Program Evaluation (Causal Inference) 2: Difference-in-Differences (DID)

Yuta Toyama

Last update: July 9, 2019

Introduction

- Difference-in-differences (DID)
  - Exploit the panel data structure to estimate the causal effect.
- Consider that
  - Treatment and control group comparison: selection bias
  - Before v.s. After comparison: time trend
- DID combines those two comparisons to draw causal conclusion.

DID in Figure (on screen)

Plan of the Lecture

- Formal Framework
- Implementation in a regression framework
- Parallel Trend Assumption

Reference

- Angrist and Pischke "Mostly Harmless Econometrics" Chapter 5
- Marianne Bertrand, Esther Duflo, Sendhil Mullainathan, How Much Should We Trust Differences-In-Differences Estimates?, *The Quarterly Journal of Economics*, Volume 119, Issue 1, February 2004, Pages 249–275, https://doi.org/10.1162/003355304772839588
  - Discuss issues of calculating standard errors in the DID method.
- Hiro Ishise, Shuhei Kitamura, Masa Kudamatsu, Tetsuya Matsubayashi, and Takeshi Murooka (2019) "Empirical Research Design for Public Policy School Students: How to Conduct Policy Evaluations with Difference-in-differences Estimation" February 2019
  - Slide: https://slides.com/kudamatsu/did-manual/fullscreen#
  - Paper: https://docs.google.com/viewer?a=v&pid=sites&srcid=ZGVmYXVsdGRvbW

Framework

- Consider two periods: $t = 1, 2$. Treatment implemented at $t = 2$.
- Y_{it}: observed outcome for person i in period t

- G_i: dummy for treatment group
- D_{it}: treatment status
  - $D_{it} = 1$ if $t = 2$ and $G_i = 1$

- *potential outcomes*
  - $Y_{it}(1)$: outcome for i when she is treated
  - $Y_{it}(0)$: outcome for i when she is not treated

- With this, we can write

$$Y_{it} = D_{it} Y_{it}(1) + (1 - D_{it}) Y_{it}(0)$$

Identification

- Goal: ATT at $t = 2$

$$E[Y_{i2}(1) - Y_{i2}(0)|G_i = 1] = E[Y_{i2}(1)|G_i = 1] - E[Y_{i2}(0)|G_i = 1]$$

- What we observe

	Pre-period ($t = 1$)	Post ($t = 2$)
Treatment ($G_i = 1$)	$E[Y_{i1}(0)\|G_i = 1]$	$E[Y_{i2}(1)\|G_i = 1]$
Control ($G_i = 0$)	$E[Y_{i1}(0)\|G_i = 0]$	$E[Y_{i2}(0)\|G_i = 0]$

- Under what assumptions can we the ATT?
  - Simple comparison if $E[Y_{i2}(0)|G_i = 1] = E[Y_{i2}(0)|G_i = 0]$.
  - Before-after comparison if $E[Y_{i2}(0)|G_i = 1] = E[Y_{i1}(0)|G_i = 1]$.
  - Other (more reasonable) assumption?

Parallel Trend Assumption

- Assumption:

$$E[Y_{i2}(0) - Y_{i1}(0)|G_i = 0] = E[Y_{i2}(0) - Y_{i1}(0)|G_i = 1]$$

  - Change in the outcome *in the absence of treatment* is the same across two groups.

- Then,

$$\begin{aligned}\underbrace{E[Y_{i2}(1) - Y_{i2}(0)|G_i = 1]}_{ATT} =& E[Y_{i2}(1)|G_i = 1] - E[Y_{i2}(0)|G_i = 1] \\ =& E[Y_{i2}(1)|G_i = 1] - E[Y_{i1}(0)|G_i = 1] \\ & - \underbrace{(E[Y_{i2}(0)|G_i = 1] - E[Y_{i1}(0)|G_i = 1])}_{=E[Y_{i2}(0) - Y_{i1}(0)|G_i = 0] \ (pararell \ \ trend)}\end{aligned}$$

- Thus,

$$ATT = E[Y_{i2}(1) - Y_{i1}(0)|G_i = 1] - E[Y_{i2}(0) - Y_{i1}(0)|G_i = 0]$$

which is why this is called “difference-in-differences”.

Estimation Approach

1. Plug-in estimator
2. Regression estimators

Plug-in Estimator

- Remember that the ATT is

$$ATT = E[Y_{i2}(1) - Y_{i1}(0)|G_i = 1] - E[Y_{i2}(0) - Y_{i1}(0)|G_i = 0]$$

- Replace them with the sample average.

$$\begin{aligned} A\hat{TT} = & \{\bar{y}(t = 2, G = 1) - \bar{y}(t = 1, G = 1)\} \\ & - \{\bar{y}(t = 2, G = 0) - \bar{y}(t = 1, G = 0)\} \end{aligned}$$

where $\bar{y}(t, G)$ is the sample average for group G in period t .

- Easy to make a 2×2 table!

Example: Card and Kruger (1994)

Variable	Stores by state		
	PA (i)	NJ (ii)	Difference, NJ – PA (iii)
1. FTE employment before, all available observations	23.33 (1.35)	20.44 (0.51)	–2.89 (1.44)
2. FTE employment after, all available observations	21.17 (0.94)	21.03 (0.52)	–0.14 (1.07)
3. Change in mean FTE employment	–2.16 (1.25)	0.59 (0.54)	2.76 (1.36)

Regression Estimators

- Run the following regression

$$y_{it} = \alpha_0 + \alpha_1 G_i + \alpha_2 T_t + \alpha_3 D_{it} + \beta X_{it} + \epsilon_{it}$$

  - G_i: dummy for treatment group
  - T_t :dummy for treatment period
  - $D_{it} = G_i \times T_t$. α_3 captures the ATT.

- Regression framework can incorporate covariates X_{it}, which is important to control for observed confounding factors.

Regression Estimators with FEs

- With panel data

$$y_{it} = \alpha D_{it} + \beta X_{it} + \epsilon_i + \epsilon_t + \epsilon_{it}$$

  where ϵ_i is individual FE and ϵ_t is time FE.

- Do not forget to use the cluster-robust standard errors!
  - See Bertrand, Duflo, and Mullainathan (2004, QJE) for the standard error issues.

Discussions on Parallel Trend

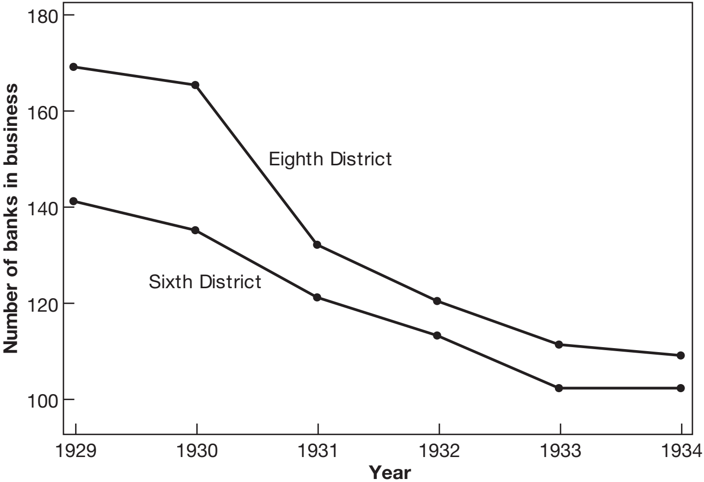
- Parallel trend assumption can be violated in various situations.
- Most critical issue: Treatment may depend on *time-varying factors*
  - DID can only deal with **time-invariant factors**.
- Self-selection: participants in worker training programs experience a decrease in earnings before they enter the program
- Targeting: policies may be targeted at units that are currently performing best (or worst).

Diagnostics for Parallel Trends: Pre-treatment trends

- Check if the trends are parallel in the pre-treatment periods
- Requires data on multiple pre-treatment periods (the more the better)
- This is very popular. You MUST do this if you have multiple pre-treatment periods.
- Note: this is only diagnostics, NEVER a direct test of the assumption!
  - You should never say "the key assumption for DID is satisfied if the pre-treatment trends are parallel.

Example (Fig 5.2 from Mastering Metrics)

FIGURE 5.2
Trends in bank failures in the Sixth and Eighth Federal Reserve Districts



Unit-Specific Time Trends

- Add group-specific time trends as

$$y_{it} = \alpha D_{it} + \beta_1 G_i \times t + \epsilon_i + \epsilon_t + \epsilon_{it}$$

- To see whether including the time trend does not change estimates that much. (robustness check)

- Note that
  - These time trends are meant to capture the trend in each group.
  - At least 3 periods of the data is needed.
  - But, these are assumed to be linear. We are not sure whether the trend is linear or not! So this is just a robustness check.

Other Diagnostics: Placebo test

- Placebo test using other period as treatment period.

$$y_{it} = \sum_{\tau} \gamma_{\tau} G_i \times I_{t,\tau} + \mu_i + \nu_t + \epsilon_{it}$$

  - The estimates of γ_τ should be close to zero up to the beggining of treatment (Fig 5.2.4 of Angrist and Pischke)

- Placebo test using different dependent variable which should not be affected by the policy.

Research Strategy using DID

- Going back to Ishise et al (2019)
  1. How to find a research question
  2. What outcome dataset to look for
  3. What policy to look for (except for example 1 and 2).