Statistics II

Week 7:

Panel Data, Diff-in-Diff, and Fixed Effects

Lecture Review

Panel Data

Time series data allow us to observe the same subject or unit in different causal states at different points in time. Data with multiple units over time are called panel data.

True panels, a.k.a. longitudinal data, record the same individuals over time in waves.

Independently pooled cross-sections are cross-sections repeated at different time periods. The observations (often respondents) are not identical over time.

Estimating Effects with Panel Data

- Assume we have a data set with random assignment into treatment and control, with **one outcome measurement before treatment** and **one outcome measurement after treatment.**
- As usual, we have a problem of not knowing **counterfactuals**: Ex. the average post-period outcome for the treated group in the absence of the treatment.
- To get around this we could:
  - Compare **before and after treatment for the treatment group**
    - This assumes no change in average potential outcome over time.
  - Compare the **treatment and control groups after treatment**
    - This assumes the PO of control group is the same as the counterfactual PO for those being treated.

Difference-in-Differences

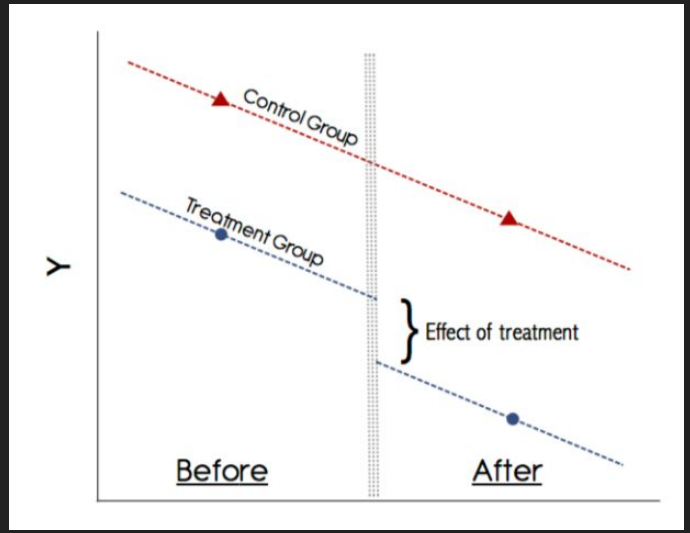
Or, compare the:

1. **Difference between the treatment and control group after treatment**

And subtract the

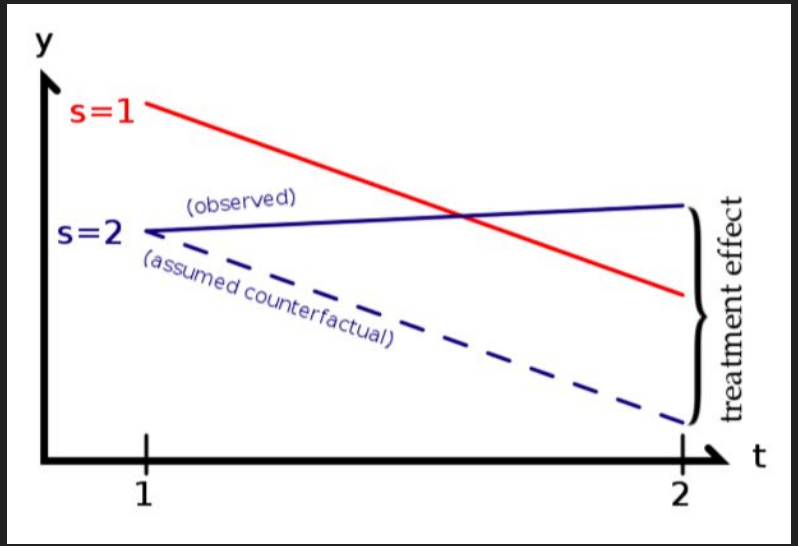
2. **Difference between the treatment and control group before treatment**

This approach uses overtime difference in control group as a counterfactual of what would have happened in the treatment group without the intervention.



Parallel Trends Assumption

Difference-in-differences estimation critically rests on the assumption that **observed overtime changes in the control group reflect, on average, unobserved changes in the treatment group** in the absence of treatment.



Parallel Trends Violations

- **Selection** and **Targeting**
  - Units may self-sort for reasons that are not random
  - Policies may be targeted at units in a non-random way
- **Compositional differences** across time
  - The composition of a sample might change in ways that confound the treatment effect.
- **Long-term effects** vs. reliability
  - Parallel trends is more likely to hold in the short term.
- **Functional form** dependence
  - DD is more reliable if the treatment and control groups are more similar at baseline.

Parallel Trends Diagnostics

1. **Pre-treatment trends** in the outcome
2. Placebo test using **previous periods**
3. Placebo test using **alternative outcomes**
4. Placebo **outcomes**

Fixed Effects

Panel estimators of causal effects exploit data with an additional time dimension to account for covariates (observed or unobserved) which are fixed:

- **Fixed within units:** Time-invariant traits of the units having an effect on the outcome. Something that affects the unit permanently (like racial identity).
- **Fixed within time**: Time-specific effects that affect all the units simultaneously (like changes in income due to national changes in the economy over time).

There are also **idiosyncratic factors** specific to a unit in a particular time. These are potential confounders.

Fixed Effects Estimation

Assume we have a common panel data setup that accounts for unit and time fixed effects:

- If we take the mean of the equation over time, the mean of the unit fixed effects is still just the original value, because it is **time-invariant.**
- If we subtract the averaged equation from the original equation, we have the **time-demeaned** equation where unit fixed effects drop out.
- If we estimate our model using this time-demeaned equation, we are left with the FE estimator (“**within estimator**”).
- Basically, we are left with a model where all the confounders that don’t vary over time just drop out.

Questions?