Causal Inference
Day 2

J.J. Naddeo

Massive Data Institute

April 19, 2023

Difference in Differences

- Finding a valid instrument is hard!!

Difference in Differences

- Finding a valid instrument is hard!!
- Let's look at another workhouse framework in economics: Difference in differences (DiD)

Difference in Differences

- Finding a valid instrument is hard!!
- Let's look at another workhouse framework in economics: Difference in differences (DiD)
- This framework is much older than IV (mid 1800s)

Difference in Differences

- Finding a valid instrument is hard!!
- Let's look at another workhouse framework in economics: Difference in differences (DiD)
- This framework is much older than IV (mid 1800s)
- A group of units are assigned some treatment and then compared to a group of units that weren't
    - Therefore mostly used when you observe a large shock/policy change in time that only impacts certain units

Difference in Differences

- Finding a valid instrument is hard!!
- Let's look at another workhouse framework in economics: Difference in differences (DiD)
- This framework is much older than IV (mid 1800s)
- A group of units are assigned some treatment and then compared to a group of units that weren't
  - Therefore mostly used when you observe a large shock/policy change in time that only impacts certain units

John Snow and cholera

- Three major waves of cholera in the early to mid 1800s in London
- John Snow believed cholera was spread through the Thames water supply which contradicted dominant theory about "dirty air" transmission
- Grand experiment: Lambeth moves its pipe between 1849 and 1854; Southwark and Vauxhall delay
- He can evaluate the effect in three ways (one of which is DiD)

Figure: Two water utility companies in London 1854

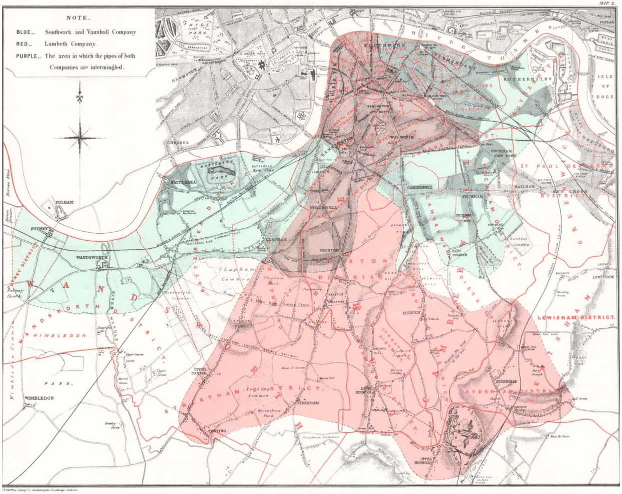


Figure: Results

Death-Rates per 1,000 of living Population in Two Epidemic Periods.		In Houses enumerated in 1854 as receiving their Water-supply— from the Lambeth Company.	from the Southwark and Vauxhall Company.
Cholera	1848-9..	12.5	11.8
	1853-4..	3.7	13.0
Diarrhœa	1848-9..	2.9	2.7
	1853-4..	2.1	3.3

1) Simple cross-sectional design

Table: Lambeth and Southwark and Vauxhall, 1854

Company	**Cholera mortality**
Lambeth	$Y = L + D$
Southwark and Vauxhall	$Y = SV$

$$\widehat{\delta}_{cs} = D + (L - SV)$$

What is L and SV?

1) Simple cross-sectional design

Table: Lambeth and Southwark and Vauxhall, 1854

Company	**Cholera mortality**
Lambeth	$Y = L + D$
Southwark and Vauxhall	$Y = SV$

$$\widehat{\delta}_{cs} = D + (L - SV)$$

This is biased if Ł $\neq SV$ (selection bias). Give an example when we're pretty sure they are equal.

2) Interrupted time series design

Table: Lambeth, 1849 and 1854

Company	Time	Cholera mortality
Lambeth	1849	$Y = L$
	1854	$Y = L + (T + D)$

$$\widehat{\delta}_{its} = D + T$$

What is required for this estimator to be unbiased?

3) Difference-in-differences

Table: Lambeth and Southwark and Vauxhall, 1849 and 1854

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

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

Treatment effect definitions

Individual treatment effect

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

Individual causal effects cannot be calculated because one of the two needed potential outcomes will always be missing.

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:

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

We can estimate the ATT, but never be sure due to missing potential outcomes for the treated group

Identification without randomization

- We may be unable to randomize – not because we lack the imagination, but because we lack the permission
- If we cannot randomize, then how does DiD identify a treatment effect, and which treatment effect?
- DiD identifies the ATT, and since we are missing Y^0 for treated group, we will restrict counterfactual Y^0 in expectation
- One of the main advantages of DiD is the hope that we can identify the ATT *without* randomization

DiD equation

DiD Equation: k and U notation for treated and untreated groups

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

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.

Potential outcomes and the switching equation

$$\widehat{\delta}^{2x2}_{kU} = \underbrace{\bigg(E[Y^1_k|Post] - E[Y^0_k|Pre] \bigg) - \bigg(E[Y^0_U|Post] - E[Y^0_U|Pre] \bigg)}_{\text{Switching equation}} + \underbrace{E[Y^0_k|Post] - E[Y^0_k|Post]}_{\text{Adding zero}}$$

Parallel trends bias

$$\widehat{\delta}^{2x2}_{kU} = \underbrace{E[Y^1_k|Post] - E[Y^0_k|Post]}_{\text{ATT}}$$

$$+ \underbrace{\Big[E[Y^0_k|Post] - E[Y^0_k|Pre]\Big] - \Big[E[Y^0_U|Post] - E[Y^0_U|Pre]\Big]}_{\text{Non-parallel trends bias in 2x2 case}}$$

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?

Parallel trends is hard

- **There is no guaranteed "science" in parallel trends**
- This makes DiD a "hard" design – *because* it doesn't rely on randomization (it relies on parallel trends), there are no "slam dunks" and evidence tends to be multi-layered
- Before we move into regression, let's go through a simple exercise to really pin down these core ideas with simple calculations

Simple Exercise

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 (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

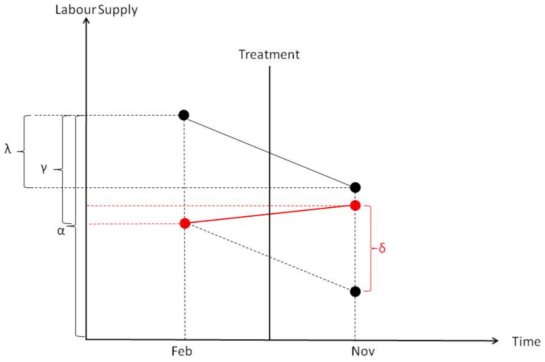
OLS specification of the DiD equation

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

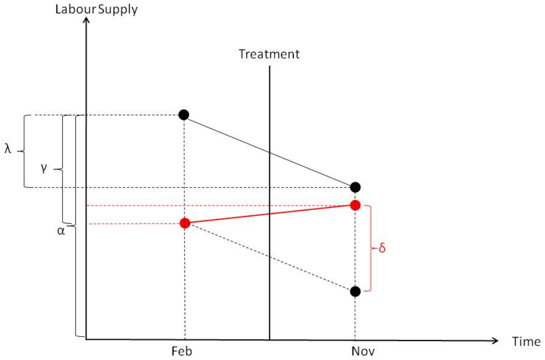
$$Y_{its} = \alpha + \gamma NJ_s + \lambda d_t + \delta(NJ \times d)_{st} + \varepsilon_{its}$$

  - 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) $= \delta$

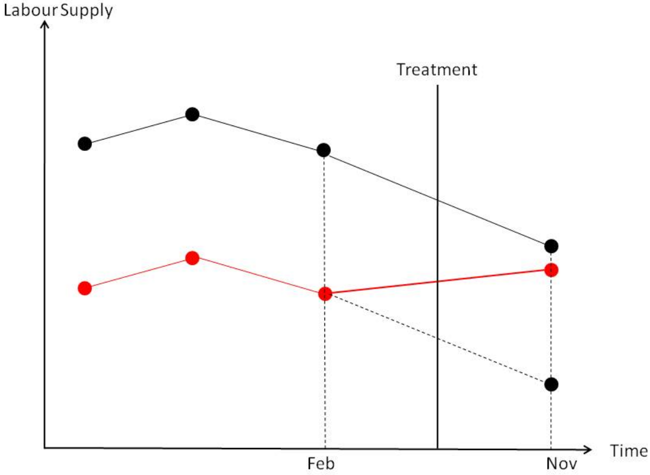
$$Y_{ist} = \alpha + \gamma NJ_s + \lambda d_t + \delta(NJ \times d)_{st} + \varepsilon_{ist}$$



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



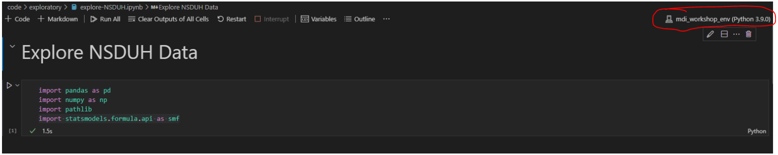
Notice how OLS is "imputing" $E[Y^0|D = 1, Post]$ for the treatment group in the post period? It is only "correct", though, if parallel trends is a good approximation


Labour Supply
Treatment
Feb
Nov
Time

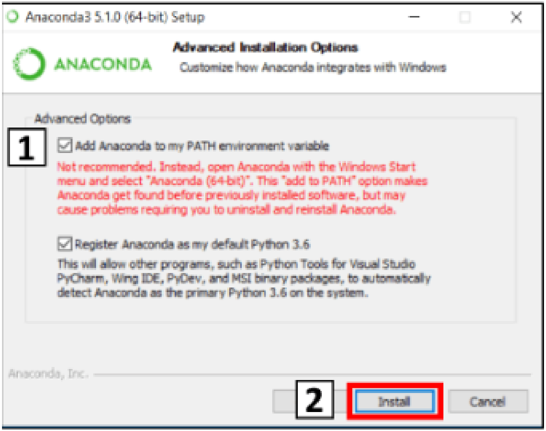
START OF APPENDIX SLIDES

Select Virtual Environment

Go Back

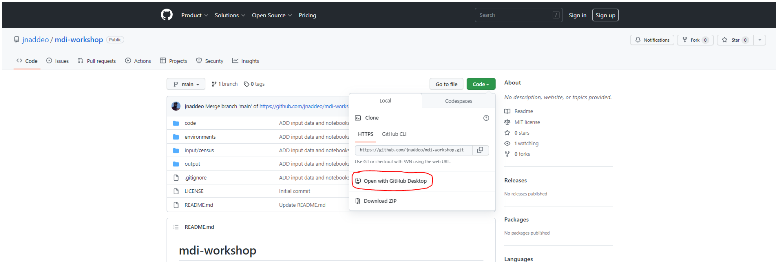


Add conda to PATH Go Back



Open Repo with Desktop

Go Back



Random Assignment Black

Go back

figures/randomization_check_dismissal_black.png

Random Assignment White

Go back

figures/randomization_check_dismissal_white.png

Monotonicity

Go back

- **?** provides joint test for null hypothesis that the exclusion and monotonicity assumptions hold, I fail to reject the null
- Also show assumption can be relaxed to average monotonicity $\Longrightarrow$ first stage should hold in all subsamples

figures/first_stage_subsamples.png