GOV1368 Section 4

Mohit Karnani

Harvard

Fall 2024

Recap

Last time we covered:

- Estimands and Estimators
- Bias and Consistency
- Potential Outcomes and Counterfactuals
- Randomized Controlled Trials
- Imperfect Compliance

We know that RCTs are the gold standard, but they are often not feasible to conduct. Today we'll learn about a technique that's very simple to apply using data "in the wild": **Difference-in-Differences**.

Agenda

Average Treatment Effect on the Treated

Difference-in-Differences

Parallel Trends and Identification

Application

Contents

Difference in Means and Selection Bias (copy-pasted from last week)

What if we take the average difference in test scores between students who receive the tutoring program and students who do not?

(No need to follow the math)

$$
\begin{aligned}
\sum_{D_i=1} Y_i/n_T - \sum_{D_i=0} Y_i/n_C \to & \mathbb{E}[Y_i|D_i = 1] - \mathbb{E}[Y_i|D_i = 0] \\
= & \mathbb{E}[Y_i(1)|D_i = 1] - \mathbb{E}[Y_i(0)|D_i = 0] \\
= & \mathbb{E}[Y_i(1)|D_i = 1] - \mathbb{E}[Y_i(0)|D_i = 0] \pm \mathbb{E}[Y_i(0)|D_i = 1] \\
= & \underbrace{\mathbb{E}[Y_i(1) - Y_i(0)|D_i = 1]}_{ATET} + \underbrace{\mathbb{E}[Y_i(0)|D_i = 1] - \mathbb{E}[Y_i(0)|D_i = 0]}_{\text{Selection Bias}}
\end{aligned}
$$

We end up with *selection bias*: students that received tutoring might be fundamentally different from those who didn't (they can have different average potential outcomes).

What is the ATET/ATT?

The *Average Treatment Effect on the Treated* (ATET or ATT) is an *estimand* defined as

$$ATT := \mathbb{E}[Y_i(1) - Y_i(0)|D_i = 1]$$

(It is slightly different from the ATE because it is conditioned on the treatment group $D_i = 1$. In an RCT this doesn't matter, because the assignment of D_i is independent of *everything*, including the potential outcomes. Thus, in a "perfect" RCT, $ATT = ATE$.)

How can we estimate the ATT in absence of an RCT? Which estimator could we use? We know a difference in means is not enough because we end with selection bias... how about another difference?

Contents

Difference-in-Differences

One way of eliminating selection bias is using a *Difference-in-Differences* (DiD) estimator, equipped with the *parallel trends* assumption.

First, we need to extend our notation to include *time* as an index. We will do this by using Pre for the pre-treatment period and $Post$ for the post-treatment period. Now we have 4 potential outcomes:

$$Y_i^{Pre}(0) \quad Y_i^{Pre}(1) \quad Y_i^{Post}(0) \quad Y_i^{Post}(1)$$

Of course, we won't observe all these potential outcomes. Just as before, in the $Post$ period we observe

$$Y_i^{Post} = D_i Y_i^{Post}(1) + (1 - D_i) Y_i^{Post}(0),$$

while in the Pre period we always observe $Y_i^{Pre} = Y_i^{Pre}(0)$, because no one is treated yet.

Question: What's the point of defining $Y_i^{Pre}(1)$?

Example: Real World Dataset

	id	score1	score0	treatment...
1	176976	289.35	247.77	1
2	213797	254.29	199.19	1
3	62302	302.46	248.05	1
4	65348	259.53	270.33	1
5	223054	243.82	269.6	1
6	78639	138.01	219.45	1
7	203358	246.87	221.3	1
8	24717	229.24	197.43	0
9	46077	248.2	271.57	0
10	98674	208.48	207.47	1
11	108932	260.29	185.73	1
12	47310	214.19	188.52	0
13	100701	215.73	298.19	1
14	113026	341.71	300.57	1
15	31906	234.7	208.38	0
16	219528	212.08	232.86	1
17	140792	217.23	322.35	1
18	18474	212.6	259.65	0

Difference-in-Differences Estimator

Now that we have outcomes for Treatment/Control units and Pre/Post periods, we can define our DiD estimator as

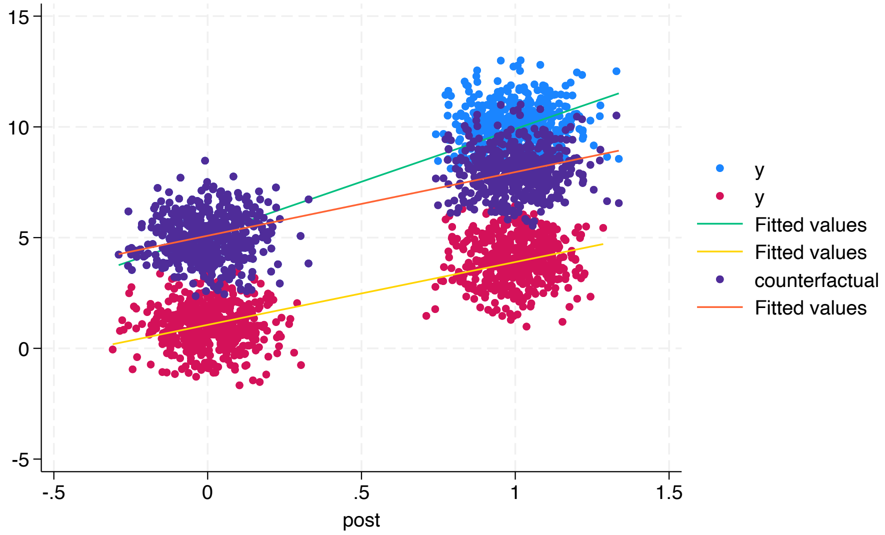
$$(\bar{Y}_T^{Post} - \bar{Y}_T^{Pre}) - (\bar{Y}_C^{Post} - \bar{Y}_C^{Pre}) := \left(\sum_{i \in T} \frac{Y_i^{Post}}{n_T} - \sum_{i \in T} \frac{Y_i^{Pre}}{n_T} \right) - \left(\sum_{i \in C} \frac{Y_i^{Post}}{n_C} - \sum_{i \in C} \frac{Y_i^{Pre}}{n_C} \right).$$

As you can see, it is indeed a *difference in differences*!

Instead of having the simple difference-in-means estimator we introduced last week, now we also subtract another difference in means: one for the treatment group -before and after the treatment takes place- and another for the control group -also comparing the average outcomes before and after the treatment.

We will see that this estimator works (i.e. is an unbiased and consistent estimator of the *ATT*) if the *parallel trends* assumption holds.

Figure: Difference-in-Differences



Contents

Parallel Trends Assumption

Our *identification assumption* for the DiD estimator is the *parallel trends* assumption:

$$\mathbb{E}[Y_i^{Post}(0) - Y_i^{Pre}(0)|D_i = 1] = \mathbb{E}[Y_i^{Post}(0) - Y_i^{Pre}(0)|D_i = 0]$$

In words, parallel trends (PT) means that the potential outcome without treatment $Y_i(0)$ would be expected to evolve in the same way for both the treatment and control group.

That is, we expect the same *trend* in outcomes for both groups in an alternative world where the treatment never happened.

It is an *assumption*, because in reality the treatment did occur, and we are assuming something about *what would have happened* in a *counterfactual* scenario.

DiD+PT Estimates the ATT

If we assume PT, we can eliminate selection bias with a DiD estimator:

(No need to follow the math)

$$
\begin{aligned}
DiD \rightarrow & (\mathbb{E}[Y_i^{Post}|D_i = 1] - \mathbb{E}[Y_i^{Pre}|D_i = 1]) - (\mathbb{E}[Y_i^{Post}|D_i = 0] - \mathbb{E}[Y_i^{Pre}|D_i = 0]) && \text{(LLN)} \\
= & (\mathbb{E}[Y_i^{Post}(1)|D_i = 1] - \mathbb{E}[Y_i^{Pre}(0)|D_i = 1]) - (\mathbb{E}[Y_i^{Post}(0)|D_i = 0] - \mathbb{E}[Y_i^{Pre}(0)|D_i = 0]) && \text{(PO)} \\
= & (\mathbb{E}[Y_i^{Post}(1)|D_i = 1] - \cancel{\mathbb{E}[Y_i^{Pre}(0)|D_i = 1]}) - (\mathbb{E}[Y_i^{Post}(0)|D_i = 1] - \cancel{\mathbb{E}[Y_i^{Pre}(0)|D_i = 1]}) && \text{(PT)} \\
= & \mathbb{E}[Y_i^{Post}(1) - Y_i^{Post}(0)|D_i = 1] && \text{(ATT!)}
\end{aligned}
$$

This is cool: we can estimate causal effects without an RCT! Moreover, it is super simple: just compute 4 means, and subtract them the right way. But there's no free lunch: we have to assume parallel trends for this to work.

Contents

Application: The Long-Run Effects of Teacher Collective Bargaining

`Stata`