Class 19 Natural Experiment II: Difference-in-Differences Design

Dr. Wei Miao

UCL School of Management

December 6, 2024

Section 1

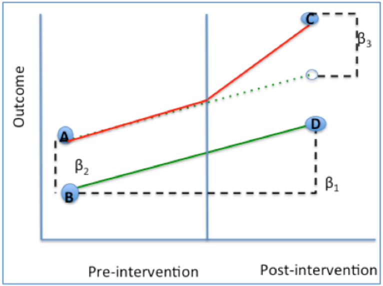
Difference-in-Differences Design

Learning Objectives

- Understand the economic intuition of the difference-in-differences (DiD) design.
- Learn how to estimate the treatment effect using DiD design.
- The intuition of Synthetic Difference-in-Differences when the parallel pre-trend assumption is violated.

Introduction to DiD

- **Difference-in-Differences Design** (**DiD**, **DD**, **Diff-in-Diff**) is a statistical technique used in economics and business that attempts to mimic an experimental research design using secondary data by comparing the changes in the outcomes of the treatment group with the changes in the outcomes of the control group.



Requirement for Using Difference-in-Differences

- **Control group**: We need to find a group of units who are unaffected by the natural experiment
- **Treatment group**: We need to find a treatment group of units who are affected by the natural experiment
- **No cross-over and spill-over**: There is no interference between the treatment and control group that can cause cross-over or spill-over effects.
- **Parallel trend assumption**: The treatment and control groups must have similar trends before treatment occurs.

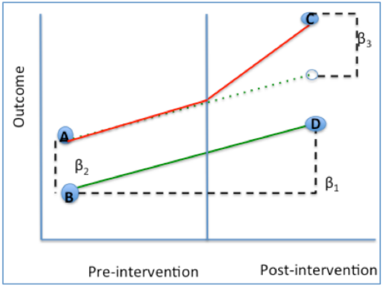
Note

The first 3 requirements apply to A/B testings as well.
Parallel trend is new for DiD analysis, and is the fundamental assumption that must be satisfied.

DiD Estimation: Linear Regression

$$Outcome_{i,t} = \beta_0 + \beta_1 Post_t + \beta_2 Treated_i + \beta_3 Treated_i \times Post_t + \mu_{i,t}$$

- $Post_t$ controls for the seasonality for both treatment and control groups
- $Treated_i$ controls for the pre-existing difference between the treatment group and control group.
- After accounting for (1) seasonality (β_1) and (2) pre-existing across-group differences (β_2), the interaction term (β_3) measures the treatment effect on iPhone users.[1]



[1]DiD method can only give average treatment effect on the treated (ATT).

Section 2

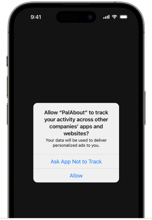
Implementation of DiD Using R

Application of DiD: The Causal Effect of Privacy Regulation

- Firms routinely collect consumer data on mobile apps and develop algorithms to target customers with personalized ads. However, consumers increasingly value privacy as an intrinsic human right: "the right to be left alone".
- Regulators and mobile ecosystems have enacted various regulations to restrict firms' collection of sensitive customer data.
  - EU's GDPR (2018), California's CCPA (2020), China's PIPL (2022)
  - Apple's App Tracking Transparency (2021), Android's Privacy Sandbox
- It's important to understand the causal effects of the privacy regulations on firms and consumers: (1) Trust-enhancing effect (2) Efficiency-decreasing effect

Apple's App Tracking Transparency Policy on Consumer Spending

- Before iOS 14.5 (26 April 2021), user data tracking on iOS lacked explicit user consent: iOS apps and advertisers could track user activities across different apps through the Identifier for Advertisers (IDFA).
- After iOS 14.5, Apple introduced the App Tracking Transparency (ATT) policy, which requires apps to obtain explicit user consent before tracking user activities across different apps.
- **Causal Question**: How does the implementation of Apple's App Tracking Transparency Policy affect consumer spending at your business?



The Causal Effect of the Introduction of ATT Policy

- y: standardized customer spending
- $x1$: standardized customer income.
- id: Identifier of the customer.
- $period$: normalized time period. 0 is the month in which the ATT policy was introduced.
- $post$: equals 1 for after ATT was introduced.
- $treat$: equals 1 for iPhone customers.

Data Preprocessing

- When you run DiD analysis, you need to construct a panel dataset similar to the following dataset.
- If the raw data are transaction data, you need to aggregate the data at the unit-time level, such as customer-month level.

```
            y         x1 id period post treat
1  2.87530627  0.5365377  1     -5    0     1
2  1.86065272 -3.0431894  1     -4    0     1
3  0.09416524  5.5768439  1     -3    0     1
4  3.78147485 -2.8300587  1     -2    0     1
5 -2.55819959 -5.0443544  1     -1    0     1
6  1.72873240 -0.6363849  1      0    1     1
```

Estimation of DiD Using Linear Regressions

- We need to run a linear regression with 3 variables: `treat`, `post`, and the interaction term `treat * post`

$$Outcome_{i,t} = \beta_0 + \beta_1 post_t + \beta_2 treat_i + \beta_3 treat_i \times post_t + \mu_{i,t}$$

```
est_did <- feols(
    fml = y ~ treat + post + treat:post, # method 1 for interactions
    # fml = y ~ treat * post # method 2 for interactions
    data = base_did
)
```

Report the DiD Results

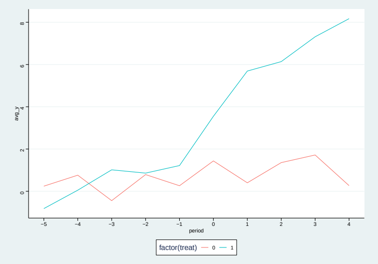
	(1)
(Intercept)	0.323
	(0.317)
treat	0.142
	(0.445)
post	0.713
	(0.449)
treat × post	4.993***
	(0.629)
Num.Obs.	1080
R2	0.183

$+$ p $<$ 0.1, * p $<$ 0.05, ** p $<$ 0.01, *** p $<$ 0.001

- The coefficient of the interaction term `treat:post` is the treatment effect of the ATT policy.
- After introducing the ATT policy, iPhone users increased their spending by 4.993 units compared to Android users.

Testing the Parallel Pre-trend Assumption

- Before we make the causal conclusion, we must test the parallel pre-trend assumption by plotting the average outcome for the treatment and control group over time.



- The graph indicates a violation of the parallel pre-trend assumption.

Section 3

Synthetic Difference-in-Differences

When the Parallel Pre-Trend is Violated

- If the parallel pre-trend assumption is violated, we can use synthetic difference-in-differences, which is a method that combines synthetic control and difference-in-differences.
- The **Synthetic Control Method** is a method that uses unit weighting to create a synthetic control group that approximates the pre-treatment outcomes of the treatment group (Abadie, Diamond, and Hainmueller 2010).
- However, Synthetic Control is very restrictive in many ways. Most importantly, it forces the treatment group to have the same level of the outcome variable as the synthetic control group, which is a very strong restriction.

Synthetic Difference-in-Differences

- Synthetic Difference-in-Differences is a method that uses synthetic control methods to estimate the treatment effect when the parallel pre-trend assumption is violated (Arkhangelsky et al. 2021; Clarke et al. 2023).

```
# devtools is required to install the synthdid package
pacman::p_load(devtools)
# install the synthdid package from GitHub
devtools::install_github("synth-inference/synthdid")
```

Prepare the Data for Synthetic DiD

- We need to prepare the data for the synthetic DiD method to the required format.

```
library(synthdid)
# Prepare the data

final_data <- panel.matrices(
    base_did %>% # treat must be treated * post
        mutate(treat = treat * post),
    unit = 3, # unit id
    time = 4, # period id
    outcome = 1, # outcome variable
    treatment = 6 # treat * post
)
```

Run SynthDiD

- After the dataset is prepared according to the `panel.matrices` function, we can run the `synthdid_estimate()` function to estimate the treatment effect.

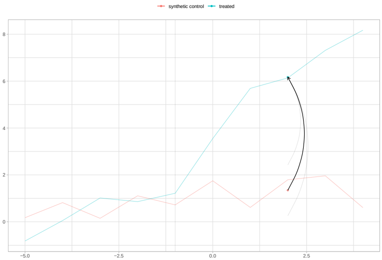
```
sdid_result <- synthdid_estimate(
    final_data$Y,
    final_data$N0,
    final_data$T0
)

print(sdid_result)
```

```
synthdid: 4.828 +- 1.025. Effective N0/N0 = 52.4/53~1.0. Effective T0/T
```

Visualization of SynthDiD

- Within the synthdid package, the `plot()` function can be used to visualize the results of the synthetic DiD method.

```
plot(sdid_result, overlay = 1, se.method = "bootstrap")
```



References

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. 2010. "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program." *Journal of the American Statistical Association* 105 (490): 493–505. https://doi.org/10.1198/jasa.2009.ap08746.

Arkhangelsky, Dmitry, Susan Athey, David A. Hirshberg, Guido W. Imbens, and Stefan Wager. 2021. "Synthetic Difference-in-Differences." *American Economic Review* 111 (12): 4088–118. https://doi.org/10.1257/aer.20190159.

Clarke, Damian, Daniel Pailañir, Susan Athey, and Guido Imbens. 2023. "Synthetic Difference In Differences Estimation." arXiv. https://arxiv.org/abs/2301.11859.