PS207 Quantitative Causal Inference, Fall 2016

Difference-in-Difference Strategies

Matto Mildenberger

UC Santa Barbara

`Special thanks to Chad Hazlett (UCLA) for select slides, used with permission`

Roadmap

1. Theory: Potential outcomes, identification, key quantities
2. Randomization
    - difference in means, variance estimation
    - covariate adjustment
    - blocking
    - cluster randomization
3. Selection on Observables
    - sub-classification
    - matching (on X)
    - weighting (on X)
    - matching and weighting on $Pr(D = 1)$ (propensity scores)
    - regression
4. Difference in Differences and Synthetic Control
5. Instrumental Variables
6. Regression Discontinuity
7. Sensitivity and Bounds

Weakening Selection on Observales

What if we think some unobservables that influence the potential outcomes may differ between treated and untreated units?

In some cases we can instead make assumptions about *how potential outcomes change over time*, and weaken slightly the SOO assumption

- differences in differences (DID)
- DID with covariates, fixed effects
- synthetic control

DID in the Time of Cholera

- Prevailing 19th century was that cholera was spread by “bad air”
- John Snow, a London physician thoughts otherwise
- In 1849, the Southwark and Vauxhall water company and the Lambeth company got water from the same polluted part of the Thames in central London. Snow assessed cholera mortality in both areas.
- In 1952, Lambeth moved its waterworks up stream.
- Shortly after, cholera death rates fell sharply in areas served by Lambeth but not in the other area

Intuitions:

- What assumptions allow us to draw a causal conclusion?
- Why not just compare the two in 1852?
- Why not just compare Lambeth’s service area post-minus-pre?

Two Groups, 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 $Y_{di}(t)$

- $Y_{1i}(t)$ 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 when non-treated between t and $t-1$

Two Groups and Two Periods

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

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

Estimand (ATT post-treatment)

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

Two Groups and Two Periods

Estimand (ATT)

$$\tau_{ATT} = \mathbb{E}[Y_1(1) - Y_0(1)|D = 1]$$

Observed Moments:

	Post-Period (T=1)	Pre-Period (T=0)
Treated D=1	$\mathbb{E}[Y_1(1)\|D = 1]$	$\mathbb{E}[Y_0(0)\|D = 1]$
Control D=0	$\mathbb{E}[Y_0(1)\|D = 0]$	$\mathbb{E}[Y_0(0)\|D = 0]$

Two Groups and Two Periods

Estimand (ATT)

$$\tau_{ATT} = \mathbb{E}[Y_1(1) - Y_0(1)|D = 1]$$

Observed Moments:

	Post-Period (T=1)	Pre-Period (T=0)
Treated D=1	$\mathbb{E}[Y_1(1)\|D = 1]$	$\mathbb{E}[Y_0(0)\|D = 1]$
Control D=0	$\mathbb{E}[Y_0(1)\|D = 0]$	$\mathbb{E}[Y_0(0)\|D = 0]$

Two Groups and Two Periods

Estimand (ATT)

$$\tau_{ATT} = \mathbb{E}[Y_1(1) - Y_0(1)|D = 1]$$

Observed Moments:

	Post-Period (T=1)	Pre-Period (T=0)
Treated D=1	$\mathbb{E}[Y_1(1)\|D = 1]$	$\mathbb{E}[Y_0(0)\|D = 1]$
Control D=0	$\mathbb{E}[Y_0(1)\|D = 0]$	$\mathbb{E}[Y_0(0)\|D = 0]$

Problem

Missing potential outcome: $\mathbb{E}[Y_0(1)|D = 1]$*, ie. what is the average non-treatment outcome in the post-treatment period, for the treated?*

Two Groups and Two Periods

Estimand (ATT)

$$\tau_{ATT} = \mathbb{E}[Y_1(1) - Y_0(1)|D = 1]$$

Observed Moments:

	Post-Period (T=1)	Pre-Period (T=0)
Treated D=1	$\mathbb{E}[Y_1(1)\|D = 1]$	$\mathbb{E}[Y_0(0)\|D = 1]$
Control D=0	$\mathbb{E}[Y_0(1)\|D = 0]$	$\mathbb{E}[Y_0(0)\|D = 0]$

Control Strategy: Before-After Comparison

- Use: $\mathbb{E}[Y(1)|D = 1] - \mathbb{E}[Y(0)|D = 1]$
- Assumes: $\mathbb{E}[Y_0(1)|D = 1] = \mathbb{E}[Y_0(0)|D = 1]$

Two Groups and Two Periods

Estimand (ATT)

$$\tau_{ATT} = \mathbb{E}[Y_1(1) - Y_0(1)|D = 1]$$

Observed Moments:

	Post-Period (T=1)	Pre-Period (T=0)
Treated D=1	$\mathbb{E}[Y_1(1)\|D = 1]$	$\mathbb{E}[Y_0(0)\|D = 1]$
Control D=0	$\mathbb{E}[Y_0(1)\|D = 0]$	$\mathbb{E}[Y_0(0)\|D = 0]$

Control Strategy: Treated-Control Comparison in Post-Period

- Use: $\mathbb{E}[Y(1)|D = 1] - \mathbb{E}[Y(1)|D = 0]$
- Assumes: $\mathbb{E}[Y_0(1)|D = 1] = \mathbb{E}[Y_0(1)|D = 0]$

Two Groups and Two Periods

Estimand (ATT)

$$\tau_{ATT} = \mathbb{E}[Y_1(1) - Y_0(1)|D = 1]$$

Observed Moments:

	Post-Period (T=1)	Pre-Period (T=0)
Treated D=1	$\mathbb{E}[Y_1(1)\|D = 1]$	$\mathbb{E}[Y_0(0)\|D = 1]$
Control D=0	$\mathbb{E}[Y_0(1)\|D = 0]$	$\mathbb{E}[Y_0(0)\|D = 0]$

Control Strategy: Difference-in-Differences (DID)

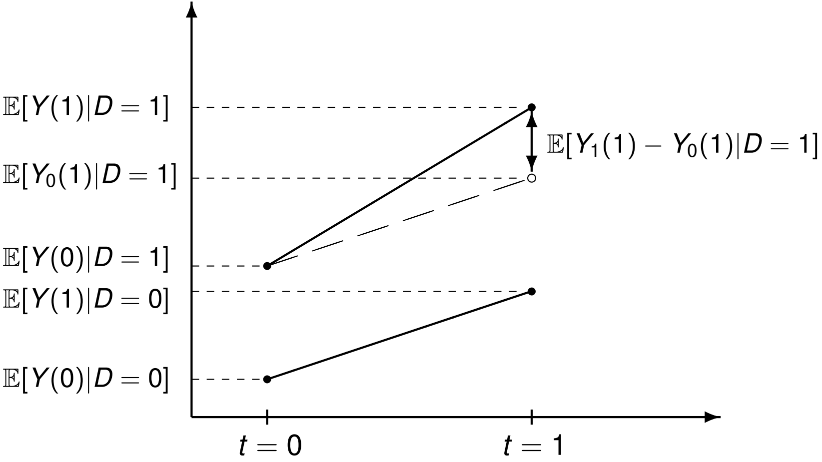
- Use:

$$\mathbb{E}[Y(1) - Y(0)|D = 1] - \mathbb{E}[Y(1) - Y(0)|D = 0]$$

- Assumes: $\mathbb{E}[Y_0(1) - Y_0(0)|D = 1] = \mathbb{E}[Y_0(1) - Y_0(0)|D = 0]$

"Parallel trends" assumption

Graphical Representation: Difference-in-Differences



Where is the parallel trends assumption?

Identification with Difference-in-Differences

Identification Assumption (Parallel Trends)

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

Identfication Result

Under parallel trends, the ATT is identified by the DID estimator:

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

Proof

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

Another view: additive effects

- Change in control unit's outcome, ΔY_0, is "effect of time", $\Delta Y_0 = \delta_0$
- Change in treated unit's outcome is effect of time and treatment, $\Delta Y_1 = \delta_1 + \tau_{ATT}$
- We estimate $DID = \Delta Y_1 - \Delta Y_0 = (\delta_1 + \tau_{ATT}) - \delta_0$
- ...equals τ_{ATT} if $\delta_0 = \delta_1$

Difference-in-Differences: Estimators

Estimand (DID)

$$\tau_{DID} = E[Y(1)|D = 1] - E[Y(1)|D = 0] - E[Y(0)|D = 1] - E[Y(0)|D = 0]$$

Estimator (Sample Means: Panel Version)

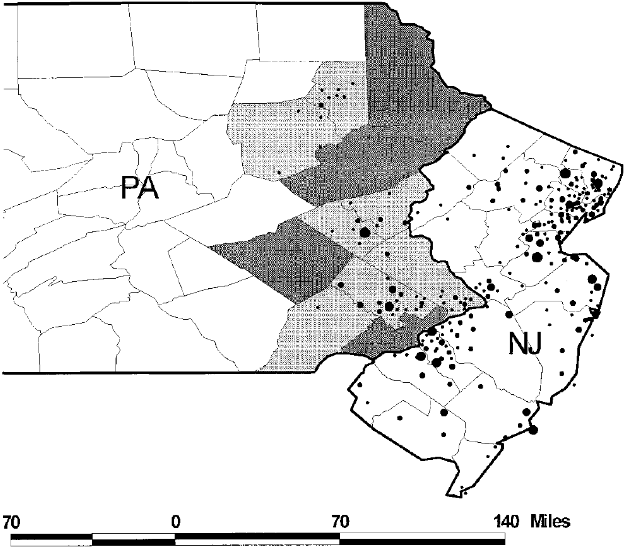
$$\frac{1}{N_1} \sum_{D_i=1} \{Y_i(1) - Y_i(0)\} - \frac{1}{N_0} \sum_{D_i=0} \{Y_i(1) - Y_i(0)\},$$

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

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 month after increase

Locations of Restaurants (Card and Krueger 2000)



Sample Means: Minimum wage laws and 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)

Difference-in-Differences: Estimators

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:

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

Difference-in-Differences: Estimators

Estimator (Regression: Repeated Cross-Sections)

Alternatively, the same estimator can be computed by regression:

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

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

Easy to show that τ estimates the DD effect:

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

Difference-in-Differences: Estimators

Estimator (Regression: Repeated Cross-Sections)

Alternatively, the same estimator can be computed by regression:

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

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

	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

Again, assumption is parallel trends $\Rightarrow$ “time-effect” δ same for treated and controls

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
```

Regression: Minimum wage laws and employment

```
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
```

DID with Covariates

Suppose you only believe parallel trends conditionally on covariates/ want to remove effect of change in some observables

Estimator (Regression: Repeated Cross-Sections)

DID regression estimator with 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*

Can interact time-invariant covariates with the time indicator:

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

Difference-in-Differences: Estimators

With panel data you have same units in pre- and post

- efficiency gains possible: unit and period fixed effects
- still relying on parallel trends or “no unobserved time-varying confounders”

Estimator (Regression: Panel Data)

$$Y_{it} = \mu + \gamma_i + \delta T + \tau D_{it} + X_{it}^\top \beta + \varepsilon_{it}$$

- *One intercept for each unit,* γ_i
- D_{it} *now coded as 1 for treated in post-period and 0 otherwise (*$D_i \cdot T_i$ *from before)*

Equivalently, first differences then simple regression:

$$\Delta Y_i = \delta + \tau \cdot D_i + \Delta X_i^\top \beta_\Delta + u_i,$$

where $\Delta Y_i = Y_i(1) - Y_i(0)$, $\Delta X_i = X_{i1} - X_{i0}$, $u_i = \Delta \varepsilon_i$.

Regression: Minimum wage laws and employment

You can setup regression with unit and time FE by hand, or use PLM:

```
library(plm)
library(lmtest)
> d$Dit=d$nj * d$postperiod #Dit=1 only in post-tx
> 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

You can do first-differencing by hand then regress, or ask PLM:

```
> 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 *
---
```

Difference-in-Differences: Threats to Validity

1. Non-parallel trends (aka unequal time-effects, unobserved time-varying confounding)
2. Compositional differences
3. Long-term effects versus reliability
4. Functional form dependence

Bias is a matter of degree:

- small violations, small bias $\Rightarrow$ may be okay
- but biases can be large, even give wrong sign

With DID, the parallel trends assumption is often tough to sell, so be cautious

Difference-in-Differences: Threats to Validity

1. **Non-parallel trends**: Often treatments/programs are targeted based on pre-existing *trend* in outcomes.
   - "Ashenfelter dip": participants in training programs often entered training *becuase* they have just experienced a dip in earnings. In absence of treatment these may recover anyway. Comparing wages of participants and non-participants using DID upwardly biases estimate of the program effect
   - More generically, regression to the mean issues when units self-select into treatment
   - Regional targeting: NGOs may target villages that appear most promising (or worst off)
   - Latent reasons why a unit that would improve in future anyway takes the treatment.

Checks for Difference-in-Differences Design

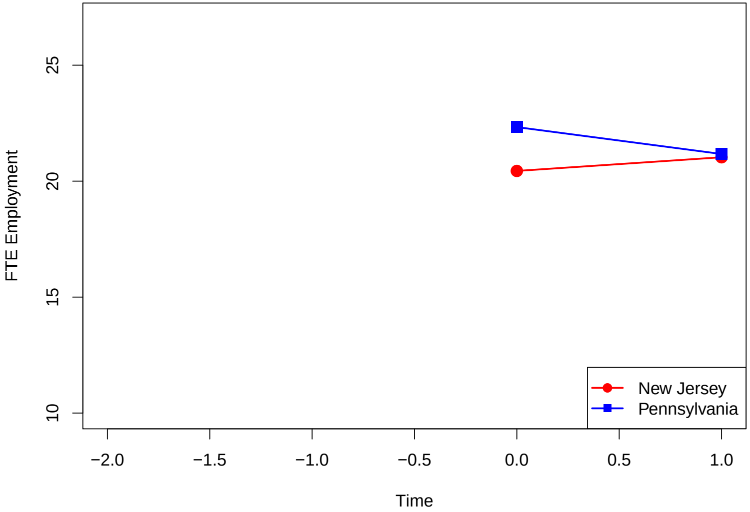
Two kinds of falsification tests for non-parallel trends

1. Falsification test using data for prior periods
   - Given parallel trends during periods $t = -1, 0, 1$, we have:

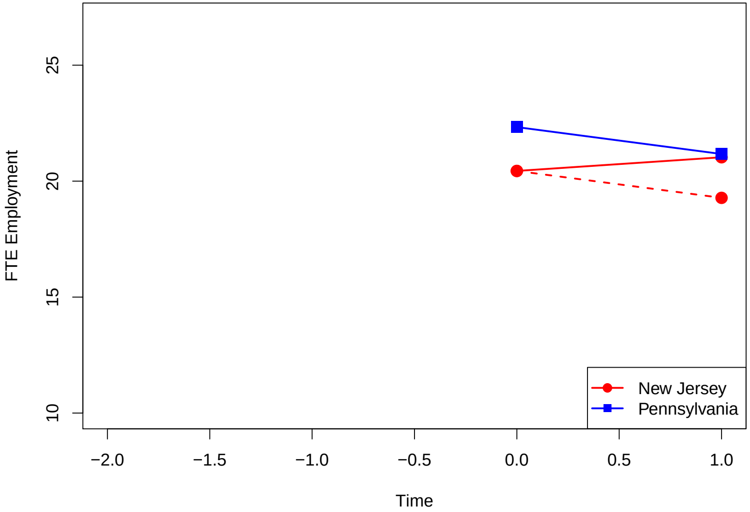
$$\mathbb{E}[Y(0) - Y(-1)|D = 1] - \mathbb{E}[Y(0) - Y(-1)|D = 0] = 0$$

   - run placebo DID on data from $t = -1, 0$.
   - what "treatment effect" should we get?
2. Falsification test using alternative outcome or treated sub-group that is not supposed to be affected by the treatment
   - troubling if DID from the placebo outcome or un-movable group is non-zero,

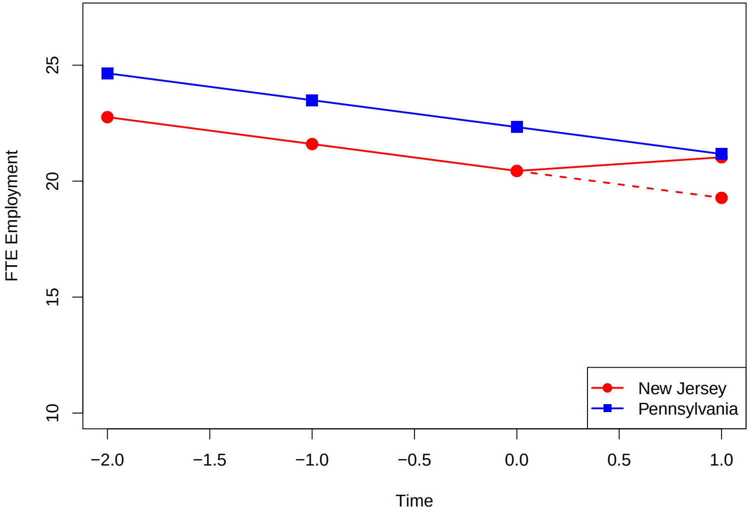
Falsification test: Data for prior periods



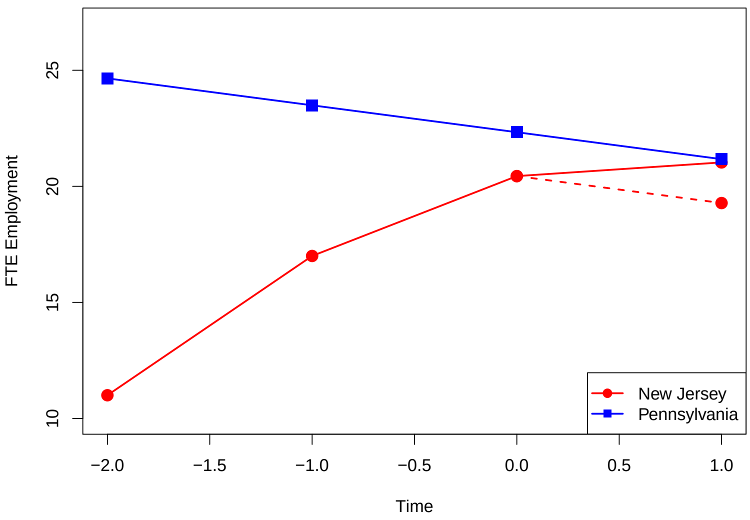
Falsification test: Data for prior periods



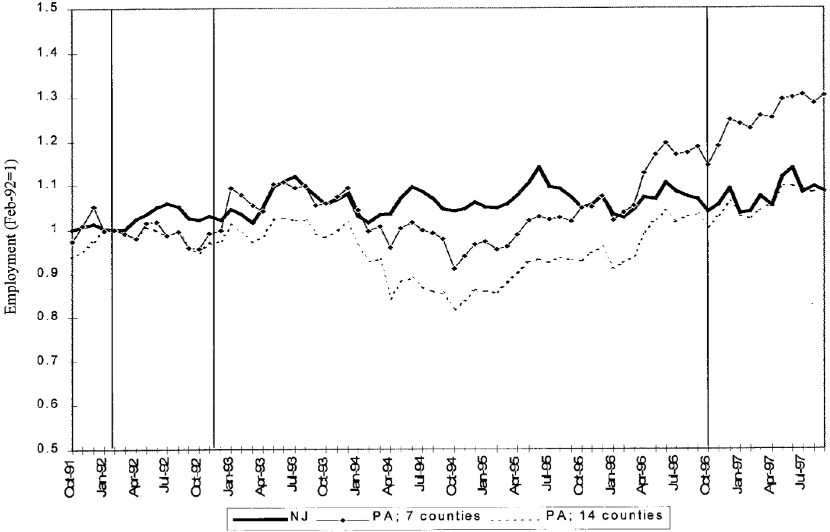
Falsification test: Data for prior periods



Falsification test: Data for prior periods



Longer Trends in Employment (Card & Krueger 2000)



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 (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)		
DDD:	**−0.054** **(0.026)**		

How useful is the Triple DDD?

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

Difference-in-Differences: Further Threats to Validity

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
        - Exploits the emergence of Napster in June 1999 as a natural experiment, comparing internet users and internet non-users, before and after emergence of Napster

Compositional differences?

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

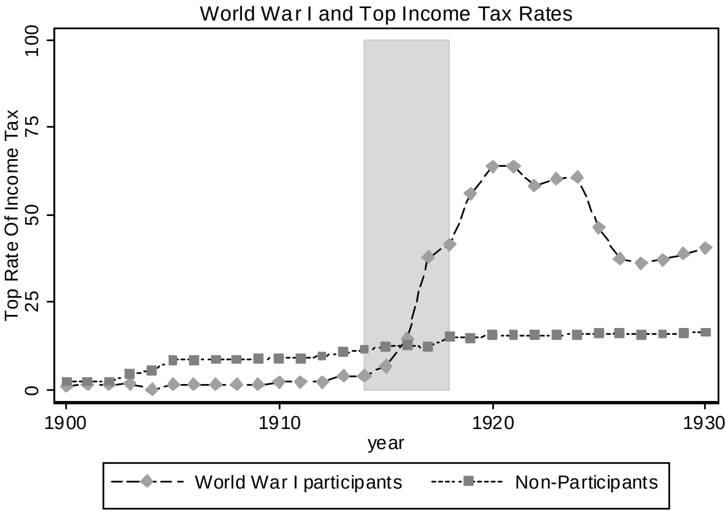
Year	1997		1998		1999	
	Internet User	Non-user	Internet User	Non-user	Internet User	Non-user
Average Expenditure						
Recorded Music	\$25.73	\$10.90	\$24.18	\$9.97	\$20.92	\$9.37
Entertainment	\$195.03	\$96.71	\$193.38	\$84.92	\$182.42	\$80.19
Zero Expenditure						
Recorded Music	.56	.79	.60	.80	.64	.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

You can check for DID effect on the demographic outcomes as a test for compositional differences

Difference-in-Differences: Further Threats to Validity

3. **Long-term effects versus reliability**:
   - Often you may want to know what happens over longer period of times (e.g. effect of polity shift, a conflict, adoption of a policy)
   - But parallel trends assumption is more viable over a short time-window. In the long-run, many other things may happen that could confound the effect of the treatment
   - And short-term effects generally can't be "extrapolated" to estimate long-term effects

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



Difference-in-Differences: Further Threats to Validity

4. **Functional form dependence**: The differencing is not-invariant to changes of scale. Magnitude or sign of the DID effect may be sensitive to this.
   - Higher risk when average outcomes for controls and treated are very different at baseline.
   - Example: Cash transfers for refugees
     - income of those receiving vouchers rises from 5 to 10
     - over same time, income for comparably skilled group in host community (no vouchers) changes from 20 to 25
     - on original scale, $DID = (10 - 5) - (25 - 20) = 0\%$ effect of voucher
     - but in terms of percentage changes or logs,
       $[100] - [25] = 75$ percentage points
       $[log(10) - log(5)] - [log(25) - log(20)] = log(1.5) - log(2) = .47$
   - In general, more similarity in original outcomes is better both for this reason and to make parallel trends assumption more believable

Matching and difference-in-differences

- Parallel trends is more plausible among units that are more similar (both on X and pre-tx Y). One strategy: combine matching (or weighting) and DID
  - 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)
- Randomized experiment, natural experiment, or a discontinuity can improve parallel trends assumption, so DID can be used in combination
- With few treated units: synthetic control method

How do newspaper endorsement affect vote choice?

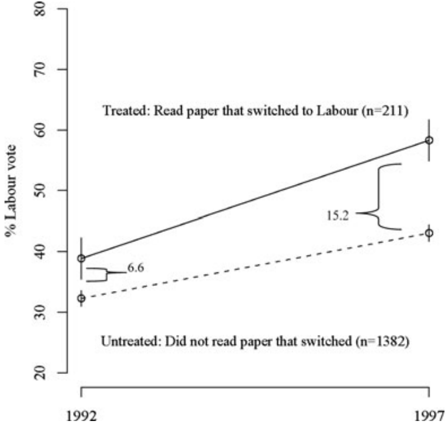
Lenz and Ladd (2009) consider effect of shift in newspaper endorsements to Tony Blair in the 1997 U.K. general election


THE Sun
MAY 1st
An historic announcement from Britain's No1 newspape
28p
THE PAPER OF THE PEOPLE
THE SUN BACKS BLAIR
Give change a chance

Sun, 18 March 1997

Difference-in-Differences Estimates

- Find people who do and do not read the Sun, see if there is a change among those who read.
- Since those who read it could have changed anyway (no SOO) anyway, compare change to those who don't read it (DID):



This figure shows that reading a paper that switched to Labour

Matching on 1992 Characteristics

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