Week 5: Panel Data and Difference-in-Differences

PUBL0050 Causal Inference

Julia de Romémont

Academic Year 24-25

UCL Departement of Political Science

1992

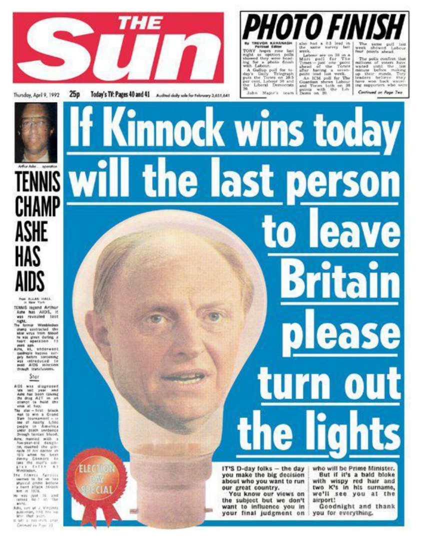

THE Sun

PHOTO FINISH

By TREVOR KAVANAGH Political Editor

TORY hopes rose last night as opinion polls showed they were heading for a photo finish with Labour.

A Gallup poll for today's Daily Telegraph puts the Tories on 38.5 per cent, Labour 38 and the Liberal Democrats 20.

John Major's team also had a 0.5 lead in the same survey last week.

Labour are on 39 in a Mori poll for The Times — just one point ahead of the Tories after having a seven-point lead last week.

An ICM poll for The Guardian shows Labour and Tories both on 38 points with the Lib Dems on 20.

The same poll last week showed Labour four points ahead.

The polls confirm that millions of voters have waited until the last minute before making up their minds. Tory leaders believe they have won back wavering supporters who were

Continued on Page Two

Thursday, April 9, 1992 25p Today's TV: Pages 40 and 41

TENNIS CHAMP ASHE HAS AIDS

If Kinnock wins today will the last person to leave Britain please turn out the lights

ELECTION DAY SPECIAL

IT'S D-day folks – the day you make the big decision about who you want to run our great country.

You know our views on the subject but we don't want to influence you in your final judgment on who will be Prime Minister.

But if it's a bald bloke with wispy red hair and two K's in his surname, we'll see you at the airport!

Goodnight and thank you for everything.

1997


THE Sun

MAY 1st

An historic announcement from Britain's No1 newspaper

THE PAPER OF THE PEOPLE

THE SUN BACKS BLAIR

Give change a chance

Was it “the Sun wot won it”?

1992

1997

Was it "the Sun wot won it"?

THE Sun

25p

TV Guide

FREE TODAY

PLUS THE GREAT AMERICAN Comic

IT'S THE SUN WOT WON IT

Was it “the Sun wot won it”?

Does the media influence vote choice?

- **Randomized experiment?**
  - Hard to persuade newspapers to randomly endorse political candidates
  - Hard to randomly allocate citizens to read certain newspapers
- **Selection on observables?**
  - The types of individual who read certain newspapers (i.e. The Sun) are likely different **in many ways** from those who read other newspapers
- **Difference-in-differences**
  - Collect data on vote choice **before** & **after** change in endorsement
  - Did people who read The Sun **change** their vote choice more than people who did not read The Sun?

Running example

Persuasive Power of the News Media

Did the change in support for the Labour Party by the Sun newspaper increase the number of people voting Labour? Ladd and Lenz (2009) use the British Election Panel Survey, which includes information on whether individuals voted for Labour in 1992, whether they voted Labour in 1997, and which newspapers they read.

- Outcome (Y, `voted_lab`): 1 if individual i voted Labour at time t
- Treatment Group (D, `reads_sun`): 1 if individual i read the Sun (in 1992)
- **Time (T, `year`)**: Election year (1992 or 1997)
- $N = 1593$, $N_1 = 211$, $N_0 = 1382$
- Note that this is panel data (repeated observations on the same individuals over time)

This Lecture

Identification with Difference-in-Differences

Estimation of Difference-in-Differences

Threats to Validity

Multiple Periods

Data requirements

Identification with Difference-in-Differences

Difference-in-differences setup

Definition

Two groups:

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

Two periods:

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

Potential outcome $Y_{di}(t)$

- $Y_{1i}(t)$ outcome of unit i in period t when treated (at $D_i = 1$)
- $Y_{0i}(t)$ outcome of unit i in period t when control (at $D_i = 0$)

Difference-in-differences setup

Definition

Causal effect for unit i at time t is $\tau_i(t) = Y_{1i}(t) - Y_{0i}(t)$

For a given unit, in a given time period, the observed outcome $Y_i(t)$ is:

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

If treatment occurs only after $t = 0$ we have:

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

$\rightarrow$ Fundamental problem of causal inference.

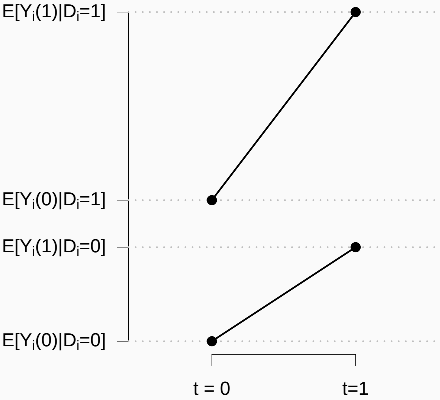
Estimand (ATT)

$\hat{\tau}_{\mathit{ATT}} = E[Y_{1i}(1) - Y_{0i}(1) | D_i = 1]$

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?*

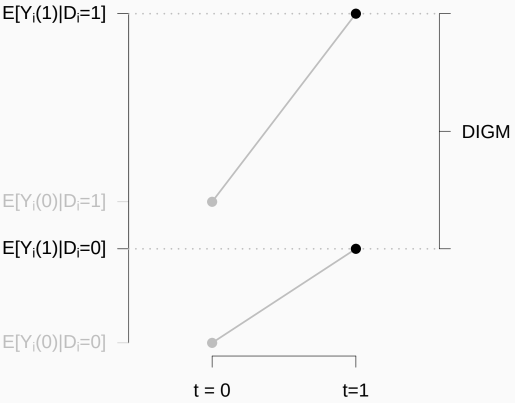
Illustration



Problem: Missing potential outcome: $E[Y_{0i}(1)|D_i = 1]$

Illustration

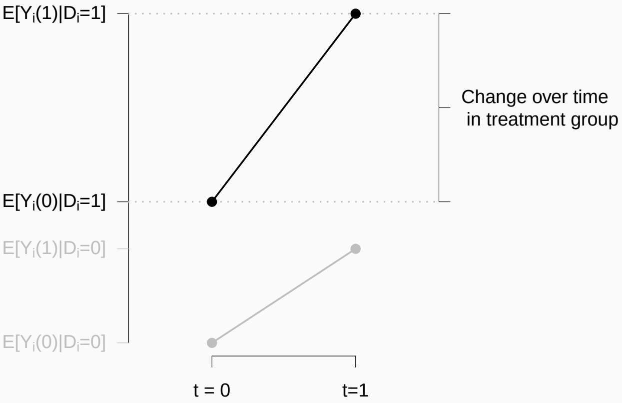
Strategy 1: Treated vs. control in post-treatment period



Assumption: No selection bias

Illustration

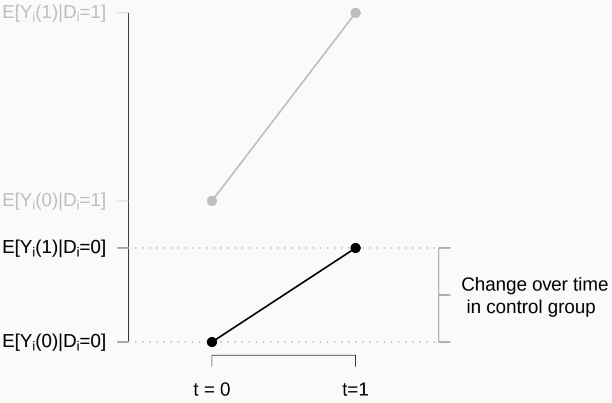
Strategy 2: Before vs. after for treatment units



Assumption: No effect of time independent of treatment

Illustration

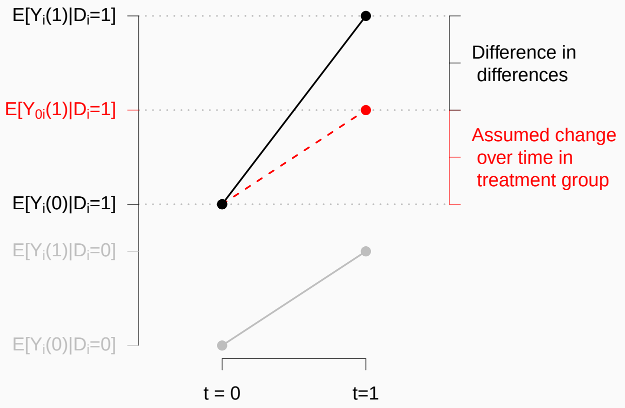
Difference-in-differences (1)



No effect of time independent of treatment

Illustration

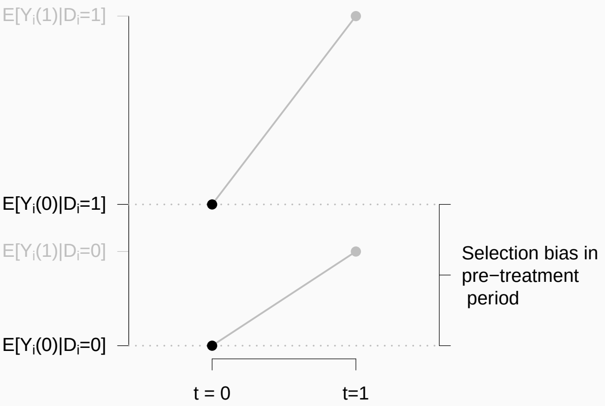
Strategy 3: Difference-in-differences (1)



Assumption: Trend over time is the same for treatment and control.

Illustration

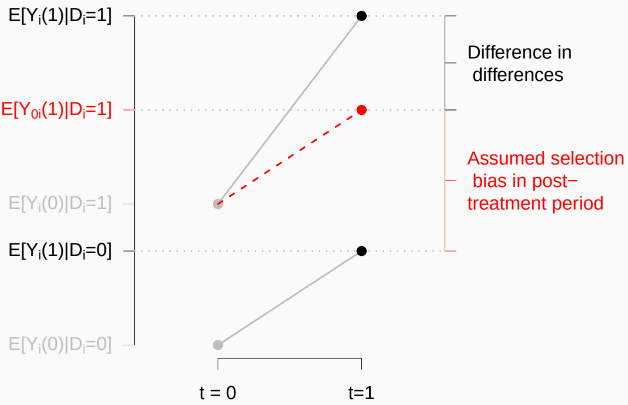
Strategy 3: Difference-in-differences (2)



Trend over time is the same for treatment and control.

Illustration

Strategy 3: Difference-in-differences (2)



Assumption: Selection bias is stable over time.

Identification with DiD

Two ways of stating the **same** identifying assumption:

- **Parellel trends**
  - If treated units did not receive the treatment, they would have followed the same trend as the control units
- **No time-varying confounders (stable selection bias)**
  - Omitted variables related both to treatment and outcome must be fixed over time

Identification with DiD

Estimand (ATT)

$\hat{\tau}_{\mathit{ATT}} = E[Y_{1i}(1) - Y_{0i}(1) | D_i = 1]$

Identification Assumption

$E[Y_{0i}(1) - Y_{0i}(0) | D_i = 1] = E[Y_{0i}(1) - Y_{0i}(0) | D_i = 0]$ *(parallel trends)*

Identification Result

$$
\begin{aligned}
E[Y_{1i}(1) - Y_{0i}(1) | D_i = 1] &= E[Y_{1i}(1) | D_i = 1] - E[Y_{0i}(1) | D_i = 1] \\
&= E[Y_{1i}(1) | D_i = 1] - \Big\{ E[Y_{0i}(0) | D_i = 1] + \\
\text{(Parallel trends)} \quad & \qquad E[Y_{0i}(1) | D_i = 0] - E[Y_{0i}(0) | D_i = 0] \Big\} \\
&= \Big\{ E[Y_i(1) | D_i = 1] - E[Y_i(1) | D_i = 0] \Big\} - \\
& \qquad \Big\{ E[Y_i(0) | D_i = 1] - E[Y_i(0) | D = 0] \Big\}
\end{aligned}
$$

Identification with DiD

Estimand (ATT)

$\hat{\tau}_{ATT} = E[Y_{1i}(1) - Y_{0i}(1) | D_i = 1]$

Identification Assumption

$E[Y_{0i}(1) - Y_{0i}(0) | D_i = 1] = E[Y_{0i}(1) - Y_{0i}(0) | D_i = 0]$ *(parallel trends)*

Identification Result

In other words:

$$\hat{\tau}_{ATT} = \left\{\textit{Difference in means in post-treatment period}\right\} - \left\{\textit{Difference in means in pre-treatment period}\right\}$$

Estimation of Difference-in-Differences

Data structure

Typically useful to store data for DiD analysis in ‘long’ format (here, 2 rows per unit):

```r
str(sun)
```

```
## 'data.frame':    3186 obs. of  4 variables:
##  $ reads_sun: int  0 0 0 0 0 0 0 0 0 1 ...
##  $ voted_lab: int  1 1 0 1 1 1 1 1 1 1 ...
##  $ year     : num  1992 1992 1992 1992 1992 ...
##  $ id       : int  1 2 3 4 5 6 7 8 9 10 ...
```

```r
table(sun$reads_sun, sun$year)
```

```
##
##      1992 1997
##    0 1382 1382
##    1  211  211
```

Question: Which observations are ‘treated’? **Answer:** Sun readers in 1997.

Four group means in R

```r
# Untreated, pre-treatment
y_d0_t0 <- mean(sun$voted_lab[sun$reads_sun == 0 & sun$year == 1992])
y_d0_t0
```

```
## [1] 0.3227207
```

```r
# Treated, pre-treatment
y_d1_t0 <- mean(sun$voted_lab[sun$reads_sun == 1 & sun$year == 1992])
y_d1_t0
```

```
## [1] 0.3886256
```

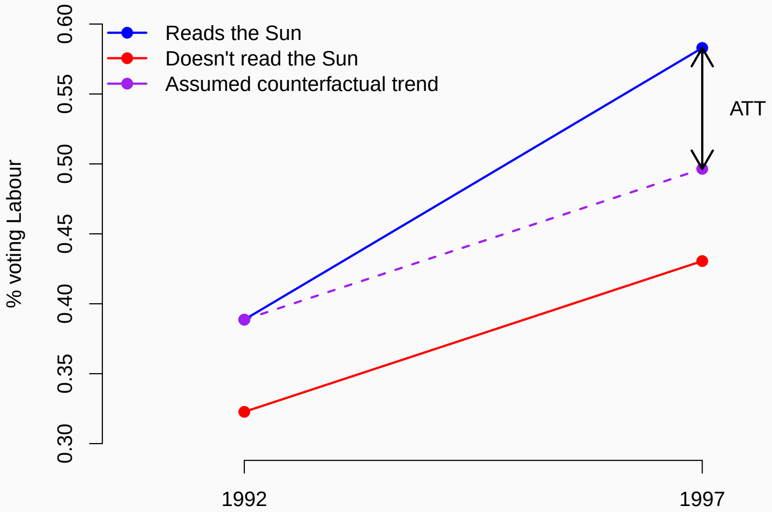
```r
# Untreated, post-treatment
y_d0_t1 <- mean(sun$voted_lab[sun$reads_sun == 0 & sun$year == 1997])
y_d0_t1
```

```
## [1] 0.4305355
```

```r
# Treated, post-treatment
y_d1_t1 <- mean(sun$voted_lab[sun$reads_sun == 1 & sun$year == 1997])
y_d1_t1
```

```
## [1] 0.5829384
```

Example



Reads the Sun
Doesn't read the Sun
Assumed counterfactual trend
ATT
% voting Labour
0.30
0.35
0.40
0.45
0.50
0.55
0.60
1992
1997

DiD in R

```r
# Parallel trend calculation
(y_d1_t1 - y_d1_t0) - (y_d0_t1 - y_d0_t0)
```

```
## [1] 0.08649803
```

```r
# Stable selection bias calculation
(y_d1_t1 - y_d0_t1) - (y_d1_t0 - y_d0_t0)
```

```
## [1] 0.08649803
```

Implication: The change in endorsement caused Labour support to increase by 8.6 percentage points more, on average, amongst readers of The Sun.

- See this GIF for an illustration of DiD

Estimator: Regression

Alternatively, the same estimate can be obtained using regression techniques.

$$Y_i = \alpha + \beta_1 \cdot D_i + \beta_2 \cdot T_i + \delta \cdot (D_i \cdot T_i) + \varepsilon,$$

where $E[\varepsilon | D_i, T_i] = 0$. Then, it is easy to show that

$E[Y_i \mid D_i, T_i]$	$T_i = 0$	$T_i = 1$	After - Before
$D_i = 0$	α	$\alpha + \beta_2$	β_2
$D_i = 1$	$\alpha + \beta_1$	$\alpha + \beta_1 + \beta_2 + \delta$	$\beta_2 + \delta$
Treated - Control	β_1	$\beta_1 + \delta$	δ

Thus, the difference-in-differences estimate is given by:

$$\hat{\tau}_{\text{ATT}} = (\beta_2 + \delta) - \beta_2 = \delta$$

Equivalently:

$$\hat{\tau}_{\text{ATT}} = (\beta_1 + \delta) - \beta_1 = \delta$$

Regression DiD in R

```
dd_mod <- lm(voted_lab ~ reads_sun *  as.factor(year),
              data  = sun)
```

	Voted Labour
Intercept	0.32***
	(0.01)
Reads the Sun	0.07*
	(0.04)
1997	0.11***
	(0.02)
Reads the Sun x 1997	0.09*
	(0.05)
Observations	3,186
R^2	0.02

Under **stable selection bias** assumption:

- $\alpha = 0.32$: Labour support amongst non-Sun readers, 1992
- $\beta_1 = 0.07$: difference between Sun and non-Sun readers, 1992
- $\beta_1 + \delta = 0.07 + 0.09 = 0.16$: difference between Sun and non-Sun readers, 1997
- $\delta = 0.09 \Rightarrow$ ATT

Regression DiD in R

```
dd_mod <- lm(voted_lab ~ reads_sun *  as.factor(year),
             data  = sun)
```

	Voted Labour
Intercept	0.32***
	(0.01)
Reads the Sun	0.07*
	(0.04)
1997	0.11***
	(0.02)
Reads the Sun x 1997	0.09*
	(0.05)
Observations	3,186
R^2	0.02

Under **parallel trends** assumption:

- $\alpha = 0.32$: Labour support amongst non-Sun readers, 1992
- $\beta_2 = 0.11$: 1992 to 1997 difference, amongst non-Sun readers
- $\beta_2 + \delta = 0.11 + 0.09 = 0.2$: 1992 to 1997 difference, amongst Sun readers
- $\delta = 0.09 \Rightarrow$ ATT

DiD with cross-sectional data

- The calculations in the previous slides are based on panel data, i.e. repeated observations of the same units.
- A nice feature diff-in-diff is it does not *require* panel data. We can also use repeated cross-sections:
  - Y_{igt} where unit i is only measured at one t
  - Units fall into treatment based on groups g
  - Particularly useful as many 'treatments' vary at some aggregate level (e.g. law changes at the region level)
- Two options:
  - Individual-level: $Y_{igt} = \alpha + \beta_1 D_{g(i)} + \beta_2 T_{t(i)} + \beta_3 (D_g \cdot T_{t(i)}) + \varepsilon_{igt}$
  - Aggregated: $Y_{gt} = \alpha + \beta_1 D_g + \beta_2 T_t + \beta_3 (D_g \cdot T_t) + \varepsilon_{gt}$
- Both approaches will give the same result, as the treatment only varies at the group level (so long as the aggregated version is weighted by cell size).

Regression estimator advantages

1. Easy to calculate standard errors (though be careful about clustering)
2. We can control for other variables
   - Individual-level data, group-level treatment: controlling for individual covariates may increase precision
   - Time-varying covariates at the group-level may strengthen the parallel trends assumption, but beware of post-treatment bias
3. Simple to extend to multiple groups/periods (more on this later)
4. Can use multi-valued (not just binary) treatments

Estimating DiD with regression II

Estimator: First-Difference Regression

With panel data we can use regression with first differences:

$$\Delta Y_i = \alpha + \delta \cdot D_i + \mathbf{X}_i'\beta + u,$$

where $\Delta Y_i = Y_i(1) - Y_i(0)$.

- With two periods, this gives identical result as the previous estimator on slide 28

First-difference regression in R

The data now only has 1 row per unit:

```r
head(sun_diff)
```

```
##   reads_sun voted_lab_92 voted_lab_97
## 1         0            1            1
## 2         0            1            0
## 3         0            0            0
## 4         0            1            1
## 5         0            1            1
## 6         0            1            1
```

```r
sun_diff$diff <- sun_diff$voted_lab_97 - sun_diff$voted_lab_92
head(sun_diff)
```

```
##   reads_sun voted_lab_92 voted_lab_97 diff
## 1         0            1            1    0
## 2         0            1            0   -1
## 3         0            0            0    0
## 4         0            1            1    0
## 5         0            1            1    0
## 6         0            1            1    0
```

First-difference regression in R

```r
first_diff_mod <- lm(diff ~ reads_sun, data = sun_diff)
```

First difference model

	Δ Voted Labour
Intercept	0.11^{***}
	(0.01)
Reads the Sun	0.09^{***}
	(0.03)
Observations	1,593
R^2	0.01

Standard errors in regression DiD

- Many papers using a DD strategy use data from many periods
- Treatments typically vary at the group level, while outcomes normally measured at the individual level
  - E.g. Minimum wage increases (state-level) and employment data (firm-level) in Card and Krueger
- Will not bias treatment effect estimates, but will cause problems for variance estimation when errors are serially correlated
- **Implication:** traditional standard errors will tend to be too small.

Standard errors in regression DiD

Solution[1]

Use cluster-robust standard errors where clusters are defined at the level of the treatment. If the number of groups is

- ... large ($\gtrapprox 30$), use `lm_robust(..., clusters = )` in `estimatr`
- ... small ($\lessapprox 30$), use block-bootstrap

[1] See Bertrand et al (2004); see also Abadie et al (2022).

Threats to Validity

Non-parallel trends

Critical identification assumption: treatment units have similar trends to control units in the absence of treatment.

Question: Why is this assumption untestable?

Answer: because of the FPOCI $\longrightarrow$ we cannot observe potential outcome under the control condition for treated units in the post-treatment period.

Potential violations of parallel trends

- "Ashenfelter's Dip"
  - Participants in worker training programs may experience decreased earnings **before** they enter the program (why are they participating?)
  - If wages revert to the mean, comparing wages of participants and non-participants leads to an upwardly biased estimate
- Targeting
  - Policymakers may target units who are most improving

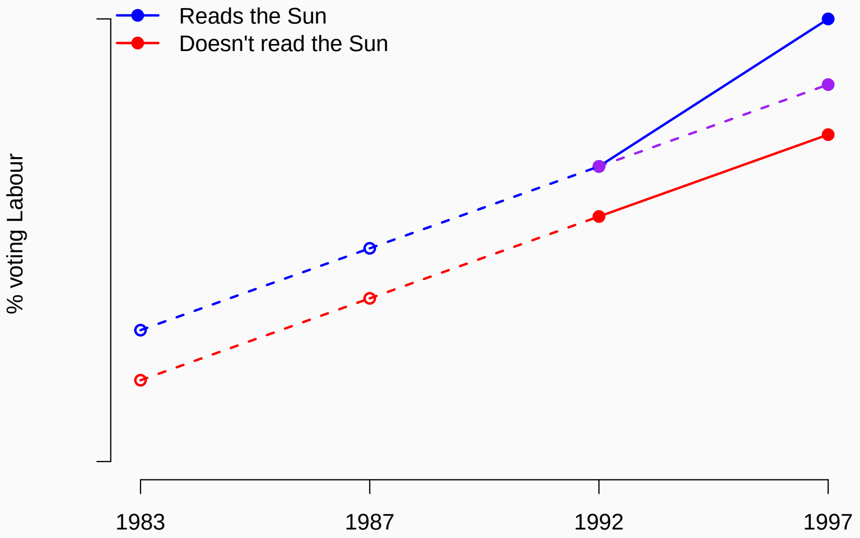
$\longrightarrow$ These are all ways in which the treatment group would **not** have had the same trend than the control group in the absence of treatment

Assessing (non-)parallel trends

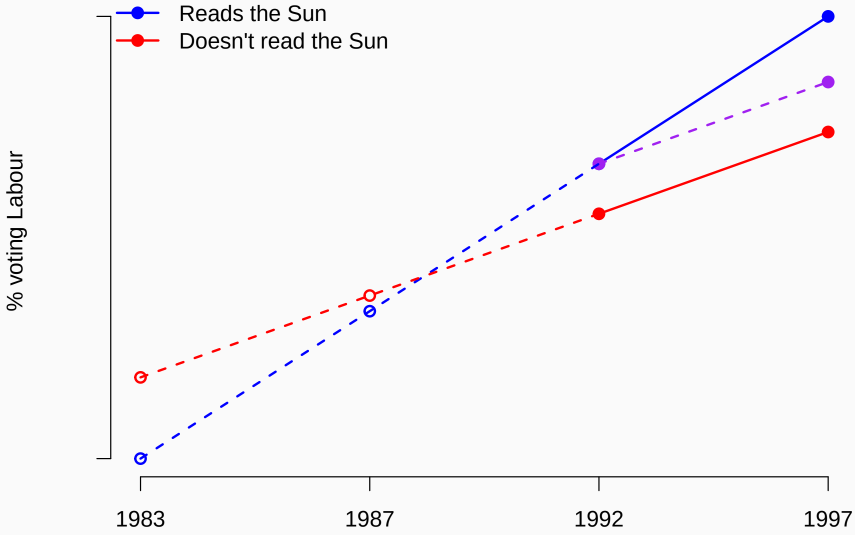
What can we do?

- One treatment/control group
  - Plot results and look at trends in periods before the treatment
  - Is the parallel trends assumption *plausible*?
- Multiple treatment/control comparisons
  - Estimate treatment effects at different time points (i.e. placebo tests) $\longrightarrow$ All estimated treatment effects before the treatment should be 0.
  - Include unit-specific time trends $\longrightarrow$ ‘relax’ parallel trends assumption

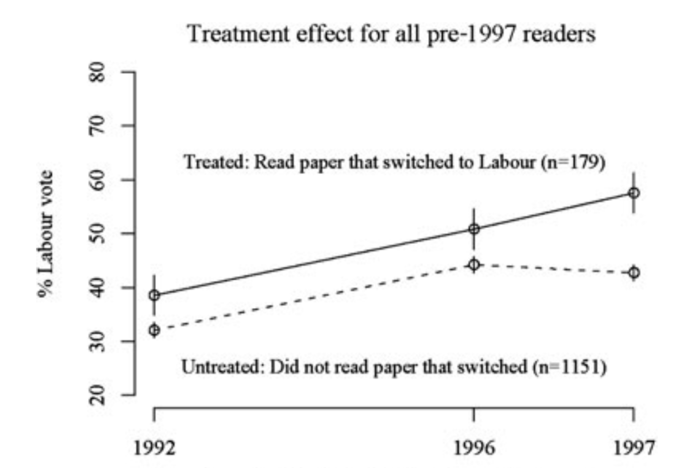
“Good” parallel trends example



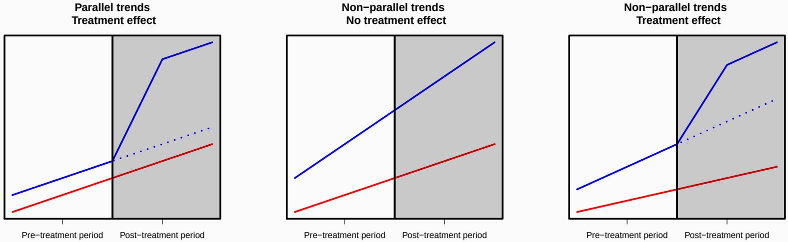
“Bad” parallel trends example



Parallel trends in Ladd and Lenz



Parallel and non-parallel trends



Multiple Periods

Estimating DiD with regression III

Estimator: Fixed-effect Regression

We can generalise to multiple groups/time periods using unit and period fixed-effects ('two-way' fixed-effect model):

$$Y_{it} = \gamma_i + \alpha_t + \delta D_{it} + \varepsilon_{it}$$

- γ_i is a fixed-effect for groups (dummy for each group)
- α_t is a fixed-effect for time periods (dummy for each time period)
- δ is the diff-in-diff estimate based on D_{it}, which is 1 for treated unit-period observations, and 0 otherwise

Very flexible:

- can replace D_{it} with almost any type of treatment (not only binary)
- can extend easily to multiple periods (i.e. more than 2)
- can have different units treated at different times

Two-way fixed-effect regression in R

```r
sun$treat  <- sun$reads_sun == 1  & sun$year == 1997
fe_model <- lm(voted_lab ~ treat + as.factor(id) + as.factor(year),
    data = sun)
```

Fixed-effect model

	Voted Labour
Intercept	0.946***
	(0.204)
Reads the Sun in 1997	0.086***
	(0.030)
Unit fixed effects?	Yes
Time fixed effects?	Yes
Observations	3,186
R^2	0.826

Note that unit dummies lead to smaller standard errors on our treatment effect.

Why not always use unit dummies?

- Only works with panel data, where we have the same units over time.
- But *not* with cross sectional data, where the units are different each time.

Why does FE regression estimate the DiD?

- With **unit/group FEs** we are holding the *unit/group* 'constant'
  - We are only using *within group* variation in Y to calculate the effect of D
  - Removes any omitted variable bias that is constant over time
- With **time FEs** we are holding *time* 'constant'
  - We are only using *within time* variation in Y to calculate the effect of D
  - Removes the effect of any changes to the outcome variable that affect all units at the same time
- **Unit and time FEs** mean that we are simultaneously adjusting for time-specific and unit-specific unobserved confounders
  - We are only using variation[2] in the changes[1] within units
  - $\hat{\delta} \rightarrow \hat{\tau}_{\text{ATT}}$
  - See this GIF for a visualisation of what fixed effects do

[2] i.e. differences

Example: Multiperiod diff-in-diff

Does lockdown prevent COVID-19 transmission?

Many countries worldwide ordered citizens to stay at home to prevent the spread of COVID-19. In the US, shelter-in-place orders (SIPO) required residents to remain in their homes for all but essential activities. How effective were these orders? Dave et. al. (2020) use data on the implementation of SIPOs between March and April 2020 at the state level in the US to study the effectiveness of local lockdowns on COVID-19 case prevalence.

- Outcome (Y): % of residents at home full time (from smartphone tracking)
- Outcome (Y): Number of confirmed COVID-19 cases (logged)
- Treatment (D): 1 if SIPO in place in state s and time t, 0 otherwise
- Time measured at the day level

Shelter in Place orders

TABLE 1
Enactment Dates of Statewide SIPOs

State	Date	State	Date
Alabama	April 4	Mississippi	April 3
Alaska	March 28	Missouri	April 6
Arizona	March 31	Montana	March 28
California	March 19	Nevada	April 1
Colorado	March 26	New Hampshire	March 28
Connecticut	March 23	New Jersey	March 21
Delaware	March 24	New Mexico	March 24
District of Columbia	April 1	New York	March 22
Florida	April 3	North Carolina	March 30
Georgia	April 3	Ohio	March 24
Hawaii	March 25	Oregon	March 23
Idaho	March 25	Pennsylvania	April 1
Illinois	March 21	Rhode Island	March 28
Indiana	March 25	South Carolina	April 7
Kansas	March 30	Tennessee	April 1
Louisiana	March 23	Texas	April 2
Maine	April 2	Vermont	March 25
Maryland	March 30	Virginia	March 30
Michigan	March 24	Washington	March 23
Minnesota	March 28	West Virginia	March 24
		Wisconsin	March 25

SIPO model specification

$$ln(COVIDCASE)_{st} = \delta * SIPO_{st} + \gamma_s + \alpha_t + \varepsilon_{st}$$

- $\gamma_s \rightarrow$ state fixed-effect
  - Controls for unobserved state-level characteristics that stay the same over time, i.e. are **time invariant**
- $\alpha_t \rightarrow$ day fixed-effect
  - Controls for (daily) changes in COVID rates over time that are common to all states, i.e. **unit invariant**
- $\delta \rightarrow$ average effect of switching from no SIPO in place to SIPO in place, among those states that see a SIPO imposed (i.e. τ_{ATT})

```r
fe_mod <- lm(covid_cases ~ sipo +
                as.factor(state) + as.factor(day),
             data = sipo_data)
```

Parallel trends in multiperiod DiD

- It is hard to provide a *visual* inspection of the parallel trends assumption here as treatment switches on at different times in different states.
- Nevertheless, we are still assuming that treated/control states would have evolved identically over time in absence of treatment.
- One way forward, test for “lags” and “leads” of the treatment

SIPO model specification with lags and leads

$$
\begin{aligned}
ln(COVIDCASE)_{st} \quad = \quad & \delta_1 * SIPO\_7\_DaysBefore_{st} + \\
& \delta_2 * SIPO\_5/6\_DaysBefore_{st} + \\
& \delta_3 * SIPO\_3/4\_DaysBefore_{st} + \\
& \delta_4 * SIPO\_1/2\_DaysBefore_{st} + \\
& \delta_5 * SIPO\_0/5\_DaysAfter_{st} + \\
& \delta_6 * SIPO\_6/9\_DaysAfter_{st} + \\
& \delta_7 * SIPO\_10/14\_DaysAfter_{st} + \\
& \delta_8 * SIPO\_15/19\_DaysAfter_{st} + \\
& \delta_9 * SIPO\_20\_DaysAfter_{st} + \\
& \gamma_s + \alpha_t + \varepsilon_{st}
\end{aligned}
$$

SIPO model specification with lags and leads

- $\delta_4 * SIPO\_1/2\_DaysBefore_{st} \rightarrow$ placebo effect
  - Measures the average difference between treatment and control before the treatment occured
- $\delta_5 * SIPO\_0/5\_DaysAfter_{st} \rightarrow$ treatment effect
  - Measures the average difference between treatment and control after the treatment occured

Implications:

1. Coefficients associated with the `DaysBefore` dummies should be zero
2. We can measure how the effect of the treatment evolved by looking at the coefficients associated with the `DaysAfter` treatment

Lags and leads SIPO model in R

```
fe_mod <- lm(covid_cases ~ sipo_7_before +
               sipo_56_before +
               sipo_34_before +
               sipo_12_before +
               sipo_05_after +
               sipo_69_after +
               sipo_1014_after +
               sipo_1519_after +
               sipo_20_after +
               as.factor(state) +  as.factor(day),
             data = sipo_data)
```

Shelter in place orders and % staying at home

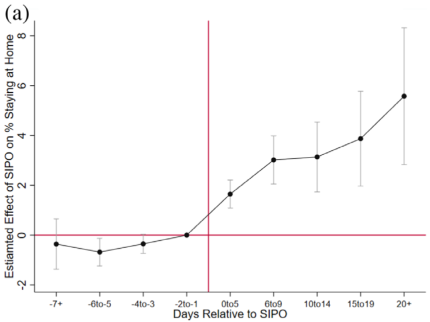
TABLE 2

Difference-in-Difference Estimates of the Effect of SIPOs on Percent of State Residents Who Remain at Home Full-Time

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel I: SIPO effect							
SIPO	2.075***	2.181***	2.264***	2.200***	2.129***	1.986***	1.995***
	(0.433)	(0.351)	(0.339)	(0.291)	(0.282)	(0.381)	(0.312)
Panel II: Lagged SIPO effect							
0–5 days after SIPO	1.935***	1.731***	1.898***	1.885***	1.795***	1.529***	1.688***
	(0.498)	(0.382)	(0.405)	(0.359)	(0.325)	(0.355)	(0.395)
6–9 days after SIPO	3.287***	2.538***	2.787***	2.849***	2.686***	2.562***	2.463***
	(0.705)	(0.490)	(0.485)	(0.400)	(0.349)	(0.473)	(0.400)
10–14 days after SIPO	3.283***	1.837***	2.125***	2.241***	2.087***	1.852***	1.822***
	(0.917)	(0.630)	(0.616)	(0.529)	(0.407)	(0.588)	(0.496)
15–19 days after SIPO	3.877***	1.423*	1.771**	1.911**	1.728***	1.346*	1.550**
	(1.164)	(0.837)	(0.827)	(0.744)	(0.542)	(0.695)	(0.695)
20 days or more after SIPO	5.364***	0.836	1.269	1.468	1.289	0.917	1.259
	(1.569)	(1.305)	(1.279)	(1.215)	(0.906)	(0.843)	(1.118)
N	2,091	2,091	2,091	2,091	2,091	2,091	2,050
State and day fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State specific linear time trend	No	Yes	Yes	Yes	Yes	Yes	Yes
Business closure order and partial SIPOs	No	No	Yes	Yes	Yes	Yes	Yes
Travel restrictions and disaster declaration	No	No	No	Yes	Yes	Yes	Yes
Weather controls	No	No	No	No	Yes	Yes	Yes
CA included?	Yes	Yes	Yes	Yes	Yes	No	Yes
NY and NJ included?	Yes	Yes	Yes	Yes	Yes	Yes	No

Shelter in place orders and % staying at home

FIGURE 3
Event-Study Analysis of Shelter in Place Orders (SIPOs) and Percent Staying at Home Full-Time



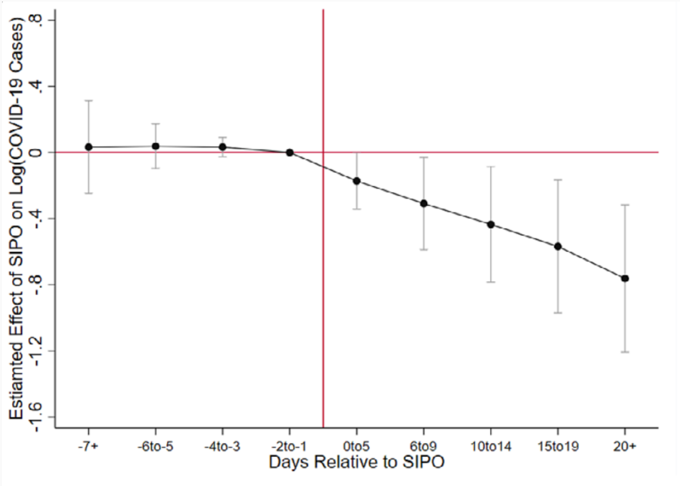
Shelter in place orders and log COVID cases

TABLE 3

Difference-in-Difference Estimates of the Effect of SIPOs on Log (COVID-19 Cases)

	(1)	(2)	(3)	(4)	(5)
1–5 days after SIPO	0.030	−0.171	−0.209*	−0.201*	−0.191*
	(0.118)	(0.116)	(0.121)	(0.109)	(0.103)
6–9 days after SIPO	−0.056	−0.319*	−0.369**	−0.341**	−0.324**
	(0.207)	(0.161)	(0.166)	(0.150)	(0.135)
10–14 days after SIPO	−0.130	−0.440**	−0.495***	−0.465***	−0.447***
	(0.275)	(0.181)	(0.183)	(0.170)	(0.154)
15–19 days after SIPO	−0.230	−0.567***	−0.628***	−0.601***	−0.577***
	(0.346)	(0.200)	(0.200)	(0.189)	(0.170)
20+ days after SIPO	−0.497	−0.740***	−0.811***	−0.791***	−0.765***
	(0.516)	(0.214)	(0.219)	(0.211)	(0.201)
N	2,100	2,100	2,100	2,100	2,100
State and day fixed effects	Yes	Yes	Yes	Yes	Yes
State specific linear time trend	No	Yes	Yes	Yes	Yes
Business closure order and partial SIPOs	No	No	Yes	Yes	Yes
Travel restrictions and disaster declaration	No	No	No	Yes	Yes
Weather controls	No	No	No	No	Yes

Shelter in place orders and log COVID cases



Threats to inference in the Covid-era[3]

1. Packaged policies/compound treatments
   - Governments often implement several policies at the same time
   - Are we identifying the effect of lockdown, or social distancing?
2. Voluntary precautions
   - Citizens may have isolated without government instruction
   - Type of omitted variable bias: the crisis could itself change behavior *and* cause government to take action
3. Spillovers
   - Do lockdowns really only affect single states?
   - People in neighbouring states may change their behaviour in response
   - Biases the DD estimate towards 0 because lockdowns affect both treatment and control.

[3] Goodman-Bacon and Marcus (2020)

Data requirements

Data requirements for Diff-in-diff

Data structure:

- Panel data or repeated cross-section
- Single or multiple treatments
- Continuous or binary treatments
- Works both at individual/aggregate level

Does this require **more** data?

- Adding a time dimension can **increase** the amount of data you need
- No need to control for extensive covariates (so long as they are fixed within units over time) which might mean **decreased** data collection

Examples of diff-in-diff designs

1. Card & Krueger, 1994
    - RQ: Do increases in the minimum wage reduce employment?
    - Outcome: Employment growth in fast-food restaurants
    - Treatment: Increased minimum wage in New Jersey; no change in Pennsylvania
    - Time: Before/after minimum wage changed
2. Dinas et al., 2019
    - RQ: What is the effect of refugee arrivals on support for the far right?
    - Outcome: Municipal support for far right party
    - Treatment: Refugee arrivals in Greek islands
    - Time: Elections before/after refugee crisis

Examples of diff-in-diff designs

3. Bechtel & Heinmueller, 2011
    - RQ: What is the effect of good policy on government support?
    - Outcome: Support for the German SPD in parliamentary constituencies
    - Treatment: Flooded German regions close to the River Elbe
    - Time: Elections before/after 2002, when the Elbe flooded
4. Hainmueller & Hangartner, 2019
    - RQ: What is the effect of direct democracy on immigrant assimilation?
    - Treatment: Whether municipality decides on naturalisation requests via expert or citizen councils
    - Time: Decisions before/after legal changes to decision making in municipalities

Conclusion

- The DiD design allows for a comparison over time in the treatment group, controlling for concurrent time trends using a control group.
- DiD requires data on multiple units in multiple periods, but can be applied to panel data or repeated cross-sectional data.
- DiD is **very** widely used, as it is a powerful conditioning strategy that doesn't require endless lists of covariates to strengthen the identifying assumption.
- The identification assumption – that treatment and control units would follow parallel trends in the absence of treatment – should be investigated with every application!