10. Panel Data and Diff-in-Diffs

ISS5096 || ECI
Jaewon ("Jay-one") Yoo
National Tsing Hua University

Where are we? Where are we going?

- Where we have found good controls:
    - Units randomized to receive control
    - Units with similar values of covariates
    - Units with opposite value of some instrument
    - At a discontinuity in treatment assignment (will cover in W13)
- What if we have repeated measurements of the same units?
- Now there are two possible sources of variation to exploit:
    - Exploit **cross-sectional** variation in treatment.
    - Exploit variation in treatment **within a unit over time** (before/after)

Source: *Chapter 5* of Mostly Harmless Econometrics (Textbook 1) by J. Angrist & J. Pischke

1/ Difference in Differences

Minimum Wages (Card & Krueger, 1994)

- Does increasing the minimum wage affect employment?
- Classical economic theory tends to point to negative effects.
- But difficult to randomize changes to the minimum wage.
- In 1992, NJ minimum wage increased from $4.25 to $5.05
    - Neighboring PA stays at $4.25
    - We observe employment in both states before and after increase
- Look at eastern PA and NJ fast food restaurants.
    - Similar prices, wages, products, etc.
    - Most likely to be affected by the change.

Minimum Wage Law in PA and NJ

Source: https://www.alamy.com/stock-photo/pennsylvania-map.html

Difference in Differences Design

- Basic setup: two groups, two time periods.
  - Pre-period ($t = 0$): neither group is treated.
  - Post-period ($t = 1$): one group is treated, the other group remains untreated.
- Groups defined by treatment status in post-period:
  - $G_i = 1$ denote those that are treated at $t = 1$.
  - $G_i = 0$ denote those that are always untreated.
- Treatment status in each period:
  - No treatment in the first period for either group: $D_{i0} = 0$.
  - In treated group: $G_i = 1 \quad \rightsquigarrow \quad D_{i1} = 1$
  - In control group: $G_i = 0 \quad \rightsquigarrow \quad D_{i1} = 0$

	Time period	
	Pre-period ($t = 0$)	Post-period ($t = 1$)
Control group ($G_i = 0$)	$D_{i0} = 0$	$D_{i1} = 0$
Treated group ($G_i = 1$)	$D_{i0} = 0$	$D_{i1} = 1$

Potential Outcomes Approach to DID

- $Y_{it}(d)$ is the potential outcome under treatment d and time t.
- Again, the individual causal effect is just $Y_{it}(1) - Y_{it}(0)$.
- Consistency: $Y_{it} = D_{it} \cdot Y_{it}(1) + (1 - D_{it}) \cdot Y_{it}(0)$
  - We observe control PO for all units in the first period: $Y_{i0}(0) = Y_{i0}$
  - In treated group: $G_i = 1 \quad \rightsquigarrow \quad Y_{i1} = Y_{i1}(1)$
  - In control group: $G_i = 0 \quad \rightsquigarrow \quad Y_{i1} = Y_{i1}(0)$

Identification Problem

- Average treatment effect on the treated (i.e., ATT):

$$\begin{aligned}\tau_{\text{ATT}} &= \mathbb{E}[Y_{i1}(1) - Y_{i1}(0)|G_i = 1] \\ &= \mathbb{E}[Y_{i1}(1)|G_i = 1] - \mathbb{E}[Y_{i1}(0)|G_i = 1] \\ &= \underbrace{\mathbb{E}[Y_{i1}|G_i = 1]}_{\text{(a)}} - \underbrace{\mathbb{E}[Y_{i1}(0)|G_i = 1]}_{\text{(b)}}\end{aligned}$$

- Part (a) is just a conditional average of observed data $\rightsquigarrow$ identified.
- Part (b) is a counterfactual: what would the average outcome in the treated group have been if it have been in control?

Three Control Strategies

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

	Time period	
	Pre-period ($t = 0$)	Post-period ($t = 1$)
Control group ($G_i = 0$)	$\mathbb{E}[Y_{i0}(0)|G_i = 0]$	$\mathbb{E}[Y_{i1}(0)|G_i = 0]$
Treated group ($G_i = 1$)	$\mathbb{E}[Y_{i0}(0)|G_i = 1]$	$\mathbb{E}[Y_{i1}(1)|G_i = 1]$

- **#1: Cross-Sectional Design**
  - Assumption: mean independence of treatment.

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

  - Use post-treatment control group as control:

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

Three Control Strategies

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

	Time period	
	Pre-period ($t = 0$)	Post-period ($t = 1$)
Control group ($G_i = 0$)	$\mathbb{E}[Y_{i0}(0)\|G_i = 0]$	$\mathbb{E}[Y_{i1}(0)\|G_i = 0]$
Treated group ($G_i = 1$)	$\mathbb{E}[Y_{i0}(0)\|G_i = 1]$	$\mathbb{E}[Y_{i1}(1)\|G_i = 1]$

- **#2: Before-and-After Design**
    - Assumption: no trends.

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

    - Use pre-period outcome in treated group:

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

Three Control Strategies

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

	Time period	
	Pre-period ($t = 0$)	Post-period ($t = 1$)
Control group ($G_i = 0$)	$\mathbb{E}[Y_{i0}(0)\|G_i = 0]$	$\mathbb{E}[Y_{i1}(0)\|G_i = 0]$
Treated group ($G_i = 1$)	$\mathbb{E}[Y_{i0}(0)\|G_i = 1]$	$\mathbb{E}[Y_{i1}(1)\|G_i = 1]$

- **#3: Difference-in-Differences**
    - Assumption: parallel trends.

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

- Use pre-period treated outcome plus trend in control group:

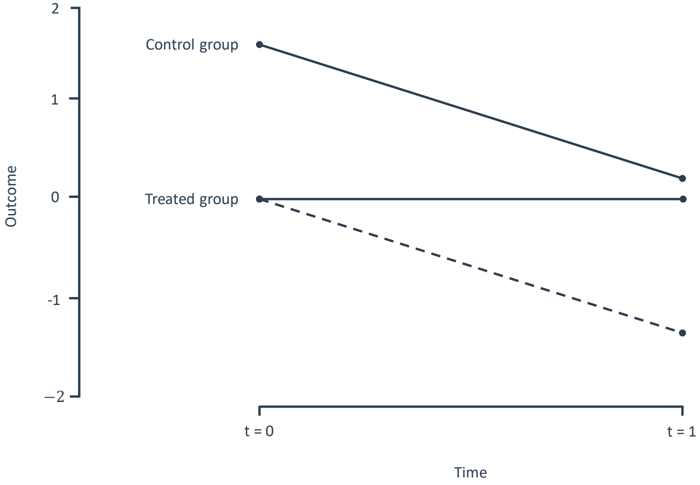
$$\begin{aligned}\tau_{ATT} &= (\mathbb{E}[Y_{i1}|G_i = 1] - \mathbb{E}[Y_{i0}|G_i = 1]) \\ &\quad - (\mathbb{E}[Y_{i1}|G_i = 0] - \mathbb{E}[Y_{i0}|G_i = 0])\end{aligned}$$

Parallel Trends

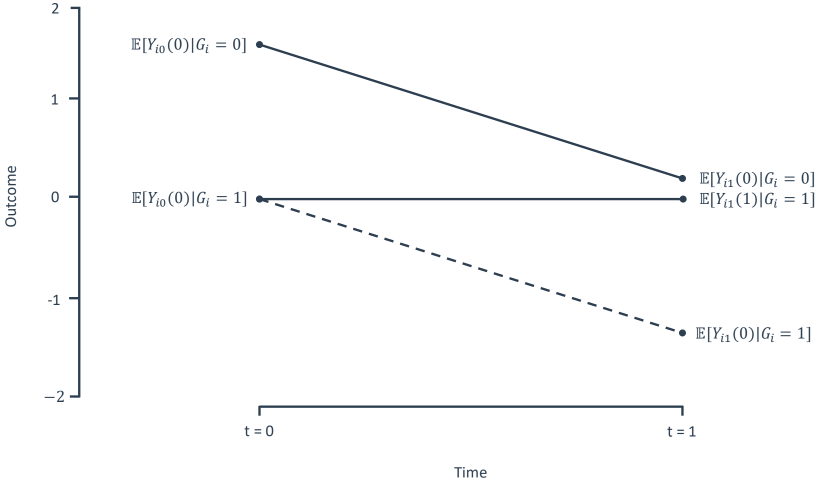
$$\mathbb{E}[Y_{i1}(0) - Y_{i0}(0)|G_i = 0] = \mathbb{E}[Y_{i1}(0) - Y_{i0}(0)|G_i = 1]$$

- Key assumption of difference-in-differences: **parallel trends**
- Interpretation:
  - Secular trends in the control group is a good proxy for how the treated group would have changed over time without treatment.
    - p.s., a secular trend = a general trend built up over the years.
- Why is this weaker than other assumptions?
  - It allows for time-constant unmeasured confounding b/w Y_{it} and G_i.
  - It allows for (common) secular trends in the outcome over time.
- Not invariant to non-linear transformations:
  - Parallel trends for Y_{it} implies non-parallel trends for $\log(Y_{it})$ vice versa.

Parallel Trends in a Graph



Parallel Trends in a Graph



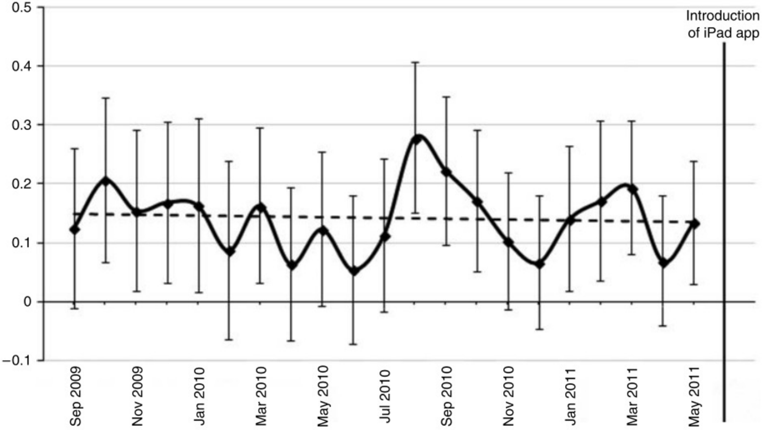
Identification

- Identification result:

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

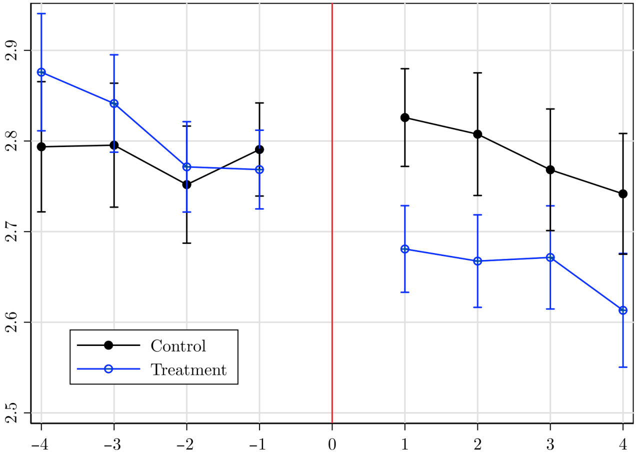
- Threat to identification: non-parallel trends
  - **Unmeasured time-variant confounders**
  - **Ashenfelter's dip**: empirical finding that people who enroll in job training programs see their earnings decline prior to that training (link).
- Falsification test: check pre-treatment parallel trends.
  - Does's imply parallel trends hold for the post-period, however!

Checking Parallel Trends (Xu et al., 2017)



Source: Xu, Kaiquan, Jason Chan, Anindya Ghose, and Sang Pil Han. "Battle of the channels: The impact of tablets on digital commerce." Management Science 63, no. 5 (2017): 1469-1492.

Checking Parallel Trends (Kretschmer & Peukert, 2020)



Source: Kretschmer, Tobias, and Christian Peukert. "Video killed the radio star? Online music videos and recorded music sales." Information Systems Research 31, no. 3 (2020): 776-800.

Estimation

- Estimation with panel data:

$$\widehat{\tau}_{\text{ATT}} = \underbrace{\frac{1}{n_1}\sum_{i=1}^{n} G_i\{Y_{i1} - Y_{i0}\}}_{\text{average trend in treated group}} - \underbrace{\frac{1}{n_0}\sum_{i=1}^{n}(1 - G_i)\{Y_{i1} - Y_{i0}\}}_{\text{average trend in control group}}$$

  - Standard errors from standard difference in means.

- Regression implementation:
  - Regress $\Delta Y_i = Y_{i1} - Y_{i0}$ on G_i.
  - Use (cluster) robust SEs.

- Also possible to use DID on repeated cross sections.

DID and Linear Two-Way Fixed Effects

- Linear two-way (group and time) fixed effects model:

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

  - Fixed effect for group and time.
  - D_{it} is treatment status for unit i at time t (i.e., $G_i \times t$).
  - Be sure to cluster by unit (or level of treatment assignment).
- Coefficient on D_{it} equivalent to DID estimation.
- Only holds for the 2 group, 2 period case!
  - Large new literature on interpretation of TWFE in more general cases.
  - Basically, TWFE is an odd weighted average of DID effects with sometimes negative weights.

Alternative Identification Assumption

- Up until now, we assumed unconditional parallel trends. What if this doesn't hold?
- Alternative identification: **conditional parallel trends**

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

- What does this assumption say?
  - Potential trend under control is the same for the control & treated groups, conditional on covariates.
    - $\rightsquigarrow$ Units that are similar at baseline will follow similar paths under no treatment.
- **Matching**: conduct DiD analysis on units with similar values of $\mathbf{X}_i$
  - *We will discuss more when we cover matching methods (e.g., PSM, CEM)!*

Example R Codes

```
> minwage <- read_csv(url("https://bit.ly/3QZmodH")) # Don't forget to load tidyverse!
> minwage <- minwage |>
    mutate(
      state = if_else(location == "PA", "PA", "NJ"),
      treated = if_else(location == "PA", 0, 1) ## PA is control
    ); head(minwage)

# A tibble: 6 × 10
  chain      location wageBefore wageAfter fullBefore fullAfter partBefore partAfter state treated
  <chr>      <chr>         <dbl>     <dbl>      <dbl>     <dbl>      <dbl>     <dbl> <chr>   <dbl>
1 wendys     PA             5         5.25         20         0         20        36 PA          0
2 wendys     PA             5.5       4.75          6        28         26         3 PA          0
3 burgerking PA             5         4.75         50        15         35        18 PA          0
4 burgerking PA             5         5            10        26         17         9 PA          0
5 kfc        PA             5.25      5             2         3          8        12 PA          0
6 kfc        PA             5         5             2         2         10         9 PA          0

> minwage |>
    count(location)
# A tibble: 5 × 2
  location      n
  <chr>     <int>
1 PA           67
2 centralNJ    45
3 northNJ     146
4 shoreNJ      33
5 southNJ      67
```

Example R Codes

```
# Simple DID
> minwage <- minwage |>
    mutate(part_prop_after = partAfter / (fullAfter + partAfter),
           part_prop_before = partBefore / (fullBefore + partBefore),
           trend = part_prop_after - part_prop_before
    )
> n1 = sum(minwage$treated); n0 = sum(1 - minwage$treated); n1; n0
[1] 291
[1] 67
> did_estimate <- mean(minwage$trend[minwage$treated==1]) -
    mean(minwage$trend[minwage$treated==0]); did_estimate # Our DID estimate is -0.06155831
[1] -0.06155831

# Regression Implementation
> require(estimatr)
> estimatr::lm_est <- lm_robust(trend ~ treated, data = minwage, se_type = "HC2")
> cat("did estimate: ", lm_est$coefficients[2], "; robust SE estimate:", lm_est$std.error[2])
did estimate:  -0.06155831 ; robust SE estimate: 0.04551265
```

Example R Codes

```
# Two-Way Fixed Effect (TWFE) Regression
> require(fixest)
> minwage_long <- minwage |>
    pivot_longer(
      cols = c(part_prop_after, part_prop_before),
      names_to = "post",
      values_to = "part_prop"
    ) |>
    mutate(post = if_else(post == "part_prop_before", 0, 1),
           year = if_else(post == 0, 1991, 1992))

> twfe_est <- feols(part_prop ~ treated:post|state + year, data = minwage_long)
> summary(twfe_est, cluster = "chain")

OLS estimation, Dep. Var.: part_prop
Observations: 716
Fixed-effects: state: 2,  year: 2
Standard-errors: Clustered (chain)
              Estimate Std. Error  t value Pr(>|t|)
treated:post -0.061558   0.021945 -2.80517 0.067562 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
RMSE: 0.240757     Adj. R2: -2.787e-4
                 Within R2:  0.00248
```

On to the Presentations & Discussions!

Contact Information:
jaewon.yoo@iss.nthu.edu.tw
https://j1yoo4.github.io/

Appendix

Basic Idea Behind Fixed Effects

- Given a panel data (repeated measurements of the same units),

$$\begin{aligned} Y_{it} &= \alpha + \tau D_{it} + \mathbf{X}'_{it}\beta + \varepsilon_{it} \\ &= \alpha + \tau D_{it} + \mathbf{X}'_{it}\beta + \underbrace{\gamma_i + \lambda_t + \upsilon_{it}}_{\varepsilon_{it} \text{ decomposed}} \end{aligned}$$

  - Individual fixed effects (i.e., γ_i): unobserved innate differences between individuals that are time-invariant.
    - Unobserved factors captured by γ_i, e.g.,:
      1) ability or preferences (at the individual-level),
      2) atmosphere or management style (at the firm-level), or
      3) brand image or fashion trend (at the product-level).
    - Observed factors? $\rightsquigarrow$ captured with the covariates (i.e., $\mathbf{X}'_{it}$)!
  - Time fixed effects (i.e., λ_t): Impacts common to all groups but vary by year (e.g., seasonality)
  - Idiosyncratic error term (i.e., υ_{it})

Basic Idea Behind Fixed Effects

- $\rightsquigarrow$ FE model takes the original econometric error term (i.e., ε_{it}) and decomposes it into three pieces (i.e., γ_i, λ_t, and υ_{it})!
- Alternative to FE specification is random effects (e.g., Wooldridge, 2006) which "ignores" the indiv. & time FE (in cases where it will not create any bias from ignoring it).
  - In this case, the FEs remain as part of the econometric error term.
  - Wooldridge, Jeffrey M. "Introductory econometrics: A modern approach, 3rd." New York: Thomson 53 (2006).

Semiparametric Estimation w/ Repeated Outcomes

- How to estimate regression DID without strong linearity assumptions?
- Abadie (2005) derives **weighting estimators** in this setting:

$$\mathbb{E}\left[Y_i(1) - Y_i(0) | G_i = 1\right] = \mathbb{E}\left[\frac{(Y_{i1} - Y_{i0})}{\mathbb{P}(G_i = 1)} \cdot \frac{G_i - \mathbb{P}(G_i = 1 \mid \mathbf{X}_i)}{1 - \mathbb{P}(G_i = 1 \mid \mathbf{X}_i)}\right]$$

- Reweights control group to have the same distribution of $\mathbf{X}_i$ as treated group.
- Have to estimate the propensity score $\mathbb{P}(G_i = 1 \mid \mathbf{X}_i)$
    - Issue: possible model misspecification!