Difference in Differences and Event Study

Long Hong

Nov. 12, 2021

DiD summary

Standard setting

- At time t = 0, an exogenous "policy" happens → "before and after"
- The policy affects certain groups but not the other groups → "treatment and control"

Example: Merit aid programs

- A merit aid programs in Georgia, Helping Outstanding Pupils Education, started in 1993
- Georgia is affected (treatment)
- 41 other states *never* had any merit aid program (control)

DiD in practice

Assumptions:

1. Pre-policy parallel trend
   - In the absence of the policy change, control and treatment would follow the same time trend
2. No other policies going on at the same time affecting the outcome
3. Exogeneity: people weren't able to manipulate the timing or effect of the policy
4. No selection
   - People switch from the treatment to the control group or vice verse in response to policy
5. No anticipation effect
   - People might change outcome in response to (future) policy before policy is implemented

Can test: pre-policy parallel trend is satisfied (science part)

Cannot test: the rest of the assumption

- Use your background knowledge from the institutional setting (art part)

Event-study specification

- Repeated cross-sectional data (Chris's notation *modified*)

$$y_i = \sum_{k \in \{-3,-4,\ldots,7\}} \beta_k T_s \mathbb{I}\{t = k\} + \delta_s + \mu_t + \epsilon_i$$

  - $T_s = 1$ if the state is treated (i.e., Georgia)
  - **Event-time indicator**: $\mathbb{I}\{t = k\}$ is a dummy: year t = k-th year relative to **start year**

$$\text{e.g.,} \quad \mathbb{I}\{t = -1\} = \begin{cases} 1 & \text{if the year is 1992 (1 year before program)} \\ 0 & \text{otherwise} \end{cases}$$

- Note: $T_s \mathbb{I}\{t = k\}$ is equivalent to Chris's $T_{g(i)t(i)}$ notation

- For panel data notation, simply change y_i to y_{it} and ϵ_i to ϵ_{it}

Event study: One tricky point

Collinearity issue: $\sum_{k\in\{-3,-4,\ldots,7\}} \mathbb{I}\{t=k\} = \mathbb{I} \quad \Rightarrow \quad$ 10 years with 10 dummies

- The same reason for any dummy variables:
- If you generate an indictor/dummy for race
  - generate white = (race == "white")
  - generate black = (race == "black")
  - generate other = (rate == "other")
- you can really put two in the regression due to collinearity.

Solution:

- Throw away one variable in the regression
- The literature typically normalize the year before the policy, i.e., $\beta_{-1} = 0$

$$y_i = \sum_{k\neq -1} \beta_k T_s \mathbb{I}\{t=k\} + \delta_s + \mu_t + \epsilon_i \qquad \text{(correct equation)}$$

Event study: extension

Example extended:

- There are 9 states also have a similar merit aid programs
- Treatment: 9 states + Georgia
- Control: rest of the 41 states that never had any merit aid program

Tricky, but the same logic follows

- Starting year is different for each treated state

Let's see how to do it in Stata

- Specifically, the effect of the merit aid programs on college attendance