150C Causal Inference

Difference in Differences

Jonathan Mummolo

Stanford

Problem

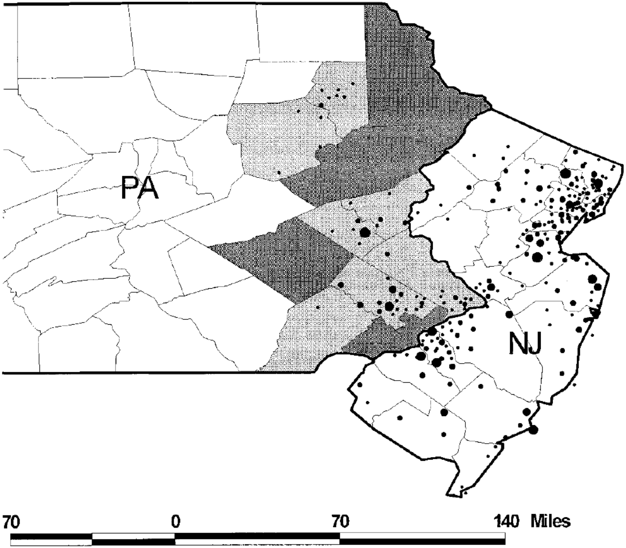
Often there are reasons to believe that treated and untreated units differ in unobservable characteristics that are associated with potential outcomes even after controlling for differences in observed characteristics.

In such cases, treated and untreated units are not directly comparable. What can we do then?

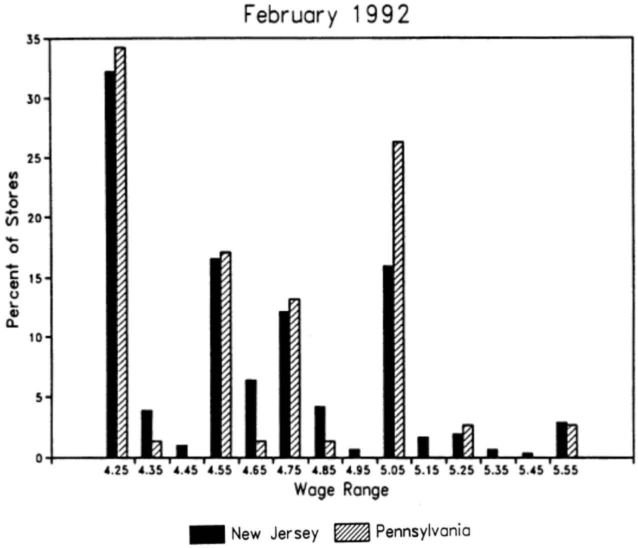
Example: Minimum wage laws and employment

- Do higher minimum wages decrease low-wage 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 months after increase

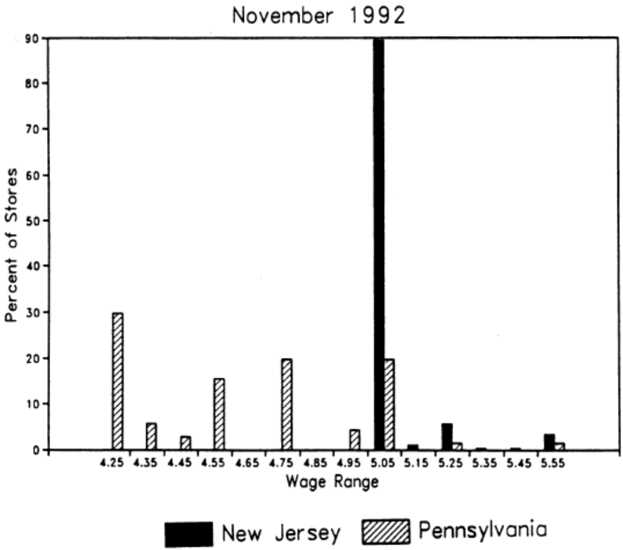
Locations of Restaurants (Card and Krueger 2000)



Wages Before Rise in Minimum Wage



Wages After Rise in Minimum Wage



Outline

Two Groups and Two Periods

Definition

Two groups:

- $D = 1$ Treated units
- $D = 0$ Control units

Two periods:

- $T = 0$ Pre-Treatment period
- $T = 1$ Post-Treatment period

Potential outcomes $Y_d(t)$:

- $Y_{1i}(t)$ potential outcome unit i attains in period t when treated between t and $t-1$
- $Y_{0i}(t)$ potential outcome unit i attains in period t with control between t and $t-1$

Definition

Causal effect for unit i at time t is

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

Observed outcomes $Y_i(t)$ are realized as

- $Y_i(t) = Y_{0i}(t) \cdot (1 - D_i(t)) + Y_{1i}(t) \cdot D_i(t)$

Fundamental problem of causal inference:

- If D only occurs when $t = 1$, then for the treatment group ($D_i = 1$) we have: $Y_i(1) = Y_{0i}(1) \cdot (1 - D_i(1)) + Y_{1i}(1) \cdot D_i(1)$
- $= Y_{1i}(1)$ (i.e. we don't get to see $Y_{0i}(1)$ for the treated)

Estimand (ATT)

Focus on estimating the average effect of the treatment on the treated: $\tau_{ATT} = E[Y_{1i}(1) - Y_{0i}(1)|D = 1]$

Estimand (ATT)

$\tau_{ATT} = E[Y_{1i}(1) - Y_{0i}(1)|D = 1]$

	Post-Period (T=1)	**Pre-Period (T=0)**
Treated D_i=1	$E[Y_{1i}(1)\|D_i = 1]$	$E[Y_{0i}(0)\|D_i = 1]$
Control D_i=0	$E[Y_{0i}(1)\|D_i = 0]$	$E[Y_{0i}(0)\|D_i = 0]$

Problem

Missing potential outcome: $E[Y_{0i}(1)|D = 1]$, *ie. what is the average post-period outcome for the treated in the absence of the treatment?*

Estimand (ATT)

$\tau_{ATT} = E[Y_{1i}(1) - Y_{0i}(1)|D = 1]$

	Post-Period (T=1)	**Pre-Period (T=0)**
Treated D_i=1	$E[Y_{1i}(1)\|D_i = 1]$	$E[Y_{0i}(0)\|D_i = 1]$
Control D_i=0	$E[Y_{0i}(1)\|D_i = 0]$	$E[Y_{0i}(0)\|D_i = 0]$

Control Strategy: Before vs. After

- Use: $E[Y_i(1)|D_i = 1] - E[Y_i(0)|D_i = 1]$

Estimand (ATT)

$\tau_{ATT} = E[Y_{1i}(1) - Y_{0i}(1)|D = 1]$

	Post-Period (T=1)	**Pre-Period (T=0)**
Treated D_i=1	$E[Y_{1i}(1)\|D_i = 1]$	$E[Y_{0i}(0)\|D_i = 1]$
Control D_i=0	$E[Y_{0i}(1)\|D_i = 0]$	$E[Y_{0i}(0)\|D_i = 0]$

Control Strategy: Before vs. After

- Use: $E[Y_i(1)|D_i = 1] - E[Y_i(0)|D_i = 1]$
- Assumes: $E[Y_{0i}(1)|D_i = 1]$= $E[Y_{0i}(0)|D_i = 1]$

Estimand (ATT)

$\tau_{ATT} = E[Y_{1i}(1) - Y_{0i}(1)|D = 1]$

	Post-Period (T=1)	**Pre-Period (T=0)**
Treated D_i=1	$E[Y_{1i}(1)\|D_i = 1]$	$E[Y_{0i}(0)\|D_i = 1]$
Control D_i=0	$E[Y_{0i}(1)\|D_i = 0]$	$E[Y_{0i}(0)\|D_i = 0]$

Control Strategy: Treated-Control Comparison in Post-Period

- Use: $E[Y_i(1)|D_i = 1] - E[Y_i(1)|D_i = 0]$

Estimand (ATT)

$\tau_{ATT} = E[Y_{1i}(1) - Y_{0i}(1)|D = 1]$

	Post-Period (T=1)	**Pre-Period (T=0)**
Treated D_i=1	$E[Y_{1i}(1)\|D_i = 1]$	$E[Y_{0i}(0)\|D_i = 1]$
Control D_i=0	$E[Y_{0i}(1)\|D_i = 0]$	$E[Y_{0i}(0)\|D_i = 0]$

Control Strategy: Treated-Control Comparison in Post-Period

- Use: $E[Y_i(1)|D_i = 1] - E[Y_i(1)|D_i = 0]$
- Assumes: $E[Y_{0i}(1)|D_i = 1] = E[Y_{0i}(1)|D_i = 0]$

Estimand (ATT)

$\tau_{ATT} = E[Y_{1i}(1) - Y_{0i}(1)|D = 1]$

	Post-Period (T=1)	**Pre-Period (T=0)**
Treated D_i=1	$E[Y_{1i}(1)\|D_i = 1]$	$E[Y_{0i}(0)\|D_i = 1]$
Control D_i=0	$E[Y_{0i}(1)\|D_i = 0]$	$E[Y_{0i}(0)\|D_i = 0]$

Control Strategy: Difference-in-Differences (DD)

- Use:
$$\Big\{E[Y_i(1)|D_i = 1] - E[Y_i(1)|D_i = 0]\Big\} - \Big\{E[Y_i(0)|D_i = 1] - E[Y_i(0)|D_i = 0]\Big\}$$

Estimand (ATT)

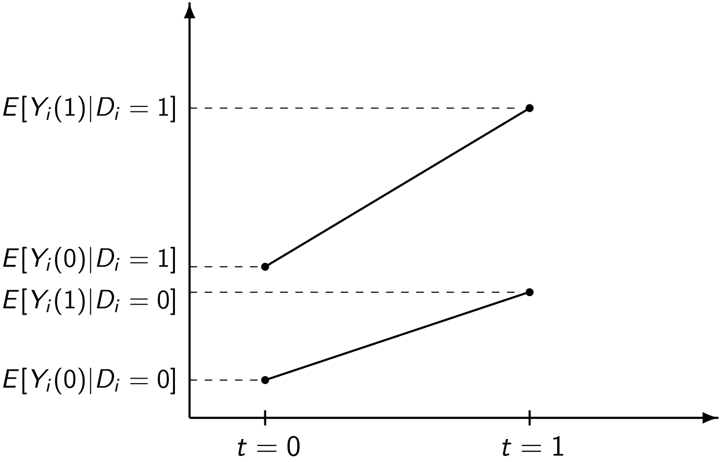
$\tau_{ATT} = E[Y_{1i}(1) - Y_{0i}(1)|D = 1]$

	Post-Period (T=1)	**Pre-Period (T=0)**
Treated D_i=1	$E[Y_{1i}(1)\|D_i = 1]$	$E[Y_{0i}(0)\|D_i = 1]$
Control D_i=0	$E[Y_{0i}(1)\|D_i = 0]$	$E[Y_{0i}(0)\|D_i = 0]$

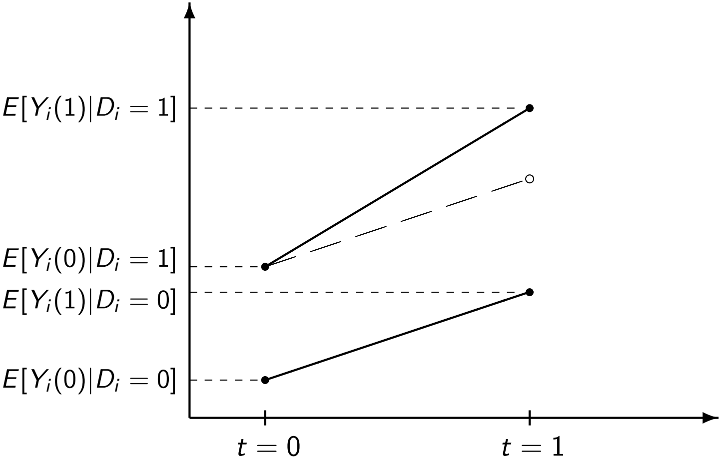
Control Strategy: Difference-in-Differences (DD)

- Use:
$$\Big\{E[Y_i(1)|D_i = 1] - E[Y_i(1)|D_i = 0]\Big\} -$$
$$\Big\{E[Y_i(0)|D_i = 1] - E[Y_i(0)|D_i = 0]\Big\}$$
- Assumes: $E[Y_{0i}(1) - Y_{0i}(0)|D_i = 1] = E[Y_{0i}(1) - Y_{0i}(0)|D_i = 0]$

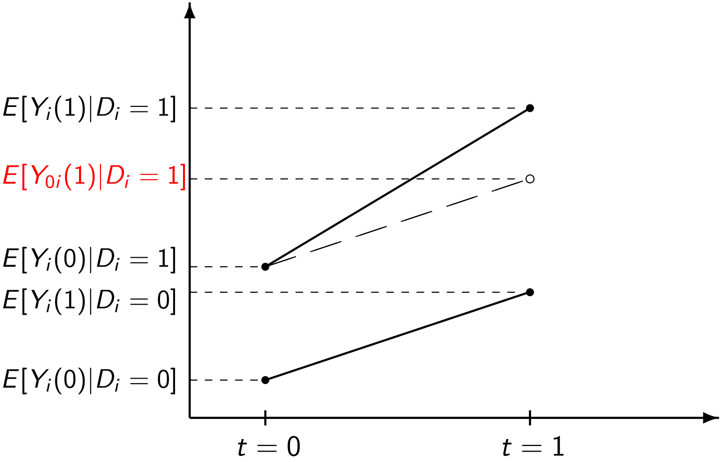
Graphical Representation: Difference-in-Differences



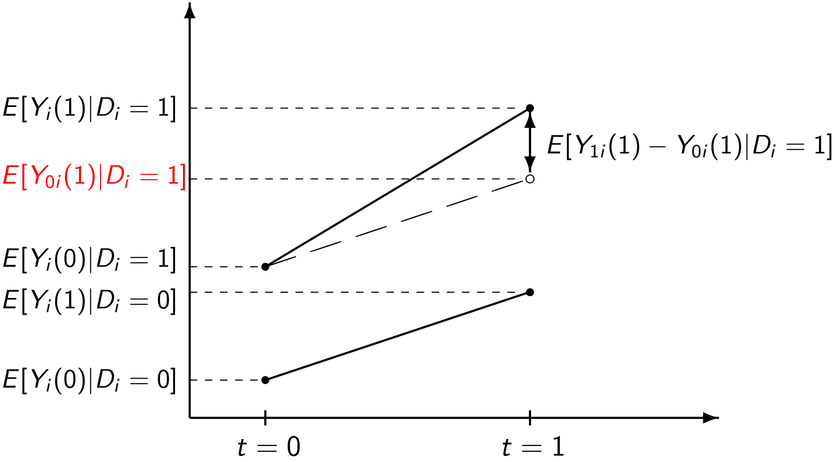
Graphical Representation: Difference-in-Differences



Graphical Representation: Difference-in-Differences



Graphical Representation: Difference-in-Differences



Outline

Identification with Difference-in-Differences

Identification Assumption (parallel trends)

$E[Y_0(1) - Y_0(0)|D = 1] = E[Y_0(1) - Y_0(0)|D = 0]$

Identification Result

Given parallel trends the ATT is identified as:

$$\begin{aligned} E[Y_1(1) - Y_0(1)|D = 1] &= \Big\{E[Y(1)|D = 1] - E[Y(1)|D = 0]\Big\} \\ &- \Big\{E[Y(0)|D = 1] - E[Y(0)|D = 0]\Big\} \end{aligned}$$

Identification Assumption (parallel trends)

$$E[Y_0(1) - Y_0(0)|D = 1] = E[Y_0(1) - Y_0(0)|D = 0]$$

Proof.

Note that the identification assumption implies
$E[Y_0(1)|D = 0] = E[Y_0(1)|D = 1] - E[Y_0(0)|D = 1] + E[Y_0(0)|D = 0]$
plugging in we get

$$\begin{aligned}
& \{E[Y(1)|D = 1] - E[Y(1)|D = 0]\} - \{E[Y(0)|D = 1] - E[Y(0)|D = 0]\} \\
= & \{E[Y_1(1)|D = 1] - E[Y_0(1)|D = 0]\} - \{E[Y_0(0)|D = 1] - E[Y_0(0)|D = 0]\} \\
= & \{E[Y_1(1)|D = 1] - (E[Y_0(1)|D = 1] - E[Y_0(0)|D = 1] + E[Y_0(0)|D = 0])\} \\
- & \{E[Y_0(0)|D = 1] - E[Y_0(0)|D = 0]\} \\
= & E[Y_1(1) - Y_0(1)|D = 1] + \{E[Y_0(0)|D = 1] - E[Y_0(0)|D = 0]\} \\
- & \{E[Y_0(0)|D = 1] - E[Y_0(0)|D = 0]\} \\
= & E[Y_1(1) - Y_0(1)|D = 1]
\end{aligned}$$

□

Outline

Estimand (ATT)

$$\begin{aligned} E[Y_1(1) - Y_0(1)|D = 1] &= \Big\{E[Y(1)|D = 1] - E[Y(1)|D = 0]\Big\} \\ &- \Big\{E[Y(0)|D = 1] - E[Y(0)|D = 0]\Big\} \end{aligned}$$

Estimator (Sample Means: Panel)

$$\left\{\frac{1}{N_1}\sum_{D_i=1} Y_i(1) - \frac{1}{N_0}\sum_{D_i=0} Y_i(1)\right\} - \left\{\frac{1}{N_1}\sum_{D_i=1} Y_i(0) - \frac{1}{N_0}\sum_{D_i=0} Y_i(0)\right\}$$

$$= \left\{\frac{1}{N_1}\sum_{D_i=1}\{Y_i(1) - Y_i(0)\} - \frac{1}{N_0}\sum_{D_i=0}\{Y_i(1) - Y_i(0)\}\right\},$$

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

Sample Means: Minimum wage laws and employment

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)

Estimator (Sample Means: Repeated Cross-Sections)

Let $\{Y_i, D_i, T_i\}_{i=1}^{n}$ *be the pooled sample (the two different cross-sections merged) where* T *is a random variable that indicates the period (0 or 1) in which the individual is observed.*

The difference-in-differences estimator is given by:

$$\left\{ \frac{\sum D_i \cdot T_i \cdot Y_i}{\sum D_i \cdot T_i} - \frac{\sum (1 - D_i) \cdot T_i \cdot Y_i}{\sum (1 - D_i) \cdot T_i} \right\} - \left\{ \frac{\sum D_i \cdot (1 - T_i) \cdot Y_i}{\sum D_i \cdot (1 - T_i)} - \frac{\sum (1 - D_i) \cdot (1 - T_i) \cdot Y_i}{\sum (1 - D_i) \cdot (1 - T_i)} \right\}$$

Estimator (Regression: Repeated Cross-Sections)

Alternatively, the same estimator can be obtained using regression techniques. Consider the linear model:

$$Y = \mu + \gamma \cdot D + \delta \cdot T + \tau \cdot (D \cdot T) + \varepsilon,$$

where $E[\varepsilon|D, T] = 0$.

Easy to show that τ estimates the DD effect:

$$\begin{aligned} \tau = \ & \{E[Y|D=1, T=1] - E[Y|D=0, T=1]\} \\ - \ & \{E[Y|D=1, T=0] - E[Y|D=0, T=0]\} \end{aligned}$$

Estimator (Regression: Repeated Cross-Sections)

Alternatively, the same estimator can be obtained using regression techniques. Consider the linear model:

$$Y = \mu + \gamma \cdot D + \delta \cdot T + \tau \cdot (D \cdot T) + \varepsilon,$$

where $E[\varepsilon | D, T] = 0$.

	After (T=1)	**Before (T=0)**	**After - Before**
Treated D=1	$\mu + \gamma + \delta + \tau$	$\mu + \gamma$	$\delta + \tau$
Control D=0	$\mu + \delta$	μ	δ
Treated - Control	$\gamma + \tau$	γ	τ

Regression: Minimum wage laws and employment

```
> d <- read.dta("CK1994_longformat.dta",convert.
    factors = FALSE)
> head(d[, c('ID', 'nj', 'postperiod', 'emptot')])
  ID nj postperiod emptot
1  1  0          0  40.50
2  1  0          1  24.00
3  2  0          0  13.75
4  2  0          1  11.50
5  3  0          0   8.50
6  3  0          1  10.50
```

```
with(d,
 (
  mean(emptot[nj == 1 & postperiod == 1], na.rm = TRUE
      ) -
  mean(emptot[nj == 1 & postperiod == 0], na.rm = TRUE
      )
  )  -
  (mean(emptot[nj == 0 & postperiod == 1], na.rm =
      TRUE) -
  mean(emptot[nj == 0 & postperiod == 0], na.rm = TRUE
      )
  )
 )
[1] 2.753606
```

Regression: Minimum wage laws and employment

```
> ols <- lm(emptot ~ postperiod * nj, data = d)
> coeftest(ols)

t test of coefficients:

              Estimate Std. Error t value Pr(>|t|)
(Intercept)    23.3312     1.0719 21.7668  < 2e-16 ***
postperiod     -2.1656     1.5159 -1.4286  0.15351
nj             -2.8918     1.1935 -2.4229  0.01562 *
postperiod:nj   2.7536     1.6884  1.6309  0.10331
```

Note: Should adjust standard errors to account for temporal dependence

Estimator (Regression: Repeated Cross-Sections)

Can use regression version of the DD estimator to include covariates:

$$Y = \mu + \gamma \cdot D + \delta \cdot T + \tau \cdot (D \cdot T) + X'\beta + \varepsilon.$$

- *introducing time-invariant X's is not helpful (they get differenced-out)*
- *be careful with time-varying X's: they are often affected by the treatment and may introduce endogeneity (e.g. price of meal)*
- *always correct standard errors to account for temporal dependence*

Can interact time-invariant covariates with the time indicator:

$$Y = \mu + \gamma \cdot D + \delta \cdot T + \alpha \cdot (D \cdot T) + X'\beta_0 + (T \cdot X')\beta_1 + \varepsilon$$

$\Rightarrow$ X is used to explain differences in trends.

Estimator (Regression: Panel Data)

With panel data we can estimate the difference-in-differences effect using a fixed effects regression with unit and period fixed effects:

$$Y_{it} = \mu + \gamma_i + \delta T + \tau D_{it} + X_{it}'\beta + \varepsilon_{it}$$

- *One intercept for each unit* γ_i
- D_{it} *coded as 1 for treated in post-period and 0 otherwise*

Or equivalently we can use regression with the dependent variable in first differences:

$$\Delta Y_i = \delta + \tau \cdot D_i + u_i,$$

where $\Delta Y_i = Y_i(1) - Y_i(0)$ *and* $u_i = \Delta\varepsilon_i$.

Regression: Minimum wage laws and employment

```
library(plm)
library(lmtest)

> d$Dit <- d$nj * d$postperiod

> d <- plm.data(d, indexes = c("ID", "postperiod"))

> did.reg <- plm(emptot ~ postperiod + Dit, data = d,
                          model = "within")

> coeftest(did.reg, vcov=function(x)
                 vcovHC(x, cluster="group", type="HC1"))

t test of coefficients:

            Estimate Std. Error t value Pr(>|t|)
postperiod1  -2.2833     1.2465 -1.8319  0.06775 .
Dit           2.7500     1.3359  2.0585  0.04022 *
```

Regression: Minimum wage laws and employment

```
> firstdiff.mod <- plm(emptot ~ postperiod * nj,
                                     data = d, model = "fd")
> coeftest(firstdiff.mod, vcov=function(x) vcovHC(x, type="HC0"))

t test of coefficients:

               Estimate Std. Error t value Pr(>|t|)
postperiod1     -2.2833     1.2465 -1.8319  0.06775 .
postperiod1:nj   2.7500     1.3359  2.0585  0.04022 *
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

Outline

Difference-in-Differences: Threats to Validity

1. Non-parallel dynamics
2. Compositional differences
3. Long-term effects versus reliability
4. Functional form dependence

Bias is a matter of degree. Small violations of the identification assumptions may not matter much as the bias may be rather small. However, biases can sometimes be so large that the estimates we get are completely wrong, even of the opposite sign of the true treatment effect.

Helpful to avoid overly strong causal claims for difference-in-differences estimates.

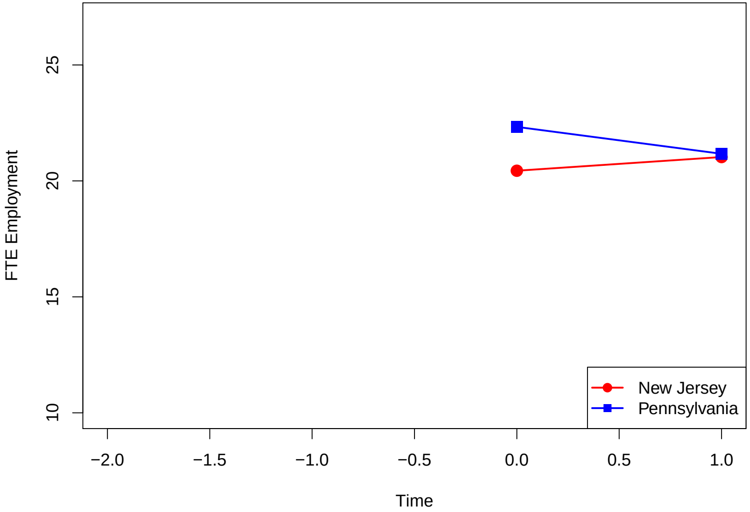
Difference-in-Differences: Threats to Validity

1. *Non-parallel dynamics*: Often treatments/programs are targeted based on pre-existing differences in outcomes.
   - "Ashenfelter dip": participants in training programs often experience a dip in earnings just before they enter the program (that may be *why* they participate). Since wages have a natural tendency to mean reversion, comparing wages of participants and non-participants using DD leads to an upward biased estimate of the program effect
   - Regional targeting: NGOs may target villages that appear most promising (or worst off)

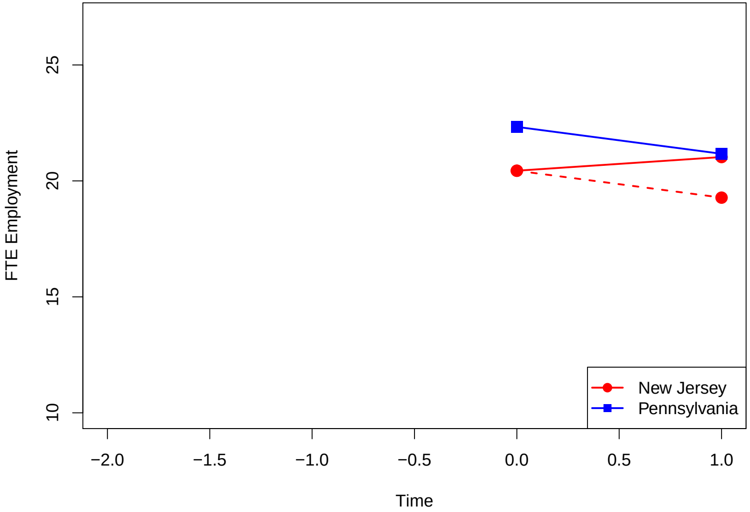
Checks for Difference-in-Differences Design

1. Falsification test using data for prior periods
2. Falsification test using data for alternative control group
3. Falsification test using alternative placebo outcome that is not supposed to be affected by the treatment

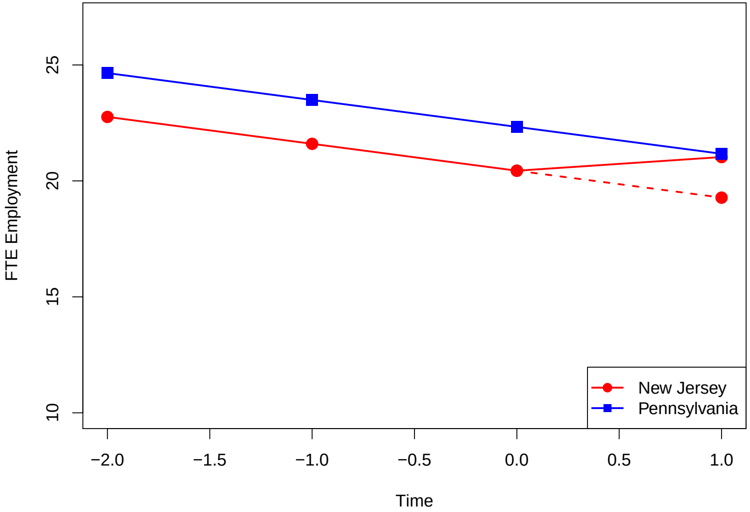
Falsification test: Data for prior periods



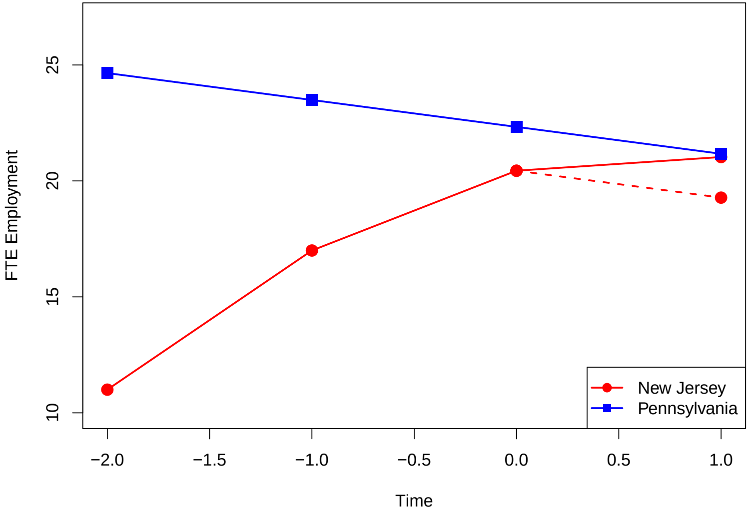
Falsification test: Data for prior periods



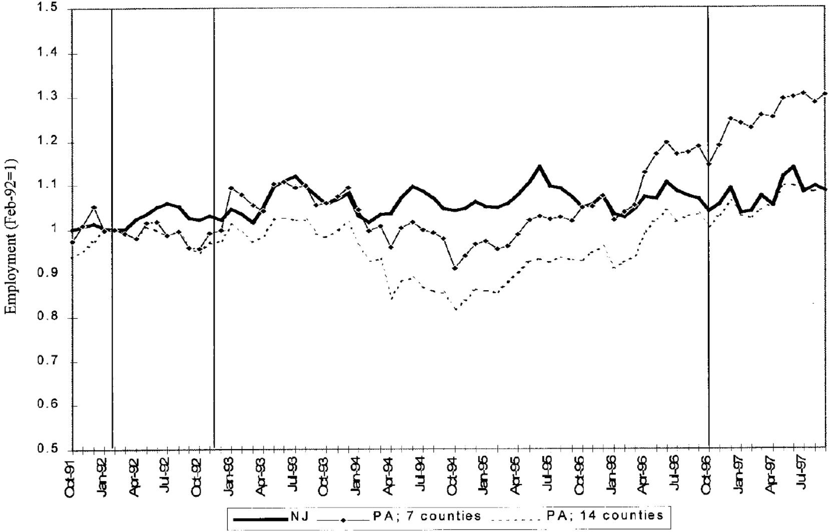
Falsification test: Data for prior periods



Falsification test: Data for prior periods



Longer Trends in Employment (Card and Krueger 2000)



Falsification test: Alternative control group

Variable	Stores by state			Stores in New Jersey[a]			Differences within NJ[b]	
	PA (i)	NJ (ii)	Difference, NJ − PA (iii)	Wage = \$4.25 (iv)	Wage = \$4.26–\$4.99 (v)	Wage ≥ \$5.00 (vi)	Low–high (vii)	Midrange–high (viii)
1. FTE employment before, all available observations	23.33 (1.35)	20.44 (0.51)	−2.89 (1.44)	19.56 (0.77)	20.08 (0.84)	22.25 (1.14)	−2.69 (1.37)	−2.17 (1.41)
2. FTE employment after, all available observations	21.17 (0.94)	21.03 (0.52)	−0.14 (1.07)	20.88 (1.01)	20.96 (0.76)	20.21 (1.03)	0.67 (1.44)	0.75 (1.27)
3. Change in mean FTE employment	−2.16 (1.25)	0.59 (0.54)	2.76 (1.36)	1.32 (0.95)	0.87 (0.84)	−2.04 (1.14)	3.36 (1.48)	2.91 (1.41)

If placebo DD between original and alternative control group is not zero, then the original DD may be biased

Triple DDD: Mandated Maternity Benefits (Gruber, 1994)

TABLE 3—DDD ESTIMATES OF THE IMPACT OF STATE MANDATES ON HOURLY WAGES

Location/year	Before law change	After law change	Time difference for location
A. *Treatment Individuals: Married Women, 20–40 Years Old:*			
Experimental states	1.547 (0.012) [1,400]	1.513 (0.012) [1,496]	−0.034 (0.017)
Nonexperimental states	1.369 (0.010) [1,480]	1.397 (0.010) [1,640]	0.028 (0.014)
Location difference at a point in time:	0.178 (0.016)	0.116 (0.015)	
Difference-in-difference:	−0.062 (0.022)		

Triple DDD: Mandated Maternity Benefits (Gruber, 1994)

TABLE 3—DDD ESTIMATES OF THE IMPACT OF STATE MANDATES ON HOURLY WAGES

Location/year	Before law change	After law change	Time difference for location
A. *Treatment Individuals: Married Women, 20–40 Years Old:*			
Experimental states	1.547 (0.012) [1,400]	1.513 (0.012) [1,496]	−0.034 (0.017)
Nonexperimental states	1.369 (0.010) [1,480]	1.397 (0.010) [1,640]	0.028 (0.014)
Location difference at a point in time:	0.178 (0.016)	0.116 (0.015)	
Difference-in-difference:	−0.062 (0.022)		
B. *Control Group: Over 40 and Single Males 20–40:*			
Experimental states	1.759 (0.007) [5,624]	1.748 (0.007) [5,407]	−0.011 (0.010)
Nonexperimental states	1.630 (0.007) [4,959]	1.627 (0.007) [4,928]	−0.003 (0.010)
Location difference at a point in time:	0.129 (0.010)	0.121 (0.010)	
Difference-in-difference:	−0.008: (0.014)		

Triple DDD: Mandated Maternity Benefits (Gruber, 1994)

TABLE 3—DDD ESTIMATES OF THE IMPACT OF STATE MANDATES ON HOURLY WAGES

Location/year	Before law change	After law change	Time difference for location
A. *Treatment Individuals: Married Women, 20–40 Years Old:*			
Experimental states	1.547	1.513	−0.034
	(0.012)	(0.012)	(0.017)
	[1,400]	[1,496]	
Nonexperimental states	1.369	1.397	0.028
	(0.010)	(0.010)	(0.014)
	[1,480]	[1,640]	
Location difference at a point in time:	0.178	0.116	
	(0.016)	(0.015)	
Difference-in-difference:	−0.062		
	(0.022)		
B. *Control Group: Over 40 and Single Males 20–40:*			
Experimental states	1.759	1.748	−0.011
	(0.007)	(0.007)	(0.010)
	[5,624]	[5,407]	
Nonexperimental states	1.630	1.627	−0.003
	(0.007)	(0.007)	(0.010)
	[4,959]	[4,928]	
Location difference at a point in time:	0.129	0.121	
	(0.010)	(0.010)	
Difference-in-difference:	−0.008:		
	(0.014)		
DDD:	**−0.054**		
	(0.026)		

How useful is the Triple DDD?

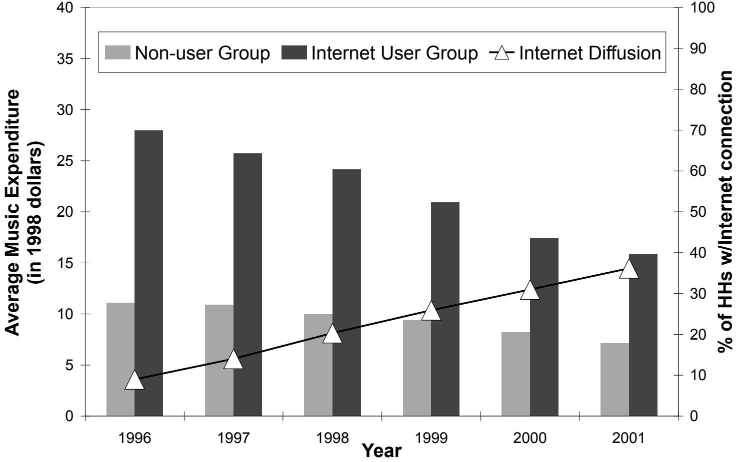
- The DDD estimate is the difference between the DD of interest and the placebo DD (that is supposed to be zero)
  - If the placebo DD is non zero, it might be difficult to convince reviewers that the DDD removes all the bias
  - If the placebo DD is zero, then DD and DDD give the same results but DD is preferable because standard errors are smaller for DD than for DDD

2. *Compositional differences*

- In repeated cross-sections, we do not want the composition of the sample to change between periods.
- Example:
  - Hong (2011) uses repeated cross-sectional data from Consumer Expenditure Survey (CEX) containing music expenditures and internet use for random samples of U.S. households
  - Study exploits the emergence of Napster (the first sharing software widely used by Internet users) in June 1999 as a natural experiment.
  - Study compares internet users and internet non-users, before and after emergence of Napster

Compositional differences?

Figure 1: Internet Diffusion and Average Quarterly Music Expenditure in the CEX



Compositional differences?

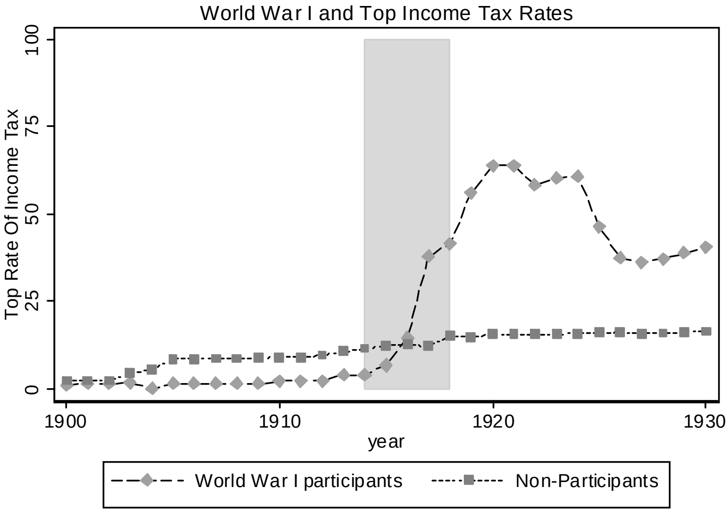
Table 1: Descriptive Statistics for Internet User and Non-user Groups[a]

Year	1997		1998		1999	
	Internet User	Non-user	Internet User	Non-user	Internet User	Non-user
Average Expenditure						
Recorded Music	$25.73	$10.90	$24.18	$9.97	$20.92	$9.37
Entertainment	$195.03	$96.71	$193.38	$84.92	$182.42	$80.19
Zero Expenditure						
Recorded Music	.56	.79	.60	.80	.64	.81
Entertainment	.08	.32	.09	.35	.14	.39
Demographics						
Age	40.2	49.0	42.3	49.0	44.1	49.4
Income	$52,887	$30,459	$51,995	$28,169	$49,970	$26,649
High School Grad.	.18	.31	.17	.32	.21	.32
Some College	.37	.28	.35	.27	.34	.27
College Grad.	.43	.21	.45	.21	.42	.20
Manager	.16	.08	.16	.08	.14	.08

Diffusion of the internet changes samples (e.g. younger music fans are early adopters)

3. *Long-term effects versus reliability*:
    - Parallel trends assumption for DD is more likely to hold over a shorter time-window
    - In the long-run, many other things may happen that could confound the effect of the treatment
    - Should be cautious to extrapolate short-term effects to long-term effects

Effect of War on Tax Rates (Scheve and Stasavage 2010)



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

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

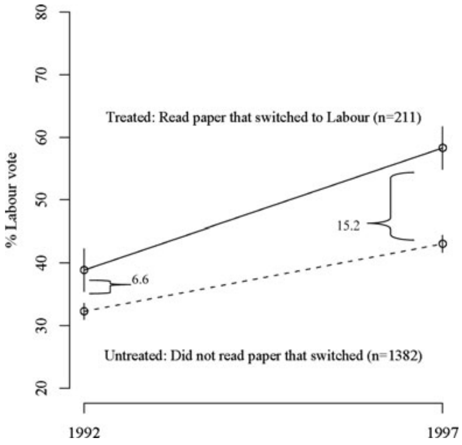
- Training program for the young:
  - Employment for the young increases from 20% to 30%
  - Employment for the old increases from 5% to 10%
  - Positive DD effect: $(30 - 20) - (10 - 5) = 5\%$ increase
  - But if you consider log changes in employment, the DD is, $[log(30) - log(20)] - [log(10) - log(5)] = log(1.5) - log(2) < 0$
- DD estimates may be more reliable if treated and controls are more similar at baseline
- More similarity may help with parallel trends assumption

- Combine matching and difference-in-differences:
  - Match on pre-treatment covariates and (lagged) outcomes
  - Run difference-in-differences regression in matched data-set
  - Can also use inverse-propensity score weighting (Hirano, Imbens, and Ridder 2003; Imai and Kim 2012)
- Can also combine difference-in-differences with regression discontinuity design or randomized experiment

Lenz and Ladd (2009) consider effect of shift in newspaper endorsements to Tony Blair on Labour Vote Choice in the 1997 U.K. general election


THE Sun
MAY 1st
An historic announcement from Britain's No1 newspape
THE PAPER OF THE PEOPLE
THE SUN BACKS BLAIR
Give change a chance

Sun, 18 March 1997



This figure shows that reading a paper that switched to Labour is associated with an $(15.2 - 6.6 =)$ 8.6 percentage point shift to Labour between the 1992 and 1997 UK elections. Paper readership is measured in the 1996 wave, before the papers switched, or, if no 1996 interview was conducted, in an earlier wave. Confidence intervals show one standard error.

TABLE 3 Comparing Covariates among the Treated and Untreated Groups

	All		*Difference* (Treated Minus Untreated)		
Covariates (Measured in 1992)	**Treated**	**Untreated**	**All**	**Exact**	**Genetic**
Prior Labour Vote (Labour 1, Other 0)	0.389	0.323	0.066	0.000	0.000
Prior Conservative Vote (Conservative 1, Other 0)	0.389	0.404	−0.015	0.000	0.000
Prior Liberal Vote (Liberal 1, Other 0)	0.156	0.188	−0.032	0.000	0.000
Prior Labour Party Identification (Labour 1, Other 0)	0.337	0.314	0.022	0.000	−0.005
Prior Conservative Party Identification (Conservative 1, Other 0)	0.412	0.418	−0.007	0.000	0.005
Prior Liberal Party Identification (Liberal 1, Other 0)	0.133	0.154	−0.021	0.000	0.005
Prior Labour Party Support (Strongly Favor 1 to Strongly Oppose 0)	0.488	0.462	0.025	0.000	−0.005
Prior Conservative Party Support (Strongly Favor 1 to Strongly Oppose 0)	0.524	0.522	0.003	0.000	0.005
Prior Political Knowledge (High 1, Mid .5, Low 0)	0.545	0.671	−0.126	0.000	−0.007

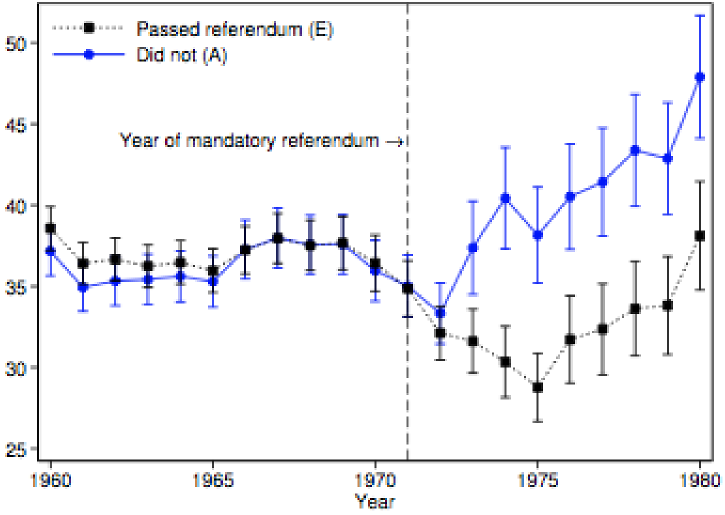
Difference-in-Differences in Matched Data

			Preprocessed with Matching			
			Exact on Selected Variables		**Genetic on All Variables**	
	Bivariate	**Multivariate (Probit)**	**Bivariate**	**Multivariate (Probit)**	**Bivariate**	**Multivariate (Probit)**
Among All Readers						
Treatment Effect (%)	8.6	12.2	10.9	14.0	10.4	9.6
(Standard error)	(3.0)	(3.6)	(4.1)	(6.0)	(4.3)	(4.9)
n Treated / *n* Control	211/1382	211/1382	192/192	192/192	211/211	211/211
Among Habitual Readers						
Treatment Effect (%)	12.7	23.1	17.9	23.4	15.8	25.7
(Standard error)	(4.1)	(6.4)	(5.4)	(11.3)	(6.6)	(9.0)
n Treated / *n* Control	102/1382	102/1382	95/95	95/95	102/102	102/102

Sances (2013):

- Uses original dataset of 920 towns in New York state, over a period where about 400 towns changed their method of choosing property tax officials.
- Plausibly exogenous changes: in 1970, state passes law requiring towns to switch to appointed tax assessors *unless* they proactively pass a local law via referendum to keep elected assessors.
- Clear measure of performance — how much is the official over or under valuing property? (Compare assessment value to sale value.)

Sances (2013)


Passed referendum (E)
Did not (A)
Year of mandatory referendum →
50
45
40
35
30
25
1960
1965
1970
1975
1980
Year

Sances (2013)