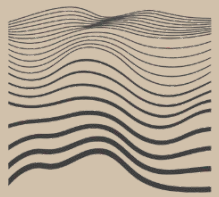
Difference in Differences



WORK IN PROGRESS

notebook link:

GitHub Pages

Tools: R in Google Colab
Data Source: TBD

Table of Contents

Intro

Difference-in-Differences

Difference in Differences

- Estimates **causal effect** of a treatment/ intervention
- Compares changes between a treatment vs. control group **over time:**
  - **"How much more did the treatment group change vs. the control group, after controlling for common changes?"**

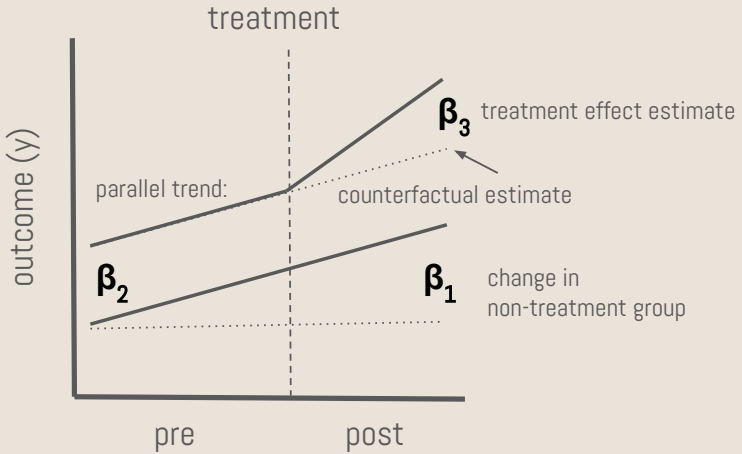
Estimated treatment effect =

(over time, before & after treatment)

Differences in treatment group - Differences in control group

Difference in Differences

$$Y = \text{intercept} + \beta_1 \text{Post} + \beta_2 \text{Treated} + \beta_3 (\text{Post} \times \text{Treated}) + \epsilon$$



Coefficients:

- **β_1: time effect:** time-related changes regardless of treatment
- **β_2: group effect:** pre-existing differences in treatment vs. control groups *(stable over time)*
- **β_3, δ: interaction term:** causal effect of the treatment *(needs significance)*

Regression Formulas

Rev in Location_1 Pre Price change = Intercept (Period, Location, Interaction all 0)

Rev in Location_2 Pre Price change = Intercept + Location (Period, Interaction all 0)

Rev in Location_1 Post Price change = Intercept + Period (Location, Interaction all 0)

Rev in Location_2 Post Price change = Intercept + Period + Location + Interaction

Diff in Diff = (Rev in Location_2 Post Price change - Rev in Location_2 Pre Price change) - (Rev in Location_1 Post Price change - Rev in Location_1 Pre Price change)

= (Intercept + Period + Location + Interaction - (Intercept + Location)) - (Intercept + Period - Intercept)

= Interaction (causal effect)

Assumptions

Assumptions

Parallel Trends

Output in control location with no treatment is similar and very correlated with output in treatment location:

Check that the slopes of the two groups are the same (with regression).

If single control group does not satisfy parallel trends, then use synthetic control (weighted combo of controls)

Model

Model

```
model <- lm(output~period+location+period:location,data=data)
```

Results

Model Output

Constant	____*** (positive)	
PeriodPost.Price.Change	____*** (positive)	revenue rose naturally over time
Location_2	____*** (positive)	revenue in location 2 was higher than in location 1
PeriodPost.Price.Change:Location_2	____*** (positive)	**interaction term coefficient:** estimated causal impact of the price change; only the interaction term needs to be significant

WORK IN PROGRESS

Appendix

Thank you!