CodeChella Madrid 2024

Roadmap

Welcome to CodeChella

Diff-in-Diff Core
- Origins of diff-in-diff
- Potential outcomes
- Identification, Estimation and Inference

Parallel Trends Violations
- Main Results, Proof and Evidence
- Parallel Trends and Event Studies
- Problems with Pre-trends: Low Power and Bounding
- Triple differences
- Falsifications

Introductions

- Thank you coming to Mixtape Sessions and CUNEF first annual CodeChella Madrid workshop
- Organizers and Speakers are:
  - → Scott Cunningham, Baylor University
  - → Kyle Butts, University of Arkansas
  - → Dan Rees, Universidad Carlos III de Madrid, IZA, NBER
  - → Mark Anderson, Montana State University, IZA, NBER
  - → Agustin Casa, CUNEF
- Thank you to beautiful Madrid, CUNEF, UC3M and all of you!

Our schedule

Coffee and Food

Each morning of the workshop at 10:30am, fresh coffee and pastries will be provided outside the auditorium. Lunch will take place from 1 — 2:30pm each day. You grab a bite either at the campus cafeteria or head to the surrounding area. On Wednesday, food is generously being provided by CUNEF Universidad for us. A lot of restaurants are to the south on Av. de la Reina Victoria and to the East.

Schedule

This is our tentative schedule. Note that on Wednesday, May 29th, we will be going to a different lecture hall. The reason is that on this day, lunch will be provided for attendees by CUNEF Universidad.

	Day 1 Monday, May 27 CUNEF Auditorium	**Day 2** Tuesday, May 28 CUNEF Auditorium	**Day 3** Wednesday, May 29 **Aula Magna**	**Day 4** Thursday, May 30 CUNEF Auditorium
9 — 1pm	*Scott Cunningham* Core DID	*Scott Cunningham* Covariates	*Scott Cunningham* Callaway and Sant'anna & Sun and Abraham	*Kyle Butts* Imputation DID & Synth
1 — 2:30pm	Lunch	Lunch	*CUNEF* Provides Lunch	Lunch
2:30 — 3:30pm	*Scott Cunningham* Core DID	*Scott Cunningham* Bacon Decomposition & Callaway and Sant'anna	*Scott Cunningham* Callaway and Sant'anna & Sun and Abraham	*Kyle Butts* Imputation DID & Synth
3:30 — 5pm	Coding Lab	Coding Lab	*Dan Rees & Mark Anderson* Doing Applied Research Workshop	*Dan Rees & Mark Anderson* Doing Applied Research Workshop

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 nearly the cutting edge
- Ask questions at any point; I'll do my best to answer them

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 and synthetic control, both in its basics and some more contemporary issues, so that by the end of the week it a very intuitive, friendly, and useful tool
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 and synth context
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: the Core

The core of difference-in-differences

- 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 core assumptions
- Parallel trends, pre-trends (event studies) and falsifications

Day 2 outline: Covariates

Violations of parallel trends that can be fixed with covariates

- Incorporating covariates
  - → Outcome regressions: Heckman, Ichimura and Todd (1997)
  - → Weighting: Abadie (2005)
  - → Doubly robust: Sant'Anna and Zhao (2020)
  - → Twoway fixed effects with time varying controls
- Lalonde coding exercise

Day 3 outline: Differential Timing

Pathologies of and Solutions to Twoway Fixed Effects with Differential Timing

- Bacon decomposition (Goodman-Bacon 2021)
- Aggregate ATT(g,t) method (Callaway and Sant'Anna 2021)
- Decomposition event study leads and lags (Sun and Abraham 2021)
- Discussion of two contemporary papers using these methods and the presentation of results
- Simple illustrations in simulations
- Walk through R shiny app

Day 4 outline: Imputation DiD and Synthetic Control

Imputation estimators

- Imputation estimators
  - → Borusyak, et al. (2023) "robust efficient imputation estimator"
  - → Gardner, et al. (2023) two stage difference in differences
- Synthetic control methods
- Continuous treatment difference-in-differences if there is time (Callaway, Goodman-Bacon and Sant'Anna 2024)

Roadmap

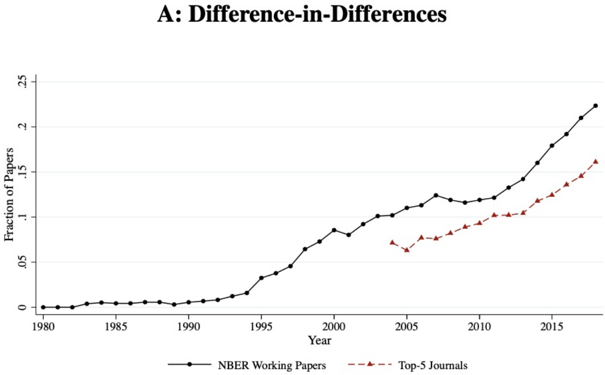
Welcome to CodeChella

Diff-in-Diff Core
- Origins of diff-in-diff
- Potential outcomes
- Identification, Estimation and Inference

Parallel Trends Violations
- Main Results, Proof and Evidence
- Parallel Trends and Event Studies
- Problems with Pre-trends: Low Power and Bounding
- Triple differences
- Falsifications

Growing popularity in economics

Figure: Currie, et al. (2020)



Origins of diff-in-diff)

- Its modern usage originates with Orley Ashenfelter and David Card in late 1970s and early 1980s work on job training programs
- The phrase “difference-in-differences” is coined in a 1985 article by Ashenfelter and Card, but it was used before then
- It seems to be the first design – it predates the RCT by over 80 years in fact
- It was also used in two famous public health debates in Vienna and London in the early to mid 19th century

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
- Does not require a linear regression model, but for reasons we will see, linear regression accommodates it (just not all specifications equally)

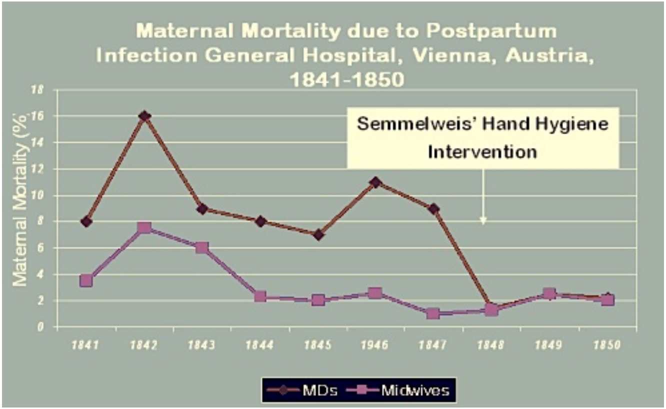
Ignaz Semmelweis and washing hands

- 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 we 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



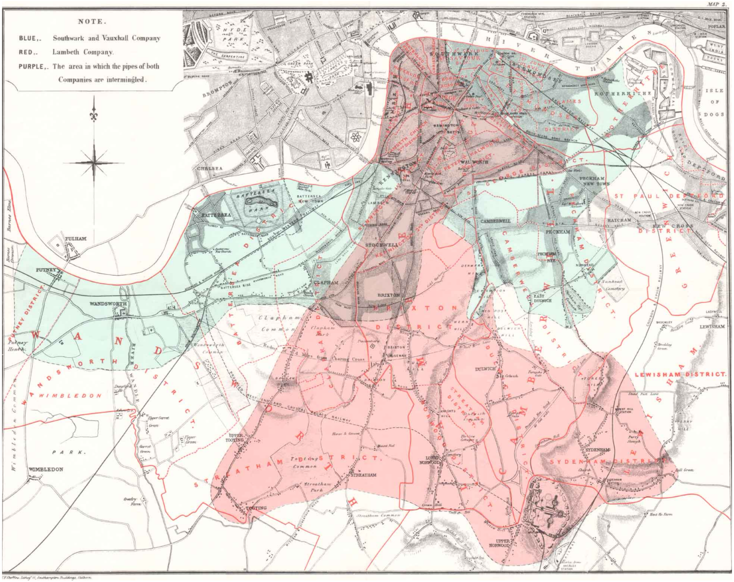
Evidence Rejected

- 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
- Let's look at an illustration using a table and another story

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	$Y = L + L_t + D$	$L_t + D$	
				$D + (L_t - SV_t)$
Southwark and Vauxhall	Before	$Y = SV$		
	After	$Y = SV + SV_t$	SV_t	

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

If $L_t = SV_t$, then this calculation equals the effect of moving the pipe on cholera mortality.

When is diff-in-diff causal?

- Causal questions require causal notation and we will use potential outcomes
- Potential outcomes notation is model-free explicit causal notation
- Rooted in the experimental design tradition and linked to statisticians Ronald Fisher, Jerzy Neyman and Don Rubin

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, at a particular point in time

Potential outcomes notation

- Potential outcomes:

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

- Potential outcomes are *a priori* real but unknown descriptions of j states of the world under different treatment exposures
- Given the treatment is binary, there are two potential outcomes for unit i at time t: $Y_{i,t}^{1}$ and $Y_{i,t}^{0}$
- But there are not observed until a treatment assignment is made

Realized outcomes and treatment assignment

- Treatment assignment mechanisms and treatment assignment are two very different things
- Treatment assignment is represented with the switching equation

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

- Switching equations only show which potential outcome you're looking at, not why it and not another
- Switching equation is not a treatment assignment mechanism – randomization is, rationality is, running variables are

Treatment effect definitions

Individual treatment effect

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

Causal effects, or treatment effects, are simple comparisons between the two potential outcomes.

Missing Potential Outcomes

- Recall the switching equation:

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

- "Fundamental problem of causal inference" is you need two potential outcomes but only have one
- Some treatment assignment mechanisms let you use the untreated units as a replacement for the missing one

Conditional Average Treatment Effects

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:

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

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

Simple spreadsheet exercise

Let’s review basic concepts here at ”WEIGHTS” tab:
https://docs.google.com/spreadsheets/d/1ODuQqGtH_Ewea7zQoLTFYHbnvqaTVDhn2GDzq3Oa6EQ/edit?usp=sharing

Orley Ashenfelter and difference-in-differences

- Orley Ashenfelter "popularized" difference-in-differences in economics in two papers – 1978 article and a 1985 article with David Card – both studying job training programs
- After graduating from Princeton, he took a job in Washington DC to work for the government
- Got his hands on large micro data and was analyzing complex fixed effects regression models of job training program participation on wages
- But explaining regression to normal people was difficult, so he called it "difference-in-differences" instead

https://youtu.be/WnB3EJ8K7lg?si=BpU4Xv5p71vwHPvP&t=120

OLS Measures Four Averages and Three Subtractions

- Here is the canonical DiD regression specification, sometimes called the 2x2

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

- Orley notes that the OLS estimator of δ actually calculates "four averages and three subtractions" calculation are numerically identical

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

- Review these two calculations using `equivalence.do` in Stata to illustrate the point

DiD equation is the 2x2

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

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

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

Three DiD assumptions

When it is the simple 2x2, there are three assumptions

- **No Anticipation**: the "pre" or baseline is untreated (i.e., $Y_{t-1} = Y^0_{t-1}$ for treated units)
- **SUTVA**: When the treatment group is treated, it does not cause the comparison group to become treated
- **Parallel trends**: evolution of mean Y^0 is the same for the treatment and comparison groups

No Anticipation

- "No anticipation" means that the unit is not treated until it is treated
- Rational, forward looking agents can "turn on" a future treatment simply by recognizing that it is coming *and* changing their behavior, but not always
  - → **Example 1**: Tomorrow I win the lottery, but don't get paid yet. I decide to buy a new house today and a banker gives me a loan. 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.
- It means $Y_{i,t} = Y^0_{i,t}$ and doesn't switch to $Y^1_{i,t}$ until the treatment "turns on" (i.e., switching equation)

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

Role of assumptions

- Parallel trends is the most commonly known assumption of those three
- But without no anticipation and SUTVA, the interpretation of difference-in-differences becomes very complex – even with parallel trends
- But let's start with the simple case with NA and SUTVA so that we can see parallel trends

No Anticipation

- Post-treatment outcome for the treated group is treated so $Y = Y^1$ in the post period for the treatment group

$$\widehat{\delta} = \bigg(E[Y_k^1|Post] - E[Y_k^0|Pre]\bigg) - \bigg(E[Y_U|Post] - E[Y_U|Pre]\bigg)$$

- But if the baseline is untreated, then $Y_i = Y_i^0$ at baseline
- No Anticipation simply means that even though the treatment will occur, it has not yet occurred at baseline for all practical purposes (foreknowledge or not)

SUTVA and never treated

- When the treatment occurs to treated group in the post period, SUTVA means that did not cause the potential outcome to change for the comparison group

$$\widehat{\delta} = \bigg(E[Y_k^1|Post] - E[Y_k^0|Pre]\bigg) - \bigg(E[Y_U^0|Post] - E[Y_U^0|Pre]\bigg)$$

- SUTVA violations will have similar issues to when we use an already-treated group as a comparison, as we'll see

Replace with potential outcomes and add a zero

- Use the switching equation and replace all realized outcomes with potential outcomes assuming NA and SUTVA
- Then add zero

$$\widehat{\delta} = \underbrace{\Big(E[Y_k^1|Post] - E[Y_k^0|Pre]\Big) - \Big(E[Y_U^0|Post] - E[Y_U^0|Pre]\Big)}_{\text{Switching equation}}$$

$$+ \underbrace{E[Y_k^0|Post] - E[Y_k^0|Post]}_{\text{Adding zero}}$$

Parallel trends bias

- Rearrange the terms into the ATT plus the parallel trends bias term
- Find the units satisfying NA and SUTVA and notice that both assumptions are nested inside the parallel trends assumption

$$\begin{aligned}\widehat{\delta} \;&=\; \underbrace{E[Y_k^1|Post] - E[Y_k^0|Post]}_{\text{ATT}} \\ &+ \underbrace{\Big[E[Y_k^0|Post] - E[Y_k^0|Pre]\Big] - \Big[E[Y_U^0|Post] - E[Y_U^0|Pre]\Big]}_{\text{Non-parallel trends bias in 2x2 case}}\end{aligned}$$

Notice that the parallel trends bias term is a difference-in-differences calculation

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"

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)

No Anticipation Violation

- Assume NA is violated; simply replace Y_k^0 at baseline with Y_k^1 instead

$$\widehat{\delta} \;=\; \Big(E[Y_k^1|Post] - E[Y_k^1|Pre]\Big) - \Big(E[Y_U^0|Post] - E[Y_U^0|Pre]\Big)$$

- What does losing NA cost us?

No Anticipation Violation

$$\widehat{\delta} = \underbrace{\bigg(E[Y^1_k|Post] - E[Y^1_k|Pre]\bigg) - \bigg(E[Y^0_U|Post] - E[Y^0_U|Pre]\bigg)}_{\text{Switching equation}}$$

$$+ \underbrace{E[Y^0_k|Post] - E[Y^0_k|Post]}_{\text{Adding zero}}$$

No Anticipation Violation

$$\widehat{\delta} = \underbrace{E[Y^1_k|Post] - E[Y^0_k|Post]}_{\text{ATT}} + \underbrace{\left[E[Y^0_k|Post] - E[Y^1_k|Pre]\right] - \left[E[Y^0_U|Post] - E[Y^0_U|Pre]\right]}_{\text{This is not the parallel trends term}}$$

"Parallel trends" refers to $\Delta E[Y^0_k]$ and $\Delta E[Y^0_U]$. But that's not what is in that line. So let's add another zero to try and get a standard parallel trends term

No Anticipation Violation

$$
\begin{aligned}
\widehat{\delta} \quad = \quad & \underbrace{E[Y^1_k|Post] - E[Y^0_k|Post]}_{\text{ATT}} \\
& + \underbrace{\Big[E[Y^0_k|Post] - E[Y^1_k|Pre] \Big] - \Big[E[Y^0_U|Post] - E[Y^0_U|Pre] \Big]}_{\text{This is not the parallel trends term}} \\
& + \underbrace{E[Y^0_k|Pre] - E[Y^0_k|Pre]}_{\text{Add another zero}}
\end{aligned}
$$

Now rearrange the second and third row

No Anticipation Violation

$$\begin{aligned}
\widehat{\delta} \quad = \quad & \underbrace{E[Y^1_k|Post] - E[Y^0_k|Post]}_{\text{ATT}} \\
& + \underbrace{\left[E[Y^0_k|Post] - E[Y^0_k|Pre] \right] - \left[E[Y^0_U|Post] - E[Y^0_U|Pre] \right]}_{\text{Parallel trends term}} \\
& + \underbrace{E[Y^0_k|Pre] - E[Y^1_k|Pre]}_{\text{Reversed treatment effect?}}
\end{aligned}$$

. Now let's switch the order of the third row

No Anticipation Violation

$$\begin{aligned}
\widehat{\delta} \quad = \quad & \underbrace{E[Y^1_k|Post] - E[Y^0_k|Post]}_{\text{Post treatment ATT}} \\
& + \underbrace{\Big[E[Y^0_k|Post] - E[Y^0_k|Pre] \Big] - \Big[E[Y^0_U|Post] - E[Y^0_U|Pre] \Big]}_{\text{Parallel trends term}} \\
& - \Big[\underbrace{E[Y^1_k|Pre] - E[Y^0_k|Pre]}_{\text{Baseline ATT}} \Big]
\end{aligned}$$

.

No Anticipation Violation

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

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

If you use a treated unit at baseline, then DiD will be attenuated towards zero, and if treatment effects are constant, DiD will equal zero

Let's look in `na.do`.

Using already treated comparison group

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

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

Replace realized outcomes with potential outcomes and rewrite using the “add zero” trick we did. We will review the answer tomorrow morning.

Hint: You know you’re not done until you see a standard parallel trends assumption

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

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 and SUTVA does DiD equal ATT + PT
- But it's crucial to remember: DiD and ATT are not the same thing

OLS Specification

- Simple DiD equation will identify ATT under parallel trends
- But so will a particular OLS specification (two groups and no covariates)
- OLS was historically 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/MOtbuRX4eyQ?t=1882

Binyamin Appelbaum
@BCAppelbaum

Replying 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

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

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."

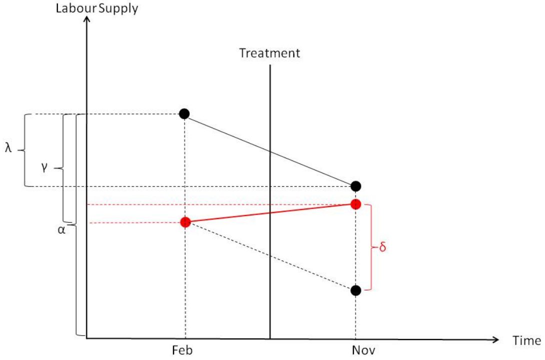
OLS specification of the DiD equation

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

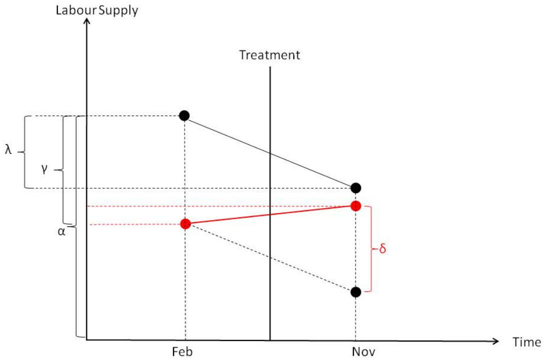
$$Y_{its} = \alpha + \gamma NJ_s + \lambda d_t + \delta(NJ \times d)_{st} + \varepsilon_{its}$$

  → 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: α
  → PA Post: $\alpha + \lambda$
  → NJ Pre: $\alpha + \gamma$
  → NJ Post: $\alpha + \gamma + \lambda + \delta$
- DiD equation: (NJ Post - NJ Pre) - (PA Post - PA Pre) $= \delta$

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



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



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 in DID

- When dealing with clustered data, a crucial issue is the calculating the standard errors associated with the sampling variance of your estimator
- Correlated errors occur when the unobserved errors are correlated within a cluster.
- Failing to account for correlated errors can lead to misleading inference, biased standard errors and higher over rejection

Serial correlation creates problems

- Bertrand, Duflo and Mullainathan (2004) show that conventional standard errors will often severely understate the standard deviation of the estimators
- They proposed three solutions: bootstrapping, allowing for arbitrary clustered correlations, and a third approach that is very strange
- Clustering is typically recommended at the aggregate level where the entire treatment occurred at the aggregate level
- It is considered a more conservative approach to inference

Summarizing

- Diff-in-diff isn't necessarily causal: it requires SUTVA, NA, parallel trends and an untreated comparison group
- But if you have all of those, then difference-in-differences identifies the ATT
- We will assume for now you have SUTVA, NA and an untreated comparison group as they are the ones easier to rationalize
- But parallel trends is untestable since we never observe $E[Y_k^0|Post]$ so how do approach this?
- We'll discuss some general practical advice and propose some solutions

Roadmap

Welcome to CodeChella

Diff-in-Diff Core

- Origins of diff-in-diff
- Potential outcomes
- Identification, Estimation and Inference

Parallel Trends Violations

- Main Results, Proof and Evidence
- Parallel Trends and Event Studies
- Problems with Pre-trends: Low Power and Bounding
- Triple differences
- Falsifications

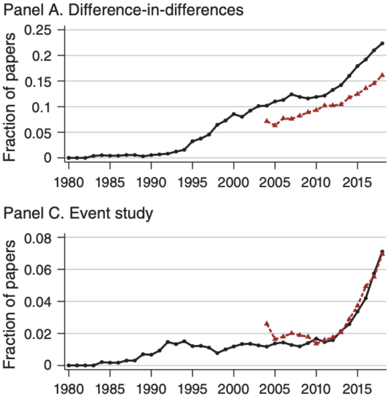
Main Results Depend on Untestable Assumption

- You put your regression results in a table and call it your main results and you claim they're causal under parallel trends (which is true)
- But what can you do to provide evidence that parallel trends is believable?
- Unfortunately, parallel trends is not testable because you are missing $E[Y_k^0|Post]$
- There's a few standard things that are always done, plus some new things, and we will discuss them all

Evidence for Parallel Trends

- Think a prosecutor arguing against a defense attorney to convince a judge and jury
- The claim the defendant is guilty but the claim is not the evidence – it's more like an assertion
- The evidence 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
- Our evidence will be bite, falsifications, mechanisms and event study data visualization

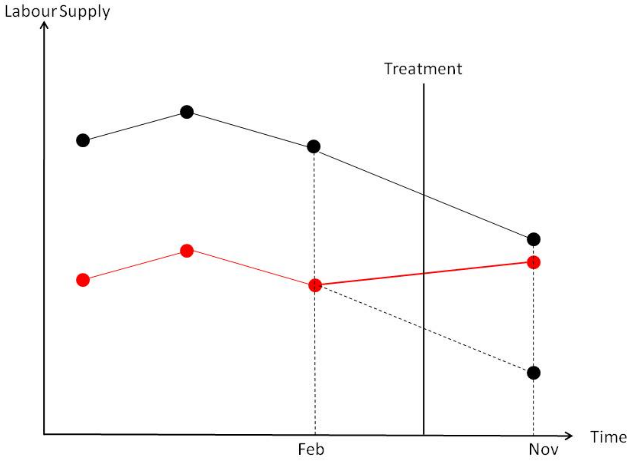
Event studies have become mandatory in DiD



Creating event studies

- Event studies mean different things to different people – in finance and accounting, they are methods for evaluating abnormal stock market returns and historically were fairly parametric
- At some point in the early 2000s, economists start plotting the treatment and control outcomes both before and after the point of treatment
- Reasoning is "maybe if the two groups had comparable trends before treatment, they would have after the treatment had the treatment not happened"
- People have done it different ways, but regardless of the way, the reason is the same

Plot the raw data when there's only two groups



Event study regression

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

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

  → With a simple 2x2, event study regressions simply interact the treatment dummy (D_i) with calendar year dummies (τ_t)

  → There are q pre-treatment leads and m post treatment lags

  → Each coefficient is a separate "four averages and three subtractions" for that period relative to the dropped baseline

- Typically you'll plot the coefficients and 95% CI on all leads and lags, usually with a vertical line marking just before treatment happened

Pre-treatment DiD coefficient

$$\widehat{\mu}_{t-2} \quad = \quad \underbrace{\Big[E[Y_k^0|t-2] - E[Y_k^0|t-1] \Big] - \Big[E[Y_U^0|t-2] - E[Y_U^0|t-1] \Big]}_{\text{Differential trends in pre-treatment } \Delta E[Y^0]}$$

- Under NA, all pre-treatment periods are untreated, therefore by definition $Y = Y^0$ for all pre-treatment periods
- If NA then ATT$=$0 in all pre-periods, therefore the pre-treatment coefficients *only* equal the differential trend
- Unlike parallel trends (which uses post-treatment missing potential outcomes), this is testable
- And if the pre-treatment coefficients are zero, interpreting the post-treatment coefficients as ATT becomes more plausible

Two types of evidence for parallel trends

1. **Event study**: Only examines whether the two groups were comparable on trends pre-treatment
2. **Falsification**: Only examines whether a similar group who was not eligible for the treatment shows similar patterns as found with the treatment group (more on this later)

Let’s look at an example involving a diff-in-diff application involving near elderly mortality in the United States and health insurance, but we will just focus on the event studies

Example US Healthcare

- United States does not have universal healthcare except for the elderly and the very poor
- Elderly get Medicare and must be 65 or older; the very poor get Medicaid so long as their income is below some threshold
- Barack Obama signature policy achievement was at the Affordable Care Act which gave incentives to states to raise that threshold
- Some states did; some did not

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 nonexpanders. 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

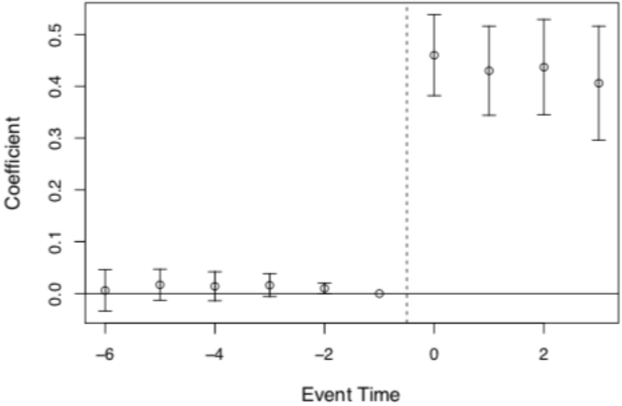
Miller, Johnson and Wherry used difference-in-differences to evaluate its effect on mortality on the “near elderly” (younger than 65) and provide strong evidence supporting the credibility of their main results

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

Bite

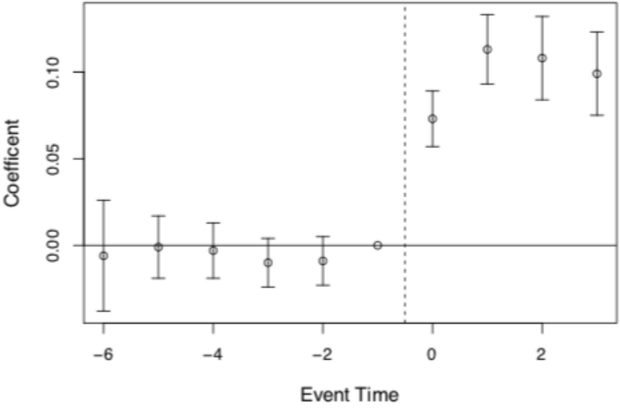
- Bite in this context means when US states made Medicaid more generous, people got on Medicaid who would not have been on it otherwise
- And as a bonus, would not have been insured at all without it
- Bite does not provide evidence for parallel trends, but provides credibility overall because if there is no evidence people enrolled in a program, then how can the program have any effect on them?
- Their bite evidence is presented entirely with event studies

Bite evidence #1



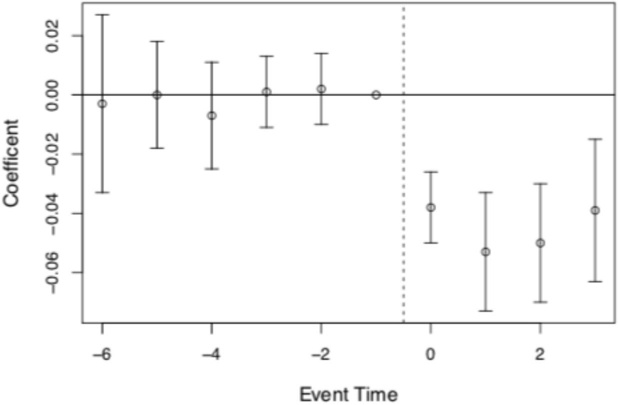
(a) Medicaid Eligibility

Bite evidence #2



(b) Medicaid Coverage

Bite evidence #3



(c) Uninsured

Summarizing bite

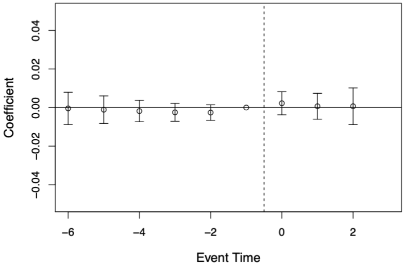
- Seems clear that when states made healthcare insurance more generous, more people enrolled (around 5-6pp more)
- Some probably switched from private insurance to public insurance
- But not all – the last figure showed a 4-6pp reduction in uninsured status
- Pre-trends look great, too – the treatment and comparison states appear to have been following *identical* trends, probably because this subpopulation of poor people are not changing or changing behavior a lot anywhere

Falsification

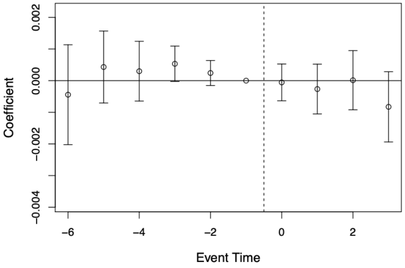
- Event studies tell us something about whether two groups were on similar trends before the treatment, but they don't say what would have happened after
- Miller, Johnson and Wherry use a falsification in addition to event studies to see if unobserved confounders might be happening in the post-treatment period in the expanding states
- They'll look at 65 and older coverage and mortality
- They're already on Medicare so they shouldn't get on Medicaid, and they're very similar to the "near elderly" so they should have similar confounders

Falsifications on 65 year old and older

Age 65+ in 2014



(c) Medicaid Coverage

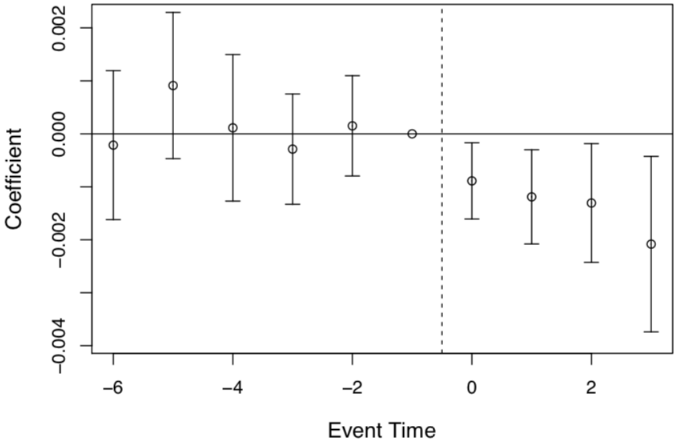


(d) Annual Mortality

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 verfiable

Near elderly mortality and Medicaid expansion



Main results and mechanism

- Main results: They translate that percentage point reduction into a 9.2% reduction in mean mortality among the near-elderly
- Mechanism: They attribute this effect to a reduction in disease-related deaths that grew over time
- Notice how they did so much with simple data visualization (event studies); let's just look at how those are created now

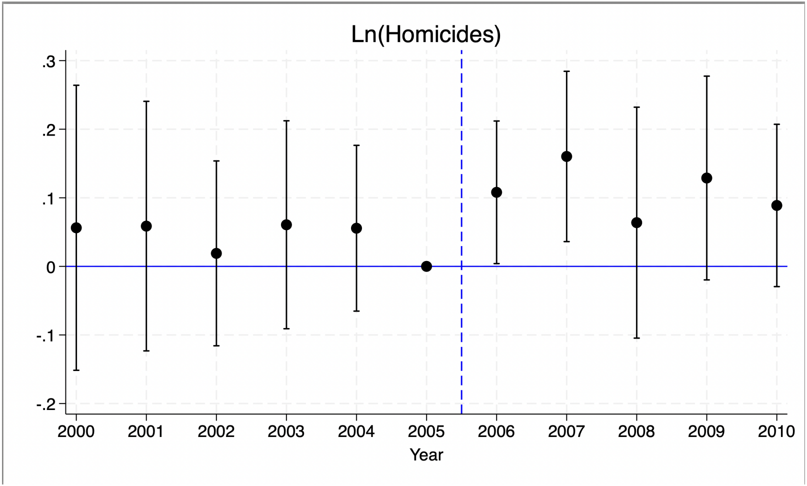
Making event study

- Recall the specification when there is only one treatment group and one comparison group

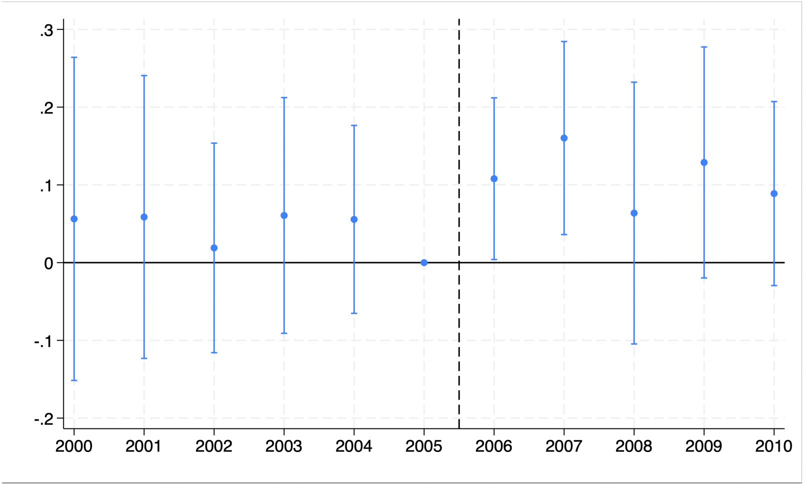
$$Y_{it} = \alpha + \sum_{\tau=-2}^{-q} \mu_\tau (D_i \times \tau_t) + \sum_{\tau=0}^{m} \delta_\tau (D_i \times \tau_t) + \tau_t + D_i + \varepsilon_{it}$$

- Regression fully interacts treatment group indicator with calendar time indicators
- Be sure to use an untreated group-period as the baseline (e.g., $t-1$) otherwise NA is violated and DID will be attenuated (minus baseline ATT)
- There is some Stata and R code (`simple_eventstudy.do`) at the repo to illustrate how to create this using a dataset on homicides and a treatment indicator

Manually creating the event study



Creating the event study with Ben Jann's `coefplot`



Biased diff-in-diff

- But what if your pre-trends look horrible? What can you do?
- There are two options I'll discuss now: one uses bounding (Rambachan and Roth 2024) and the other uses triple differences
- Tomorrow we will discuss using covariates to address a biased diff-in-diff

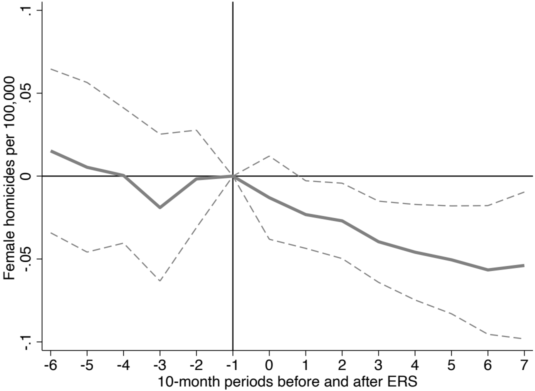
Low Power

- Parallel trends plus NA and SUTVA is sufficient to identify the ATT using diff-in-diff
- And testing for whether differential trends between the two groups existed in the pre-period is a common sense way of assessing the plausibility of the parallel trends assumption
- But has limitations: one of them has to do with power and one has to do with a kind of selection bias created by only analyzing cases without statistically significant pre-trends
- This is from Jon Roth's work, which I will briefly summarize

Low Power

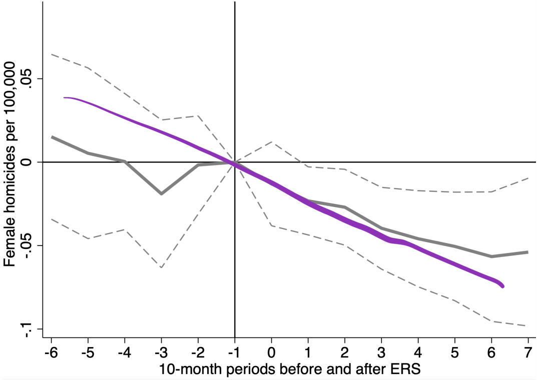
- Even if the pre-trends were not zero, you may not have the power needed to detect it statistically
- Jon in his 2022 AER: Insights, “Pre-test with Caution: Event-study Estimates After Testing for Parallel Trends” discusses this in detail
- Let me illustrate with one of my recent papers the problem to consider: Cunningham, DeAngelo and Tripp (2024), “Did Craigslist’s Erotic Services Reduce Female Homicide and Rape?”, *Journal of Human Resources*
- Study uses the staggered rollout of Craigslist’s “erotic services” section of its website, used for matching sex workers and clients, to identify its effect on female victimization with twoway fixed effects and matrix completion

Event study from my research



- We weren't able to reject the null that there were no pre-trends (p-value of 0.5).

Event study from my research



- However, we couldn't have rejected the null of a linear trend either due to large CI.
- So even if pre-trends were non-zero, we may fail to detect it statistically

Simulations

- Roth (2022) had simulations calibrated to papers published in the AER, AEJ: Applied and AEJ: Policy between 2014 and 2018
- 70 papers contained and event study plot; he focused on 12 with available data
- He evaluated properties of standard estimates/CIs under linear violations of parallel trends against which conventional tests have limited power and found:
    1. Bias was often of the same magnitude as the estimated treatment effect
    2. CIs substantially undercovered in many cases
- There are limitations to pre-trends testing due to low power – you may not find significant pre-trends even if parallel trends itself is being violated

Summary of Pretrends Package

- The `pretrends` package offers tools for power calculations for pre-trends tests and visualization of potential parallel trends violations, based on Roth (2022, AER:Insights) (`https://github.com/jonathandroth/pretrends`)
- It helps assess the power of pre-trends tests by calculating the ex ante power to detect violations and visualizing these on an event-study plot.
- For a more comprehensive solution to parallel trends violations, consider using the HonestDiD package, which forms confidence intervals for treatment effects accounting for pre-treatment violations.

Quote from Rambachan and Roth (2024, REstud)

2.4.1. Bounding relative magnitudes. In empirical applications, researchers may be willing to assume that the confounding factors which produce non-parallel trends in the post-treatment periods are not too much larger in magnitude than the confounding factors in the pre-treatment periods. In their empirical application to right-to-carry gun laws, Manski and Pepper (2018) operationalize this intuition by calibrating bounds on $|\delta_1|$ to the largest violations of parallel trends in the pre-treatment period (see their Table 3).[6] Such a restriction can be formalized in our framework by imposing that $\delta \in \Delta^{RM}(\bar{M})$ for $\bar{M} \geq 0$, where

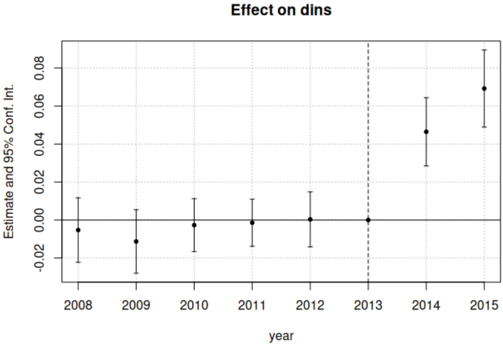
$$\Delta^{RM}(\bar{M}) = \{\delta : \forall\, t \geq 0,\ |\delta_{t+1} - \delta_t| \leq \bar{M} \cdot \max_{s<0} |\delta_{s+1} - \delta_s|\}.$$

$\Delta^{RM}(\bar{M})$ bounds the maximum post-treatment violation of parallel trends between consecutive periods by $\bar{M}$ times the maximum pre-treatment violation of parallel trends. We use the abbreviation RM for "relative magnitudes." The choice $\Delta^{RM}(\bar{M})$ may be reasonable if the researcher suspects that possible violations of parallel trends are driven by confounding economic shocks that are of a similar magnitude to confounding economics shocks in the pre-period. When the number of pre-treatment and post-treatment periods is similar, a natural benchmark may be $\bar{M} = 1$, which bounds the worst-case post-treatment difference in trends by the equivalent maximum in the pre-treatment period.[7]

Bounds on Relative Magnitudes

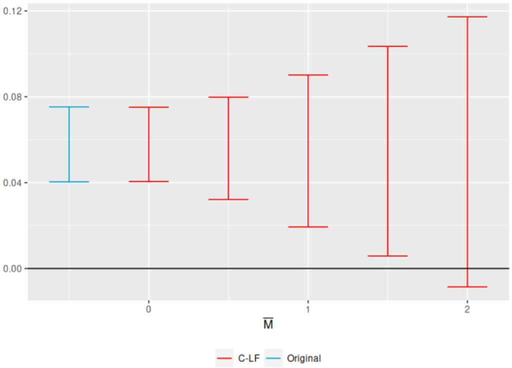
- Rambachan and Roth (2024), “A More Credible Approach to Parallel Trends”, *Review of Economic Studies*
- Intuition, recall, in pre-trends testing is that the pre-trends are informative about counterfactual post-treatment trends
- How different would the counterfactual trend have to be from the pre-trends to erase your conclusion about the treatment effects?
- They use this restriction to bound the treatment effect under these imposed restrictions

Medicaid's Effect on Insurance Coverage



- Note the negative effect in 2009 of around -1pp

Event study from my research



- Impose that the post-treatment violation of parallel trends is no more than some constant $\overline{M}$ which you can set to 0 (original) or some factor (e.g., 2 or no more than twice as large)
- We can rule out a null effect unless we allow for violations of PT of 2x larger than the max in the pre-period

Triple differences

- Another core methodology is the triple differences design (Gruber 1994)
- Many people equate triple differences with falsification exercise, but actually it isn't that – it is it's own design
- You use triple differences when diff-if-diff is biased and parallel trends is violated
- Is is *not* falsification; rather it is a research design with assumptions which are new

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	

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

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	$Y = NJ$		
	After	$Y = NJ + NJ_t$	NJ_t	
				$(NJ_t - PA_t)$
Non-experimental states	Before	$Y = PA$		
	After	$Y = PA + PA_t$	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, $(NJ_t \neq PA_t)$, but what if that's the same bias in our placebo DiD?
- Then we can subtract the second from the first:

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

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

$$({NJ_t}^{true} - PA_t^{true}) = ({NJ_t}^{placebo} - PA_t^{placebo})$$

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
A. *Treatment Individuals: Married Women, 20–40 Years Old:*			
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. *Control Group: Over 40 and Single Males 20–40:*			
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)		
DDD:	**−0.054** **(0.026)**		

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	$NJ + MW + NJ_t + MW_t + D$	$NJ_t + MW_t + D$	$D + NJ_t - PA_t$	D
		Before	$NJ + MW$			
	Non-experimental states	After	$PA + MW + PA_t + MW_t$	$PA_t + MW_t$		
		Before	$PA + MW$			
Single men Older women	Experimental states	After	$NJ + SO + NJ_t + SO_t$	$NJ_t + SO_t$	$NJ_t - PA_t$	
		Before	$NJ + SO$			
	Non-experimental states	After	$PA + SO + PA_t + SO_t$	$PA_t + SO_t$		
		Before	$PA + SO$			

Triple diff assumption

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

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

DDD in Regression

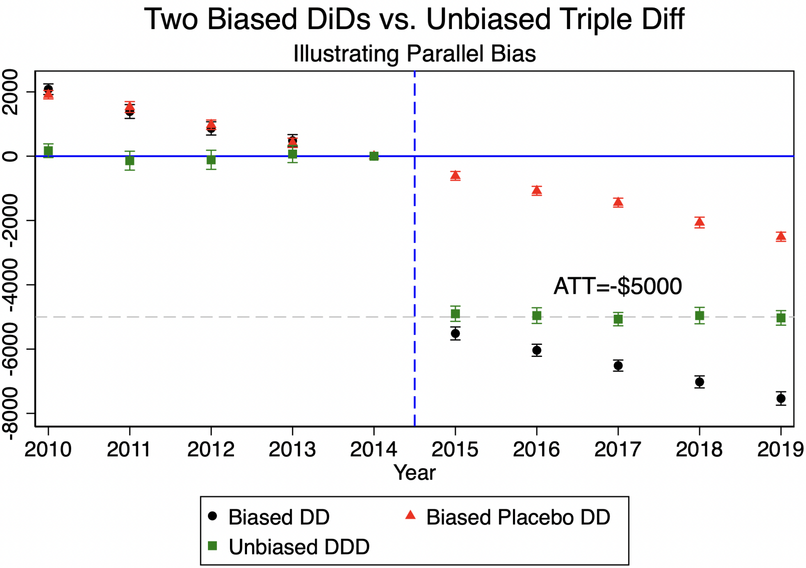
$$
\begin{aligned}
Y_{ijt} &= \alpha + \beta_2\tau_t + \beta_3\delta_j + \beta_4 D_i + \beta_5(\delta \times \tau)_{jt} \\
&\quad + \beta_6(\tau \times D)_{ti} + \beta_7(\delta \times D)_{ij} + \beta_8(\delta \times \tau \times D)_{ijt} + \varepsilon_{ijt}
\end{aligned}
$$

- 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



Great new paper to learn more

The Econometrics Journal

Econometrics Journal (2022), volume **00**, pp. 1–23.
https://doi.org/10.1093/ectj/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, 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.[1]

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

[1] 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

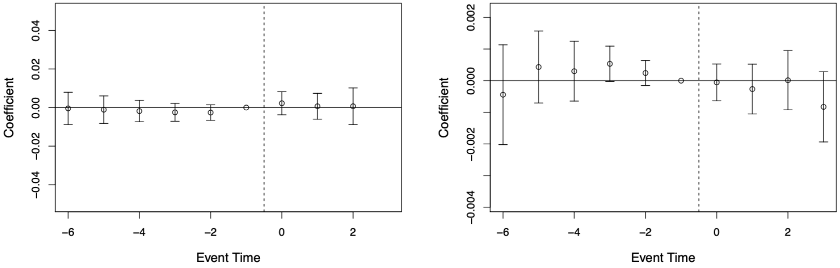
- Used to be people thought DDD required two parallel trends assumptions but it does not – it is a real design and requires one parallel trends assumption
- Parallel trends assumption is “parallel bias” – that the bias of the true DiD is the same as the bias of the placebo DiD
- The ladder of evidence 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 with similar groups

- Recall that Miller, Johnson and Wherry (2021) found a very similar group to their near elderly population (i.e., the 65 and older) as a falsification to provide evidence for parallel trends
- Evidence, again, is not proof – could still be that the confounder is unique to the treatment group (i.e., the near elderly)
- Ideal falsification is a group that is "almost" the treatment group and use the same outcome and the same model
- Another example is Cunningham, DeAngelo and Tripp (2024) looked at the effect of Craigslist's erotic services on male homicides

Falsifications on 65 year old and older

Age 65+ in 2014



(c) Medicaid Coverage (d) Annual Mortality

Falsifications with Different Outcomes

- Usually you have in mind a general confounder affecting many outcomes, not just your outcome.
- Falsifications are helped when there are specific confounders you have in mind that are consistent with your hypothesis but affect other things your hypothesis is unrelated to
- 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
- Cunningham, DeAngelo and Tripp (2024) looked at assaults as a falsification against our main result of female victimization

Example 1: Falsifications as a Critique of Rational Addiction

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.

Example 2: Falsification 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

Concluding the basics

- That concludes the core of diff-in-diff
- A lot of what we just went through is pretty standard and common to any diff-in-diff, but some of it even other studies too
- But what if actually doesn't hold and we can't fix it with triple differences or if the bounds are too large to be informative?
- Next we look at one very common example – the use of covariates to fix parallel trends violations