Business Analytics

Session 8a. Difference-in-Differences

Renyu (Philip) Zhang

New York University Shanghai

Spring 2019

Midterm Feedback

- More explanations on coding.
- Sample questions for final.
- Slow down the pace.
- Proof read code and lecture notes.

Some Questions from Office Hours

- No coding in the final exam.
  - Though you may be asked to recognize a short piece of code.
- *t*-test: To examine whether the mean value of a variable in two samples is the same.
  - 95% confidence interval: $[-1.96\widehat{SE}, 1.96\widehat{SE}]$
  - If $\hat{\mu}_1 - \hat{\mu}_2$ is out of the 95% confidence interval, the mean values of a variable in two samples are *statistically different*.
- Balance check: Make sure that the treatment assignment mechanism is well randomized.
  - The distribution of each covariate is the same for the treatment group and the control group.
- Including interactions of treatment and covariate in causal inference: When you believe the causal effect depends on the covariate.

Selection on Unobservables

- Often treated and untreated subjects differ in unobservable characteristics that are associated with potential outcomes even after controlling for differences in observed characteristics.

- In this case, treated and untreated subjects are not directly comparable. What can we do then?

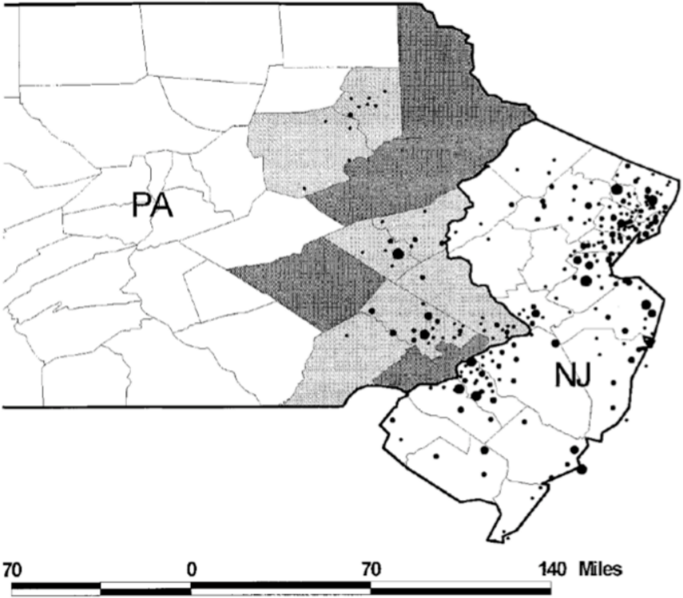
- If we can trace the subjects over-time, we can say more about the causal effect of treatment.

Do higher minimum wages reduce employment?

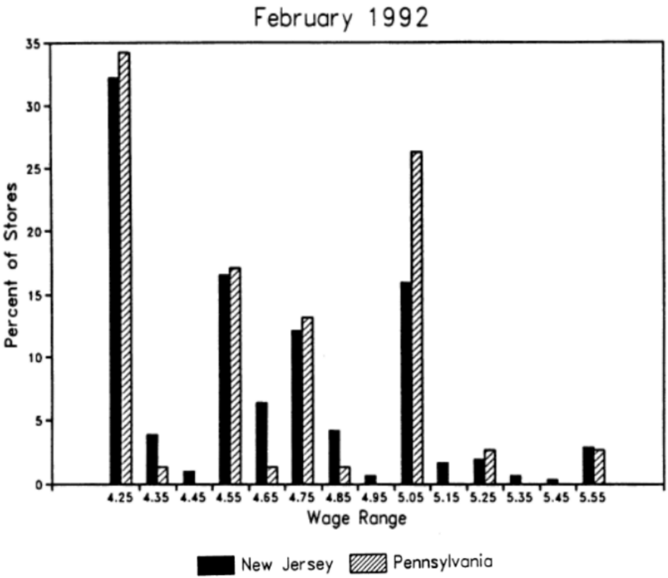
Example: Minimum Wage and Employment

- Difficult to measure the impact of minimum wage on employment.
- Card and Krueger (1994) consider impact of New Jersey's 1992 minimum wage increase from $4.25 to $5.05 per hour.
- Compare employment in 410 fast-food restaurants in New Jersey and eastern Pennsylvania before and after the rise.
- Survey data on wages and employment from two waves:
  - Wave 1: March 1992, one month before the minimum wage increase.
  - Wave 2: December 1992, eight month after the increase.

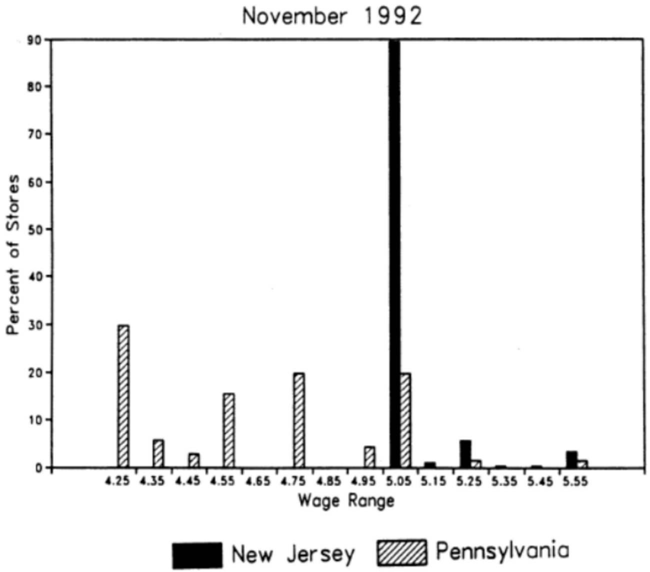
Restaurant Locations



Before Minimum Wage Rise



After Minimum Wage Rise



Diff-in-Diff Setup

- Subjects: $i = 1, 2, , \cdots , N$ (N is the sample size)
- Treatment: $W_i \in \{0, 1\}$, not necessarily randomly assigned
- Two periods:
  - $t = 0$: Pre-treatment period
  - $t = 1$: Post-treatment period

- Potential outcomes:
  - $Y_{it}(1)$: Potential outcome of subject i in period t if treated.
  - $Y_{it}(0)$: Potential outcome of subject i in period t if not treated.

- # of treated/untreated subjects: $N_1 = \sum_{i=1}^{N} W_i$ and $N_0 = N - N_1$
- Causal effect of treatment for subject i in period t:

$$Y_{it}(1) - Y_{it}(0)$$

Diff-in-Diff Identification Strategy

- Want to estimate:

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

	Post-Period ($t = 1$)	Pre-Period ($t = 0$)
Treated ($W_i = 1$)	$\mathbb{E}[Y_{i1}(1)\|W_i = 1]$	$\mathbb{E}[Y_{i0}(0)\|W_i = 1]$
Control ($W_i = 0$)	$\mathbb{E}[Y_{i1}(0)\|W_i = 0]$	$\mathbb{E}[Y_{i0}(0)\|W_i = 0]$

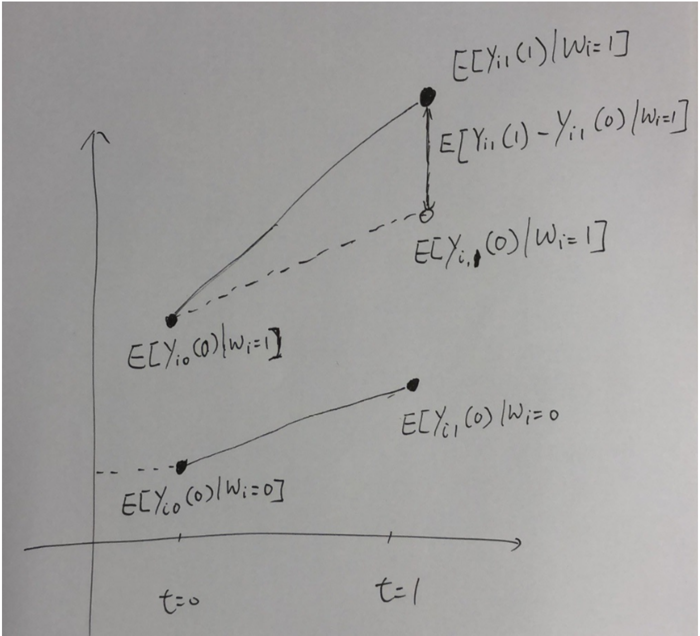
- Missing potential outcome: $\mathbb{E}[Y_{i1}(0)|W_i = 1]$, i.e., the average post-period outcome for treated in the absence of the treatment.
- Paralell trends assumption:

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

- Under the parallel trends assumption,

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

Diff-in-Diff Identification Strategy



Estimation of Causal Effect using Diff-in-Diff

Diff-in-Diff Estimator

$$\widehat{ATT} = \left\{ \frac{1}{N_1} \sum_{W_i=1} Y_{i1}(1) - \frac{1}{N_0} \sum_{W_i=0} Y_{i1}(0) \right\} - \left\{ \frac{1}{N_1} \sum_{W_i=1} Y_{i0}(0) - \frac{1}{N_0} \sum_{W_i=0} Y_{i0}(0) \right\}$$

$$= \left\{ \frac{1}{N_1} \sum_{W_i=1} (Y_{i1}(1) - Y_{i0}(0)) - \frac{1}{N_0} \sum_{W_i=0} (Y_{i1}(0) - Y_{i0}(0)) \right\}$$

- Theorem. If the parallel trends assumption holds, $\widehat{ATT} \approx ATT$.
- Estimated standard error of $\widehat{ATT}$:

$$\widehat{SE} = \sqrt{\frac{\hat{\sigma}_1^2}{N_1} + \frac{\hat{\sigma}_0^2}{N_2}},$$

where $\hat{\sigma}_1$ is the sample standard error for $Y_{i1}(1) - Y_{i0}(0)$ in the treatment group and $\hat{\sigma}_0$ is the sample standard error for $Y_{i1}(0) - Y_{i0}(0)$ in the control group.

Example: Impact of Minimum Wage on Employment

	Stores by state		
Variable	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)

Diff-in-Diff Estimator using Regression

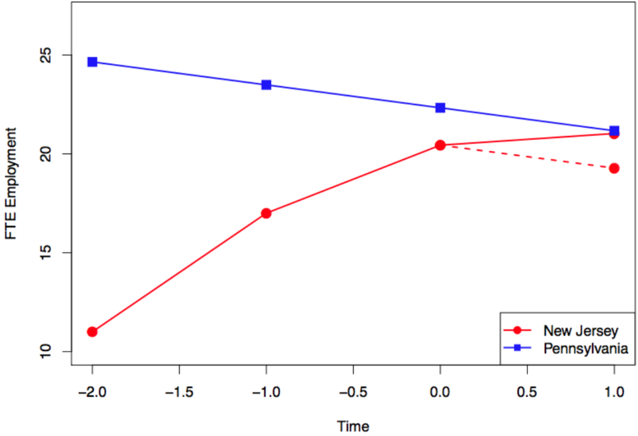
- Regression: $Y \approx \hat{\beta}_0 + \hat{\beta}_1 W + \hat{\beta}_2 t + \hat{\beta}_3 (W \cdot t)$.
  - $\hat{\beta}_3$ is the causal effect we are interested in.

	Post-Period ($t = 1$)	Pre-Period ($t = 0$)	After-Before
Treated ($W_i = 1$)	$\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_3$	$\hat{\beta}_0 + \hat{\beta}_1$	$\hat{\beta}_2 + \hat{\beta}_3$
Control ($W_i = 0$)	$\hat{\beta}_0 + \hat{\beta}_2$	$\hat{\beta}_0$	$\hat{\beta}_2$
Treatment-Control	$\hat{\beta}_1 + \hat{\beta}_3$	$\hat{\beta}_1$	$\hat{\beta}_3$

- Question: How do we interpret $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\beta}_3$?
- With other covariates: Diff-in-Diff + Propensity Score Matching (make the treatment and control groups comparable)
- Using multiple periods before and after treatment.

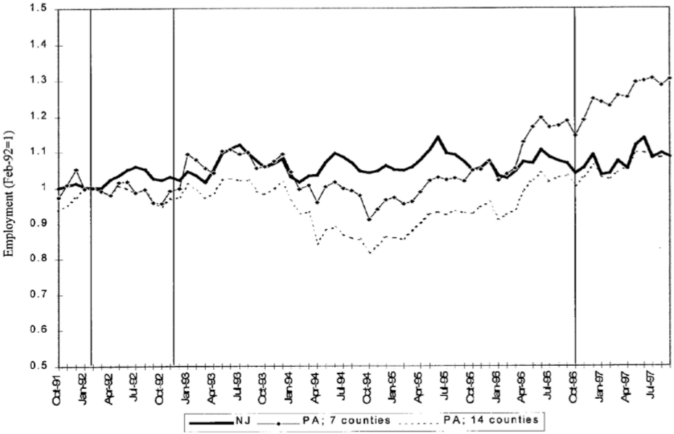
Violation of Parallel Trends Assumption

- Self-selection and targeting:
  - 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).



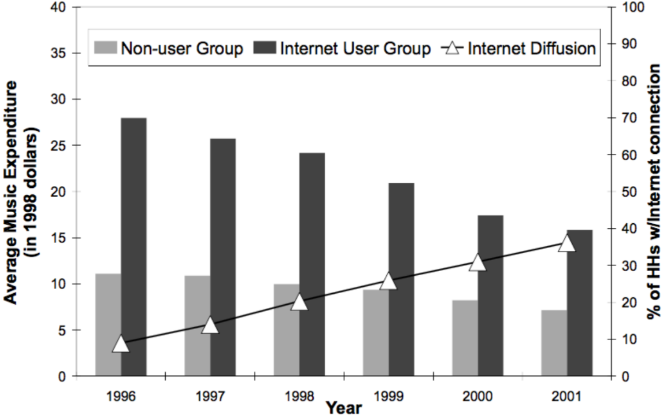
Violation of Parallel Trends Assumption

- Self-selection and targeting:
  - 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).



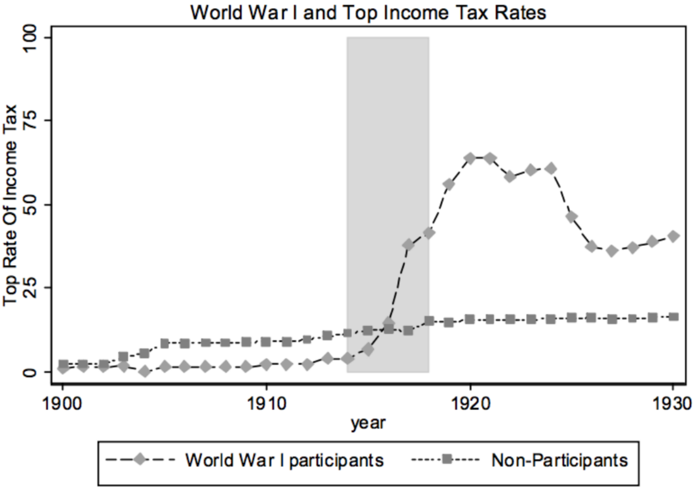
Violation of Parallel Trends Assumption

- Compositional Differences Across Time:
  - The composition of the sample may change between periods, i.e. due to migration.
  - This may confound any DiD estimate since the "effect" may be attributable to change in population.



Violation of Parallel Trends Assumption

- Long-term effects:
  - Parallel trends assumption is hardly true in the long run.



Violation of Parallel Trends Assumption

- Functional form dependence:
  - Training program for young workers: Employment for young workers increases from 20% to 30%.
  - Control: Employment for old workers increases from 5% to 10%.
  - DiD effect: $(30\% - 20\%) - (10\% - 5\%) = 5\%$.
  - If we take log: $(\log(0.3) - \log(0.2) - (\log(0.1) - \log(0.05)) < 0$
  - DiD estimates are more reliable if treated and control are more similar in period 0.

Life-Saving Diff-in-Diff

- The cholera epidemic in London in 1849 killed over 14,000 lives.
- John Snow (the father of modern epidemiology and biostatistics) believed cholera was spread by cotaminated water.
  - How to prove it?

Life-Saving Diff-in-Diff

- First Difference: Water provided by two companies, (1) the Lambeth and (2) the Southwark and Vauxhall. Both got water from the Thames.
- Second Difference: Before and after 1852. In 1852, Lambeth moved their intake upriver.
- Before moving:
  - The same death rates for the housholds under the two companies.
- After moving:
  - Southwark and Vauxhall: 71 cholera deaths/10,000 homes.
  - Lambeth after moving water source: 5 cholera deaths/10,000 homes.
- As a result, Southwark and Vauxhall moved their intake upriver in 1855 and the epidemic subsided.