Section 10

Panel Data

Sooahn Shin

GOV 2003

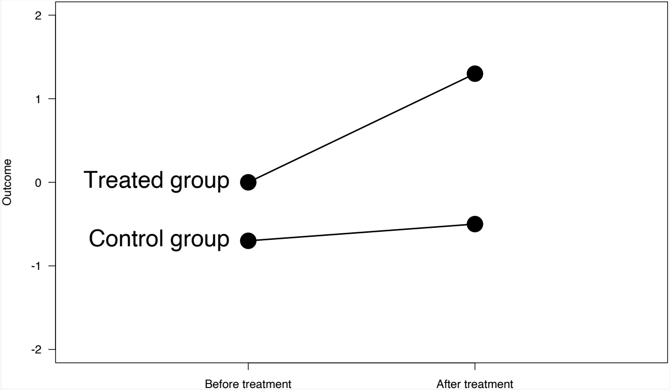
Nov 18, 2021

Overview

- Logistics:
  - **No problem set!**
  - **November 19th**: Submit a brief (no longer than 5 page) page memo of your main results, including tables, figures, and brief analysis. For methodological projects, this should include a description of the method and any analytical/simulation results. You will be required to give feedback on another group's project, which will be counted toward the overall grade based on attentiveness and usefulness of the feedback provided.
- Today's topics:
  - Difference-in-Differences design

Motivation

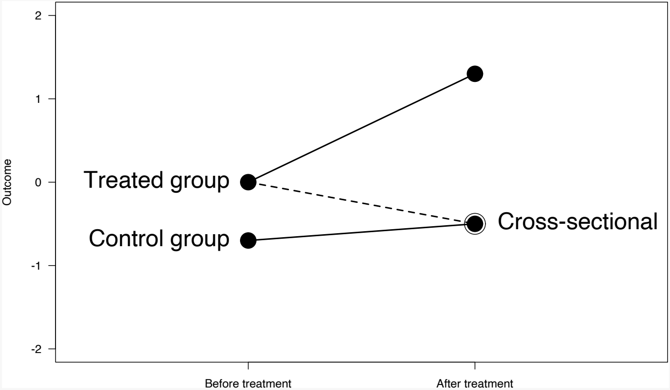
- What if we have repeated measurements of the same units before and after the treatment?



- Setup: two groups (binary G_i), two time periods (binary t)
  - $Y_{it}(d)$ is the potential outcome under treatment d at time t.
  - Estimand: $\tau_{\text{ATT}} = \mathbb{E}[Y_{i1}(1) - Y_{i1}(0) | G_i = 1]$

Identification problem

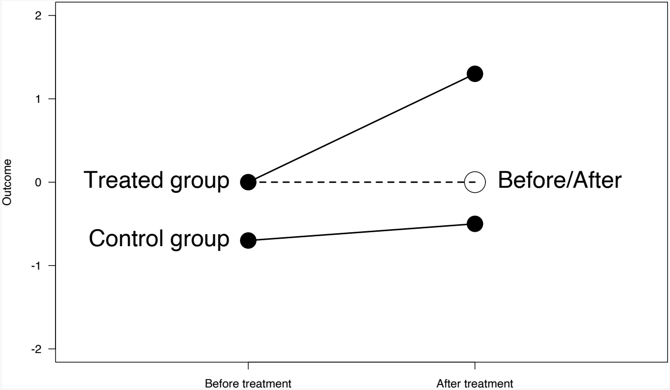
- Identifying counterfactual $\mathbb{E}[Y_{i1}(0) \mid G_i = 1]$



1. **Cross sectional variation**: At time $t = 1$ (post-period), some units received the treatment $(G_i = 1)$ while others didn't $(G_i = 0)$.

Identification problem

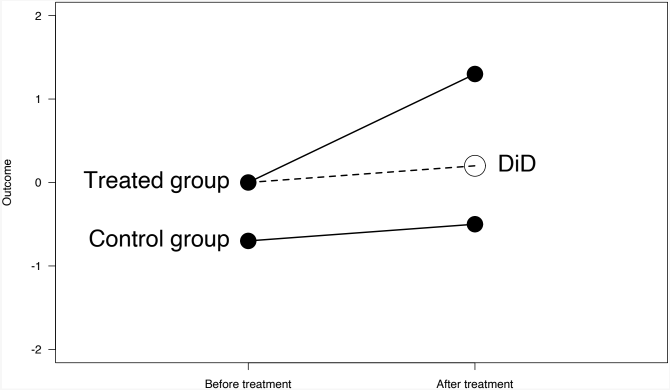
- Identifying counterfactual $\mathbb{E}[Y_{i1}(0) \mid G_i = 1]$



2. **Over time variation**: A unit (i) in the treated group didn't receive the treatment at time $t = 0$ (pre-period).

DiD Identification

- Identifying counterfactual $\mathbb{E}[Y_{i1}(0) \mid G_i = 1]$



- Key assumption: **parallel trends** (PT)
  $\mathbb{E}[Y_{i1}(0) - Y_{i0}(0)|G_i = 0] = \mathbb{E}[Y_{i1}(0) - Y_{i0}(0)|G_i = 1]$

  $\tau_{ATT} = (\mathbb{E}[Y_{i1}|G_i = 1] - \mathbb{E}[Y_{i0}|G_i = 1]) - (\mathbb{E}[Y_{i1}|G_i = 0] - \mathbb{E}[Y_{i0}|G_i = 0])$

Example

- Bechtel, Hangartner, and Schmid (2015)
  - The effect of compulsory voting on support for leftist policies?
  - Starting with the election of 1925, one Swiss canton (Vaud) introduced compulsory voting for its districts

- Data:
  - Outcome: Support for left-wing platforms at the district level (smaller than cantons)
  - Treatment: Compulsory voting
  - Two groups: District belongs to Vaud or not
  - Two time periods: 1924 (pre-period) and 1925 (post-period)

- Strategy: Compare voters' support for left-wing platforms across districts in Vaud vs. other cantons (which had no voting change policies), before and after the compulsory voting rule

- Q: What does parallel trends assumption imply in this context?

Estimation

```r
# Simple DiD
dat <- swiss.wide %>%
  mutate(trend = support_left_1925 - support_left_1924)
n1 = sum(dat$treated); n0 = sum(1-dat$treated)
did_estimate <- mean(dat$trend[dat$treated==1]) -
  mean(dat$trend[dat$treated==0])
did_estimate
```

```
## [1] 0.1551693
```

```r
# Regression implementation
library(estimatr)
lm_est = lm_robust(trend ~ treated, data = dat, se_type = "HC2")
cat(lm_est$coefficients[2], lm_est$std.error[2])
```

```
## 0.1551693 0.02876936
```

Linear two-way fixed effects model

- Alternatively,

$$Y_{it} = \alpha + \gamma G_i + \beta t + \tau D_{it} + \varepsilon_{it}$$

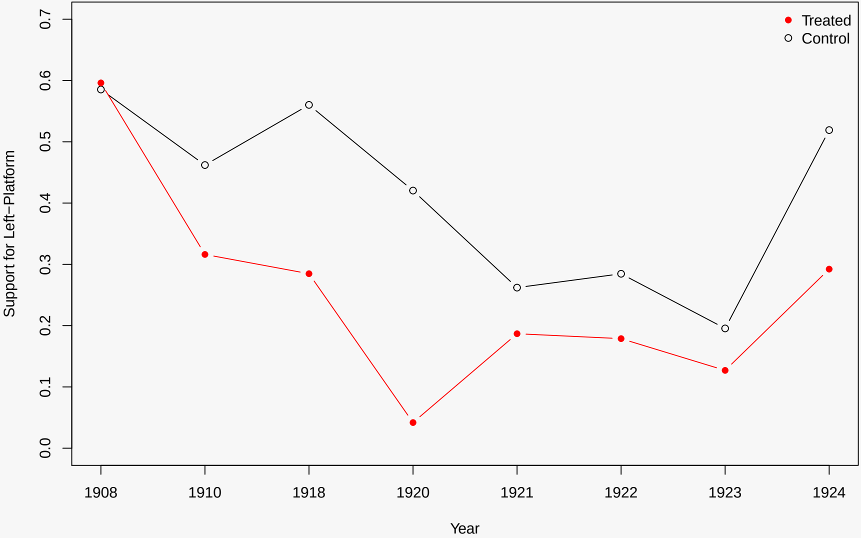
  - PT assumption: $\mathbb{E}[Y_{i1}(0) - Y_{i0}(0)|G_i = g] = \beta$ for $g = 0, 1$
  - Or equivalently, $\mathbb{E}[\varepsilon_{i1} - \varepsilon_{i0}|G_i = g] = 0$
  - $\mathbb{E}[Y_{i1}(1) \mid G_i = 1] - \mathbb{E}[Y_{i1}(0) \mid G_i = 1] = (\alpha+\gamma+\beta+\tau) - (\alpha+\gamma+\beta) = \tau$ $\rightsquigarrow$ DiD estimation
  - Only holds for the 2 group, 2 period case

Linear two-way fixed effects model

```
# Two-way fixed effect regression
library(fixest)
twfe_est = feols(support_left ~ treated:post|district_id + year, swiss)
summary(twfe_est, cluster = "district_id")
```

```
## OLS estimation, Dep. Var.: support_left
## Observations: 206
## Fixed-effects: district_id: 103,  year: 2
## Standard-errors: Clustered (district_id)
##              Estimate Std. Error t value   Pr(>|t|)
## treated:post 0.155169   0.028521 5.44055 3.6628e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.077775     Adj. R2: 0.568304
##                  Within R2: 0.130209
```

Falsification test: Check pre-treatment trends



Weighted DiD

- Standard DiD: **unconditional** parallel trends.
- This assumption may not be plausible – what if groups are unbalanced on characteristics that are associated with outcome?
- Alternative identification: **conditional** parallel trends

$$\mathbb{E}[Y_{i1}(0) - Y_{i0}(0) \mid G_i = 1, \mathbf{X}_i] = \mathbb{E}[Y_{i1}(0) - Y_{i0}(0) \mid G_i = 0, \mathbf{X}_i]$$

- Abadie (2015) derives weighting estimators

Weighted DiD: Estimation

- Weighted DiD estimator is similar to the IPW estimator for ATT

$$\widehat{\tau}_w = \frac{1}{n_1} \sum_{i=1}^{n} \left\{ G_i(Y_{i1} - Y_{i0}) - \frac{\pi(\mathbf{X}_i)(1 - G_i)(Y_{i1} - Y_{i0})}{1 - \pi(\mathbf{X}_i)} \right\}$$

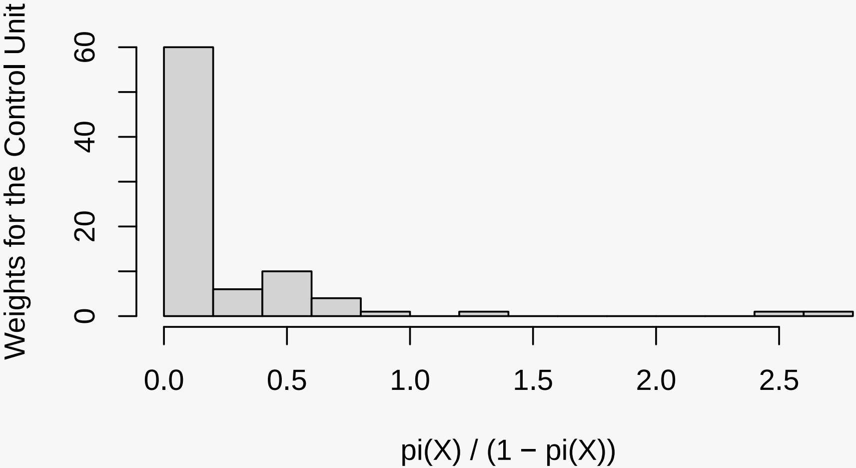
  - Review: Bonus Q2 of Pset7
- Here $\pi(\mathbf{X}_i) = \Pr(G_i = 1 \mid \mathbf{X}_i)$ is the propensity score
- Intuition: Weighting control observations such that
  - $1 - \pi(\mathbf{X}_i)$ is high $\rightsquigarrow$ *overrepresented* in the control $\rightsquigarrow$ downweight
  - $\pi(\mathbf{X}_i)$ is high $\rightsquigarrow$ looks like treated group $\rightsquigarrow$ upweight
- In practice, replace $\pi(\mathbf{X}_i)$ with its estimate $\widehat{\pi}(\mathbf{X}_i)$

Weighted DiD: Example

```r
# Estimate propensity score
swiss.wide$prop.score <- glm(treated ~ turnout * prop_kath +
  prop_sector1 + prop_sector2, data = swiss.wide, family = "binomial")$fitted
# Estimate ATT
weighted_did_fun <- function(dat, indices = NULL) {
  if (is.null(indices)) indices <- 1:nrow(dat)
  dat <- dat[indices,]; n <- nrow(dat); n1 <- sum(dat$treated)
  Y10 <- with(dat, support_left_1925 - support_left_1924);
  Gi  <- dat$treated
  weights <- with(dat,
                  ifelse(treated==1, 1, prop.score / (1 - prop.score)))
  weighted_did <- sum(Gi * Y10 - weights * (1 - Gi) * Y10) / n1
  attr(weighted_did, 'weights') <- weights; return(weighted_did)
}
set.seed(1234)
weighted_did_boot <- boot::boot(swiss.wide, weighted_did_fun, R = 200)
weighted_did_att  <- weighted_did_boot$t0
weighted_did_se   <- sd(weighted_did_boot$t)
cat(weighted_did_att, weighted_did_se)
```

```
## 0.1395537 0.06657771
```

Distribution of Weights



Weighted DiD: Pre-Treatment Trends

- Treated: $\overline{Y}_{t,\text{treated}} = \sum_{i=1}^{n} G_i Y_{it}/n_1$
- Control: $\overline{Y}_{t,\text{control}} = \sum_{i=1}^{n} (1 - G_i) w_i Y_{it}/n_1$

Other strategies

- Linear one way fixed effects generalizing the before/after design
    - Identification under strict exogeneity (no feedback!) + Estimation via:
        - within estimator
        - first differences
        - least squares dummy variable
    - Identification under sequential ignorability + Estimation via IV (Arellano-Bond method)

- Difference-in-Differences extensions
    - PT assumption of standard DiD not invariant to a nonlinear transformation of outcome (e.g., log)
    - $\rightsquigarrow$ Nonlinear DiD using quantile treatment effect (Athey and Imbens 2006)
    - Inappropriate for the ordinal outcome
    - $\rightsquigarrow$ Assumption on the quantile of the latent continuous variable (Yamauchi 2021+)

Other strategies

- Matching methods for panel data (Imai, Kim and Wang 2019; `panelmatch`)
  - Choose the number of lags L and leads F
  - ATE of policy change for the treated

$$\mathbb{E}[Y_{i,t+F}(D_{it} = 1, D_{i,t-1} = 0, \{D_{i,t-l}\}_{l=2}^{L}) - Y_{i,t+F}(D_{it} = 0, D_{i,t-1} = 0, \{D_{i,t-l}\}_{l=2}^{L}) \mid D_{it} = 1, D_{i,t-1} = 0]$$

  - Estimation: construct a matched set for each treated unit that consists of control units with the identical treatment history up to L time periods

- Synthetic Control Method (next class)