Causal Inference II

Mixtape Session

MIXTAPE SESSIONS

Roadmap

- Background
  - Introduction
  - Potential outcomes
- Identification and Estimation
  - Parallel trends
  - Estimation with OLS specification
  - Inference
- Parallel trends violations
  - How parallel trends can get violated
  - Types of evidence
  - Triple difference
  - Placebo outcomes

Introduction

- Welcome to a day long workshop on difference-in-differences
- Lecture, discussion, exercises, application

Workshop outline

Introduction to DiD basics

- Potential outcomes review
- DiD equation and estimation with OLS
- Evaluating parallel trends with falsifications, event studies
- Triple differences

Important caveat

- Causal inference can feel overwhelming for even the most seasoned person
- Don’t let econometrics ever become a substitute for hard work – they aren’t the same
- Learn everything you can about the subject, not just the econometrics
- Talk to people, read books, read articles, learn the history – don’t just read econometrics as there’s no magic inside an econometrics estimator as we saw
- Econometrics is a tool, but it’s only a tool; it doesn’t replace your brain and heart

Natural experiments

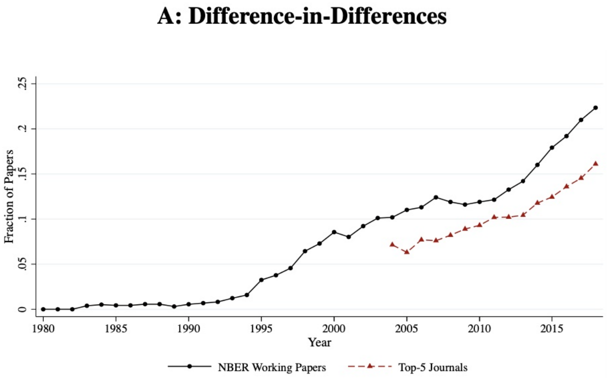
"A good way to do econometrics is to look for good natural experiments and use statistical methods that can tidy up the confounding factors that nature has not controlled for us." – Daniel McFadden (Nobel Laureate recipient with Heckman 1992)


I ♥ federalism
(for the natural experiments)

What is difference-in-differences (DiD)

- DiD is a very old, relatively straightforward, intuitive research design
- A group of units are assigned some treatment and then compared to a group of units that weren't
- One of the most widely used quasi-experimental methods in economics and even used in industry
- Historically used with "big shocks" happening in space over time but there are exceptions

Figure: Currie, et al. (2020)



Origins in Economics and Public Health

- In economics, David Card and Orley Ashenfelter are often associated with its origin – used in the 1970s and 1980s, with mixed success, to study job training programs
- Their dissatisfaction with it led to a call for more randomized controlled trials because, as we will see, it is not able to solve every kind of problem
- But it predates economics by 150 years when two health scientists (separately) used it to prove disease transmission mechanisms

Case I: Ignaz Semmelweis and washing hands

- 1840s, Vienna maternity wards had high postpartum infections in one wing compared to other wings
- One division had doctors and trainee doctors, but another had midwives and trainee midwives

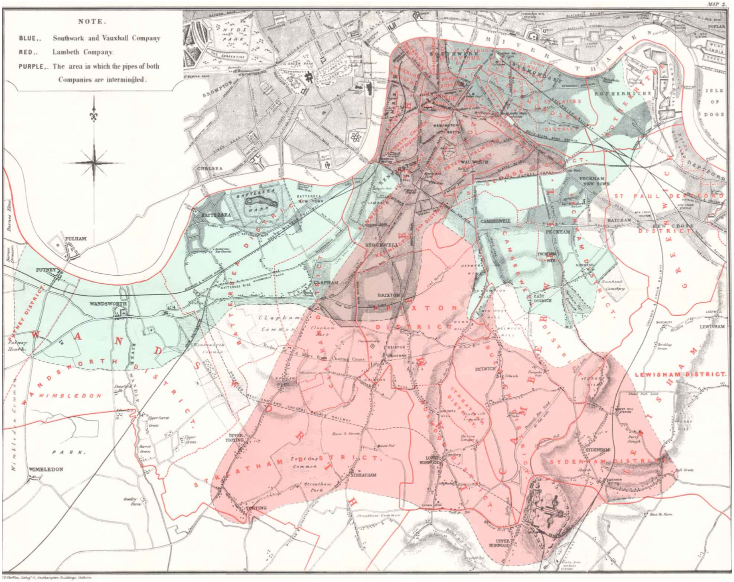
Case I: Ignaz Semmelweis and washing hands

- Training hospitals of students had earlier moved to “anatomical” training involving cadavers for classes
- Semmelweis thinks the mortality is caused by working with cadavers
- Proposes washing hands with chlorine in 1847 in the midwives’ wing (but not the physician wing)
- Mortality converged to the same levels in the two wings

Case II: John Snow and cholera

- Three major waves of cholera in the early to mid 1800s in London and people mistakenly thought the cause was “smelly air” (or *miasma*)
- John Snow believed cholera was spread through water and food which entered the Thames river through host’s evacuations
- Provides a variety of evidence for this, including a beautiful map, but also takes advantage of a natural experiment
- Lambeth water company moves its pipe between 1849 and 1854 but Southwark and Vauxhall delays

Figure: Two water utility companies in London 1854



Three ways to study this

1. Simple cross-section: compare mortality in 1854 for the two neighborhoods
2. Before and after: also called interrupted time series. Compare Lambeth in 1854 to 1849
3. Difference-in-differences: do both

3) Difference-in-differences

Table: Lambeth and Southwark and Vauxhall, 1849 and 1854

Companies	**Time**	**Outcome**	D_1	D_2
Lambeth	Before	$Y = L$		
	After	$Y = L + T_L + D$	$T_L + D$	
				D
Southwark and Vauxhall	Before	$Y = SV$		
	After	$Y = SV + T_{SV}$	T_{SV}	

$$\widehat{\delta}_{did} = D + (T_L - T_{SV})$$

D is the "treatment effect". It's the effect of moving the water on Lambeth mortality and it's in blue because we can't see it

3) Difference-in-differences

Table: Lambeth and Southwark and Vauxhall, 1849 and 1854

Companies	**Time**	**Outcome**	D_1	D_2
Lambeth	Before	$Y = L$		
	After	$Y = L + T_L + D$	$T_L + D$	
				D
Southwark and Vauxhall	Before	$Y = SV$		
	After	$Y = SV + T_{SV}$	T_{SV}	

$$\widehat{\delta}_{did} = D + (T_L - T_{SV})$$

T_L is a natural change in mortality that would have happened had they not moved the pipe upstream. It creates problems for us

3) Difference-in-differences

Table: Lambeth and Southwark and Vauxhall, 1849 and 1854

Companies	**Time**	**Outcome**	D_1	D_2
Lambeth	Before	$Y = L$		
	After	$Y = L + T_L + D$	$T_L + D$	
				D
Southwark and Vauxhall	Before	$Y = SV$		
	After	$Y = SV + T_{SV}$	T_{SV}	

$$\widehat{\delta}_{did} = D + (T_L - T_{SV})$$

But, if $T_L = T_{SV}$, which is called "parallel trends", then we can identify D using DiD. And that's DiD in a nutshell.

Potential outcomes review

- Previous table was how I learned DiD the first time, but now we need to become more formalized
- The following notation is not often taught in our introductory econometrics courses which tend to focus on regressions first and causality second
  - $\rightarrow$ We will focus on causality first, regressions second
- This is a simple review of the potential outcomes model by Jerzy Neyman (1923) and Don Rubin (1973)
- Potential outcomes notation is the dominant language of causality, though there are others too (e.g., Pearl)

Potential outcomes notation

- Let the treatment be a binary variable:

$$D_{i,t} = \begin{cases} 1 \text{ if pipe inlet is upstream at time } t \\ 0 \text{ if pipe inlet is downstream at time } t \end{cases}$$

where i indexes an individual observation, such as a person

Potential outcomes notation

- Potential outcomes:

$$Y_{i,t}^{j} = \begin{cases} 1\text{: health if drank from upstream at time } t \\ 0\text{: health if drank from downstream at time } t \end{cases}$$

where j indexes a counterfactual state of the world

Potential vs realized

- Distinction between the potential outcome Y^1 and the realized outcome Y – one is hypothetical and the other is real
- Potential outcomes are "selected" to become real when people choose their treatments represented here with a "switching equation"

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

- Example: My wages if I go to college are Y^1 and my wages if I don't go to college are Y^0, but since I went to college ($D = 1$), my wages are $Y = Y^1$.
- Point here is we define causality using potential outcomes, but data is realized outcomes, which creates problems

Treatment effect definitions

Individual treatment effect

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

Core building block of causal inference is the individual treatment effect.

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:

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

Since each person has an individual treatment effect, we can summarize them any number of ways – average treatment effect for girls, for old people, for people who like trivia night. Or for people who live in the Lambeth neighborhood.

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:

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

Why is the first term black but the second term red?

Roadmap

Background

- Introduction
- Potential outcomes

Identification and Estimation

- Parallel trends
- Estimation with OLS specification
- Inference

Parallel trends violations

- How parallel trends can get violated
- Types of evidence
- Triple difference
- Placebo outcomes

Fundamental problem of causal inference

- Potential outcomes reference a causality concept called the counterfactual
- Individual treatment effect is defined by comparing your outcome today with a college degree to your outcome today without a college degree
  - → Obviously those can't both be true – either you have a degree today or you don't
- Parameters (i.e., the causal effects, δ) are different from estimates of those parameters (i.e., $\widehat{\delta}$)

Steps of your causal projects

1. Define the parameter we want (“ATT”),
2. Ask what what beliefs do you need (“identification”), and
3. Build cranks that produce the correct numbers (“estimator”)

People often skip 1 and 2 and go straight to 3 and run regressions then go back and assume exogeneity (step 2), and hope that the estimates are weighted averages of individual treatment effects (1), but that is not guaranteed

DiD is four averages and three differences

I call this the DiD equation, but Goodman-Bacon calls it the “2x2”; I’ll use his k and U notation for treated and untreated groups

$$\widehat{\delta}^{2x2}_{kU} = \bigg(E[Y_k|Post] - E[Y_k|Pre] \bigg) - \bigg(E[Y_U|Post] - E[Y_U|Pre] \bigg)$$

k index people with Lambeth, U index people with Southwark and Vauxhall, $Post$ is after Lambeth moved pipe upstream, Pre before Lambeth moved its pipe (baseline), and $E[y]$ mean cholera mortality.

DiD is four averages and three differences

"Pre" (1849) and "Post" (1854) refer to when Lambeth, k, was treated which is why it is the same for both k and U groups

$$\widehat{\delta}^{2x2}_{kU} = \bigg(E[Y_k|Post] - E[Y_k|Pre]\bigg) - \bigg(E[Y_U|Post] - E[Y_U|Pre]\bigg)$$

Since we have one treatment group, then "Pre" and "Post" reference Lambeth's treatment date

Potential outcomes and the switching equation

$$\widehat{\delta}^{2x2}_{kU} = \underbrace{\bigg(E[Y^1_k|Post] - E[Y^0_k|Pre]\bigg) - \bigg(E[Y^0_U|Post] - E[Y^0_U|Pre]\bigg)}_{\text{Replace potential outcomes with realized outcomes using switching equation}}$$

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

Parallel trends bias

Rearrange and we get this:

$$\widehat{\delta}^{2x2}_{kU} = \underbrace{E[Y^1_k|Post] - E[Y^0_k|Post]}_{\text{ATT}} + \underbrace{\left[E[Y^0_k|Post] - E[Y^0_k|Pre]\right] - \left[E[Y^0_U|Post] - E[Y^0_U|Pre]\right]}_{\text{Non-parallel trends bias in 2x2 case}}$$

Parallel trends bias

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

The left hand side is our DiD estimator (i..e, four averages, three differences); the right hand side has our parameter (top) and assumption (parallel trends, bottom). Recall from the earlier table how DiD was equal to $D + (T_L - T_{SV})$. That's this.

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 cholera mortality for Lambeth *had it kept its pipe downstream* is the same as the evolution of cholera mortality for Southwark and Vauxhall".

It's in red so you know it's a nontrivial assumption. But why? Can't we just check?

Group work

- Before we move into regression, let's go through a simple exercise to really pin down these core ideas with simple calculations
- I've passed around a worksheet. We will spend some time together doing this worksheet; use the online or use the worksheet

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

OLS Specification

- Simple DiD equation (four averages, three differences) estimates ATT under parallel trends; don't need regression
- But there is an OLS specification that is numerically identical to four averages and three differences
- OLS was historically preferred because
  - → OLS estimates the ATT under parallel trends so it is valid
  - → Easy to calculate the standard errors
  - → Easy to include multiple periods which increases power and makes estimates more precise
- This specification is not appropriate under differential timing or with the inclusion of covariates

Minimum wages

- Card and Krueger (1994) have a famous study estimating causal effect (ATT) of minimum wages on employment
- Exploited a policy change in New Jersey between February and November in mid-1990s where minimum wage was increased, but neighbor PA did not
- Using DiD, they do not find a negative effect of the minimum wage on employment which is part of its legacy today, but I mainly present it to illustrate the history and the design principles

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

Quick comment

- Buchanan’s comment gets taken out of historical context to a degree
- Empirical labor and empirical macroeconomics (e.g., Lucas Critique) had been going back to the 1970s in a bit of a “empirical crisis” much like we see sometimes today with debates about p-hacking, but theirs was more basic confusion of causality and correlation
- Consequently, the dominant paradigm in “knowing facts in economics” was theory, not empiricism
- So Buchanan’s dismissiveness probably had traces of that; quality of empirical work was sub standard so people tended to not take it very seriously

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."*

But let's listen to Orley's opinion about the paper's controversy at the time. `https://youtu.be/MOtbuRX4eyQ?t=1882`

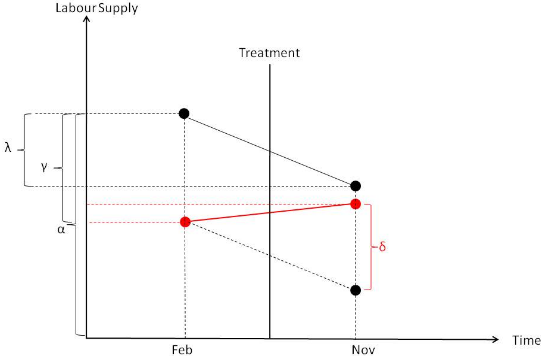
OLS specification of the DiD equation

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

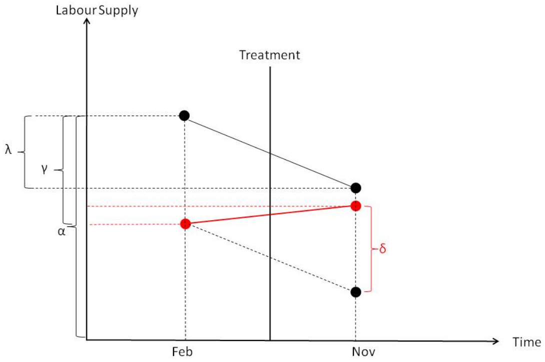
$$Y_{its} = \alpha + \gamma NJ_s + \lambda d_t + \delta(NJ \times d)_{st} + \varepsilon_{its}$$

  - → 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

- Bertrand, Duflo and Mullainathan (2004) show that conventional standard errors will often severely understate the standard deviation of the estimators
- Standard errors are biased downward (i.e., too small, over reject)
- They proposed three solutions, but most only use one of them (clustering)

Inference

1. Block bootstrapping standard errors (if you analyze states the block should be the states and you would sample whole states with replacement for bootstrapping)
2. Clustering standard errors at the group level (in Stata one would simply add `, cluster(state)` to the regression equation if one analyzes state level variation)

Most people will simply cluster, but there are issues if you have too few clusters. They mention a third way but it's only a curiosity.

Roadmap

- Background
  - Introduction
  - Potential outcomes
- Identification and Estimation
  - Parallel trends
  - Estimation with OLS specification
  - Inference
- Parallel trends violations
  - How parallel trends can get violated
  - Types of evidence
  - Triple difference
  - Placebo outcomes

Violating parallel trends exercise

- Parallel trends enable regressions to correct impute counterfactual $E[Y^0|D=1]$ using $E[Y^0|D=0]$
- OLS *always* imputes but it only is right if parallel trends is true
- Which means if parallel trends isn't true, then the imputation isn't correct and therefore estimates are biased
- To illustrate this, let's go through the document again (tab is "DID 2")

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

Violating parallel trends

- Parallel trends are in expectation only – we don't rely everybody to follow the same trend, just that the group average for Y^0 be approximately the same for treated and control
- Violations are a form of selection bias and there are two straightforward ways that parallel trends will be violated
  1. Compositional differences in samples associated with repeated cross-sections
  2. Treatment coincided with some other important event that only affected treatment group

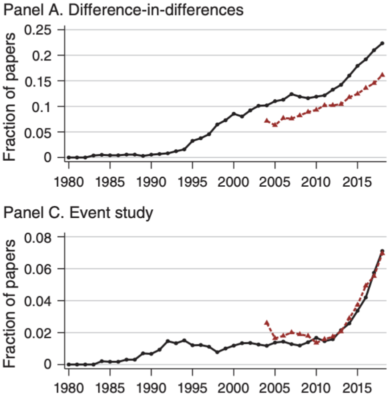
Think like a prosecutor

- You are building a case, the prosecutor before a judge and jury, always in battle with the defense attorney
- Evidence has particular broadly defined forms that can help you on the front end
- Your goal in my humble opinion should be mixing tight logic based falsifications with particular kinds of data visualization, starting with the event study

Three types of evidence

1. **Bite**: Show that the treatment impacted first order behavior before showing how it affected second order behavior
2. **Event studies**: A particular kind of data visualization focused on pre- and post-treatment DiD coefficients in a regression equation
3. **Placebos**: Ruling out reasonable competing theories using the same regression model on different outcomes; can include triple differences
4. Mechanisms: Can you a link between the bite and the outcome?

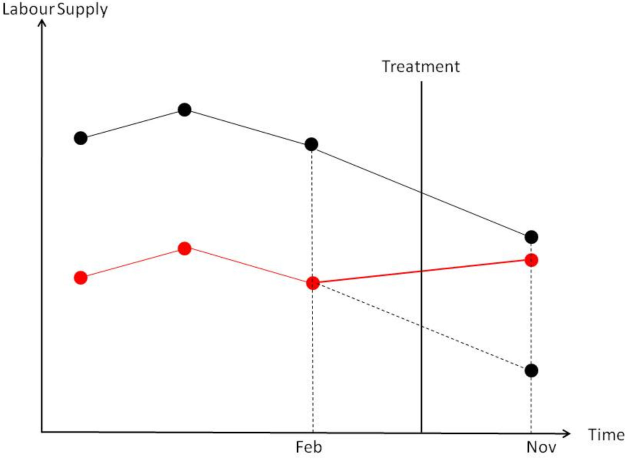
Event studies have become mandatory in DiD



Intuition behind event studies

- Princeton Industrial Relations Section seems to be behind this – this intense focus on research design but also verifying assumptions
- The identifying assumption for all DD designs is parallel trends , but since we cannot verify parallel trends, we often look at pre-trends
- It’s a type of check for selection bias, but you must understand what it is and what it isn’t to see its value but not be naive about it (it is not a silver bullet)
- 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)

Plot the raw data when there's only two groups



Evidence for parallel trends: pre-trends

Let's do the bonus questions on first and second tab now
https://docs.google.com/spreadsheets/d/1onabpc14JdrGo6NFv0zCWo-nuWDLLV2L1qNogDT9SBw/edit?usp=sharing

Pre trends as a type of evidence

- Parallel pre-trends $\neq$ parallel trends – these are often thought to be the same thing, and they aren't
- Equating them is a kind of *post hoc ergo propter hoc* fallacy
- Parallel pre-trends is more like a smoking gun based on things "looking the same" before
- Checking pre-trends is just a form of falsification, as well as giving us some assurances our groups are comparable

Event study regression

- Event studies have a simple OLS specification with only one treatment group and one never-treated group

$$Y_{its} = \alpha + \sum_{\tau=-2}^{-q} \mu_\tau D_{s\tau} + \sum_{\tau=0}^{m} \delta_\tau D_{s\tau} + \varepsilon_{ist}$$

→ where D is an interaction of the treatment group s with the calendar year τ
→ Treatment occurs in year 0, no anticipation, drop baseline $t-1$
→ Includes q leads or anticipatory effects and m lags or post treatment effects

Event study regression

$$Y_{its} = \alpha + \sum_{\tau=-2}^{-q} \mu_\tau D_{s\tau} + \sum_{\tau=0}^{m} \delta_\tau D_{s\tau} + \varepsilon_{ist}$$

Typically you'll plot the coefficients and 95% CI on all leads and lags (binned or not, trimmed or not)

Under no anticipation, then you expect $\widehat{\mu}$ coefficients to be zero, which gives you confidence that parallel trends holds (but is not a guarantee, and there are still specification issues – see Jon Roth's work)

Under parallel trends, $\widehat{\delta}$ are estimates of the ATT at points in time

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

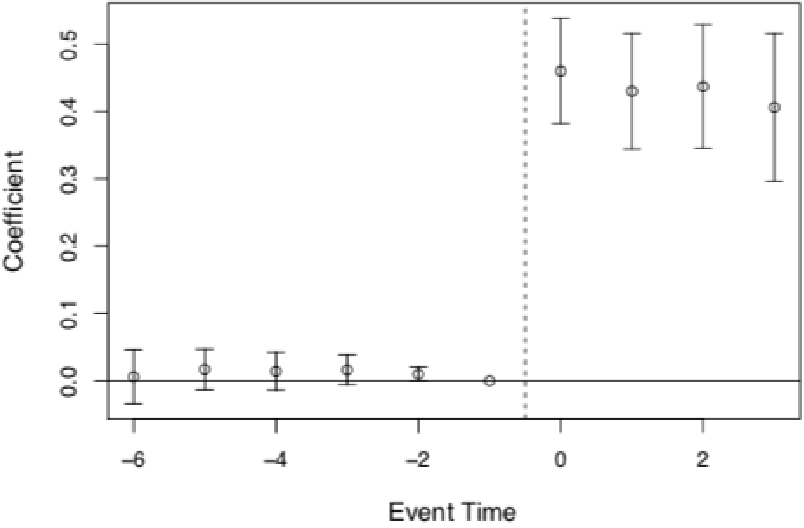
We use large-scale federal survey data linked to administrative death records to investigate the relationship between Medicaid enrollment and mortality. Our analysis compares changes in mortality for near-elderly adults in states with and without Affordable Care Act Medicaid expansions. We identify adults 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

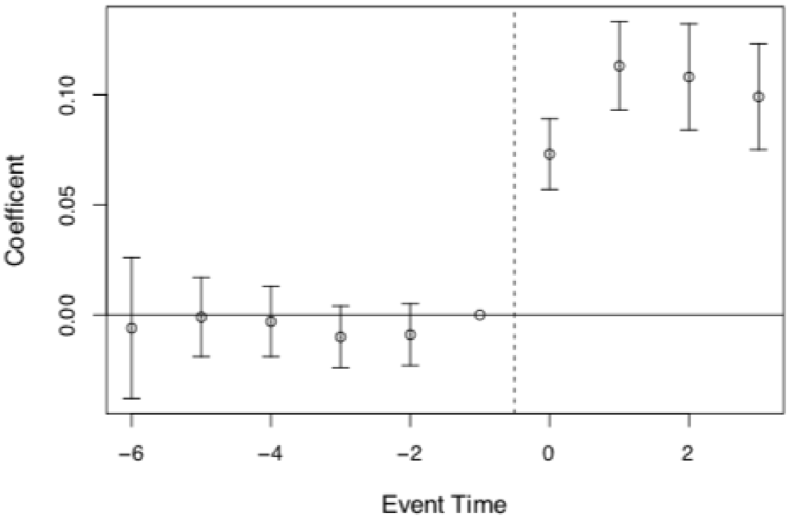
Issue Section: Article

Their three types of evidence

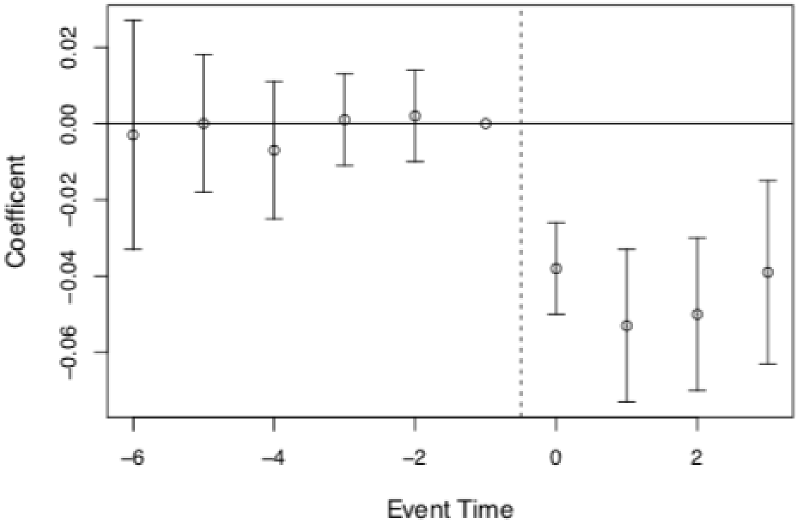
- **Bite** – show that the expansion shifted people into Medicaid and out of uninsured status
- **Placebos** – Show that there's no effect on mortality for groups it shouldn't be affecting (people 65+)
- **Event study** – Show leads and lags on mortality



(a) Medicaid Eligibility



(b) Medicaid Coverage

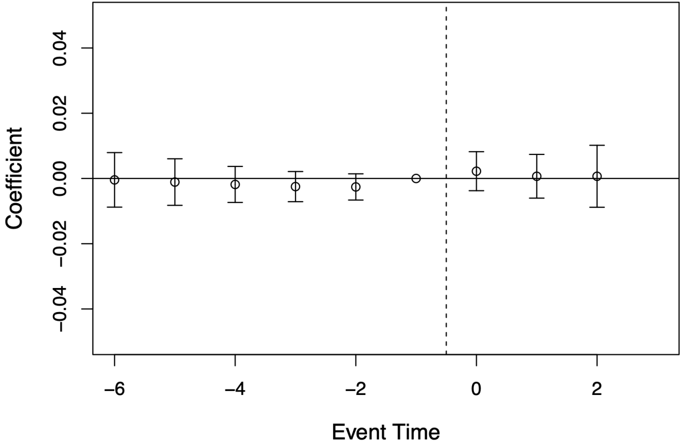


(c) Uninsured

Quick review

- **Bite**: Did the expansion of Medicaid put more people on Medicaid?
    1. 40pp increase in people eligible (but this was mechanical)
    2. 6-10pp increase in people on Medicaid (but maybe it was crowding out private insurance?)
    3. 4-6pp decrease in uninsured (at least some of the marginal Medicaid enrollees had been uninsured)
- **Placebo**: their main result is about near-elderly mortality (i.e., around age 64), but first they look at the effect on actual elderly (65+)

65 and older mortality placebo



Discussion: Why do they do this? Explain to me like I'm 5 the value of a picture like this.

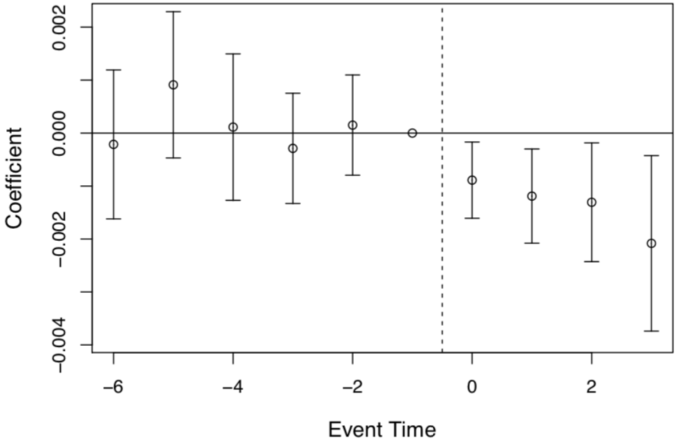


Figure: Miller, et al. (2019) estimates of Medicaid expansion’s effects on on annual mortality

Lab

- We will now use Stata and R code to implement simple DiD equations manually, with OLS, and for event studies (both manually and using OLS)
- Goal of this exercise is simply to deepen your understanding of the mechanics
- Please go to this link `https://github.com/Mixtape-Sessions/Causal-Inference-2/tree/main/Lab/Lalonde`
- Q1:a-c vs. Q2a. We won't do the covariate part yet.

Triple differences as alternative strategy

- Very common for readers and others to request a variety of "robustness checks" from a DD design
- We saw some of these just now (e.g., falsification test using data for alternative control group, the Medicare population)
- Triple differences uses a within-state untreated group; little trickier, so let's use the table again

DDD Example by Gruber

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)**		

Table: Difference-in-Difference-in-Differences numerical example

States	Group	Period	Outcomes	D_1	D_2	D_3
Experimental states	Married women, 20-40yo	After	$NJ + T + NJ_t + l_t + D$	$T + NJ_t + l_t + D$	$D + l_t - s_t$	D
		Before	NJ			
	Older 40, Single men 20-40yo	After	$NJ + T + NJ_t + s_t$	$T + NJ_t + s_t$		
		Before	NJ			
Non-experimental states	Married women, 20-40yo	After	$PA + T + PA_t + l_t$	$T + PA_t + l_t$	$l_t - s_t$	
		Before	PA			
	Older 40, Single men 20-40yo	After	$PA + T + PA_t + s_t$	$T + PA_t + s_t$		
		Before	PA			

What is our identifying assumption? Answer: just another parallel trend assumption but with a slightly different interpretation!

$l_t - s_t$ is the same for the experimental and non-experimental states. This is "change in inequality between two groups hourly wages" from pre to post. It's a new parallel trend assumption.

DDD in Regression

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

- Your panel is now a group j state i (e.g., AR high wage worker 1991, AR high wage worker 1992, etc.)
- Assume we drop τ_t but I just want to show it to you for now.
- If the placebo DD is non-zero, it might be difficult to convince the reviewer that the DDD removed all the bias

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'.

Falsification test with alternative outcome

- The within-group control group (DDD) is a form of placebo analysis using the same *outcome*
- But there are also placebos using a *different* outcome – but you need a hypothesis of mechanisms to figure out what is in fact a *different outcome*
- Figure out what those are, and test them – finding no effect on placebo outcomes tends to help people your other results interestingly enough
- 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

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
- Ockham’s razor - given social interaction endogeneity (Manski 1993), homophily more likely explanation

Concluding remarks

- So we hopefully see a few of the key elements of DiD
  - → Remember: the DiD equation and ATT equation are distinct concepts and definitions
  - → DiD designs can be implemented with OLS specifications that calculate differences in means
  - → Parallel pre-trends and parallel trends are not the same thing – the first is testable, the latter is not testable
  - → Event studies are mandatory but pre-trends are smoking guns, but can mislead nonetheless
- Now we want to move into the fixed effects work