages of individual treatment effects for a specific population of units

## Potential outcomes notation

- Let the treatment be a binary variable:

$$D_{i,t} = \begin{cases} 1 & \text{if in job training program } t \\ 0 & \text{if not in job training program at time } t \end{cases}$$

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

## Potential outcomes notation

- Potential outcomes:

$$Y_{i,t}^j = \begin{cases} 1: & \text{wages at time } t \text{ if trained} \\ 0: & \text{wages at time } t \text{ if not trained} \end{cases}$$

where  $j$  indexes a state of the world where the treatment happened or did not happen

# Treatment effect definitions

## Individual treatment effect

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

Missing data problem: No data on the counterfactual

# Average Treatment Effects for the Treated Subpopulation

## 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:

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

It's the average causal effect but only for the people exposed to some intervention; notice we can't calculate it, also, because we are missing the red term

## ATT vs ATE

- Imagine there are 100 cities – 25 of them raise the minimum wage, 75 do not
  - ATE is the average treatment effect across all 100 cities
  - ATT is the average treatment effect for the 25 cities that raised the minimum wage
- If you want to know the ATE, diff-in-diff is not appropriate as it only identifies the ATT
- Just keep in mind the *population* – you can average the treatment effects for everyone or just some people, but whichever affects the interpretation

# Potential Outcomes vs Realized Outcomes

- When you're analyzing data, you are working with the realized outcomes
- When you are thinking about your treatment parameters, you are working with potential outcomes
- When treatment does occur, then only one of the potential outcomes happens ("realized outcome") but the other one becomes "counterfactual" or missing
- We represent this with the switching equation

$$Y_{it} = D_{it}Y_{it}^1 + (1 - D_{it})Y_{it}^0$$

- How those treatments get assigned is called the treatment assignment mechanism, but let's review this together first before we discuss that

# Simple spreadsheet exercise

Let's review basic concepts here at "WEIGHTS" tab:

[https://docs.google.com/spreadsheets/d/10DuQqGtH\\_Ewea7zQoLTFYHbnvqaTVDhn2GDzq30a6EQ/edit?usp=sharing](https://docs.google.com/spreadsheets/d/10DuQqGtH_Ewea7zQoLTFYHbnvqaTVDhn2GDzq30a6EQ/edit?usp=sharing)

DiD equation is the 2x2

Orley's "four averages and three subtractions" uses two groups, two time periods, or 2x2

$$\hat{\delta} = \left( E[Y_k|Post] - E[Y_k|Pre] \right) - \left( E[Y_U|Post] - E[Y_U|Pre] \right)$$

$k$  are the people in the job training program,  $U$  are the untreated people not in the program,  $Post$  is after the trainees took the class,  $Pre$  is the period just before they took the class, and  $E[y]$  is mean earnings.

When will  $\hat{\delta}$  equal the ATT? When will it not?

Replace with potential outcomes and add a zero

$$\hat{\delta} = \underbrace{\left( E[Y_k^1|Post] - E[Y_k^0|Pre] \right) - \left( E[Y_U^0|Post] - E[Y_U^0|Pre] \right)}_{\text{Switching equation}} + \underbrace{E[Y_k^0|Post] - E[Y_k^0|Post]}_{\text{Adding zero}}$$

## Parallel trends bias

$$\hat{\delta} = \underbrace{E[Y_k^1|Post] - E[Y_k^0|Post]}_{\text{ATT}} + \underbrace{\left[ E[Y_k^0|Post] - E[Y_k^0|Pre] \right] - \left[ E[Y_U^0|Post] - E[Y_U^0|Pre] \right]}_{\text{Non-parallel trends bias in 2x2 case}}$$

# Identification through parallel trends

## Parallel trends

Assume two groups, treated and comparison group, then we define parallel trends as:

$$E(\Delta Y_k^0) = E(\Delta Y_U^0)$$

**In words:** “The evolution of earnings for our trainees *had they not trained* is the same as the evolution of mean earnings for non-trainees”.

It's in red because parallel trends is untestable and critically important to estimation of the ATT using any method, OLS or “four averages and three subtractions”

## Design stage in diff-in-diff

- There is a period of time before you ever touch your data on outcomes where you design your study
- You want to be thinking in this stage of things like this:
  - What does a treatment effect mean in my study?
  - Diff-in-diff identifies the ATT – what does that mean in my study?
  - If I ran a randomized experiment on my population, what would I be randomizing?
  - Do I think unconditional parallel trends (i.e., no covariates) holds? Why?
- Because many people come to causal inference from outside the experimental design tradition, this step is confusing, but keep this picture in mind

## Two rivers into causal inference

**Orley Ashenfelter**

↓  
Princeton Industrial Relations Section

↓  
Quasi-Experimental Design

↓  
David Card

↓  
Alan Krueger

↓

**Don Rubin**

↓  
Harvard Statistics

↓  
Experimental Design

↓  
Potential Outcomes

↓  
Treatment Effects

↓



**Harvard Economics**

# Causal Inference in Design Tradition

- So you have to learn the design tradition and adopt its approach
- Does not mean you abandon your prior human capital
- Practical implication is that you spend more time in the design stage thinking about treatment assignment mechanism, covariate imbalance
- But it also means you don't "peek at the outcomes" to avoid biasing *yourself* until you're ready
- Imagine you were pre-registering an RCT – try to mimic that even in your diff-in-diff

## Selection and Parallel Trends

- When designing a study, it's essential to think about how some units were treated and others were not.
- This means conducting thorough research on who got treated, who didn't, and who knew what during the process.
- Ghanem, Sant'Anna, and Wüthrich (2023) suggest that there are limited theoretical mechanisms that can justify parallel trends.

# Parallel Trends is Consistent with Certain Assignment Mechanisms

## 1. Selection on pre-treatment covariates

- Assignment to treatment depends on observed characteristics before treatment, allowing us to condition on these covariates when assessing trends.

## 2. Selection on fixed effects

- Treatment is assigned based on characteristics that do not change over time, such as inherent traits or long-term conditions.

## 3. Selection on imperfect beliefs about returns

- Units select into treatment based on their expectations, which may not be accurate. This can create an imperfect belief-based selection.

## 4. Martingale property

- The change in outcomes follows a martingale process, where future changes in unobservable factors are independent of past ones, under certain selection conditions.

# What We Should Be Concerned About

- **Time-varying unobservables**
  - Unobservable factors that change over time can create biases if they affect treated and control groups differently.
  - This can lead to violations of the parallel trends assumption if such unobservables are correlated with treatment.
- **Selection based on post-treatment information**
  - If individuals select into treatment based on information they gain after the treatment starts (foresight), this undermines the assumption.
  - Essential heterogeneity means that the potential outcomes vary systematically with unobserved traits, which complicates inference.

# What We Should Be Concerned About

- **Non-stationarity**
  - Non-stationarity refers to changing relationships between covariates and outcomes over time, which can lead to inconsistent treatment effect estimates.
  - Trends in the data that are not stable or stationary across time are particularly problematic.
- **Martingale property violation**
  - If the unobservable shocks affecting outcomes do not follow a martingale process, meaning future changes are not independent of past outcomes, the model's assumptions may fail.
  - This can result in time-dependent correlations in unobservable factors that distort trend comparisons.

## What is and is not parallel trends?

- Parallel trends does *not* mean treatments were randomly assigned (though random assignment guarantees parallel trends)
- Parallel trends does *not* require that the groups be similar at baseline on outcomes (though random assignment guarantees that would be)
- Parallel trends does require that the comparison group follows a trend in outcomes that is approximately the same as the counterfactual trend of the treatment group (what would have had happened had the treatment not occurred)

# Understanding parallel trends through worksheets

Before we move into regression, let's go through a simple exercise to really pin down these core ideas with simple calculations

[https://docs.google.com/spreadsheets/d/  
1onabpc14JdrGo6NFv0zCWo-nuWDLLV2L1qNogDT9SBw/edit?usp=  
sharing](https://docs.google.com/spreadsheets/d/1onabpc14JdrGo6NFv0zCWo-nuWDLLV2L1qNogDT9SBw/edit?usp=sharing)

## Three main DiD assumptions

- Parallel trends is the most common one and most well known
- But parallel trends is nested within a bundle of assumptions, and all of them are needed for traditional difference-in-differences
- Other two lesser known assumptions are "No anticipation" (or NA) and Stable Unit Treatment Value Assumption (SUTVA)

# No Anticipation

- “No anticipation” simply means that the unit is not treated until it is treated (and that can be violated with rational forward looking agents but not always)
  - **Example 1:** Tomorrow I win the lottery, but don’t get paid yet. I decide to buy a new house today. That violates NA
  - **Example 2:** Next year, a state lets you drive without a driver license and you know it. But you can’t drive without a driver license today. This satisfies NA.
- Violations in simple 2x2 where baseline is treated creates problems that I discuss here:

[https://causalinf.substack.com/p/  
difference-in-differences-no-anticipation](https://causalinf.substack.com/p/difference-in-differences-no-anticipation)

[https://causalinf.substack.com/p/  
mixtape-mailbag-7-what-happens-in](https://causalinf.substack.com/p/mixtape-mailbag-7-what-happens-in)

## No Anticipation Violation

$$\hat{\delta} = \left( E[Y_k|Post] - E[Y_k|Pre] \right) - \left( E[Y_U|Post] - E[Y_U|Pre] \right)$$

What if the  $k$  group had been treated at the baseline “pre” period in our 2x2?

Let's do this one together.

## No Anticipation Violation

If the baseline period is treated, then the simple 2x2 identifies the following three terms:

$$\begin{aligned}\delta &= ATT_k(Post) \\ &\quad + \text{Non PT bias} \\ &\quad - ATT_k(Pre)\end{aligned}$$

First row is the ATT in the post period; middle row is parallel trends; third row subtracts the baseline ATT from the calculation. If treatment effects are constant, then the DiD coefficient will be zero despite positive treatment effects. Let's look in `na.do`.

# SUTVA

- Stable Unit Treatment Value Assumption (Imbens and Rubin 2015) focuses on what happens when in our analysis we are combining units (versus defining treatment effects)
  1. **No Interference:** a treated unit cannot impact a control unit such that their potential outcomes change (unstable treatment value)
  2. **No hidden variation in treatment:** When units are indexed to receive a treatment, their dose is the same as someone else with that same index
  3. **Scale:** If scaling causes interference or changes inputs in production process, then #1 or #2 are violated
- Shifts from defining treatment effects to estimating them, which means being careful about who is the control group, how you define treatments and what questions can and cannot be answered with this method

# Setup for SUTVA

$$\begin{aligned}\hat{\delta} &= \underbrace{\left( E[Y_k|Post] - E[Y_k|Pre] \right) - \left( E[Y_U|Post] - E[Y_U|Pre] \right)}_{\text{Diff-in-diff}} \\ &= \underbrace{\left( E[Y_k^1|Post] - E[Y_k^0|Pre] \right) - \left( E[\textcolor{blue}{Y_U^0}|Post] - E[Y_U^0|Pre] \right)}_{\text{Switching equation with SUTVA in blue}} \\ &\quad + \underbrace{E[Y_k^0|Post] - E[Y_k^0|Post]}_{\text{Adding zero}}\end{aligned}$$

# Illustrate the role of SUTVA

Rearrange

$$\hat{\delta} = \underbrace{E[Y_k^1|Post] - E[Y_k^0|Post]}_{\text{ATT}} + \underbrace{\left[ E[Y_k^0|Post] - E[Y_k^0|Pre] \right] - \left[ E[Y_U^0|Post] - E[Y_U^0|Pre] \right]}_{\text{Non-parallel trends bias in 2x2 case}}$$

Notice how NA and SUTVA are key for making DiD be this simple expression.

Do not use already treated controls

$$\widehat{\delta} = \left( E[Y_k|Post] - E[Y_k|Pre] \right) - \left( E[Y_U|Post] - E[Y_U|Pre] \right)$$

What if the  $U$  group had always been treated in both periods? Is parallel trends enough to identify the ATT?

Similar to SUTVA but this time it isn't caused by a spillover – it's just an already treated group. Let's do this together now. It'll help to see it with your own eyes.

## Already Treated Control Group

If the baseline period is treated, then the simple 2x2 identifies the following three terms:

$$\begin{aligned}\delta &= ATT(Post) \\ &\quad + \text{Non PT bias} \\ &\quad - \Delta ATT_k\end{aligned}$$

Again, first row is the target parameter, plus parallel trends term, minus the changing ATT in our control group

# Summarizing

- Lots of restrictions placed on difference-in-differences
  - NA: you chose a baseline that is not treated
  - SUTVA: your comparison group is never treated during the course of the calculations
  - PT: your comparison group has a trend in  $E[Y^0]$  that is the same as the counterfactual
- Only when you have NA SUTVA and you use an untreated group as a control does DiD equal ATT + PT
- But it's crucial to remember: DiD and ATT are not the same thing

# OLS Specification

- OLS specification is the same as four averages and three subtractions
- Assume unconditional parallel trends (i.e., no covariates are necessary), then OLS might be preferred because ...
  - OLS estimates the ATT under parallel trends
  - Easy to calculate the standard errors
  - Easy to include multiple periods
- People liked it also because of differential timing, continuous treatments and covariates, but those are more complex so we address them later

# Minimum wages

- Card and Krueger (1994) have a famous study estimating causal effect of minimum wages on employment
- New Jersey raises its minimum wage in April 1992 (between February and November) but neighboring Pennsylvania does not
- Using DiD, they do not find a negative effect of the minimum wage on employment leading to complex reactions from economists
- Orley's describes his understanding of people's reaction to the paper.  
<https://youtu.be/M0tbuRX4eyQ?t=1882>



Binyamin Appelbaum

@BCAppelbaum



Replies to @BCAppelbaum

The Nobel laureate James Buchanan wrote in the Wall Street Journal that Card and Krueger were undermining the credibility of economics as a discipline. He called them and their allies "a bevy of camp-following whores."

3:49 PM · Mar 18, 2019



179



Reply



Share

[Read 18 replies](#)

## Reaction to the paper

Lots of anecdotes in this interview with Card, but here are just two. First, Card and Krueger received a lot of personal hostility from their peers (1:07 to 1:10)

[https://youtu.be/1soLdywFb\\_Q?si=laAVYf\\_E2KBZKywG&t=4020](https://youtu.be/1soLdywFb_Q?si=laAVYf_E2KBZKywG&t=4020)

Later Card says Sherwin Rosen accused them of having an agenda. But the worst is what happens to Alan Krueger maybe (1:16 to 1:17)

[https://youtu.be/1soLdywFb\\_Q?si=jsb8h50ZosGDnKrv&t=4556](https://youtu.be/1soLdywFb_Q?si=jsb8h50ZosGDnKrv&t=4556)

## Card on that study

*"I've subsequently stayed away from the minimum wage literature for a number of reasons. First, it cost me a lot of friends. People that I had known for many years, for instance, some of the ones I met at my first job at the University of Chicago, became very angry or disappointed. They thought that in publishing our work we were being traitors to the cause of economics as a whole."*

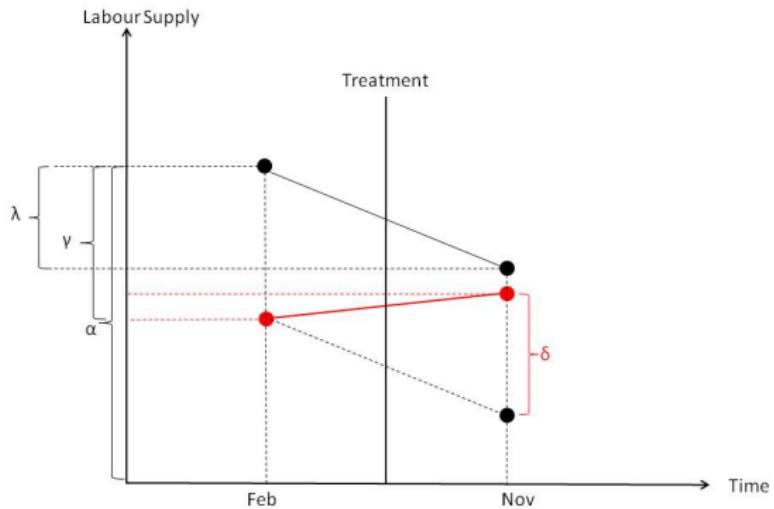
# OLS specification of the DiD equation

- The correctly specified OLS regression is an interaction with time and group fixed effects:

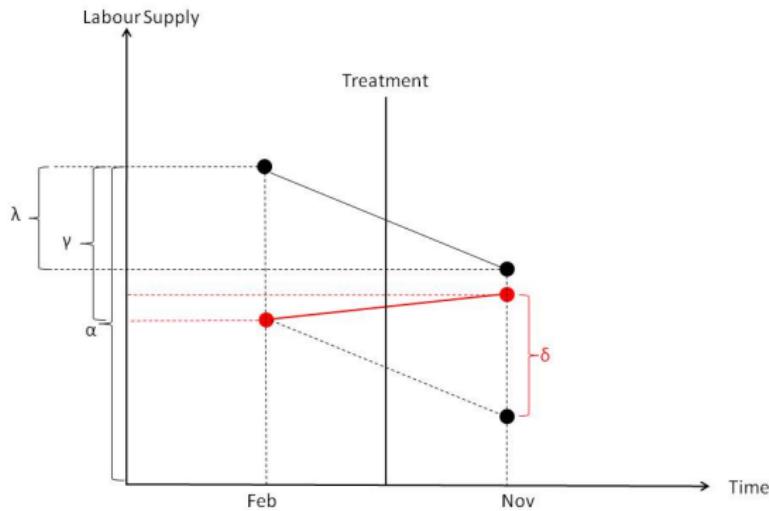
$$Y_{its} = \alpha + \gamma NJ_s + \lambda d_t + \delta(NJ \times d)_{st} + \varepsilon_{its}$$

- NJ is a dummy equal to 1 if the observation is from NJ
- d is a dummy equal to 1 if the observation is from November (the post period)
- This equation takes the following values
  - PA Pre:  $\alpha$
  - PA Post:  $\alpha + \lambda$
  - NJ Pre:  $\alpha + \gamma$
  - NJ Post:  $\alpha + \gamma + \lambda + \delta$
- DiD equation:  $(NJ \text{ Post} - NJ \text{ Pre}) - (PA \text{ Post} - PA \text{ Pre}) = \delta$

$$Y_{ist} = \alpha + \gamma N J_s + \lambda d_t + \delta (N J \times d)_{st} + \varepsilon_{ist}$$



$$Y_{ist} = \alpha + \gamma NJ_s + \lambda d_t + \delta(NJ \times d)_{st} + \varepsilon_{ist}$$



Notice how OLS is “imputing”  $E[Y^0|D = 1, Post]$  for the treatment group in the post period? It is only “correct”, though, if parallel trends is a good approximation

## Inference with correlated errors

- Correlated errors occur when the unobserved errors are correlated within a cluster.
- This violates the assumption of independent errors, leading to possibly biased standard errors and higher over rejection rates
- Failing to account for correlated errors can lead to misleading inference.

## Conservative inference in DiD

- Bertrand, Duflo and Mullainathan (2004) show that conventional standard errors will often severely underestimate the standard deviation of the estimators
- They proposed three solutions, but most only use one of them (clustering)
- Clustering standard errors accounts for this within-cluster correlation and is a more conservative approach
- Clustering is typically recommended at the aggregate unit where the entire treatment occurred

# Roadmap

## Introduction

Managing expectations

Origins of diff-in-diff in public health

Potential outcomes

Identification, Estimation and Inference

Design Stage

## Parallel Trends Violations

Results versus Evidence

Compositional Changes and Cross Sections

Event Studies

Triple differences

Falsifications

## Conclusion

# Evidence versus the Main Result

- Causal inference is about *warranted beliefs* – should you or should you not believe the *causal claim*?
- Your DiD *results* are like the claim of guilt, but your DiD results are *not* the smoking gun
  - Your table of regression coefficients *is not enough* for evidence
  - You need to do more to provide a justification for parallel trends
- You need to provide evidence for parallel trends against several well known vulnerabilities
- Evidence will be bite, falsifications, mechanisms and event study data visualization
- We will mix the parallel trends violations with the evidence concept before getting into advanced estimators

# Court metaphor

- Think of yourself as a prosecutor arguing against a defense attorney to convince a judge and jury of a defendant's guilt
- The claim the defendant is guilty is your table of main results
- But the claim is not the evidence – you have to back up that claim
- Your evidence of guilt is the smoking gun, the fingerprints, the eye witnesses, the footprints in the mud outside the house
- If your claim is supported by weak evidence, then no one *should* convict – it would be borderline corruption if they did

## Three classic parallel trends violations

1. Compositional change with repeated cross-sections
2. Policy endogeneity
3. Omitted covariates needed for parallel trends to hold

## Repeated cross-sections and compositional change

- One of the risks of a repeated cross-section is that the composition of the sample may have changed between the pre and post period in ways that are correlated with treatment
- Hong (2013) uses repeated cross-sectional data from the Consumer Expenditure Survey (CEX) containing music expenditure and internet use for a random sample of households
- Study exploits the emergence of Napster (first file sharing software widely used by Internet users) in June 1999 as a natural experiment
- Study compares internet users and internet non-users before and after emergence of Napster

# Introduction of Napster and spending on music

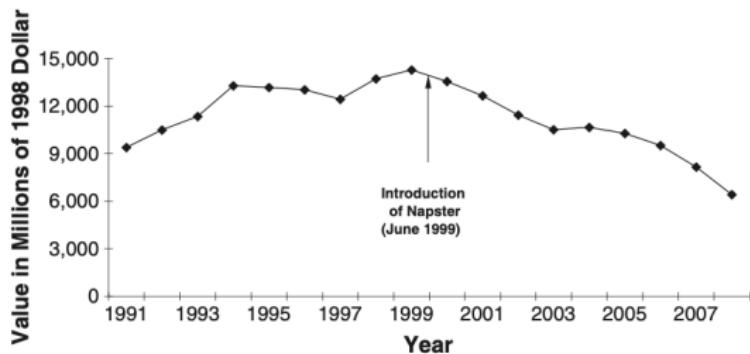


Figure 2. Total real value of record shipments in the USA. Refer to the RIAA's year-end statistics. Total sales include CDs, cassettes, LPs, and music videos. Starting from 2004, total sales also include digital formats such as legitimate download

Figure 1: Internet Diffusion and Average Quarterly Music Expenditure in the CEX

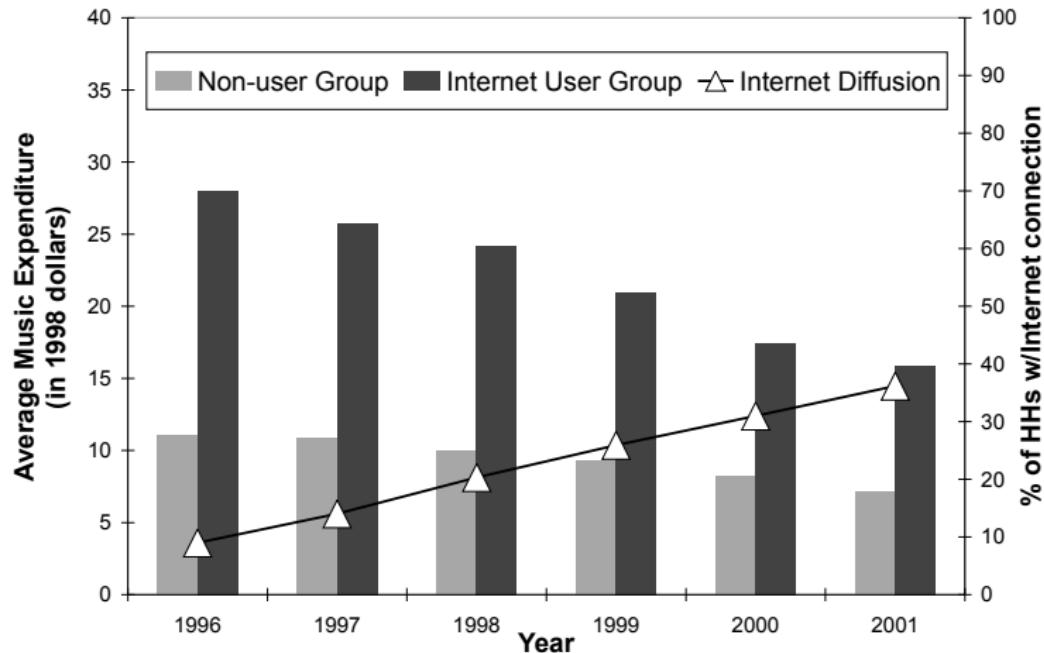


Table 1: Descriptive Statistics for Internet User and Non-user Groups<sup>a</sup>

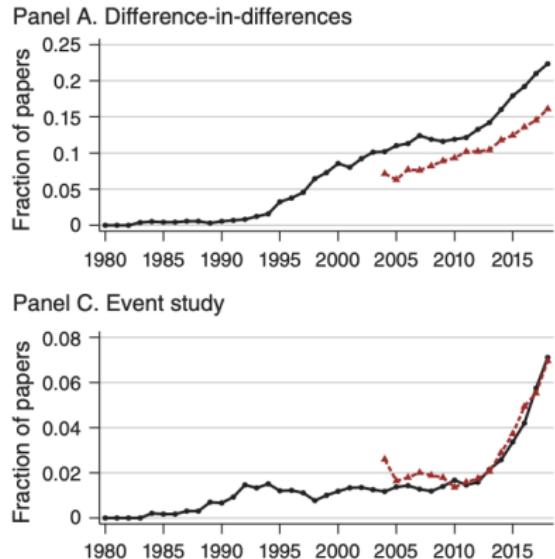
Year	1997		1998		1999	
	Internet User	Non-user	Internet User	Non-user	Internet User	Non-user
<b>Average Expenditure</b>						
Recorded Music	\$25.73	\$10.90	\$24.18	\$9.97	\$20.92	\$9.37
Entertainment	\$195.03	\$96.71	\$193.38	\$84.92	\$182.42	\$80.19
<b>Zero Expenditure</b>						
Recorded Music	.56	.79	.60	.80	.64	.81
Entertainment	.08	.32	.09	.35	.14	.39
<b>Demographics</b>						
Age	40.2	49.0	42.3	49.0	44.1	49.4
Income	\$52,887	\$30,459	\$51,995	\$28,169	\$49,970	\$26,649
High School Grad.	.18	.31	.17	.32	.21	.32
Some College	.37	.28	.35	.27	.34	.27
College Grad.	.43	.21	.45	.21	.42	.20
Manager	.16	.08	.16	.08	.14	.08

Diffusion of the Internet changes samples (e.g., younger music fans are early adopters)

## Repeated cross-sections

- Surprisingly underappreciated problem with almost no literature around it
- Replace the outcome with your time-varying covariates and estimate your DiD model
- Use covariates highly predictive of the missing  $E[Y^0|D = 1]$  for this exercise
- “Difference-in-differences with Compositional Changes” by Pedro Sant’Anna and Qi Xu (2023) is an update to Hong (2013)

# Event studies have become mandatory in DiD



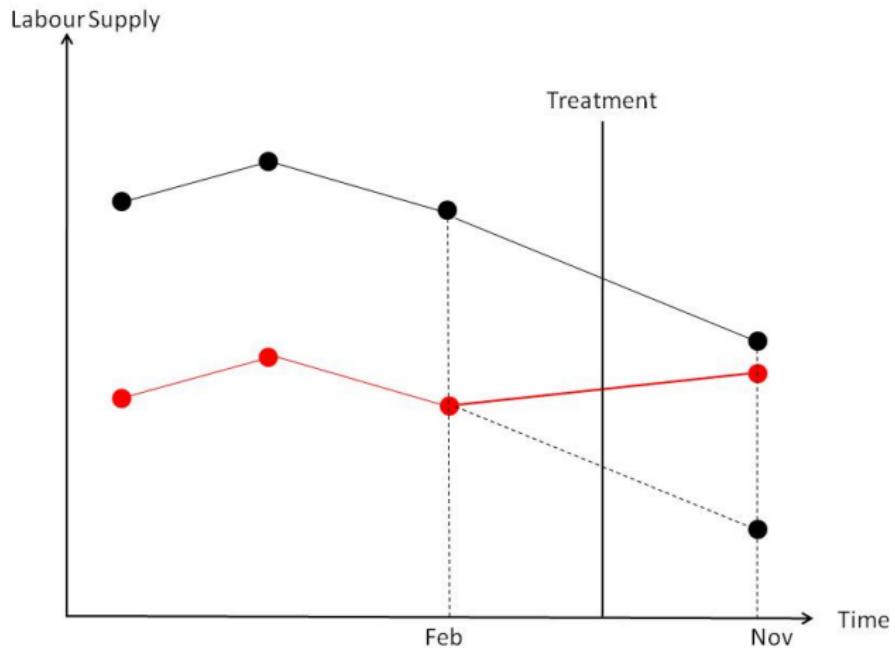
# Intuition behind event studies

- We cannot directly verify parallel trends, so for a long time researchers have focused on the pre-trends (Ashenfelter's Dip)
- Parallel pre-trends not required for parallel trends and vice versa, but this is the smoking gun we typically look for nonetheless
- Think of it as a type of check for selection bias, but imperfect with false positives and false negatives
- Even if pre-trends are the same one still has to worry about other policies changing at the same time (omitted variable bias is a parallel trends violation)

# Creating event studies

- Originally, there were no event studies (as we saw in the First Wave)
- Economists pulled from finance and took the event study concept and changed it to suit Ashenfelter Dip reasoning
- Always presented graphically, but there were different ways people went about it so we will review them and make suggestions

Plot the raw data when there's only two groups



# Event study regression

- Alternatively, present estimated coefficients from a dynamic regression specification:

$$Y_{its} = \alpha + \sum_{\tau=-2}^{-q} \mu_\tau (D_s \times \tau_t) + \sum_{\tau=0}^m \delta_\tau (D_s \times \tau_t) + \tau_t + D_s + \varepsilon_{ist}$$

- With a simple 2x2, you are interacting treatment indicator with calendar year dummies
- Includes  $q$  leads or anticipatory effects and  $m$  lags or post treatment effects
- Estimated  $\hat{\delta}$  coefficients are estimated ATT parameters assuming parallel trends and  $\hat{\mu}$  is part of your evidence for that

## Event study regression

- Typically you'll plot the coefficients and 95% CI on all leads and lags (binned or not, trimmed or not)
- Under No Anticipation, SUTVA and parallel pre-trends, then mechanically  $\widehat{\mu}_\tau$  will be zero (does not guarantee the same about post-trends)
- There are still specification and power issues that Jon Roth has written about, but I will skip that
- But also under NA, SUTVA and parallel trends (post trends), then  $\widehat{\delta}$  are estimates of the ATT at points in time

## Normal DiD coefficient

$$\hat{\delta} = \underbrace{E[Y_k^1|Post] - E[Y_k^0|Post]}_{\text{ATT}} + \underbrace{\left[ E[Y_k^0|Post] - E[Y_k^0|Pre] \right] - \left[ E[Y_U^0|Post] - E[Y_U^0|Pre] \right]}_{\text{Non-parallel trends bias in 2x2 case}}$$

But this was *post*-treatment. Still, put that aside – diff-in-diff equations always identify the sum of those terms, even in the pre-period

## Pre-treatment DiD coefficient

$$\hat{\delta}_{t-2} = \underbrace{\left[ E[Y_k^0|t-2] - E[Y_k^0|t-1] \right]}_{\text{Non-parallel trends bias in 2x2 case}} - \underbrace{\left[ E[Y_U^0|t-2] - E[Y_U^0|t-1] \right]}_{}$$

Under NA, then the  $t - 1$  period is untreated. But then so are the other pre-periods so the ATT is implicitly zero and the *only* thing that you can be measuring with pre-trend DiD coefficients is differential trends.

## Event study coefficients

- Remember that the OLS specification we discuss collapses to ATT plus parallel trends bias
- This is *always* true because it's an identity and holds even in the pre-period as much in the post
- It's just in the pre period, you do not have the missing  $E[Y^0|D = 1]$  term as no one and nothing is treated in pre-period under NA
- This means pre-period is basically an opportunity to directly verify parallel pre-trends – but it's the past's pre-trends, not the counterfactual pre-trend of the present/future
- And that's how people use the pre-period – they use the pre-period to evaluate whether they think this is a good control group

## Event study example

- The notion is really simple: if PT held then, you'll argue that it's reasonable it would've still held
- But this is an assertion, and you need to build the case as we said
- At this point, it's a lot easier to show you what I'm talking about – where the art and the science meet – with a great paper

# Medicaid and Affordable Care Act example



Volume 136, Issue 3  
August 2021

< Previous    Next >

## Medicaid and Mortality: New Evidence From Linked Survey and Administrative Data [Get access >](#)

Sarah Miller, Norman Johnson, Laura R Wherry

*The Quarterly Journal of Economics*, Volume 136, Issue 3, August 2021, Pages 1783–1829,

<https://doi.org/10.1093/qje/qjab004>

Published: 30 January 2021

[Cite](#) [Permissions](#) [Share ▾](#)

### Abstract

We use large-scale federal survey data linked to administrative death records to investigate the relationship between Medicaid enrollment and mortality. Our analysis compares changes in mortality for near-elderly adults in states with and without Affordable Care Act Medicaid expansions. We identify adults most likely to benefit using survey information on socioeconomic status, citizenship status, and public program participation. We find that prior to the ACA expansions, mortality rates across expansion and nonexpansion states trended similarly, but beginning in the first year of the policy, there were significant reductions in mortality in states that opted to expand relative to nonexpander states. Individuals in expansion states experienced a 0.132 percentage point decline in annual mortality, a 9.4% reduction over the sample mean, as a result of the Medicaid expansions. The effect is driven by a reduction in disease-related deaths and grows over time. A variety of alternative specifications, methods of inference, placebo tests, and sample definitions confirm our main result.

**JEL:** H75 - State and Local Government: Health; Education; Welfare; Public Pensions, I13 - Health Insurance, Public and Private, I18 - Government Policy; Regulation; Public Health

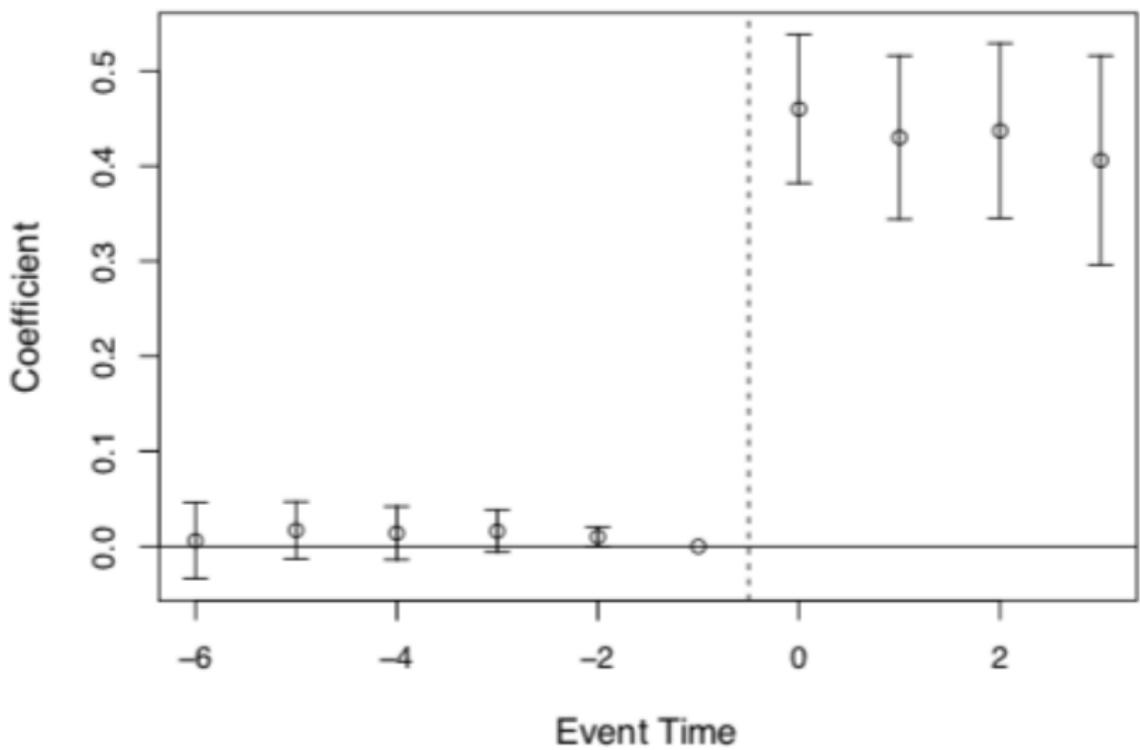
**Issue Section:** Article

# Their Evidence versus Their Result

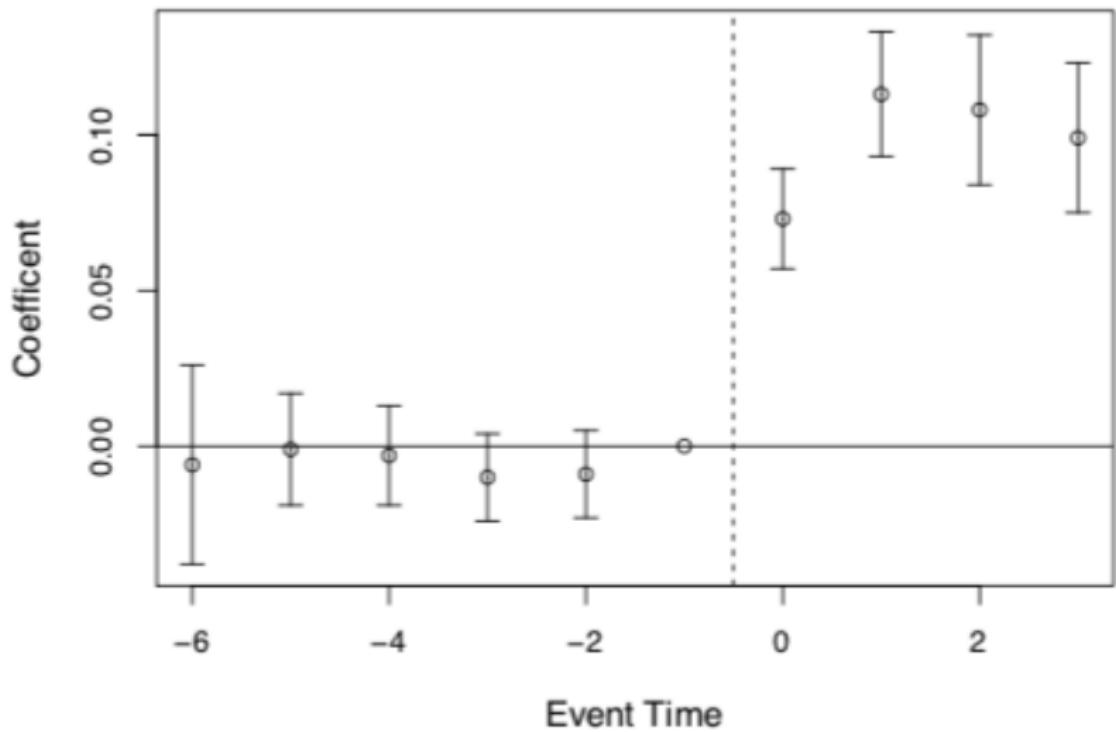
- **Bite** – they will show that the expansion shifted people into Medicaid and out of uninsured status
- **Placebos** – they show that there's no effect of Medicaid on a similar group that didn't enroll
- **Event study** – they will lean hard on those dynamic plots
- **Main results** – with all of this, they will show Medicaid expansion caused near elderly mortality to fall
- **Mechanisms** – they think they can show it's coming from people treating diseases causing mortality declines to compound over time

## Bite

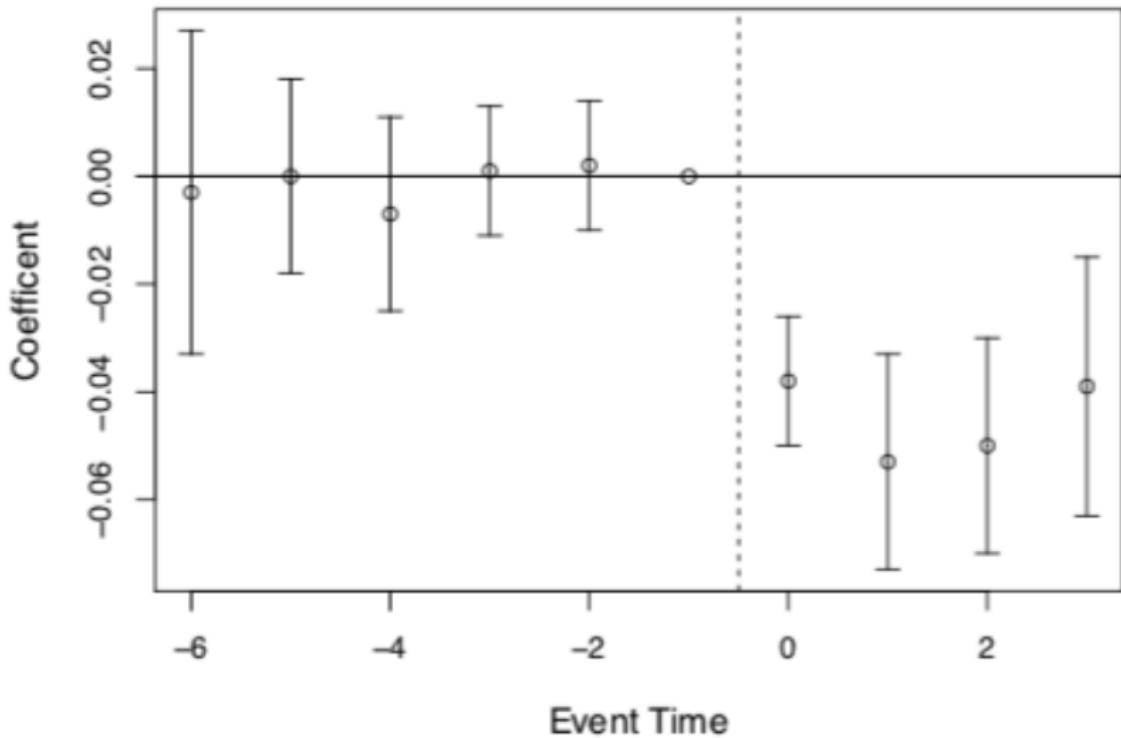
- Bite is a labor economist's phrase, often used with the minimum wage, to say that the minimum wage actually was binding in the first place
- Here it means when US states made Medicaid more generous, people got on Medicaid who would not have been on it otherwise
- And as a bonus, would not have been insured at all without it
- Not the most exciting result, but imagine if the main results on mortality were shown but there was no evidence for bite – is it believable?



(a) Medicaid Eligibility



(b) Medicaid Coverage



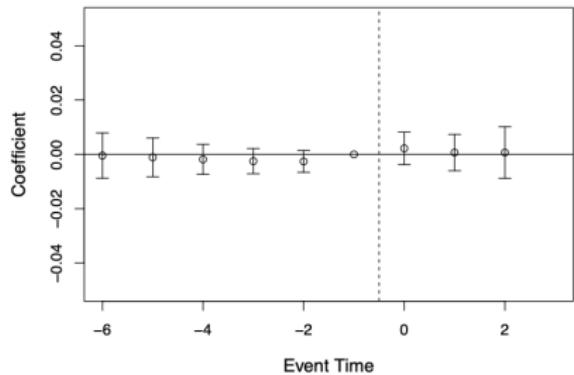
(c) Uninsured

# Falsification

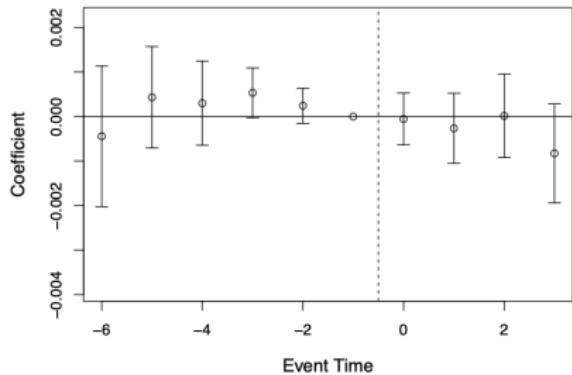
- Their study focuses on “near elderly”, which means just under 65
- They choose just under 65 because in the US, 65 and older are eligible for Medicare so more generous Medicaid is irrelevant
- *But* probably the near elderly and the elderly are equally susceptible to unobserved factors correlated with the treatment
- So they painstakingly examine the effects on elderly as a falsification as this will strengthen the parallel trends assumption on the near elderly

# Falsifications on elderly

*Age 65+ in 2014*



(c) Medicaid Coverage

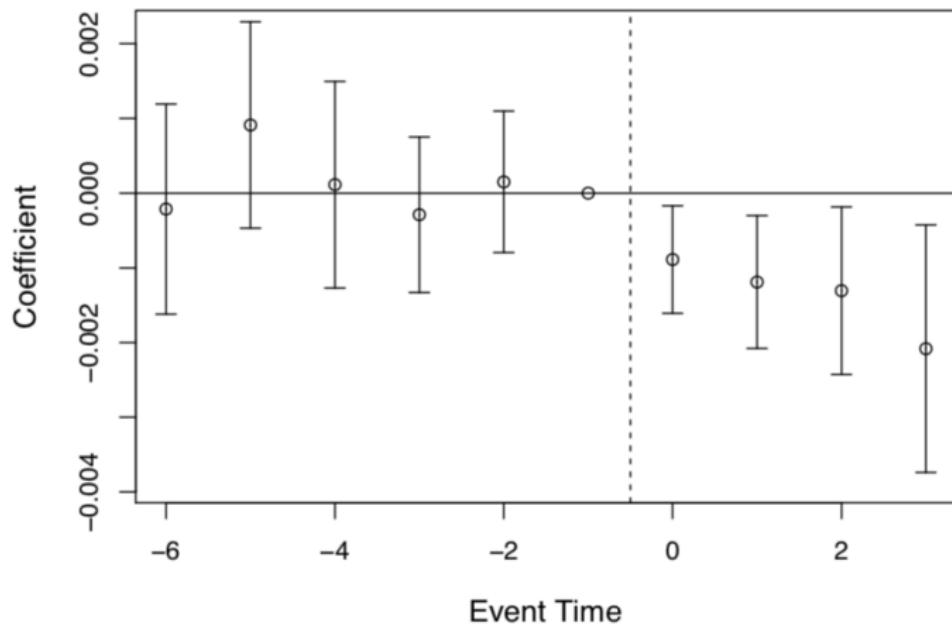


(d) Annual Mortality

## Main result

- Finally they focus on the main result – and there's more in the paper than I'm showing
- Event study plots with same specification as the rest allowing us to look at the pre-trends and the post-treatment coefficients
- If parallel trends holds, then the post-treatment coefficients are interpreted as ATT parameter estimates for each time period
- The result alone isn't nearly as strong the result in combination with the rest, but it could still be wrong as parallel trends is ultimately not verifiable

# Near elderly mortality and Medicaid expansion



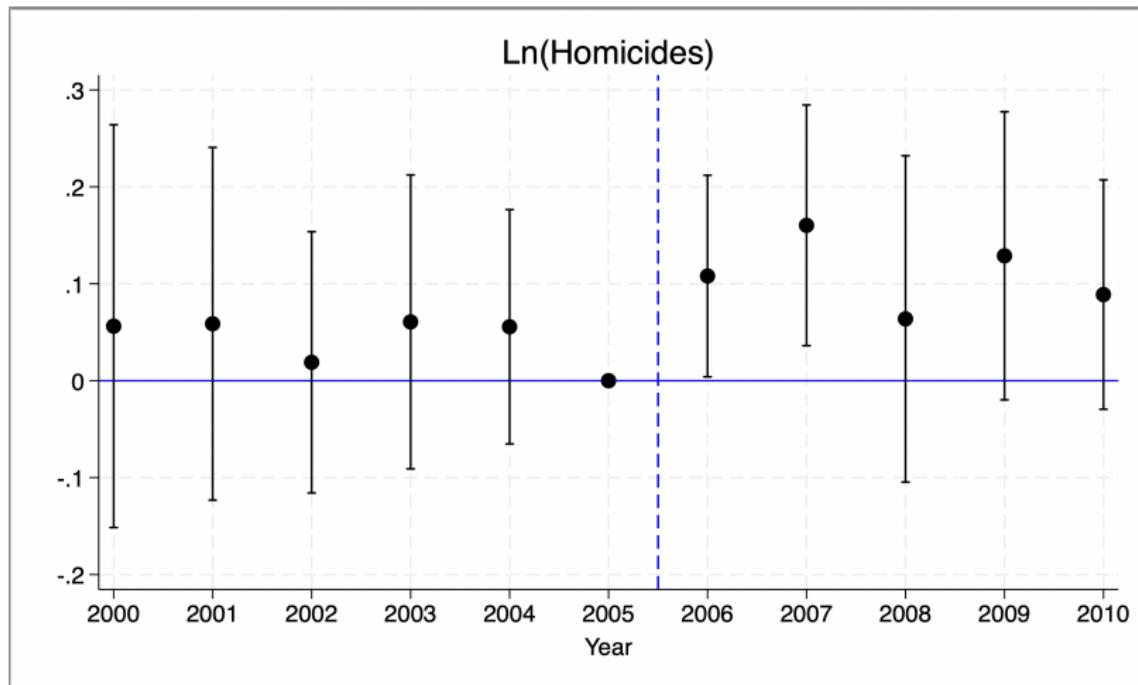
## Summarizing evidence and results

- **Bite:** Increases in enrollment and reductions in uninsured support that there is adoption of the treatment
- **Event studies:** Compelling graphics showing similarities between treatment and control
- **Falsifications:** no effect on a similar group who isn't eligible
- **Main results:** 9.2% reduction in mortality among the near-elderly
- **Mechanism:** "The effect is driven by a reduction in disease-related deaths and grows over time."

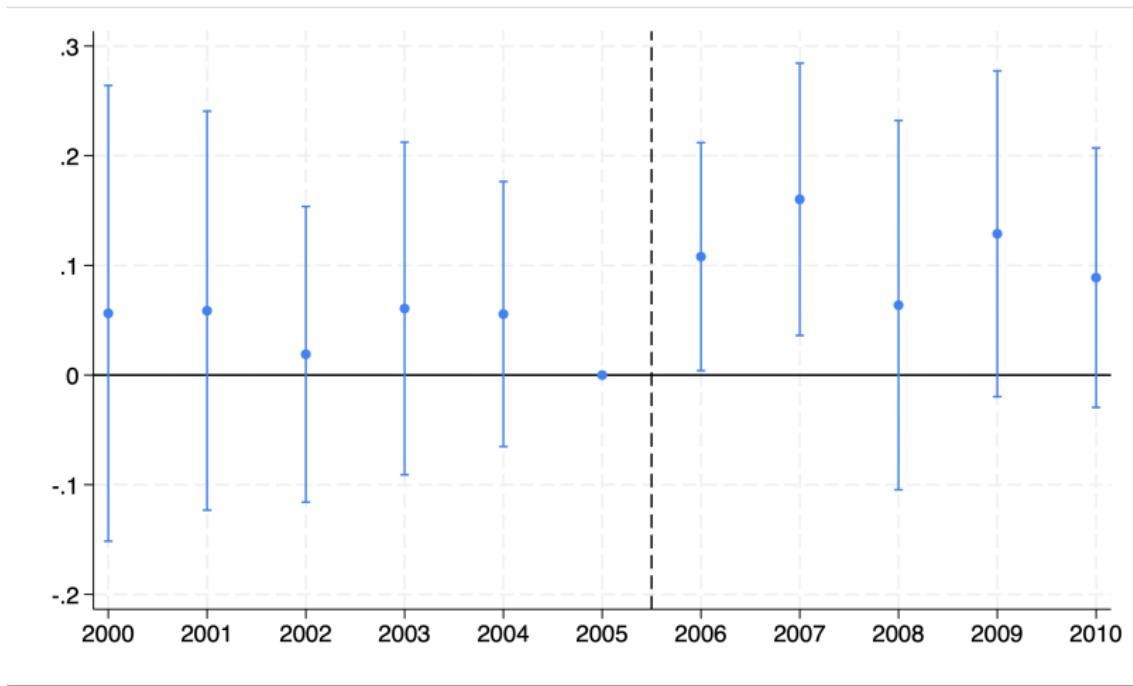
# Making event study

- When there is only one treatment group and one comparison group, then you run a regression with an interaction of the treatment group dummy and the calendar year dummies (plus both separately)
- You must drop  $t - \tau$  as the baseline (e.g.,  $t - 1$ ) and it must be  $Y^0$  untreated comparisons (No Anticipation)
- I have included in a do file that will do it for you either manually or using coefplot in `simple_eventstudy.do` at the shared github labs directory

# Manually creating the event study



# Creating the event study with Ben Jann's coefplot



## Honest DiD

- Another principled approach is to use the maximum change in the event study coefficients pre-treatment and ask yourself this question:
- If parallel trends was violated by the maximum pre-treatment violation, then can I bound my estimates of the ATT?
- Rambachan and Roth (2023) provide a way to do this and it includes code called `honestdid` at  
<https://github.com/asheshrambachan/HonestDiD>
- It doesn't mean parallel trends is violated – it's more like "what if?"

## Biased diff-in-diff

- Many people equate triple differences with falsification exercise, but actually it isn't that – it is its own design
- You use triple differences when diff-if-diff is biased
- It's used, in other words, when you have a parallel trends violation
- Miller, Johnson and Wherry (2021) didn't use it because they didn't think they had a parallel trends violation – so instead they used falsification

# Biased diff-in-diff #1

Table: Biased diff-in-diff #1: comparing states

States	Period	Outcomes	$D_1$	$D_2$
Experimental states	Before	$Y = NJ$		
	After	$Y = NJ + NJ_t + D$	$NJ_t + D$	$D + (NJ_t - PA_t)$
Non-experimental states	Before	$Y = PA$		
	After	$Y = PA + PA_t$	$PA_t$	

$$\hat{\delta}_{did}^{true} = D + (NJ_t - PA_t)$$

The ATT is  $D$ . Assume, though, that parallel trends does not hold,  
 $(NJ_t \neq PA_t)$

# Biased Placebo diff-in-diff

*Table:* Biased placebo diff-in-diff: comparing states but single men and older women

States	Period	Outcomes	$D_1$	$D_2$
Experimental states	Before	$Y = NJ$	$NJ_t$	
	After	$Y = NJ + NJ_t$		
				$(NJ_t - PA_t)$
Non-experimental states	Before	$Y = PA$	$PA_t$	
	After	$Y = PA + PA_t$		

$$\widehat{\delta}_{did}^{placebo} = (NJ_t - PA_t)$$

Assume that parallel trends does not hold, ( $NJ_t \neq PA_t$ )

## Two biased diff-in-diffs

- Parallel trends does not hold, ( $\textcolor{red}{NJ}_t \neq PA_t$ ), but what if that's the same bias in our placebo DiD?
- Then we can subtract the second from the first:

$$\hat{\delta}_{ddd} = \hat{\delta}_{did}^{true} - \hat{\delta}_{did}^{placebo}$$

- Triple differences is a “real design” with one parallel trends assumption:

$$(\textcolor{red}{NJ}_t^{true} - PA_t^{true}) = (\textcolor{red}{NJ}_t^{placebo} - PA_t^{placebo})$$

# Triple differences by Gruber (1995)

TABLE 3—DDD ESTIMATES OF THE IMPACT OF STATE MANDATES  
ON HOURLY WAGES

Location/year	Before law change	After law change	Time difference for location
<b>A. Treatment Individuals: Married Women, 20–40 Years Old:</b>			
Experimental states	1.547 (0.012) [1,400]	1.513 (0.012) [1,496]	−0.034 (0.017)
Nonexperimental states	1.369 (0.010) [1,480]	1.397 (0.010) [1,640]	0.028 (0.014)
Location difference at a point in time:	0.178 (0.016)	0.116 (0.015)	
Difference-in-difference:		−0.062 (0.022)	
<b>B. Control Group: Over 40 and Single Males 20–40:</b>			
Experimental states	1.759 (0.007) [5,624]	1.748 (0.007) [5,407]	−0.011 (0.010)
Nonexperimental states	1.630 (0.007) [4,959]	1.627 (0.007) [4,928]	−0.003 (0.010)
Location difference at a point in time:	0.129 (0.010)	0.121 (0.010)	
Difference-in-difference:		−0.008 (0.014)	
<b>DDD:</b>		<b>−0.054 (0.026)</b>	

## Triple differences commentary

- Some people think that it requires that the placebo DiD be zero, but that's incorrect
- In Gruber's 1995 article, it isn't clear why he needed triple differences in the first place – his triple differences yielded -0.054 which is almost the same as what he found with his first diff-in-diff (-0.062)
- The main value of triple differences is that you use it when you believe the parallel trends assumption doesn't hold

*Table:* Difference-in-Difference-in-Differences (Gruber version)

Groups	States	Period	Outcomes	$D_1$	$D_2$	$D_3$
Married women 20-40	Experimental states	After	$NJ + MW + \textcolor{blue}{NJ}_t + \textcolor{red}{MW}_t + D$	$\textcolor{blue}{NJ}_t + MW_t + D$	$D + \textcolor{blue}{NJ}_t - PA_t$	$D$
		Before	$NJ + MW$			
	Non-experimental states	After	$PA + MW + PA_t + MW_t$	$PA_t + MW_t$	$NJ_t - PA_t$	$D$
		Before	$PA + MW$			
Single men Older women	Experimental states	After	$NJ + SO + NJ_t + SO_t$	$NJ_t + SO_t$	$NJ_t - PA_t$	$D$
		Before	$NJ + SO$			
	Non-experimental states	After	$PA + SO + PA_t + SO_t$	$PA_t + SO_t$	$NJ_t - PA_t$	$D$
		Before	$PA + SO$			

### Triple diff assumption

$$\hat{\delta}_{DDD} = D + \underbrace{[(\textcolor{blue}{NJ}_t^{MW} - PA_t^{MW}) - (NJ_t^{SO} - PA_t^{SO})]}_{\text{Equally biased DiD #1 and #2}}$$

Triple differences requires two diff-in-diff, from different groups, with the same bias.  
 Parallel bias

# DDD in Regression

$$\begin{aligned} Y_{ijt} = & \alpha + \beta_2 \tau_t + \beta_3 \delta_j + \beta_4 D_i + \beta_5 (\delta \times \tau)_{jt} \\ & + \beta_6 (\tau \times D)_{ti} + \beta_7 (\delta \times D)_{ij} + \color{red}{\beta_8 (\delta \times \tau \times D)_{ijt}} + \varepsilon_{ijt} \end{aligned}$$

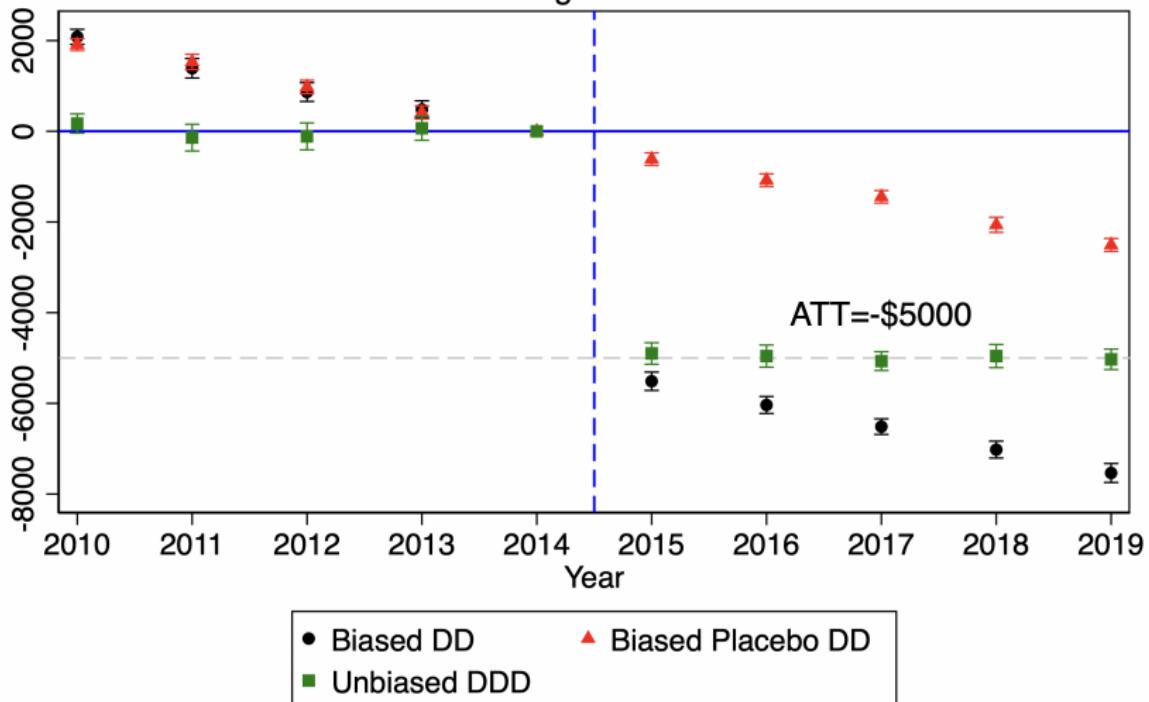
- Your dataset will be stacked by group  $j$  and state  $i$
- $\widehat{\beta}_8$  estimates the ATT
- Parallel bias, NA and SUTVA necessary and sufficient for identification

## Simulation

In /Labs/DDD I have a simulation to illustrate this for us called ddd2.do. The ATT is -\$5,000 but the biased DiD is -\$7487. The non-parallel trends bias is -\$2,487. So I replicate Gruber (with simulated data) where the placebo DiD is close (-\$2,507). I then present a triple differences which gives us -\$4,972. Let's look at the final product.

# Triple differences event study

Two Biased DiDs vs. Unbiased Triple Diff  
Illustrating Parallel Bias



# Great new paper to learn more



*Econometrics Journal* (2022), volume 00, pp. 1–23.  
<https://doi.org/10.1093/econj/utac010>

## The triple difference estimator

ANDREAS OLDEN AND JARLE MØEN

*Dept. of Business and Management Science, NHH Norwegian School of Economics, Hellevn.  
30, N-5045 Bergen, Norway.*  
Email: [andreasolden@gmail.com](mailto:andreasolden@gmail.com), [jarle.moen@nhh.no](mailto:jarle.moen@nhh.no)

First version received: 14 May 2020; final version accepted: 10 May 2021.

**Summary:** Triple difference has become a widely used estimator in empirical work. A close reading of articles in top economics journals reveals that the use of the estimator to a large extent rests on intuition. The identifying assumptions are neither formally derived nor generally agreed on. We give a complete presentation of the triple difference estimator, and show that even though the estimator can be computed as the difference between two difference-in-differences estimators, it does not require two parallel trend assumptions to have a causal interpretation. The reason is that the difference between two biased difference-in-differences estimators will be unbiased as long as the bias is the same in both estimators. This requires only one parallel trend assumption to hold.

**Keywords:** DD, DDD, DID, DiDID, difference-in-difference-in-differences, difference-in-differences, parallel trend assumption, triple difference.

**JEL Codes:** C10, C18, C21.

### 1. INTRODUCTION

The triple difference estimator is widely used, either under the name ‘triple difference’ (TD) or the name ‘difference-in-difference-in-differences’ (DDD), or with minor variations of these spellings. Triple difference is an extension of double differences and was introduced by Gruber (1994). Even though Gruber’s paper is well cited, very few modern users of triple difference credit him for his methodological contribution. One reason may be that the properties of the triple difference estimator are considered obvious. Another reason may be that triple difference was little more than a curiosity in the first ten years after Gruber’s paper. On Google Scholar, the annual number of references to triple difference did not pass one hundred until year 2007. Since then, the use of the estimator has grown rapidly and reached 928 unique works referencing it in the year 2017.<sup>1</sup>

Looking only at the core economics journals *American Economic Review* (AER), *Journal of Political Economy* (JPE), and *Quarterly Journal of Economics* (QJE), we have found 32 articles using triple difference between 2010 and 2017, see Table A1 in Appendix A. A close reading of these articles reveals that the use of the triple difference estimator to a large extent rests on

<sup>1</sup> More details on the historical development of the use of the triple difference estimator can be found in the working paper version of Olden and Møen (2020, fig. 1). In the working paper, we also analyse naming conventions and suggest that there is a need to unify terminology. We recommend the terms ‘triple difference’ and ‘difference-in-difference-in-differences’.

# Summarizing DDD

- Used to be people thought DDD required two parallel trends assumptions but it does not – it is a real design and requires one parallel trends assumption
- Parallel trends assumption is “parallel bias” – that the bias of the true DiD is the same as the bias of the placebo DiD
- The ladder of evidence still holds – you’ll want to present the event study plot, and my code provides it for you, because you need to evaluate the parallel bias assumption
- Given the lack of triple diff literacy, you may have to write this anticipating reader and maybe editor confusion and so “educate” as you go – overlaying all three plots could be help

## Falsification on outcomes

- Miller, Johnson and Wherry (2021) used the same outcome, but a placebo untreated treatment group (elderly) as a falsification to provide evidence for parallel trends
- But you can also use the same group but different outcomes
- Cheng and Hoekstra (2013) examine the effect of castle doctrine gun laws on non-gun related offenses like grand theft auto and find no evidence of an effect
- You just want to pick things that make logical sense in that they are a similar enough group (near elderly versus elderly) or an outcome that would be susceptible to some omitted variable you're worried about

## Rational addiction as a placebo critique

Sometimes, an empirical literature may be criticized using nothing more than placebo analysis

*"A majority of [our] respondents believe the literature is a success story that demonstrates the power of economic reasoning. At the same time, they also believe the empirical evidence is weak, and they disagree both on the type of evidence that would validate the theory and the policy implications. Taken together, this points to an interesting gap. On the one hand, most of the respondents claim that the theory has valuable real world implications. On the other hand, they do not believe the theory has received empirical support."*

## Placebo as critique of empirical rational addiction

- Auld and Grootendorst (2004) estimated standard “rational addiction” models (Becker and Murphy 1988) on data with milk, eggs, oranges and apples.
- They find these plausibly non-addictive goods are addictive, which casts doubt on the empirical rational addiction models.

# Placebo as critique of peer effects

- Several studies found evidence for “peer effects” involving inter-peer transmission of smoking, alcohol use and happiness tendencies
- Christakis and Fowler (2007) found significant network effects on outcomes like obesity
- Cohen-Cole and Fletcher (2008) use similar models and data and find similar network “effects” for things that aren’t contagious like acne, height and headaches
- Homophily (sorting) is probably just as likely an explanation

# Roadmap

## Introduction

Managing expectations

Origins of diff-in-diff in public health

Potential outcomes

Identification, Estimation and Inference

Design Stage

## Parallel Trends Violations

Results versus Evidence

Compositional Changes and Cross Sections

Event Studies

Triple differences

Falsifications

## Conclusion

# Spend time in the design stage

Reviewing some things

- Practice writing down your research question as a causal parameter
- Start with potential outcomes, not a regression model, because you need to be clear about your target parameter (ATT)
- Learn to go from the 2x2 DiD equation as the sum of ATT and non-parallel trends bias
- Learn what parallel trends does and does not mean (remember – it is not directly testable as it is based on a counterfactual unobserved)
- Don't peek at the outcomes until you're ready

## Shoelather and DiD

- Slow down and have a design stage which is where you answer these questions and don't yet look at the data
- Do everything you can to understand selection – shoelather dominates technical knowledge (Rubin 2008 "Design Trumps Analysis"; Freedman 1991)
- John Snow did not just do analysis – he worked hard to find comparable groups, learn the institutional details, and think carefully about selection
- Do everything you can to figure out the treatment assignment mechanism

2:52 ↗

5G E



## New message

craig newmark

decision at all. I don't want to ask you to share things you're not comfortable sharing, and understand if you can't, but I thought I'd ask. Hope you are well.

Scott

1/17/18, 12:43 PM ✓

Scott, thanks! but I haven't been a craigslist spokesman nor in management since 2000, have had no part in decision making. However, you're asking a question involving sophisticated marketing, economic analysis, and business development, and as a matter of public record, craigslist doesn't do that. Thanks!



1/17/18, 12:47 PM

Thanks Craig. That's a



Start a message



# Parallel Trends

- Of the three assumptions we discussed – SUTVA and no treated control group, unconditional parallel trends, and no anticipation – it's usually parallel trends we worry about
- We discussed two things
  1. Triple differences
  2. Event studies
- But we were focused on *unconditional parallel trends* (i.e., no covariates), but is unconditional parallel trends plausible in your study?
- Next we will explore when you think you need covariates – why do you need them and how do you choose them?