Difference-in-Differences Methods

Teppei Yamamoto

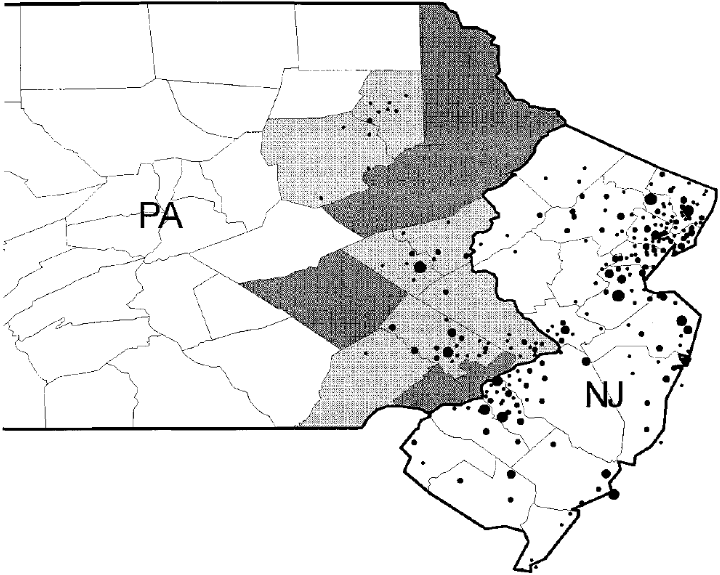
Keio University

Introduction to Causal Inference
Spring 2016

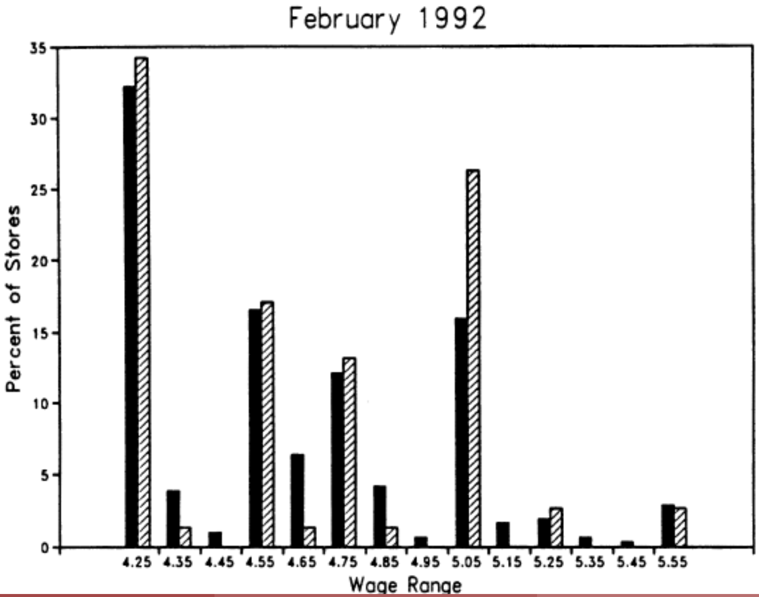
Example: Minimum Wage and Employment

- Do higher minimum wages decrease 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 increase

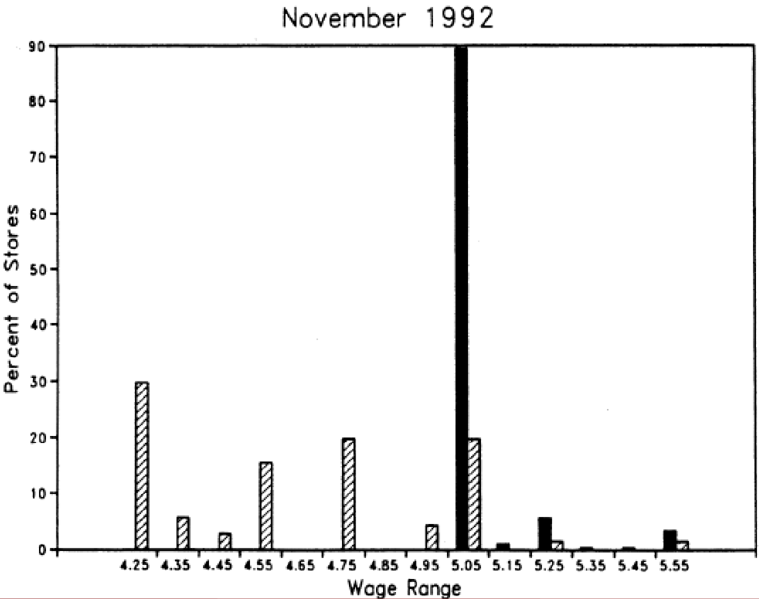
Location of Restaurants



Wages Before Rise in Minimum Wage



Wages After Rise in Minimum Wage



Setup: Groups, Periods and Treatments

Data structure:

- Two waves of randomly sampled cross-sectional observations
- Either panel or repeated cross sections

Cross-sectional units: $i \in \{1, ..., N\}$

Time periods: $t \in \{0 \text{ (pre-treatment)}, 1 \text{ (post-treatment)}\}$

Group indicator: $G_i = \begin{cases} 1 & \text{(treatment group)} \\ 0 & \text{(control group)} \end{cases}$

Treatment indicator: $Z_{it} \in \{0, 1\}$

Units in the treatment group receive treatment in $t = 1$:

	Time Period	
Group	$t = 0$	$t = 1$
$G_i = 1$ (treatment group)	$Z_{i0} = 0$ (untreated)	$Z_{i1} = 1$ (treated)
$G_i = 0$ (control group)	$Z_{i0} = 0$ (untreated)	$Z_{i0} = 0$ (untreated)

Setup: Potential Outcomes

Potential outcomes $Y_{it}(z)$:

- $Y_{it}(0)$: potential outcome for unit i in period t when not treated
- $Y_{it}(1)$: potential outcome for unit i in period t when treated

Setup: Potential Outcomes

Potential outcomes $Y_{it}(z)$:

- $Y_{it}(0)$: potential outcome for unit i in period t when not treated
- $Y_{it}(1)$: potential outcome for unit i in period t when treated

Causal effect for unit i at time t is

$$\tau_{it} = Y_{it}(1) - Y_{it}(0)$$

Setup: Potential Outcomes

Potential outcomes $Y_{it}(z)$:

- $Y_{it}(0)$: potential outcome for unit i in period t when not treated
- $Y_{it}(1)$: potential outcome for unit i in period t when treated

Causal effect for unit i at time t is

$$\tau_{it} = Y_{it}(1) - Y_{it}(0)$$

Observed outcomes Y_{it} are realized as

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

Because $Z_{i1} = G_i$ in the post-treatment period, we can also write

$$Y_{i1} = Y_{i1}(0)(1 - G_i) + Y_{i1}(1)G_i$$

Identification Strategies

Estimand: ATT in the post-treatment period

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

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

Problem: Missing potential outcome:

Identification Strategies

Estimand: ATT in the post-treatment period

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

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

Problem: Missing potential outcome: $\mathbb{E}[Y_{i1}(0)|G_i = 1]$, i.e. what is the average post-period outcome for the treated group in the absence of the treatment?

Identification Strategies

Estimand: ATT in the post-treatment period

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

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

Control Strategy: Before vs. After

- Use $\mathbb{E}[Y_{i1}|G_i = 1] - \mathbb{E}[Y_{i0}|G_i = 1]$ for τ_{ATT}
- Assumes

Identification Strategies

Estimand: ATT in the post-treatment period

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

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

Control Strategy: Before vs. After

- Use $\mathbb{E}[Y_{i1} | G_i = 1] - \mathbb{E}[Y_{i0} | G_i = 1]$ for τ_{ATT}
- Assumes $\mathbb{E}[Y_{i1}(0) | G_i = 1]$ = $\mathbb{E}[Y_{i0}(0) | G_i = 1]$
  (No change in average potential outcome over time)

Identification Strategies

Estimand: ATT in the post-treatment period

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

	Pre-Period ($t = 0$)	Post-Period ($t = 1$)
Treatment Group ($G_i = 1$)	$\mathbb{E}[Y_{i0}(0) \mid G_i = 1]$	$\mathbb{E}[Y_{i1}(1) \mid G_i = 1]$
Control Group ($G_i = 0$)	$\mathbb{E}[Y_{i0}(0) \mid G_i = 0]$	$\mathbb{E}[Y_{i1}(0) \mid G_i = 0]$

Control Strategy: Treated vs. Control in Post-Period

- Use $\mathbb{E}[Y_{i1} | G_i = 1] - \mathbb{E}[Y_{i1} | G_i = 0]$ for τ_{ATT}
- Assumes

Identification Strategies

Estimand: ATT in the post-treatment period

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

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

Control Strategy: Treated vs. Control in Post-Period

- Use $\mathbb{E}[Y_{i1} | G_i = 1] - \mathbb{E}[Y_{i1} | G_i = 0]$ for τ_{ATT}
- Assumes $\mathbb{E}[Y_{i1}(0) | G_i = 1]$ = $\mathbb{E}[Y_{i1}(0) | G_i = 0]$
  (Mean ignorability of treatment assignment)

Identification Strategies

Estimand: ATT in the post-treatment period

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

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

Control Strategy: Difference-in-Differences (DD)

- Use: $\left\{\mathbb{E}[Y_{i1}|G_i = 1] - \mathbb{E}[Y_{i1}|G_i = 0]\right\} - \left\{\mathbb{E}[Y_{i0}|G_i = 1] - \mathbb{E}[Y_{i0}|G_i = 0]\right\}$

Identification Strategies

Estimand: ATT in the post-treatment period

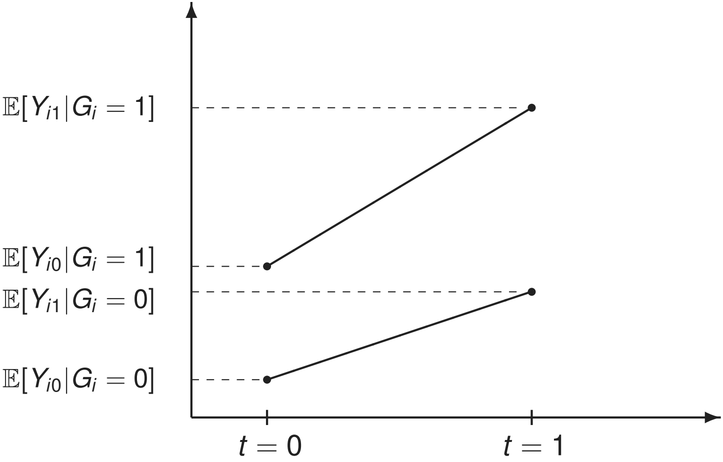
$$\begin{aligned}\tau_{ATT} &= \mathbb{E}[Y_{i1}(1) - Y_{i1}(0) | G_i = 1] \\ &= \mathbb{E}[Y_{i1}(1) | G_i = 1] - \mathbb{E}[Y_{i1}(0) | G_i = 1]\end{aligned}$$

	Pre-Period ($t = 0$)	Post-Period ($t = 1$)
Treatment Group ($G_i = 1$)	$\mathbb{E}[Y_{i0}(0) \mid G_i = 1]$	$\mathbb{E}[Y_{i1}(1) \mid G_i = 1]$
Control Group ($G_i = 0$)	$\mathbb{E}[Y_{i0}(0) \mid G_i = 0]$	$\mathbb{E}[Y_{i1}(0) \mid G_i = 0]$

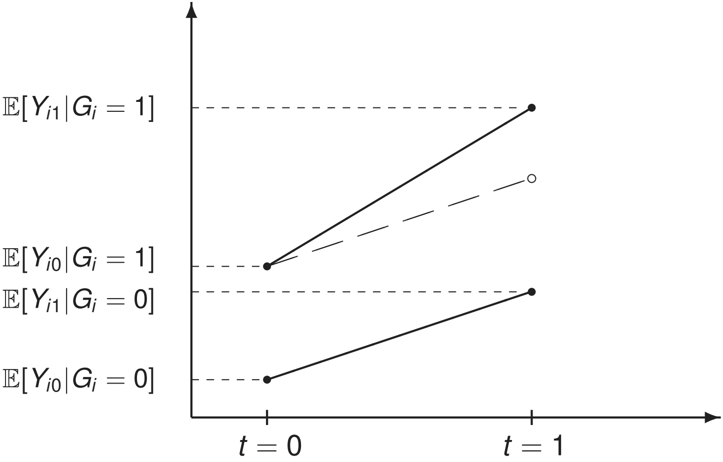
Control Strategy: Difference-in-Differences (DD)

- Use: $\left\{\mathbb{E}[Y_{i1} | G_i = 1] - \mathbb{E}[Y_{i1} | G_i = 0]\right\} - \left\{\mathbb{E}[Y_{i0} | G_i = 1] - \mathbb{E}[Y_{i0} | G_i = 0]\right\}$
- Assumes: $\mathbb{E}[Y_{i1}(0) - Y_{i0}(0) | G_i = 1] = \mathbb{E}[Y_{i1}(0) - Y_{i0}(0) | G_i = 0]$ (Parallel trends)

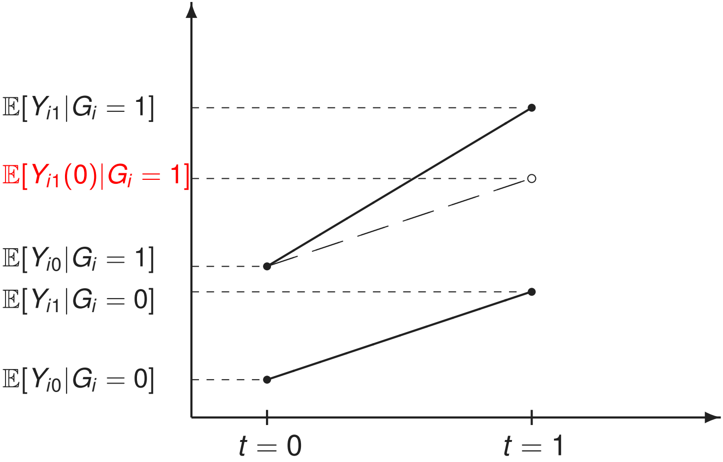
Graphical Representation: Difference-in-Differences



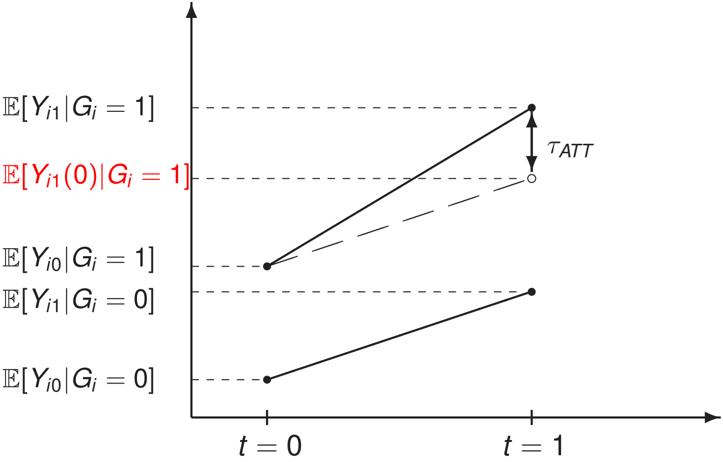
Graphical Representation: Difference-in-Differences



Graphical Representation: Difference-in-Differences



Graphical Representation: Difference-in-Differences



Identification with Difference-in-Differences

Under the parallel trends assumption:

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

The ATT can be nonparametrically identified as:

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

Identification with Difference-in-Differences

Under the parallel trends assumption:

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

The ATT can be nonparametrically identified as:

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

Proof:

$$\begin{aligned} & \{\mathbb{E}[Y_{i1} | G_i = 1] - \mathbb{E}[Y_{i1} | G_i = 0]\} - \{\mathbb{E}[Y_{i0} | G_i = 1] - \mathbb{E}[Y_{i0} | G_i = 0]\} \\ = \; & \{\mathbb{E}[Y_{i1}(1) | G_i = 1] - \mathbb{E}[Y_{i1}(0) | G_i = 0]\} - \{\mathbb{E}[Y_{i0}(0) | G_i = 1] - \mathbb{E}[Y_{i0}(0) | G_i = 0]\} \\ = \; & \underbrace{\mathbb{E}[Y_{i1}(1) | G_i = 1] - \mathbb{E}[Y_{i1}(0) | G_i = 1]}_{= \tau_{ATT}} + \mathbb{E}[Y_{i1}(0) | G_i = 1] \\ & \qquad - \mathbb{E}[Y_{i1}(0) | G_i = 0] - \mathbb{E}[Y_{i0}(0) | G_i = 1] + \mathbb{E}[Y_{i0}(0) | G_i = 0] \\ = \; & \tau_{ATT} + \underbrace{\{\mathbb{E}[Y_{i1}(0) - Y_{i0}(0) | G_i = 1] - \mathbb{E}[Y_{i1}(0) - Y_{i0}(0) | G_i = 0]\}}_{= 0 \text{ under parallel trends}} \\ = \; & \tau_{ATT} \end{aligned}$$

Notes on the Parallel Trends Assumption

- What type of confounding does DD make us robust to?

Notes on the Parallel Trends Assumption

- What type of confounding does DD make us robust to?
  - Parallel trends is satisfied if unobserved confounding is time-invariant and additive
  - Parallel trends is violated if there is unobserved time-varying confounding

Notes on the Parallel Trends Assumption

- What type of confounding does DD make us robust to?
  - Parallel trends is satisfied if unobserved confounding is time-invariant and additive
  - Parallel trends is violated if there is unobserved time-varying confounding
- Parallel trends may be more plausible with pre-treatment covariates:

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

  This assumes parallel trends within strata
- Under the conditional parallel trends assumption, the ATT is identified as

$$\tau_{ATT} \;=\; \sum_x \Big[\{\mathbb{E}[Y_{i1} | G_i = 1, X_i = x] - \mathbb{E}[Y_{i1} | G_i = 0, X_i = x]\} - \{\mathbb{E}[Y_{i0} | G_i = 1, X_i = x] - \mathbb{E}[Y_{i0} | G_i = 0, X_i = x]\} \Big] \Pr(X_i = x \mid G_i = 1)$$

Notes on the Parallel Trends Assumption

- What type of confounding does DD make us robust to?
  - Parallel trends is satisfied if unobserved confounding is time-invariant and additive
  - Parallel trends is violated if there is unobserved time-varying confounding
- Parallel trends may be more plausible with pre-treatment covariates:

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

  This assumes parallel trends within strata

- Under the conditional parallel trends assumption, the ATT is identified as

$$\tau_{ATT} \; = \; \sum_x \Big[\{\mathbb{E}[Y_{i1} | G_i = 1, X_i = x] - \mathbb{E}[Y_{i1} | G_i = 0, X_i = x]\} - \{\mathbb{E}[Y_{i0} | G_i = 1, X_i = x] - \mathbb{E}[Y_{i0} | G_i = 0, X_i = x]\} \Big] \Pr(X_i = x \mid G_i = 1)$$

- Note the parallel trends assumption is *not invariant to nonlinear transformation* of the outcome scale
- For example, parallel trends in $Y_{it}(z)$ implies non-parallel trends in $\log Y_{it}(z)$ and vice versa

Plug-in Estimation for Panel Data

Estimand:

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

Plug-in Estimation for Panel Data

Estimand:

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

A plug-in estimator ("difference in difference-in-means"):

$$\left\{\frac{1}{N_1}\sum_{i=1}^{N} G_i Y_{i1} - \frac{1}{N_0}\sum_{i=1}^{N}(1 - G_i) Y_{i1}\right\} - \left\{\frac{1}{N_1}\sum_{i=1}^{N} G_i Y_{i0} - \frac{1}{N_0}\sum_{i=1}^{N}(1 - G_i) Y_{i0}\right\}$$

$$= \left\{\frac{1}{N_1}\sum_{i=1}^{N} G_i\{Y_{i1} - Y_{i0}\} - \frac{1}{N_0}\sum_{i=1}^{N}(1 - G_i)\{Y_{i1} - Y_{i0}\}\right\},$$

where N_1 and N_0 are the number of treated and control units respectively

Plug-in Estimation for Panel Data

Estimand:

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

A plug-in estimator ("difference in difference-in-means"):

$$\left\{\frac{1}{N_1}\sum_{i=1}^{N} G_i Y_{i1} - \frac{1}{N_0}\sum_{i=1}^{N}(1 - G_i) Y_{i1}\right\} - \left\{\frac{1}{N_1}\sum_{i=1}^{N} G_i Y_{i0} - \frac{1}{N_0}\sum_{i=1}^{N}(1 - G_i) Y_{i0}\right\}$$

$$= \left\{\frac{1}{N_1}\sum_{i=1}^{N} G_i \{Y_{i1} - Y_{i0}\} - \frac{1}{N_0}\sum_{i=1}^{N}(1 - G_i)\{Y_{i1} - Y_{i0}\}\right\},$$

where N_1 and N_0 are the number of treated and control units respectively

Standard errors can be estimated by extending the diff-in-means variance formula using the same logic, assuming no clustering

Example: Card and Krueger

Variable	PA (i)	NJ (ii)	Difference, NJ − PA (iii)
	Stores by state		
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)

Plugin-Estimation for Repeated Cross Sections

Repeated cross-sectional data require slight change in notation:

- Period indicator is now a variable: $T_i \in \{0, 1\}$
- Estimand: $\tau_{ATT} = \mathbb{E}[Y_i(1) - Y_i(0) \mid G_i = 1, T_i = 1]$
- Identified as: $$\tau_{ATT} = \mathbb{E}[Y_i \mid G_i = 1, T_i = 1] - \mathbb{E}[Y_i \mid G_i = 0, T_i = 1] - \{\mathbb{E}[Y_i \mid G_i = 1, T_i = 0] - \mathbb{E}[Y_i \mid G_i = 0, T_i = 0]\}$$
- N now refers to the size of the pooled sample

Plugin-Estimation for Repeated Cross Sections

Repeated cross-sectional data require slight change in notation:

- Period indicator is now a variable: $T_i \in \{0, 1\}$
- Estimand: $\tau_{ATT} = \mathbb{E}[Y_i(1) - Y_i(0) \mid G_i = 1, T_i = 1]$
- Identified as: $\tau_{ATT} = \mathbb{E}[Y_i \mid G_i = 1, T_i = 1] - \mathbb{E}[Y_i \mid G_i = 0, T_i = 1] - \{\mathbb{E}[Y_i \mid G_i = 1, T_i = 0] - \mathbb{E}[Y_i \mid G_i = 0, T_i = 0]\}$
- N now refers to the size of the pooled sample

The plug-in estimator is then written as:

$$\hat{\tau}_{ATT} = \left\{ \frac{\sum_{i=1}^N G_i T_i Y_i}{\sum_{i=1}^N G_i T_i} - \frac{\sum_{i=1}^N (1 - G_i) T_i Y_i}{\sum_{i=1}^N (1 - G_i) T_i} \right\} - \left\{ \frac{\sum_{i=1}^N G_i (1 - T_i) Y_i}{\sum_{i=1}^N G_i (1 - T_i)} - \frac{\sum_{i=1}^N (1 - G_i)(1 - T_i) Y_i}{\sum_{i=1}^N (1 - G_i)(1 - T_i)} \right\}$$

Covariates X_i can be incorporated via subclassification

Regression Estimator for Repeated Cross Sections

Because G_i and T_i are both binary, the same estimator can be calculated via regression:

$$\hat{Y}_i = \hat{\mu} + \hat{\gamma} G_i + \hat{\delta} T_i + \hat{\tau} G_i T_i$$

where $\hat{\mu}$, $\hat{\gamma}$, $\hat{\delta}$ and $\hat{\tau}$ are OLS regression estimates

Regression Estimator for Repeated Cross Sections

Because G_i and T_i are both binary, the same estimator can be calculated via regression:

$$\hat{Y}_i = \hat{\mu} + \hat{\gamma} G_i + \hat{\delta} T_i + \hat{\tau} G_i T_i$$

where $\hat{\mu}$, $\hat{\gamma}$, $\hat{\delta}$ and $\hat{\tau}$ are OLS regression estimates

Easy to show that $\hat{\tau} = \hat{\tau}_{ATT}$:

	After ($T_i = 1$)	Before ($T_i = 0$)	After - Before
Treated $G_i = 1$	$\hat{\mu} + \hat{\gamma} + \hat{\delta} + \hat{\tau}$	$\hat{\mu} + \hat{\gamma}$	$\hat{\delta} + \hat{\tau}$
Control $G_i = 0$	$\hat{\mu} + \hat{\delta}$	$\hat{\mu}$	$\hat{\delta}$
Treated - Control	$\hat{\gamma} + \hat{\tau}$	$\hat{\gamma}$	$\hat{\tau}$

Regression Estimator for Repeated Cross Sections

Because G_i and T_i are both binary, the same estimator can be calculated via regression:

$$\hat{Y}_i = \hat{\mu} + \hat{\gamma} G_i + \hat{\delta} T_i + \hat{\tau} G_i T_i$$

where $\hat{\mu}$, $\hat{\gamma}$, $\hat{\delta}$ and $\hat{\tau}$ are OLS regression estimates

Easy to show that $\hat{\tau} = \hat{\tau}_{ATT}$:

	After ($T_i = 1$)	Before ($T_i = 0$)	After - Before
Treated $G_i = 1$	$\hat{\mu} + \hat{\gamma} + \hat{\delta} + \hat{\tau}$	$\hat{\mu} + \hat{\gamma}$	$\hat{\delta} + \hat{\tau}$
Control $G_i = 0$	$\hat{\mu} + \hat{\delta}$	$\hat{\mu}$	$\hat{\delta}$
Treated - Control	$\hat{\gamma} + \hat{\tau}$	$\hat{\gamma}$	$\hat{\tau}$

- Covariates (X_i) can be added to the right-hand side, with the risk of possible misspecification bias
- Don't include X_i that can be affected by the treatment! (post-ttt bias)
- Cluster standard errors at the level the treatment is assigned (Card & Krueger get this wrong)

Regression Estimator for Panel Data

For panel data, consider an additive linear model for potential outcomes:

$$Y_{it}(z) = \alpha_i + \gamma t + \tau z + \varepsilon_{it}$$

where α_i is a time-invariant unobserved effect for unit i that may be correlated with treatment.

Regression Estimator for Panel Data

For panel data, consider an additive linear model for potential outcomes:

$$Y_{it}(z) \;=\; \alpha_i + \gamma t + \tau z + \varepsilon_{it}$$

where α_i is a time-invariant unobserved effect for unit i that may be correlated with treatment.

We can show:

- $\tau = \tau_{ATE} = \tau_{ATT}$
- Parallel trends imply:

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

$$\iff \mathbb{E}[\varepsilon_{i1} - \varepsilon_{i0} \mid G_i = d] \;=\; 0 \quad \text{for} \quad d \in \{0, 1\}$$

Regression Estimator for Panel Data

For panel data, consider an additive linear model for potential outcomes:

$$Y_{it}(z) \ = \ \alpha_i + \gamma t + \tau z + \varepsilon_{it}$$

where α_i is a time-invariant unobserved effect for unit i that may be correlated with treatment.

We can show:

- $\tau = \tau_{ATE} = \tau_{ATT}$
- Parallel trends imply:

$$\mathbb{E}[Y_{i1}(0) - Y_{i0}(0) | G_i = 1] \ = \ \mathbb{E}[Y_{i1}(0) - Y_{i0}(0) | G_i = 0]$$
$$\iff \ \mathbb{E}[\varepsilon_{i1} - \varepsilon_{i0} \mid G_i = d] \ = \ 0 \quad \text{for} \quad d \in \{0, 1\}$$

Therefore, the first-differenced regression of $Y_{i1} - Y_{i0}$ on G_i can unbiasedly estimate $\tau_{ATT} = \tau_{ATE}$

Notice that panel data allow for *unit-level* unobserved confounding unlike the repeated cross-section case, but it must be additive and time-invariant

Can include time-varying covariates (X_{it}) with possible risk of post-ttt bias

Example: Minimum Wage and Employment

```
ck_data <- plm.data(ck_data, indexes = c("ID", "postperiod"))
firstdiff.mod <- (plm(emptot ~ postperiod * nj,
                      data = ck_data,
                      model = "fd"))
summary(firstdiff.mod)

Oneway (individual) effect First-Difference Model

Call:
plm(formula = emptot ~ postperiod * nj, data = ck_data, model = "fd")

Unbalanced Panel: n=410, T=1-2, N=794

Coefficients :
              Estimate Std. Error t-value Pr(>|t|)
(intercept)    -2.2833     1.0355 -2.2050  0.02805 *
postperiod1:nj  2.7500     1.1544  2.3823  0.01769 *
```

Parallel Trends Violations: Selection

The parallel trends assumption can be violated for various reasons

1. **Selection and targeting**: Treatment assignment may depend on time-varying factors

Examples:

- 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).

More Parallel Trends Violations

2. **Compositional Differences Across Time**

   - In repeated cross-sections, the composition of the sample may change between periods, i.e. due to migration.
   - This may confound any DD estimate since “effect” may be attributable to change in population.

3. **Long-term effects versus reliability**

   - Parallel trends assumption most likely to hold over shorter time-period
   - In the long run, many things may happen that could confound effect of treatment.

Functional form dependence

4. **Functional form dependence**: Magnitude or even sign of the DD effect may be sensitive to the functional form, when average outcomes for controls and treated are very different at baseline

- Training program for the young
  - Employment for the young increases from 20% to 30%
  - Employment for the old increases from 5% to 10%
  - Positive DD effect: (30-20) - (10 -5) = 5% increase

Functional form dependence

4. **Functional form dependence**: Magnitude or even sign of the DD effect may be sensitive to the functional form, when average outcomes for controls and treated are very different at baseline

- Training program for the young
    - Employment for the young increases from 20% to 30%
    - Employment for the old increases from 5% to 10%
    - Positive DD effect: (30-20) - (10 -5) = 5% increase
    - But if you consider log changes in employment, the DD is

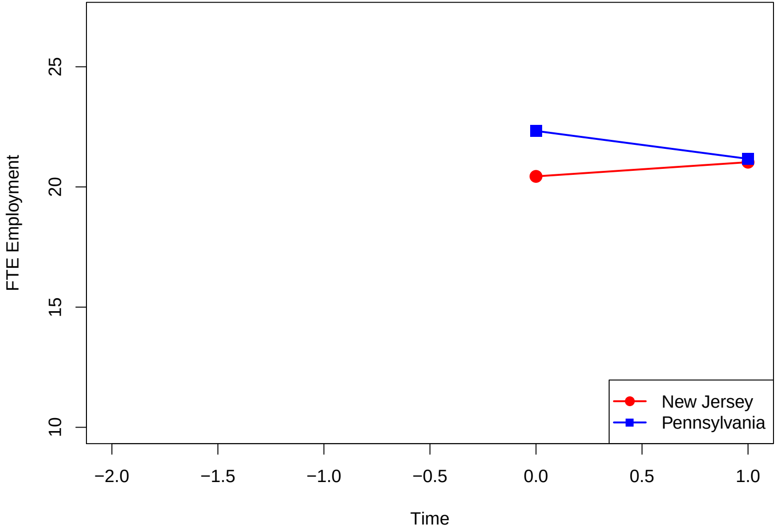
$$[log(30) - log(20)] - [log(10) - log(5)] = log(1.5) - log(2) < 0$$

- DD estimates may be more reliable if treated and control are more similar at baseline

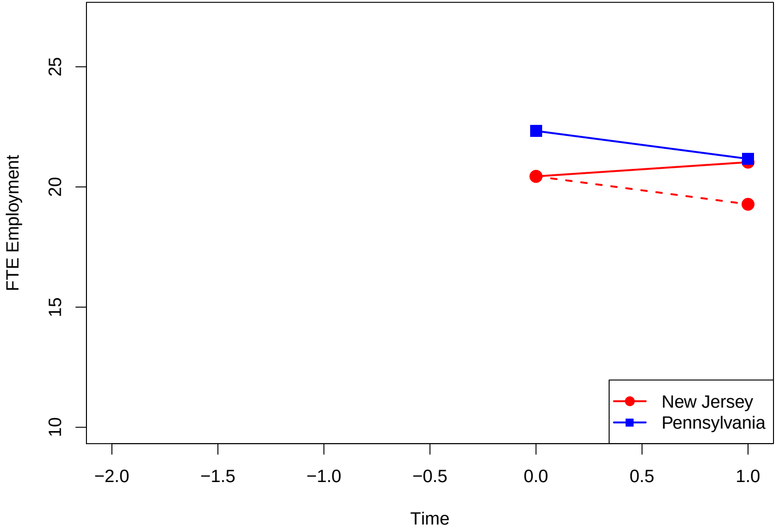
Diagnostics for Parallel Trends

1. Pre-treatment trends in the outcome:
   - Check if the trends are parallel in the pre-treatment periods
   - Requires data on multiple pre-treatment periods (the more the better)
   - Note this is only diagnostics, not a direct test of the assumption!
2. Placebo test using previous periods:
   - Suppose DD with time periods t_1, t_2, t_3, where treatment occurs in t_3
   - Exclude data from t_3, assign t_2 as “placebo" treatment period, and re-estimate DD
3. Placebo test using alternative groups:
   - Re-code some control groups as treated
   - Re-estimate DD with the placebo treated units & without actual treated units
4. Placebo outcomes:
   - Find outcomes that, theoretically, should be unaffected by the treatment, but might
   - Re-estimate DD on these outcomes

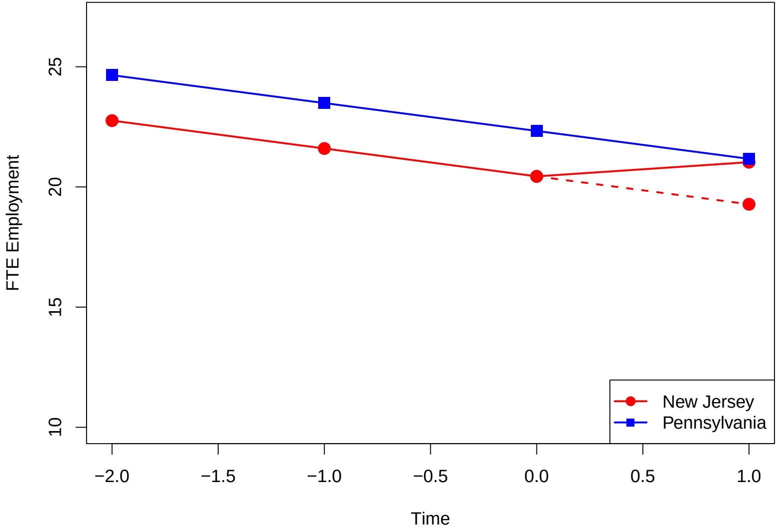
Checking for Pre-Treatment Trends



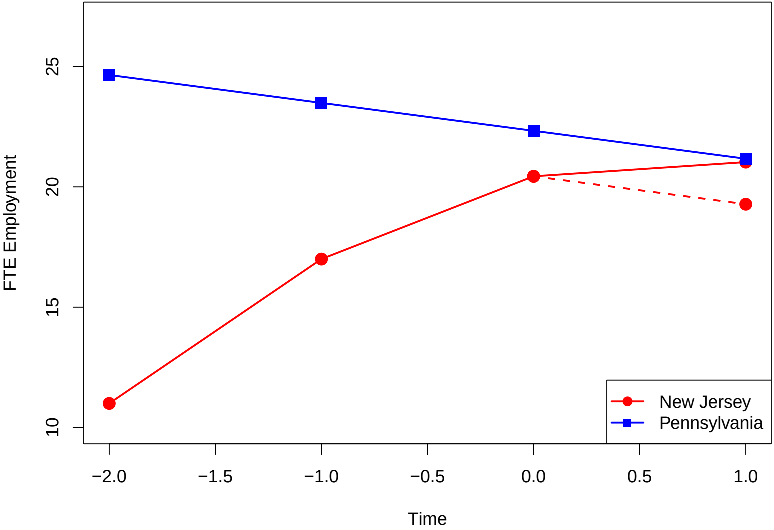
Checking for Pre-Treatment Trends



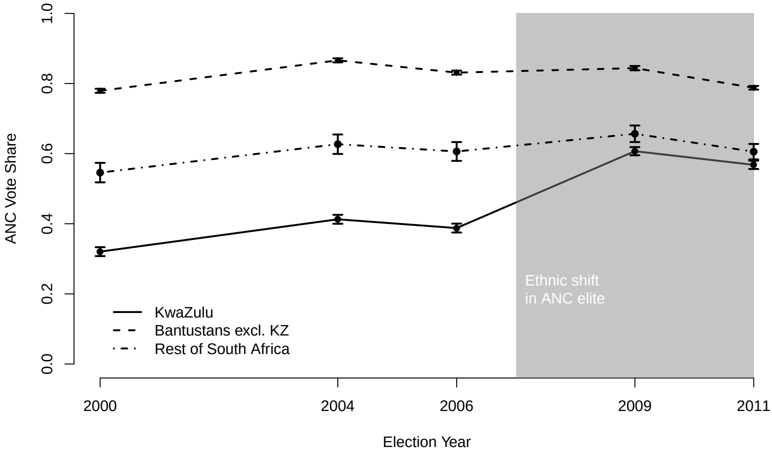
Checking for Pre-Treatment Trends ☺



Checking for Pre-Treatment Trends ☹



Chiefs and Voting (de Kadt and Larreguy 2015)



Extension: Triple Differences

Triple differences (or "difference in differences in differences"): Use a placebo DD to make parallel trends more plausible

Extension: Triple Differences

Triple differences (or "difference in differences in differences"): Use a placebo DD to make parallel trends more plausible

Example: In state A, all 9th-grade girls are given a free bicycle.

- 1st DD: $G_i \in \{\text{girl}, \text{boy}\}$ and $T_i \in \{9, 10\}$ in state A
- 2nd DD: $G_i \in \{\text{girl}, \text{boy}\}$ and $T_i \in \{9, 10\}$ in state B (placebo)
- The DDD estimator:

$$\begin{aligned}\hat{\tau}_{DDD} &= \hat{\mathbb{E}}[Y_i|\text{girl}, 10, A] - \hat{\mathbb{E}}[Y_i|\text{girl}, 9, A] \\ &\qquad - \left\{\hat{\mathbb{E}}[Y_i|\text{boy}, 10, A] - \hat{\mathbb{E}}[Y_i|\text{boy}, 9, A]\right\} \\ &\quad - \Big(\hat{\mathbb{E}}[Y_i|\text{girl}, 10, B] - \hat{\mathbb{E}}[Y_i|\text{girl}, 9, B] \\ &\qquad - \left\{\hat{\mathbb{E}}[Y_i|\text{boy}, 10, B] - \hat{\mathbb{E}}[Y_i|\text{boy}, 9, B]\right\}\Big)\end{aligned}$$

- Can use regression with a triple interaction
- May eliminate time-varying confounding that are common in states A and B (e.g. girls change more from grade 9 to 10 than boys)

Muralidharan and Prakash (2013)

	Dependent variable: Log (9th Grade Enrollment)
	PANEL A: Testing Parallel Trends for the Difference-in-Difference (DD)
Female Dummy×Year	**0.0518*** **
	(0.00)
Female Dummy	-0.870***
	(0.06)
Year (time trend)	0.0852***
	(0.01)
Constant	4.235***
	(0.05)
Observations	20,266
R-squared	0.167

Muralidharan and Prakash (2013)

PANEL B: Testing Parallel Trends for the Triple Difference (DDD)	
Female Dummy×Year×Bihar dummy	**-0.0100**
	(0.01)
Female Dummy×Year	0.0618***
	(0.01)
Female Dummy×Bihar dummy	0.175
	(0.11)
Bihar dummy×Year	0.0290**
	(0.01)
Female dummy	-1.045***
	(0.09)
Year (time trend)	0.0562***
	(0.01)
Bihar dummy	-0.123
	(0.12)
Constant	4.358***
	(0.11)
Observations	22,279
R-squared	0.171

Summary and Remarks

- DD: An extremely popular strategy when there is longitudinal data (panel or repeated cross-sections) and the treatment is one-shot
- Parallel trends $=$ a form of ignorability assumption, i.e., unobserved confounding must be additive and time-invariant

- An important generalization of DD is fixed effects regression with both time and unit fixed effects:

$$y_{it} = \boldsymbol{x}_{it}^{\top} \boldsymbol{b} + \alpha_i + \eta_t + \varepsilon_{it}$$

  where $\boldsymbol{x}_{it}$ includes the treatment variable Z_{it}

- Other extensions include:
  - Nonlinear and semi-/nonparametric DD
  - Difference in discontinuities: Combine RD with DD
  - Matching before DD: Make estimates more robust to functional form misspecification