# DS-UA 201: Causal Inference: Differences-in-differences



Parijat Dube



New York University Center for Data Science

August 8, 2022



Acknowledgement: Slides including material from DS-US 201 Fall 2021 offered by Marco Morucci.

# Fixed-effects regression

**Last lecture:** we talked about regression with **grouped data**

- Grouped data is any data whose units of analysis can be divided into groups
- and **unobserved confounding** is constant within groups
- even if we don't observe all confounders causal inference is still possible



We can use linear regression to analyze grouped data

- We can de-mean outcomes and treatments and then use regression on the de-meaned variables ("within" estimator)
- or we can one-hot encode the group indicators and then run a linear regression on the new variables (Fixed effects estimator)
- we need the OLS assumptions in both cases
- . . . but we can account for correlation of units within groups
- . . . and deal with multiple groupings at once

# Today: time-series data

**Today** and next week, we will talk about one specific kind of grouping: **time-series** data.

- This is data in which the same unit, $i$, was observed over multiple time-periods $t$
- Intuitively, there are two groups in this data
  - The **unit** itself
  - and the **time** of the observation
- We will see that, in this case, the assumption of constant unobserved confounding within unit can be relaxed even further in this setting
- FE regression is going to be our tool of choice for this setting as well

pictures/nobel-economics.jpeg

## Card and Krueger (1994)

**Question:** Does increasing the minimum wage reduce employment?

- Classical theoretical models predict that wage floors $\rightsquigarrow$ people unemployed that would otherwise be employed.
- But what does the empirical evidence tell us?

**Card and Krueger** (1994) analyze a policy change that occurred in New Jersey in 1992 raising MW from \$4.25 to \$5.05.

- They survey 410 fast food restaurants in New Jersey and neighboring Eastern Pennsylvania.
- Compare change in employment in NJ before and after the policy to the change in employment in PA restaurants (which experienced no minimum wage change) before and after NJ's policy.
- **Finding:** The minimum wage increase didn't decrease employment (in fact, there was a slight but statistically not significant increase).

## Card and Krueger: Research design

**Key Idea**: not only do Card and Krueger compare "treated" (NJ) restaurants to "control" (PA) restaurants, but also address concerns that there is something about NJ restaurants that affects both the MW change, and employment



Underlying assumption



Here the key assumption is that whatever was special about NJ **did not change** in the period before and after the policy was implemented.



- In other words, if there was truly no effect of the MW change, the difference in employment before and after the MW policy should be the same in the two states.
- Alternatively, if there was something special about the NJ restaurants, we should see those districts have a different employment rate before the policy was implemented.

# Difference-in-differences

Two groups (treated/control); two time periods $(0, 1)$.

- $D_i = 1$: treated in time 1; $D_i = 0$: control in time 1. All units under control in time 0. Can also think in terms of a treatment indicator for each time period: $D_{i1} = D_i$, $D_{i0} = 0$
- Observe two outcomes for each unit $i$: $Y_{i1}$: outcome in period 1, $Y_{i0}$: outcome in period 0.



**Potential outcomes**:

$$Y_{i1}(d) = Y_{i1} \text{ if } D_i = d$$

Treatment in time 1 **has no effect** on the outcome in time 0 (everyone is under control in time 0)

$$Y_{i0}(1) = Y_{i0}(0) = Y_{i0}.$$

# Types of datasets for DiD

The most typical dataset used for DiD is one with **repeated observations of the same unit**



- Known in economics as "panel" data
- We will work with this throughout the rest of this lecture



Alternatively, DiD works also for repeated cross-sections sampled from treated/untreated units.

- Here we have **different samples** of units in the two time periods
- Inferences still valid **as long as the samples are from the same population**
- Example: Two different samples of restaurants from NJ pre-and post NJ minimum wage change, Two samples from PA during the same period

# Identifying assumptions

Causal effect of interest is the **Average Treatment Effect on the Treated** (ATT) in time 1:

$$\tau_1 = \mathbb{E}[Y_{i1}(1)|D_i = 1] - \mathbb{E}[Y_{i1}(0)|D_i = 1]$$

The first part we can estimate directly from the data (observed outcome among the treated group)

$$\tau_1 = \mathbb{E}[Y_{i1}|D_i = 1] - \mathbb{E}[Y_{i1}(0)|D_i = 1]$$

The second part we don't observe directly and **need additional assumptions** to identify from the observed data.

# Identifying assumptions

Remember the **selection bias** formula for the ATT:



$$\tau_1 = \underbrace{\{\mathbb{E}[Y_{i1}|D_i = 1] - \mathbb{E}[Y_{i1}|D_i = 0]\}}_{\text{Difference-in-means in time 1}} -$$

$$\underbrace{\{\mathbb{E}[Y_{i1}(0)|D_i = 1] - E[Y_{i1}(0)|D_i = 0]\}}_{\text{Selection bias}}$$



- Can we estimate the selection bias?



Key Assumption: Parallel trends

The selection bias in time 1 (difference in $Y_{i1}(0)$ between treated and control) is the same as the selection bias in time 0 (difference in $Y_{i0}(0)$ between treated and control).

# Identifying assumptions

Our identifying assumption lets us write



$$\mathbb{E}[Y_{i1}(0)|D_i = 1] - E[Y_{i1}(0)|D_i = 0] =$$
$$\mathbb{E}[Y_{i0}(0)|D_i = 1] - E[Y_{i0}(0)|D_i = 0]$$

Since $Y_{i0}(0) = Y_{i0}(1)$ (no effect of future on past)



$$\mathbb{E}[Y_{i1}(0)|D_i = 1] - E[Y_{i1}(0)|D_i = 0] =$$
$$\mathbb{E}[Y_{i0}(1)|D_i = 1] - E[Y_{i0}(0)|D_i = 0]$$



Then consistency yields

$$\mathbb{E}[Y_{i1}(1)|D_i = 1] - E[Y_{i1}(0)|D_i = 0] =$$
$$\mathbb{E}[Y_{i0}|D_i = 1] - E[Y_{i0}|D_i = 0]$$

# Identifying assumptions

Substituting back into the ATT formula yields the **differences-in-differences** formula



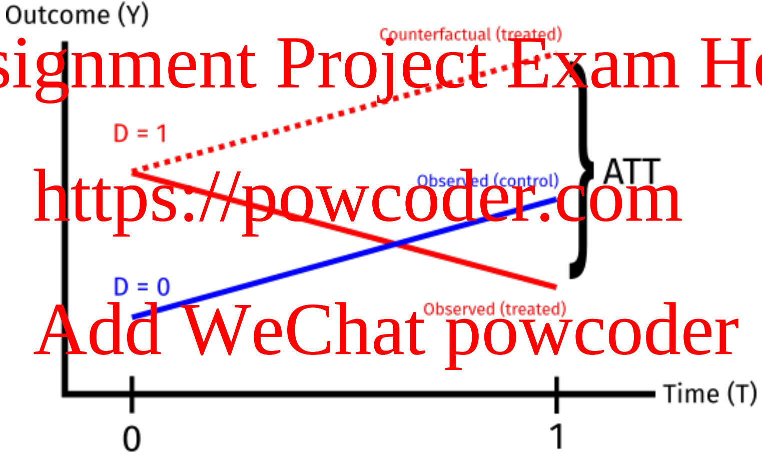
$$\tau_1 = \underbrace{\{\mathbb{E}[Y_{i1}|D_i = 1] - \mathbb{E}[Y_{i1}|D_i = 0]\}}_{\text{Difference-in-means in time 1}} - \underbrace{\{\mathbb{E}[Y_{i0}|D_i = 1] - \mathbb{E}[Y_{i0}|D_i = 0]\}}_{\text{Difference-in-means in time 0}}$$





- Each of these can be estimated non-parametrically using the conditional sample means.
- But: the Neyman variance won't work. We will need a special type of bootstrap.

# Visualizing DiD



# Identifying assumptions

The key identifying assumption of DiD is often referred to as the "parallel trends" assumption.



- We cannot assume treatment is ignorable with respect to the potential outcomes at time 1.
- We instead assume that the **trend** in the potential outcomes under control from time 0 to time 1 in the treated group is the same as the observed trend in the control group.



**Parallel trends assumption**:



$$\underbrace{E[Y_{i1}(0)|D_i = 1] - E[Y_{i0}(1)|D_i = 1]}_{\text{Trend in control counterfactual for treated}} = \underbrace{E[Y_{i1}(0)|D_i = 0] - E[Y_{i0}(0)|D_i = 0]}_{\text{Trend in control counterfactual for control}}$$

## Identifying assumptions

Another way of phrasing the parallel trends assumption is that the **trends** are independent of treatment assignment (but not the levels):

$$\{Y_{i1}(0) - Y_{i0}(0)\} \perp\!\!\!\perp D_i$$

But it **is not true** that:

$$Y_{i1}(0) \perp\!\!\!\perp D_i, \text{ and } Y_{i0}(0) \perp\!\!\!\perp D_i.$$

Parallel trends is equivalent to saying that confounding is **constant** over time.

- So parallel trends is a special case of the more general assumption we made for grouped data

# Estimation

In the most general setting, we can just estimate the four different means used in DiD using the sample averages:

- $\hat{E}[Y_{i1}|D_i = 1] = \frac{1}{N_{1,1}} \sum_{i:t_i=1} Y_{i1} D_i$
- $\hat{E}[Y_{i0}|D_i = 1] = \frac{1}{N_{1,0}} \sum_{i:t_i=0} Y_{i0} D_i$
- $\hat{E}[Y_{i1}|D_i = 0] = \frac{1}{N_{0,1}} \sum_{i:t_i=1} Y_{i1} (1 - D_i)$
- $\hat{E}[Y_{i0}|D_i = 0] = \frac{1}{N_{0,0}} \sum_{i:t_i=0} Y_{i0} (1 - D_i)$

and then take the difference of the differences to estimate $\tau_1$:

$$\hat{\tau}_1 = \hat{E}[Y_{i1}|D_i = 1] - \hat{E}[Y_{i1}|D_i = 0]$$
$$- \hat{E}[Y_{i0}|D_i = 1] - \hat{E}[Y_{i0}|D_i = 0].$$

# Estimation

If the data consists of **repeated observations** of the same unit, then we can use a simple difference-in-means estimator on the differenced outcomes:

$$\hat{\tau}_1 = \frac{1}{N_t}\sum_{i=1}^{n}(Y_{i1} - Y_{i0})D_i - \frac{1}{N_c}\sum_{i=1}^{n}(Y_{i1} - Y_{i0})(1 - D_i)$$

No **parametric assumptions** required!

- We did not have to assume that we know the functional form of our outcome!

The SE of this estimator can be estimated with the **Neyman variance estimator** applied to the differenced outcomes.

## Connection to fixed-effects estimators

Suppose our dataset is organized where each row is a unit/time period, *it* (just like we just did).



We can write a model with "two-way fixed effects":

$$Y_{it} = \gamma_i + \delta_t + \tau D_{it} + \epsilon_{it}$$



- Estimating this regression and obtaining $\hat{\tau}$ is mathematically equivalent to the nonparametric DiD estimator in the two-period/two-treatment case.
  - This means that this specific regression is valid even without the OLS assumption as it's just a shortcut to another estimator.
- More complicated when we have many time periods and treatment initiation at different times (additional hidden assumptions to estimating the "two-way fixed effects" model)

## Standard Errors

In the two-way FE model, we have correlated errors ($Cov(\epsilon_{i1}, \epsilon_{i0}) \neq 0$)



- This is because the same unit appears at multiple times
- It is unrealistic to believe that there will be no error correlation **within the same unit**!



**Two Solutions**:

- "Cluster-robust" standard errors: the estimator we saw during our last lecture.



- Block-bootstrap (bootstrapping but resampling all observations within a cluster rather than just *it* rows)

# The block bootstrap

The **block bootstrap** is a version of the bootstrap for grouped data.

- The key is that we sample **whole groups** instead of individual observations

**Algorithm:** For $b = 1, \ldots, B$ bootstrap iterations:

1. From the $n$ units, randomly sample $n$ **with replacement**
2. For each unit sampled, store **both** $Y_{i1}$ and $Y_{i0}$, and $D_i$.
3. On the bootstrapped data, estimate $\hat{\tau}_1^{(b)}$ using either a difference in de-meaned outcomes or a two-way FE regression
4. Store $\hat{\tau}_1^{(b)}$

The **standard deviation** of the vector $(\hat{\tau}_1^{(1)}, \ldots, \hat{\tau}_1^{(B)})$ will be a consistent estimator of the standard error of $\hat{\tau}_1$.

# Block bootstrap in R (example from Matt Blackwell)

| unit | group | D | Y |
|---|---|---|---|
| 1 | A | 0 | 0.5 |
| 2 | A | 1 | 0.06 |
| 3 | B | 1 | -0.14 |
| 4 | B | 0 | -1.3 |



```
B = 1000
tau_boot = rep(NA, B)
for (b in 1:B){
  lookup <- split(1:nrow(data), dat$group)

  gnames <- names(lookup)
  star <- sample(gnames, size = length(gnames), replace
     = TRUE)
  head(lookup[star], n = 2)

  dat.star <- dat[unlist(lookup[star]), ]
  tau_boot[b] = lm_robust(Y~D + group, data=dat)
}
```

## The Card/Krueger Minimum Wage Study



**Question:** Does raising the minimum wage raise unemployment?

**Data**:



- Fast food restaurant in PA and NJ before and after the MW policy
- treatment: restaurant is treated if is in NJ after the policy
- outcome: number of full time employees

# Analyzing the Card/Krueger data

```
### Analyze the Card-Krueger Minimum Wage study
library(tidyverse)
library(haven)
library(estimatr)

## Read in the data (missing data denoted with a .)
minwage <- read_table2("public.tab", na= ".")

## Drop units with available wage + employment data in
    wave 1 and 2
minwage <- subset(minwage, !is.na(WAGE_ST)&!is.na(WAGE_
    ST2)&!is.na(EMPFT)&!is.na(EMPFT2)&!is.na(EMPPT)&!is
    .na(EMPPT2))

## State = 1: New Jersey (treated), State = 0:
    Pennsylvania (control)
## Outcome is FT employment
```

# Analyzing the Card/Krueger data

```r
## State = 1: New Jersey (treated), State = 0: Pennsylvania (control)
## Outcome is FT employment

## Naive difference-in-means
lm_robust(EMPFT2 ~ STATE, data=minwage)

### But NJ and PA differ - we want to look at the change
minwage$CHGEMPFT <- minwage$EMPFT2 - minwage$EMPFT

## Get the DiD estimate
lm_robust(CHGEMPFT ~ STATE, data=minwage)
```



| Method | **Naive** | **DiD** |
|---|---|---|
| Estimate | 0.23 | 2.92 |
| SE | 1.16 | 1.72 |
| 95% CI | -2.05, 2.5 | -0.47, 6.32 |

# Up next



- Assessing DiD assumptions + what happens with multiple time periods.



- Instrumental variables - what happens when treatment is confounded but we have a variable that's as-good-as randomized and could only affect the outcome through the treatment?