Topic 7: Difference-in-Differences and Factor Models

ECON 5783 — University of Arkansas

Prof. Kyle Butts

November 2024

Difference-in-Differences

- Initial Difference-in-difference usage
- Classic Example: Card and Krueger (2000, AER)
- Econometric formulation to DID

What is difference-in-differences (DiD)

Difference-in-differences compares a group assigned to treatment versus a group not assigned to treatment

- The estimator compares the treated groups change in outcomes before and after the treatment to the control groups change in outcomes before and after the treatment

One of the most widely used quasi-experimental methods in economics and increasingly in industry

Difference-in-Differences

- Initial Difference-in-difference usage
- Classic Example: Card and Krueger (2000, AER)
- Econometric formulation to DID

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)

- depending on the day of week and time of day, she would be routed to either the midwife wing or the physician wing

Pregnant women died after delivery in the (male) wing at a rate of 13-18%, but only 3% in the (female) midwife wing

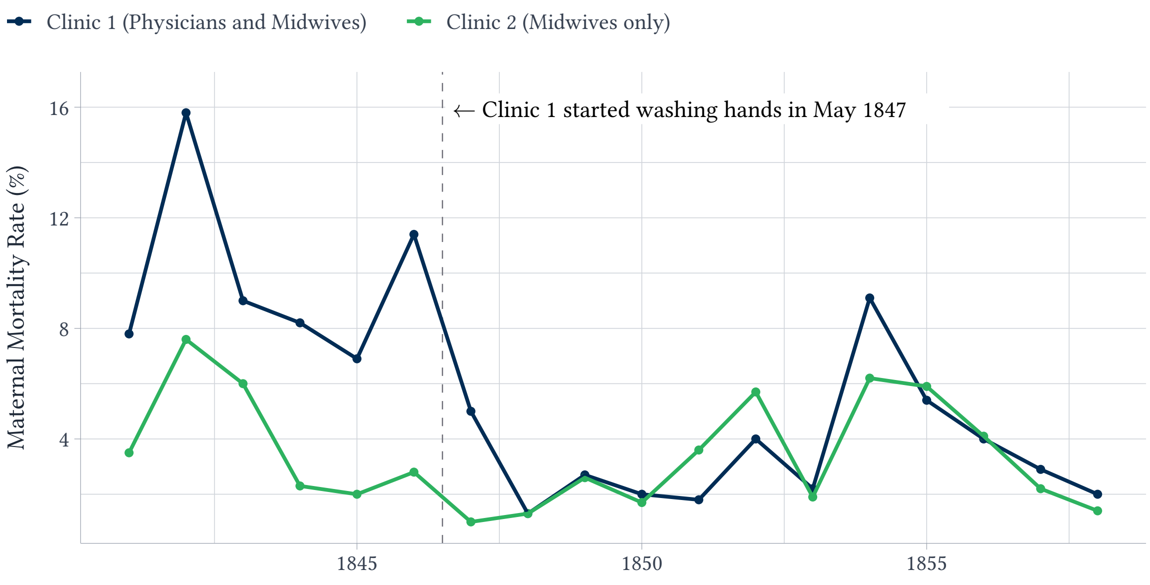
Ignaz Semmelweis and washing hands

Ignaz Semmelweis, after a lot of observation, conjectures that the cause is:

- the teaching faculty would teach anatomy using cadavers and then delivering babies without washing hands

Convinced the hospital to have physicians wash their hands in chlorine but not the midwives

- Compared mortality rates in treated Clinic 1 (Physicians and Midwives) vs. untreated Clinic 2 (Midwives only)



Clinic 1 (Physicians and Midwives)
Clinic 2 (Midwives only)
← Clinic 1 started washing hands in May 1847
Maternal Mortality Rate (%)
16
12
8
4
1845
1850
1855

Identifying assumptions

While this, today, seems like an obvious treatment effect, people at the time did not believe this result

- In fact, Semmelweis was fired about a year and a half later and his life was ruined by critics

It is worth asking for this topic, "What do we need to assume to believe this result?"

Identifying assumptions

Looking at the previous figure, we see that prior to treatment, mortality rates were way higher in the physicians clinic than midwives. Then, right when treatment starts we see a large drop in the mortality rate

The main issue is that we can not be sure what would happen had the physician clinic not been required to wash their hands

- Do not observe the post-treatment $y(0)$

Identifying assumptions

Looking at the previous figure, we see that prior to treatment, mortality rates were way higher in the physicians clinic than midwives. Then, right when treatment starts we see a large drop in the mortality rate

The main issue is that we can not be sure what would happen had the physician clinic not been required to wash their hands

- Do not observe the post-treatment $y(0)$

We, however, do not see a similar drop in the second clinic, so this rules out many shocks that would impact both clinics

Identifying assumptions

What we will come to formalize is the parallel counterfactual trends assumption:

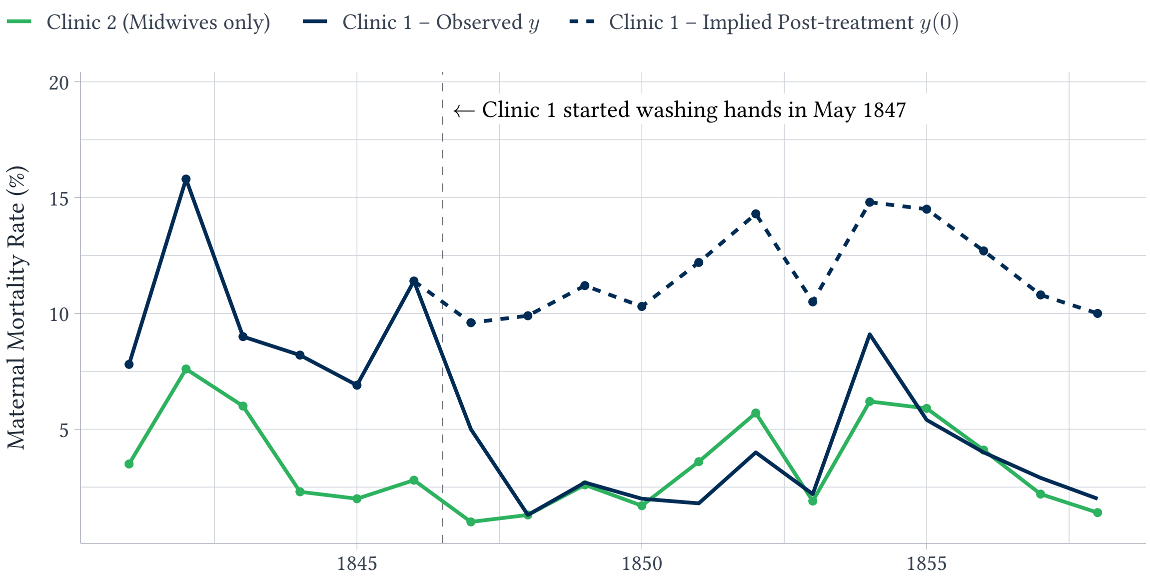
- In the absence of treatment, the treated units would be on the same counterfactual trend as we observe in the untreated units

You can imagine taking the trend from Clinic 2 and appending that onto the start of the post-period for Clinic 1

- The implied $Y(0)$ if indeed the two clinics would have the same counterfactual trends

People typically call it the *parallel trends* assumption

- But I prefer the full phrase because it emphasizes this is about trends for the treated units *had they not been treated*



Clinic 2 (Midwives only)
Clinic 1 – Observed y
Clinic 1 – Implied Post-treatment $y(0)$
← Clinic 1 started washing hands in May 1847
Maternal Mortality Rate (%)
20
15
10
5
1845
1850
1855

Difference-in-Differences

- Initial Difference-in-difference usage
- Classic Example: Card and Krueger (2000, AER)
- Econometric formulation to DID

Card and Krueger (1994, AER)

The first “modern” economics paper to use difference-in-differences

Card and Krueger studied the 1992 minimum wage increase in New Jersey from \$4.25 to \$5.05

- The story goes that they heard about the minimum wage change and *ran to the field* to start collecting data on fast-food employment prior to the minimum wage
    - $\rightarrow$

Card and Krueger (1994, AER)

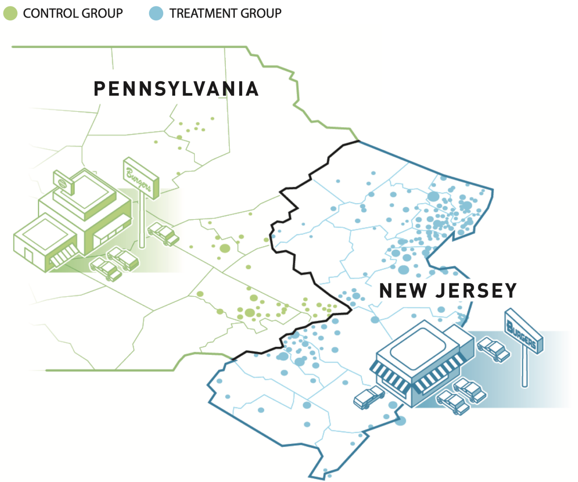
The first "modern" economics paper to use difference-in-differences

Card and Krueger studied the 1992 minimum wage increase in New Jersey from \$4.25 to \$5.05

- The story goes that they heard about the minimum wage change and *ran to the field* to start collecting data on fast-food employment prior to the minimum wage
  - $\rightarrow$

Their strategy was to compare changes to New Jersey fast-food employment to those in Eastern Pennsylavnia

- 331 in New Jersey (treated)
- 79 in Eastern Pennsylvania (untreated)

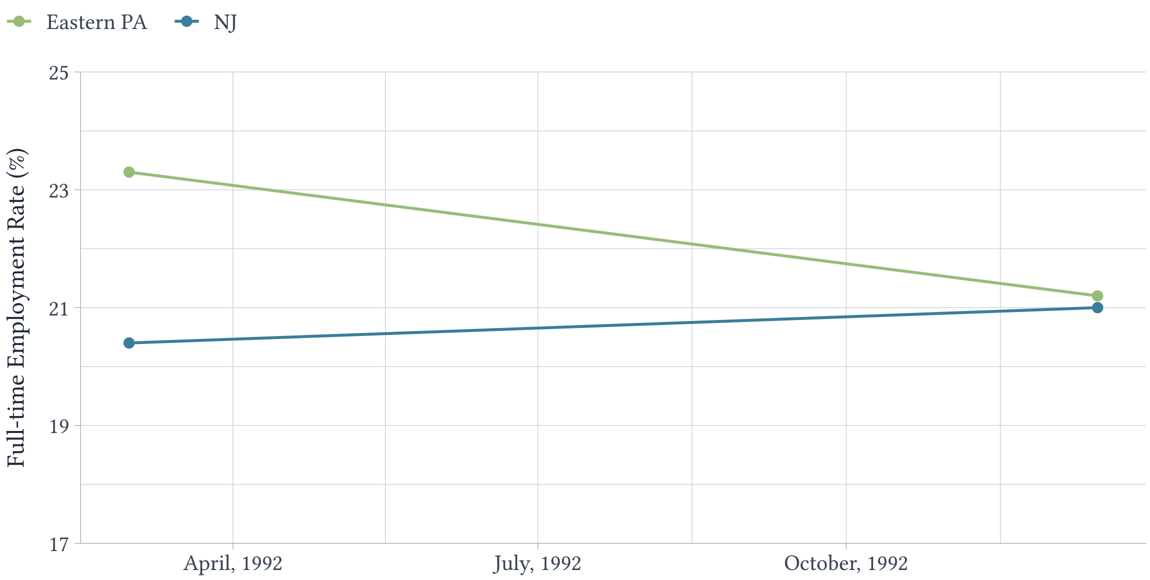


Source: Nobel Prize summary

Measurements

They measured employment before (in March 1992) and after (in December 1992) the minimum wage passed

- This is a relatively small survey, but it was novel because no one really tried to see what the actual impacts of minimum wage changes was


Eastern PA
NJ
Full-time Employment Rate (%)
25
23
21
19
17
April, 1992
July, 1992
October, 1992

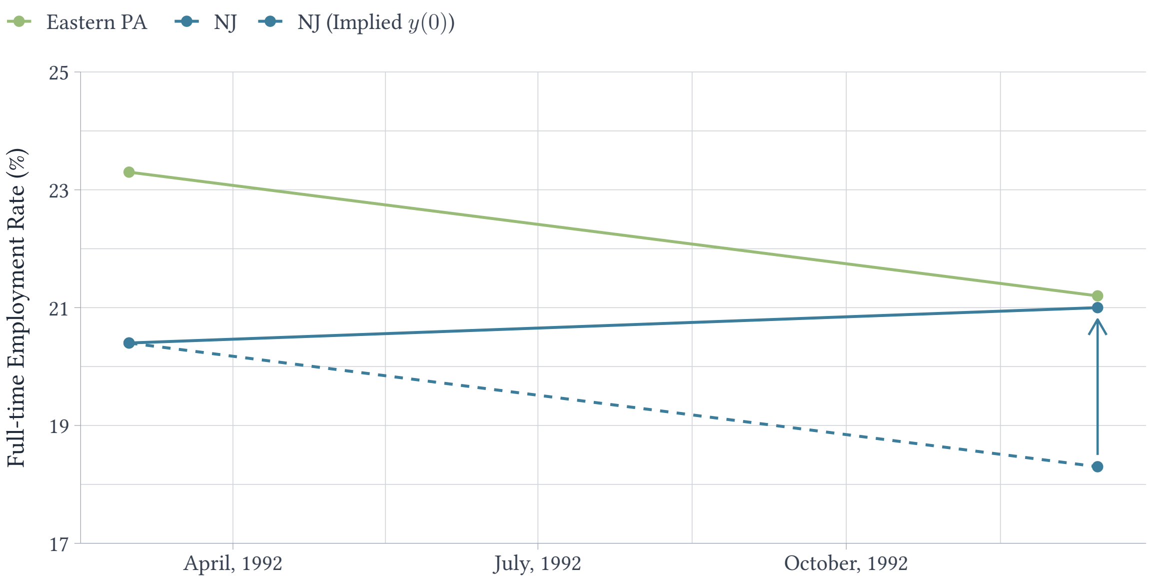
Identification

So, we see that NJ employment went up slightly and Eastern PA employment went down a bit more

- We infer that NJ would have went down by the same amount as Eastern PA had the minimum wage not passed

That is, we assume that there are "common shocks" to both areas and assume that there are no additional shocks that impact *only one* of the two regions

- They are on "parallel counterfactual trends"



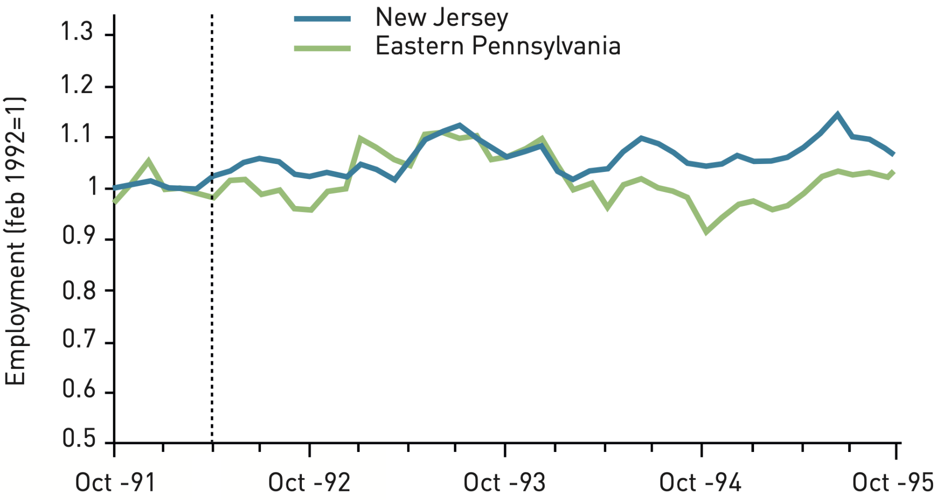
Eastern PA
NJ
NJ (Implied $y(0)$)
25
23
21
19
17
Full-time Employment Rate (%)
April, 1992
July, 1992
October, 1992

Is this believable?

From this graph, it’s not clear what to think about this assumption of parallel counterfactual trends

For this reason, it is (now) typical to compare the treated and control units *prior* to treatment uptake to see if they are on similar trends

- Using many observations before and after treatment are called ‘event-study’ estimates



Source: Nobel Prize summary

Pre-trends

The previous figure shows that for a few months prior to the minimum wage change, the employment trends of Eastern PA and New Jersey followed closely to one another

- This supports the idea that in the absence of treatment the NJ and Eastern PA trends would be similar in the post-period

Pre-trends

The previous figure shows that for a few months prior to the minimum wage change, the employment trends of Eastern PA and New Jersey followed closely to one another

- This supports the idea that in the absence of treatment the NJ and Eastern PA trends would be similar in the post-period

To be clear, parallel counterfactual trends involves the *post-treatment* y_{it}

- Having similar trends prior to treatment helps support this assumption, but does not *prove it*

Ashenfelter’s dip

Orley Ashenfelter’s 1978 paper entitled “Estimating the Effect of Training Programs on Earnings” is a great example to illustrate the different between common trends before treatment and the *parallel counterfactual trends* assumption

Ashenfelter’s dip

Orley Ashenfelter’s 1978 paper entitled “Estimating the Effect of Training Programs on Earnings” is a great example to illustrate the different between common trends before treatment and the *parallel counterfactual trends* assumption

He looks at individuals that sign-up for a work training program on their future earnings

- For many years prior to treatment, the workers that do and do not enter the training have common earnings trends
- Just prior to treatment, the workers that do enter the program face a sudden *dip* in earnings
- Then, after the program, the workers’ earnings go back up towards the original level

Ashenfelter's dip

What was happening was that workers just prior to treatment lost their job (hence trying to learn new labor force skills)

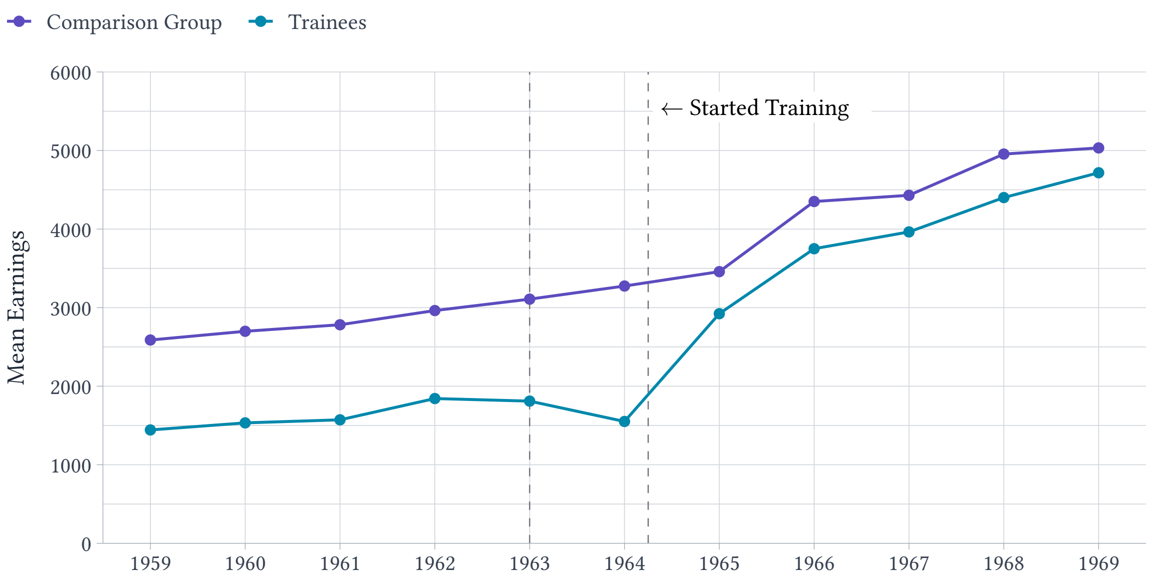
Ashenfelter’s dip

What was happening was that workers just prior to treatment lost their job (hence trying to learn new labor force skills)

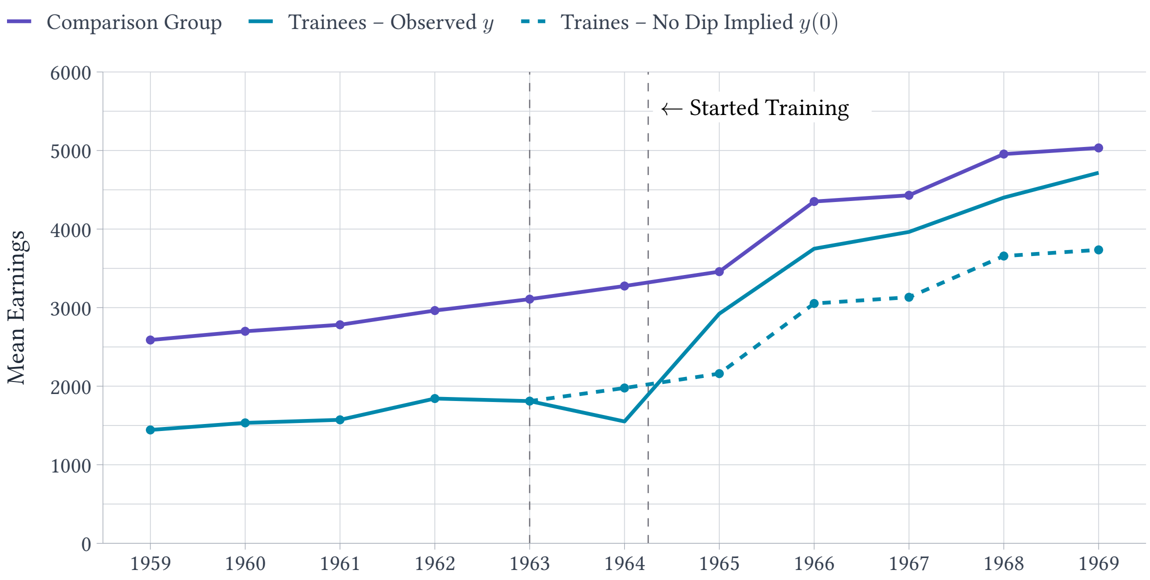
In the absence of the training, we would expect those workers to have a raise in earnings anyways because they would likely be hired somewhere

- The treated workers and the untreated workers have different earning dynamics

So even though they have similar trends prior to treatment, the parallel counterfactual trends assumption does not hold in this setting



Comparison Group
Trainees
Mean Earnings
← Started Training
6000
5000
4000
3000
2000
1000
0
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969


Comparison Group
Trainees – Observed y
Traines – No Dip Implied $y(0)$
← Started Training
Mean Earnings
0
1000
2000
3000
4000
5000
6000
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969

DID Key Ideas

Difference-in-differences compares a group assigned to treatment versus a group not assigned to treatment

- The estimator compares the treated groups change in outcomes before and after the treatment to the control groups change in outcomes before and after the treatment

The key assumption we make is the parallel counterfactual trends assumption

- The change in outcomes over time for control units are an appropriate stand-in for the treated unit's change in outcomes *if they did not receive treatment*

Difference-in-Differences

- Initial Difference-in-difference usage
- Classic Example: Card and Krueger (2000, AER)
- Econometric formulation to DID

2×2 Difference-in-Differences

The Card and Krueger minimum wage paper is an example of the canonical 2×2 DID, so we will begin there

We observe units $i \in \{1, \ldots, N\}$ for two periods (before and after), $t = 0$ and $t = 1$

- Let D_i be an indicator for which units receive treatment
- Let $\text{Post}_t = \mathbb{1}[t = 1]$ be an indicator for being in the post-period

2×2 Difference-in-Differences

The Card and Krueger minimum wage paper is an example of the canonical 2×2 DID, so we will begin there

We observe units $i \in \{1, \dots, N\}$ for two periods (before and after), $t = 0$ and $t = 1$

- Let D_i be an indicator for which units receive treatment
- Let $\text{Post}_t = \mathbb{1}[t = 1]$ be an indicator for being in the post-period

Then, we have potential outcomes for each unit in the post-period:

- $y_{i0}(D_i)$ and $y_{i1}(D_i)$
    - $\rightarrow$ We typically assume that treatment does not impact y_{i0}, i.e. $y_{i0} = y_{i0}(1) = y_{i0}(0)$. This is called the “no anticipation” assumption

Treatment effect of interest

The treatment effect of interest is the average effect of treatment in period 1 for the treated units:

$$\text{ATT}_1 = \mathbb{E}[y_{i1}(1) - y_{i1}(0) \mid D_i = 1]$$

The counterfactual compares the period 1 outcome under treatment to the period 1 outcome in the absence of treatment

- This is ***not*** the post-y minus pre-y!

Parallel Counterfactual Trends assumption

Our Parallel Counterfactual Trends imposes restrictions on the change in untreated potential outcomes:

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

This says, in the absence of treatment, the change in y is on average the same for the treated and the control group

Observed difference in y

For the treated unit, we can do an econometrician's favorite math trick (add and subtract something) to analyze the observed change in y for the treated units:

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

Observed difference in y

For the treated unit, we can do an econometrician's favorite math trick (add and subtract something) to analyze the observed change in y for the treated units:

$$\begin{aligned}\mathbb{E}[y_{i1} - y_{i0} \mid D_i = 1] &= \mathbb{E}[y_{i1}(1) - y_{i0}(0) \mid D_i = 1] \\ &= \mathbb{E}\Big[y_{i1}(1) \boxed{-y_{i1}(0) + y_{i1}(0)} - y_{i0}(0) \mid D_i = 1\Big]\end{aligned}$$

Observed difference in y

For the treated unit, we can do an econometrician's favorite math trick (add and subtract something) to analyze the observed change in y for the treated units:

$$
\begin{aligned}
\mathbb{E}[y_{i1} - y_{i0} \mid D_i = 1] &= \mathbb{E}[y_{i1}(1) - y_{i0}(0) \mid D_i = 1] \\
&= \mathbb{E}\Big[y_{i1}(1) \boxed{-y_{i1}(0) + y_{i1}(0)} - y_{i0}(0) \mid D_i = 1\Big] \\
&= \mathbb{E}[y_{i1}(1) - y_{i1}(0) \mid D_i = 1] + \mathbb{E}[y_{i1}(0) - y_{i0}(0) \mid D_i = 1]
\end{aligned}
$$

Observed difference in y

For the treated unit, we can do an econometrician's favorite math trick (add and subtract something) to analyze the observed change in y for the treated units:

$$
\begin{aligned}
\mathbb{E}[y_{i1} - y_{i0} \mid D_i = 1] &= \mathbb{E}[y_{i1}(1) - y_{i0}(0) \mid D_i = 1] \\
&= \mathbb{E}\Big[y_{i1}(1) \boxed{-y_{i1}(0) + y_{i1}(0)} - y_{i0}(0) \mid D_i = 1\Big] \\
&= \mathbb{E}[y_{i1}(1) - y_{i1}(0) \mid D_i = 1] + \mathbb{E}[y_{i1}(0) - y_{i0}(0) \mid D_i = 1] \\
&= \text{ATT}_1 + \mathbb{E}[y_{i1}(0) - y_{i0}(0) \mid D_i = 1]
\end{aligned}
$$

$\implies$ The change in outcome for the treated units is the effect of treatment plus the treated groups' counterfactual trend

Observed difference in y

$$\mathbb{E}[y_{i1} - y_{i0} \mid D_i = 1] = \textsf{ATT}_1 + \mathbb{E}[y_{i1}(0) - y_{i0}(0) \mid D_i = 1]$$

For control units, the math is simpler

$$\mathbb{E}[y_{i1} - y_{i0} \mid D_i = 0] = \mathbb{E}[y_{i1}(0) - y_{i0}(0) \mid D_i = 0]$$

$\implies$ The change in outcome for the controlunits is the effect of treatment plus the control groups' counterfactual trend

Difference-in-differences

Now, the difference-in-differences estimand is formed by subtracting the two change in outcomes:

$$\tau_{\texttt{DID}} = \mathbb{E}[y_{i1} - y_{i0} \mid D_i = 1] - \mathbb{E}[y_{i1} - y_{i0} \mid D_i = 0]$$

Difference-in-differences

Now, the difference-in-differences estimand is formed by subtracting the two change in outcomes:

$$\begin{aligned}\tau_{\texttt{DID}} &= \mathbb{E}[y_{i1} - y_{i0} \mid D_i = 1] - \mathbb{E}[y_{i1} - y_{i0} \mid D_i = 0] \\ &= \text{ATT}_1 + \mathbb{E}[y_{i1}(0) - y_{i0}(0) \mid D_i = 1] - \mathbb{E}[y_{i1}(0) - y_{i0}(0) \mid D_i = 0]\end{aligned}$$

The difference-in-differences estimand compares treated unit's change in y to control unit's change in y

- This estimates the effect of treatment plus the difference in trends between the two groups

Difference-in-differences

$$\begin{aligned}\tau_{\texttt{DID}} &= \mathbb{E}[y_{i1} - y_{i0} \mid D_i = 1] - \mathbb{E}[y_{i1} - y_{i0} \mid D_i = 0] \\ &= \mathsf{ATT}_1 + \mathbb{E}[y_{i1}(0) - y_{i0}(0) \mid D_i = 1] - \mathbb{E}[y_{i1}(0) - y_{i0}(0) \mid D_i = 0]\end{aligned}$$

For example, if the treated group had a larger counterfactual growth in y (like in Ashenfelter's dip example), then the treatment effect will be biased upwards

Difference-in-differences

However, assuming parallel counterfactual trends implies that these two counterfactual trend terms are the same and therefore cancel out

$$
\begin{aligned}
\tau_{\texttt{DID}} &= \mathbb{E}[y_{i1} - y_{i0} \mid D_i = 1] - \mathbb{E}[y_{i1} - y_{i0} \mid D_i = 0] \\
&= \textsf{ATT}_1 + \mathbb{E}[y_{i1}(0) - y_{i0}(0) \mid D_i = 1] - \mathbb{E}[y_{i1}(0) - y_{i0}(0) \mid D_i = 0]
\end{aligned}
$$

Difference-in-differences

However, assuming parallel counterfactual trends implies that these two counterfactual trend terms are the same and therefore cancel out

$$
\begin{aligned}
\tau_{\texttt{DID}} &= \mathbb{E}[y_{i1} - y_{i0} \mid D_i = 1] - \mathbb{E}[y_{i1} - y_{i0} \mid D_i = 0] \\
&= \textsf{ATT}_1 + \mathbb{E}[y_{i1}(0) - y_{i0}(0) \mid D_i = 1] - \mathbb{E}[y_{i1}(0) - y_{i0}(0) \mid D_i = 0] \\
&= \textsf{ATT}_1 + \mathbb{E}[y_{i1}(0) - y_{i0}(0) \mid D_i = 1] - \mathbb{E}[y_{i1}(0) - y_{i0}(0) \mid D_i = 1] \\
&= \textsf{ATT}_1
\end{aligned}
$$

Difference-in-differences as an imputation estimator

Remember in the selection on observables topic, we used a regression imputation estimator to explicitly estimate the treated units' $y_i(0)$.

It turns out, we can write the difference-in-differences estimator as an imputation estimator

Difference-in-differences as an imputation estimator

Our imputation for $y_{i1}(0)$ is given as:

$$\hat{y}_{i1}(0) = y_{i0} + \mathbb{E}[y_{i1}(0) - y_{i0}(0) \mid D_i = 0]$$

In words, take the unit's period $t = 0$ outcome and add to it the average change in y for the comparison group.

- This is what I was drawing in the figures at the start of the slides

2×2 DID Estimation

Our estimation strategy replaces these terms with their sample averages:

$$\hat{\tau}_{\texttt{DID}} = \hat{\mathbb{E}}[y_{i1} - y_{i0} \mid D_i = 1] - \hat{\mathbb{E}}[y_{i1} - y_{i0} \mid D_i = 0]$$

We could do this as four averages

$$\left(\hat{\mathbb{E}}[y_{i1} \mid D_i = 1] - \hat{\mathbb{E}}[y_{i0} \mid D_i = 1] \right) - \left(\hat{\mathbb{E}}[y_{i1} \mid D_i = 0] - \hat{\mathbb{E}}[y_{i0} \mid D_i = 0] \right)$$

2×2 DID Estimation

Our estimation strategy replaces these terms with their sample averages:

$$\hat{\tau}_{\texttt{DID}} = \hat{\mathbb{E}}[y_{i1} - y_{i0} \mid D_i = 1] - \hat{\mathbb{E}}[y_{i1} - y_{i0} \mid D_i = 0]$$

We could do this as four averages

$$\left(\hat{\mathbb{E}}[y_{i1} \mid D_i = 1] - \hat{\mathbb{E}}[y_{i0} \mid D_i = 1]\right) - \left(\hat{\mathbb{E}}[y_{i1} \mid D_i = 0] - \hat{\mathbb{E}}[y_{i0} \mid D_i = 0]\right)$$

Or just do a difference-in-means using $y_{i1} - y_{i0}$ as the outcome variable

- Be careful to only have one row per unit when running this regression

A note on the name ‘Difference-in-Differences’

The correct name is Difference in Difference*s*

- You are taking the difference between *two* avrages of first-differences

Personal pet-peeve, but this is the one and only name for this estimator

2×2 in regression form

Just like difference-in-means, it turns out you can use OLS regression to estimate $\hat{\tau}_{\text{DID}}$

$$y_{it} = \alpha + \gamma D_i + \lambda \text{Post}_t + \tau d_{it} + u_{it}$$

- $d_{it} = D_i \text{Post}_t$ is an indicator for when a unit is actively under treatment

2×2 in regression form

$$y_{it} = \alpha + \gamma D_i + \lambda \text{Post}_t + \tau d_{it} + u_{it}$$

Since these are just a bunch of indicator variables, we can derive what they estimate:

$$\text{E}\left[Y_{it} \mid D_i = 0, \text{Post}_t = 1\right] = \mathbb{E}[Y_{i1} \mid D_i = 0] = \alpha + \lambda$$
$$\text{E}\left[Y_{it} \mid D_i = 1, \text{Post}_t = 0\right] = \mathbb{E}[Y_{i0} \mid D_i = 1] = \alpha + \gamma$$
$$\text{E}\left[Y_{it} \mid D_i = 1, \text{Post}_t = 1\right] = \mathbb{E}[Y_{i1} \mid D_i = 1] = \alpha + \gamma + \lambda + \tau$$

2×2 in regression form

$$y_{it} = \alpha + \gamma D_i + \lambda \text{Post}_t + \tau d_{it} + u_{it}$$

First, we have $\hat{\alpha} = \mathbb{E}[y_{i0} \mid D_i = 0]$ and $\hat{\gamma} = \mathbb{E}[y_{i0} \mid D_i = 1]$

Second, we have $\hat{\lambda} = \mathbb{E}[y_{i1} - y_{i0} \mid D_i = 0]$

Last, we have

$$\hat{\tau}_{\text{OLS}} = \mathbb{E}[y_{i1} - y_{i0} \mid D_i = 1] - \mathbb{E}[y_{i1} - y_{i0} \mid D_i = 0]$$

2×2 in regression form

You can also use unit and time fixed-effects to estimate this

$$y_{it} = \mu_i + \lambda_t + \tau d_{it} + u_{it}$$

It is also true that, $\hat{\tau}_{\text{OLS}} = \hat{\tau}$DID!

- Note that one of the time fixed-effects will need to be omitted for collinearity

Users Beware !!

The equivalence between OLS and 2×2 DID only holds in this case. Using OLS in other cases will turn out to bite us in the butt later on

- People have been using OLS for DID for a long time and it turns out to create problems when treatment starts at different points in time for different units