Data Science for Economists

Difference-in-differences

Kyle Coombs
Bates College | DCS/ECON 368

Table of contents

Prologue

Attribution

- These slides are adapted from work by Nick Huntington-Klein, Ed Rubin, and Scott Cunningham
- They're both superb econometric instructors and I highly recommend their work

Check-in on causality

- We've been talking about causality for a bit now
- Popcorn style: what do we need for causality?

Check-in on causality

- We've been talking about causality for a bit now
- Popcorn style: what do we need for causality?
- We need to account for everything that is correlated with treatment
- There are two extreme strategies:

1. Control variables for everything!
    - Did we control for it all? How do we know?
2. Get a treatment that is uncorrelated with everything else
    - How do we know we got rid of all the correlation?

And then there's a middle ground...

Causal Inference Review

An old causality problem: Cholera

- DiD is an old method with its first documented use by John Snow in 1854 to identify that cholera was transmitted through water[1]
- Cholera was a big problem in London in the 1800s
- Many were convinced that *miasma* was the cause of cholera
- John Snow was a doctor in London in the 1800s
- He thought it might be due to contaminated water. But how to prove it?
- An experiment was out of the question, ethically and practically
- So he needed some other variation in water quality

[1] Unlike Jon Snow, who knows nothing, John Snow knew quite a lot.

A natural experiment

- There were three water companies in London: Lambeth, Southwark and Vauxhall (SV)
- In 1849, Parliament passed a bill requiring all water companies to move their pumps further up the Thames
- Natural experiment: Lambeth water company moves its pipes between 1849 and 1854; Southwart and Vauxhall (SV) water company delayed
- John Snow went door-to-door collecting info where people got their water in 1849 and counted cholera cases in 1849 and 1854

supplier	1849	1854	diff
Non-Lambeth Only	134.9	146.6	11.7
Lambeth + Others	130.1	84.9	-45.2

- Let's think about how John Snow could have estimated the effect of the water pump move on death rates

Let's consider the counterfactual

Let's look at observed and counterfactual death rates per 10K people (a subscript 1 indicates having a pump moved)

$\text{Death}_{1,\text{Lambeth}} = 85$

$\text{Death}_{0,\text{Lambeth}} = 145$

$\text{Death}_{1,\text{SV}} = 75$

$\text{Death}_{0,\text{SV}} = 135$

Let's consider the counterfactual

Let's look at observed and counterfactual death rates per 10K people (a subscript 1 indicates having a pump moved)

$\text{Death}_{1,\text{Lambeth}} = 85$ $\quad$ $\text{Death}_{1,\text{SV}} = 75$

$\text{Death}_{0,\text{Lambeth}} = 145$ $\quad$ $\text{Death}_{0,\text{SV}} = 135$

The change in death rates differs between the companies

Let's consider the counterfactual

Let's look at observed and counterfactual death rates per 10K people (a subscript 1 indicates having a pump moved)

$\text{Death}_{1,\text{Lambeth}} = 85$ $\quad$ $\text{Death}_{1,\text{SV}} = 75$

$\text{Death}_{0,\text{Lambeth}} = 145$ $\quad$ $\text{Death}_{0,\text{SV}} = 135$

The change in death rates differs between the companies

$\tau_{\text{Lambeth}} = \text{Death}_{1,\text{Lambeth}} - \text{Death}_{0,\text{Lambeth}} = -60$

Let's consider the counterfactual

Let's look at observed and counterfactual death rates per 10K people (a subscript 1 indicates having a pump moved)

$\text{Death}_{1,\text{Lambeth}} = 85$ $\quad$ $\text{Death}_{1,\text{SV}} = 75$

$\text{Death}_{0,\text{Lambeth}} = 145$ $\quad$ $\text{Death}_{0,\text{SV}} = 135$

The change in death rates differs between the companies

$\tau_{\text{Lambeth}} = \text{Death}_{1,\text{Lambeth}} - \text{Death}_{0,\text{Lambeth}} = -60$

$\tau_{\text{SV}} = \text{Death}_{1,\text{SV}} - \text{Death}_{0,\text{SV}} = -60$

Let's consider the counterfactual

Let's look at observed and counterfactual death rates per 10K people (a subscript 1 indicates having a pump moved)

$\text{Death}_{1,\text{Lambeth}} = 85$ $\quad$ $\text{Death}_{1,\text{SV}} = 75$

$\text{Death}_{0,\text{Lambeth}} = 145$ $\quad$ $\text{Death}_{0,\text{SV}} = 135$

The change in death rates differs between the companies

$\tau_{\text{Lambeth}} = \text{Death}_{1,\text{Lambeth}} - \text{Death}_{0,\text{Lambeth}} = -60$

$\tau_{\text{SV}} = \text{Death}_{1,\text{SV}} - \text{Death}_{0,\text{SV}} = -60$

What happens if we just compare what we can observe, Lambeth to SV?

- Well Lambeth served the upper income areas (western London), so they might be healthier

Okay, but why not compare Lambeth before vs. Lambeth after?

- London was actively fighting cholera, so attribution of deaths to the pump move is difficult

Both of these are examples of selection bias

Selection bias

- We have defined selection bias as the difference between the unobserved control outcomes for the treated and the observed outcomes for the control group

$$\begin{aligned} &E(y_i \mid D_i = 1) - E(y_i \mid D_i = 0) \\ &\quad = \tau + \underbrace{E(y_{0,i} \mid D_i = 1) - E(y_{0,i} \mid D_i = 0)}_{\text{Selection bias}} \end{aligned}$$

Practical problem: Selection bias is also difficult to observe

$$\underbrace{E(y_{0,i} \mid D_i = 1)}_{\text{Unobservable}} - E(y_{0,i} \mid D_i = 0)$$

(back to the *fundamental problem of causal inference*)

Selection bias

- We have defined selection bias as the difference between the unobserved control outcomes for the treated and the observed outcomes for the control group

$$\begin{aligned} &E(y_i \mid D_i = 1) - E(y_i \mid D_i = 0) \\ &\quad = \tau + \underbrace{E(y_{0,i} \mid D_i = 1) - E(y_{0,i} \mid D_i = 0)}_{\text{Selection bias}} \end{aligned}$$

Practical problem: Selection bias is also difficult to observe

$$\underbrace{E(y_{0,i} \mid D_i = 1)}_{\text{Unobservable}} - E(y_{0,i} \mid D_i = 0)$$

(back to the *fundamental problem of causal inference*)

Bigger problem: If selection bias is present, our estimate for τ is biased, preventing us from understanding the causal effect of treatment.

Selection bias

- We have defined selection bias as the difference between the unobserved control outcomes for the treated and the observed outcomes for the control group

$$\begin{aligned} &E(y_i \mid D_i = 1) - E(y_i \mid D_i = 0) \\ &\quad = \tau + \underbrace{E(y_{0,i} \mid D_i = 1) - E(y_{0,i} \mid D_i = 0)}_{\text{Selection bias}} \end{aligned}$$

Practical problem: Selection bias is also difficult to observe

$$\underbrace{E(y_{0,i} \mid D_i = 1)}_{\text{Unobservable}} - E(y_{0,i} \mid D_i = 0)$$

(back to the *fundamental problem of causal inference*)

Bigger problem: If selection bias is present, our estimate for τ is biased, preventing us from understanding the causal effect of treatment.

Sounds a bit like omitted-variable bias, right? That's cause they're all forms of **endogeneity!**

Selection bias

- We have defined selection bias as the difference between the unobserved control outcomes for the treated and the observed outcomes for the control group

$$\begin{aligned} &E(y_i \mid D_i = 1) - E(y_i \mid D_i = 0) \\ &\quad = \tau + \underbrace{E(y_{0,i} \mid D_i = 1) - E(y_{0,i} \mid D_i = 0)}_{\text{Selection bias}} \end{aligned}$$

Practical problem: Selection bias is also difficult to observe

$$\underbrace{E(y_{0,i} \mid D_i = 1)}_{\text{Unobservable}} - E(y_{0,i} \mid D_i = 0)$$

(back to the *fundamental problem of causal inference*)

Bigger problem: If selection bias is present, our estimate for τ is biased, preventing us from understanding the causal effect of treatment.

Sounds a bit like omitted-variable bias, right? That's cause they're all forms of **endogeneity!** Our treatment variable is correlated with something that makese the two groups different.

John Snow got pre-period data

$\text{Death}_{1,\text{Lambeth}}^{\text{Post}} = 85$

$\text{Death}_{0,\text{Lambeth}}^{\text{Pre}} = 130$

$\text{Death}_{0,\text{SV}}^{\text{Post}} = 147$

$\text{Death}_{0,\text{SV}}^{\text{Pre}} = 135$

John Snow got pre-period data

$Death_{1,Lambeth}^{Post} = 85$

$Death_{0,Lambeth}^{Pre} = 130$

$Death_{0,SV}^{Post} = 147$

$Death_{0,SV}^{Pre} = 135$

- We can take the average of observations and subtract it from each observation
- Individual fixed effect for Lambeth on their post-treatment observation:

$Death_{1,Lambeth}^{Post}$ - Avg $Death_{Lambeth} = 85 - 107.5 = -22.5$

- This is a method called difference-in-differences

1. Difference in the groups' means before treatment
2. Difference in the groups' means after treatment
3. Difference in these differences

Bridge from fixed effects

- Note that we used group averages to demean the data of *between* variation, leaving us just *within* variation
- That's a fixed effect!
    - The year fixed effect removes the variation both groups experience over time
    - The group fixed effect removes the variation within each group in each periods
- The untreated group, or control group, is our **counterfactual**
- Then, we compare the within-variation for the treated group vs. the within-variation for the untreated group
- Voila, we have an effect as long as the *within* variation left over is as good as randomly assigned, we'll have causality
- Put another way, as long as nothing else affects the outcome for the treated group between the pre- and post-periods, we'll have causality
- Note: other irrelevant things can change, but as long as the treatment is the only thing that changes, we'll be good

Difference-in-Differences

What *changes* are included in each value?

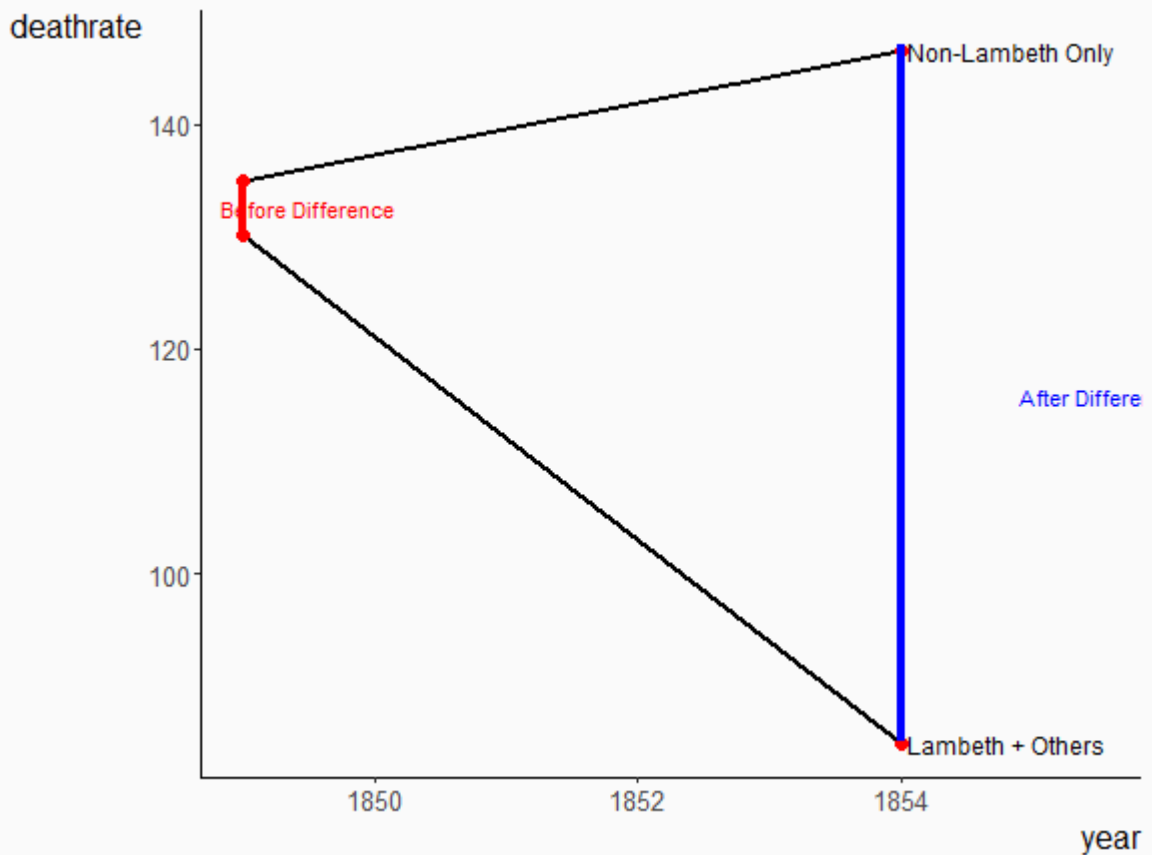
- Untreated Before: Untreated Group Mean
- Untreated After: Untreated Group Mean + Time Effect
- **Treated Before**: Treated Group Mean
- **Treated After**: Treated Group Mean + Time Effect + **Treatment Effect**
- Untreated After - Untreated Before = Time Effect
- **Treated After** - **Treated Before** = Time Effect + **Treatment Effect**

DiD = (**TA** - **TB**) - (UA - UB) = **Treatment Effect**

(Abbreviations for Untreated and Treated Before/After to save space)

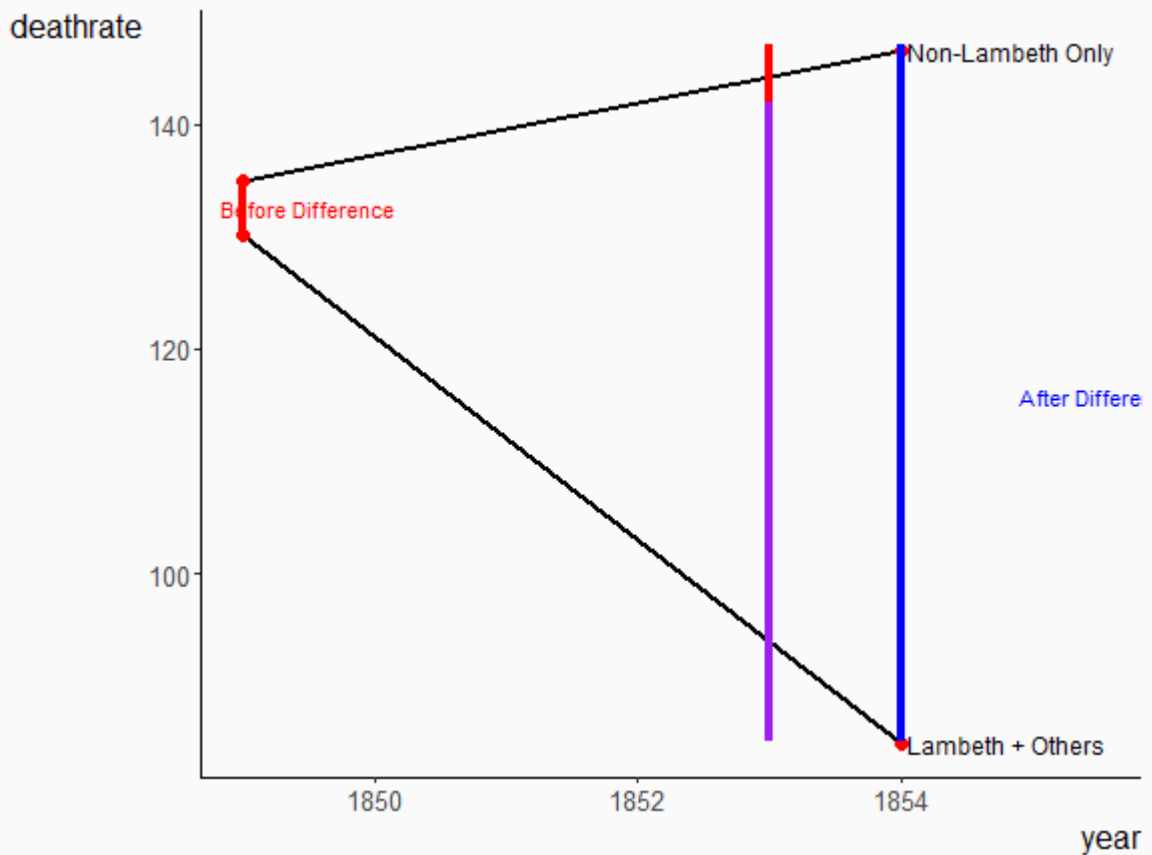
Visualizing Lambeth vs SV

supplier	1849	1854	diff
Non-Lambeth Only	134.9	146.6	11.7
Lambeth + Others	130.1	84.9	-45.2



Visualizing Lambeth vs SV

supplier	1849	1854	diff
Non-Lambeth Only	134.9	146.6	11.7
Lambeth + Others	130.1	84.9	-45.2



Difference-in-Differences

- Because the requirements to use it are so low, DiD is used a *lot*
- At its simplest, we need a treatment that *goes into effect* at a particular time, and we need a group that is *treated* and a group that is *not*
- Any time a policy is enacted but isn't enacted everywhere at once? DiD!
- Plus, the logic is pretty straightforward
- The question DiD tries to answer is "what was the effect of (some policy) on the people who were affected by it?"
- We have some data on the people who were affected both before the policy went into effect and after
- However, we can't just compare before and after, because things usually change over time for other reasons
- So we compare to people who weren't affected by the policy

Difference-in-Difference

- What if there are more than four data points?
- Usually these four points would be four means from lots of observations, not just two people in two time periods
- How can we do this and get things like standard errors, and perhaps include controls?
- Why, use OLS regression of course, just use binary variables and interaction terms to get a DiD

$$Y_{it} = \beta_0 + \beta_1 After_t + \beta_2 Treated_i + \beta_3 After_t \times Treated_i + \varepsilon_{it}$$

where $After_t$ is a binary variable for being in the post-treatment period, and $Treated_t$ is a binary variable for being in the treated group

```
## # A tibble: 4 × 6
##    year supplier         treatment           deathrate After Treated
##   <dbl> <chr>            <chr>                   <dbl> <lgl> <lgl>
## 1  1849 Non-Lambeth Only Dirty                    135.  FALSE FALSE
## 2  1849 Lambeth + Others Mix Dirty and Clean      130.  FALSE TRUE
## 3  1854 Non-Lambeth Only Dirty                    147.  TRUE  FALSE
## 4  1854 Lambeth + Others Mix Dirty and Clean       84.9 TRUE  TRUE
```

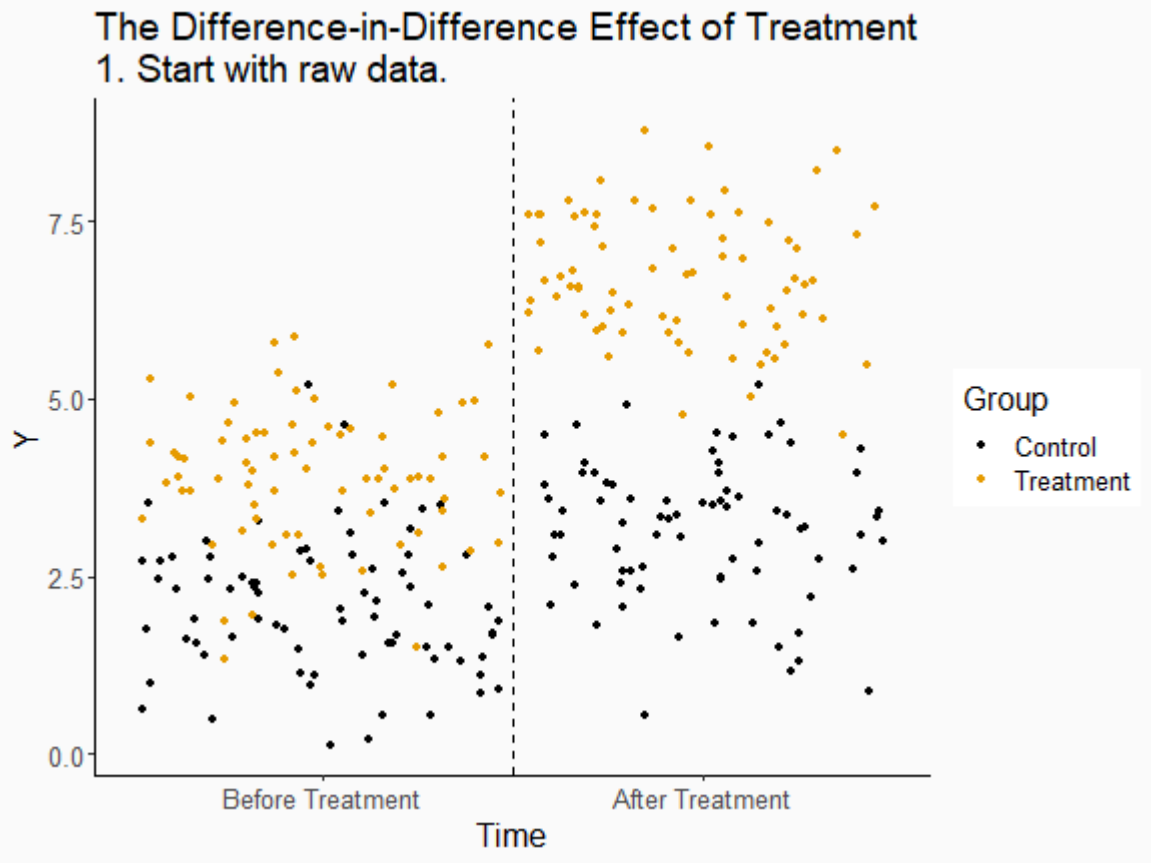
Difference-in-Differences

- How can we interpret this using what we know?

$$Y_{it} = \beta_0 + \beta_1 After_t + \beta_2 Treated_i + \beta_3 After_t \times Treated_i + \varepsilon_{it}$$

- β_0 is the prediction when $Treated_i = 0$ and $After_t = 0 \rightarrow$ the Untreated Before mean!
- β_1 is the *time difference* for $Treated_i = 0 \rightarrow$ UA - UB
- β_2 is the *treatment difference* for $After_t = 0 \rightarrow$ BT-BU
- β_3 is *how much bigger the Before-After difference* is for $Treated_i = 1$ than for $Treated_i = 0 \rightarrow$ (**TA** - **TB**) - (UA - UB) = DID!

Graphically



Design vs. Regression

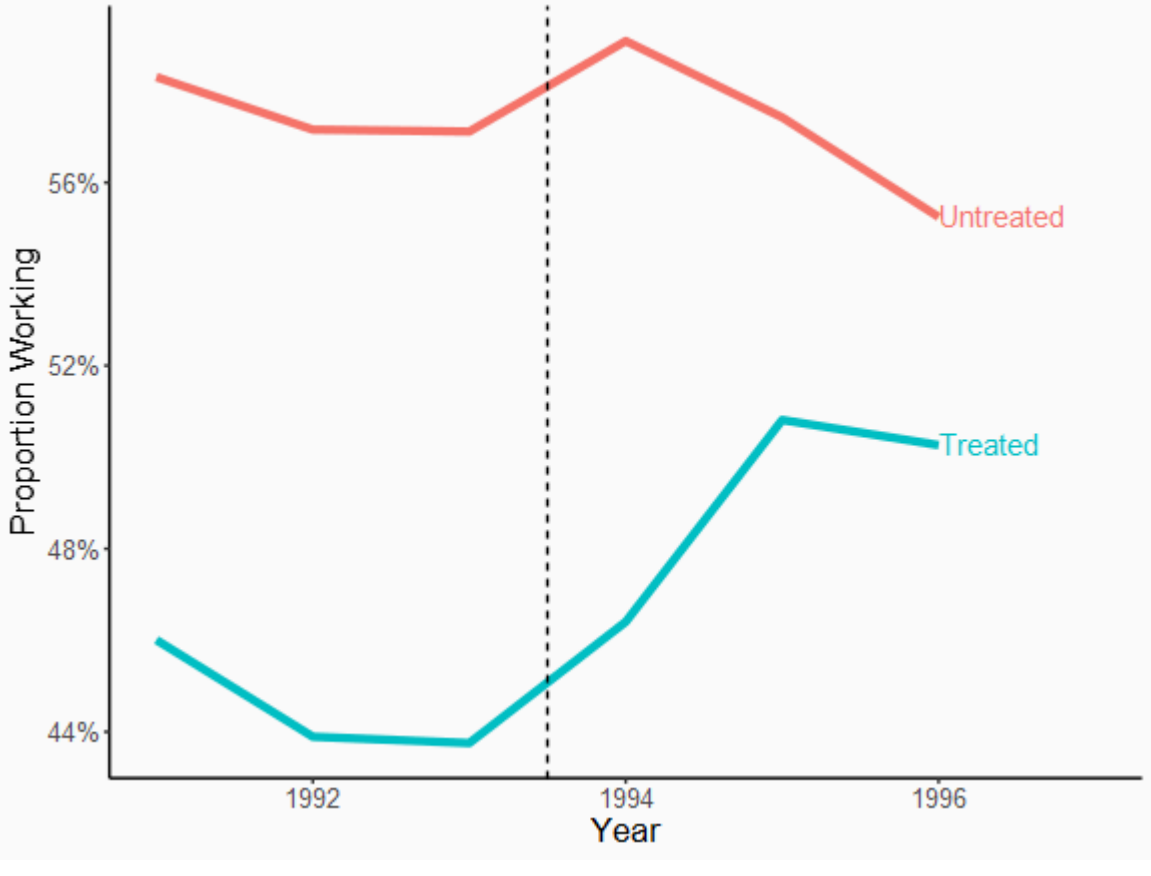
- There is a distinction between *regression model* and *research design*
    - This is also true for fixed effects
- We have a model with an interaction term
- Not all models with interaction terms are DID!
- It's DID because it's an interaction between treated/control and before/after
- If you don't have a before/after, or you don't have a control group, that same setup may tell you something interesting but it won't be DID!

Parallel Trends

- This assumption - that nothing else changes at the same time, is the poorly-named "parallel trends"
- Again, this assumes that, *if the Treatment hadn't happened to anyone*, the gap between the two would have stayed the same
- You can if *prior trends* are the same - if we have multiple pre-treatment periods, was the gap changing a lot during that period?
- There are methods to "adjust for prior trends" to fix parallel trends violations, or use related methods like Synthetic Control
    - These are beyond the scope of this class and also often snake oil

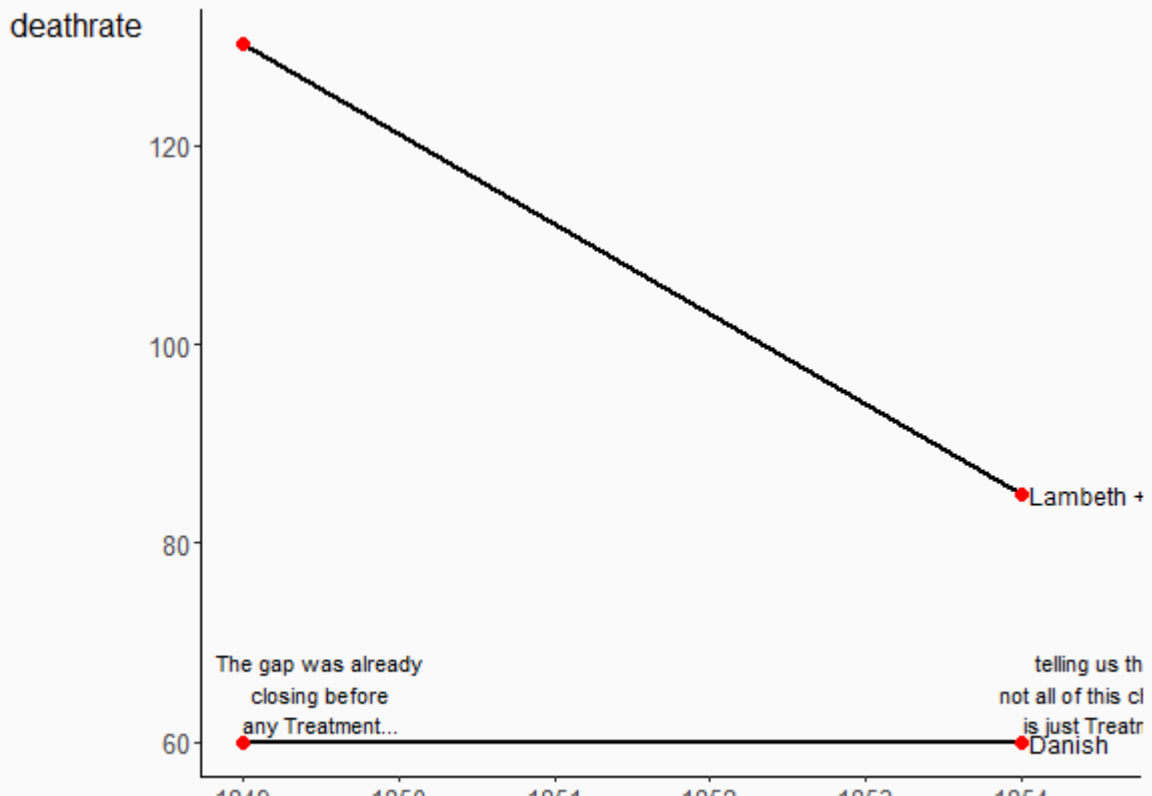
Prior Trends

- Let's look at an example involving an expasnion of the EITC in 1993
- Looks like the gap between them is pretty constant before 1994! They move up and down but the *gap* stays the same. That's good.



Prior Trends

- Formally, prior trends being the same tells us nothing about parallel trends
- But it can be suggestive if the gap was closing *anyway*
    - For example, what if you compare cholera rates between Lambeth users and Denmark, which was relatively cholera-free due to a quarantine



Parallel Trends

- Just because *prior trends* are equal doesn't mean that *parallel trends* holds.
- *Parallel trends* is about what the before-after change *would have been* - we can't see that!

Parallel Trends

- Just because *prior trends* are equal doesn't mean that *parallel trends* holds.
- *Parallel trends* is about what the before-after change *would have been* - we can't see that!
- For example, let's say we want to see the effect of online teaching on student test scores, using COVID school shutdowns to get a Before/After
- As of March/April 2020, some schools had gone online (Treated) and others hadn't (Untreated)
- Test score trends were probably pretty similar in the Before periods (Jan/Feb 2020), so prior trends are likely the same
- But LOTS of stuff changed between Jan/Feb and Mar/Apr, like, uh, Coronavirus, lockdowns, etc. not just online teaching! SO *parallel trends* likely wouldn't hold

What if there are more groups?

- You can recognize $(\mathbf{Year} \geq 1994)_t$ and $\mathbf{Has\ Kids}_i$ as *fixed effects*
- $\boldsymbol{Treated}_i$ is a fixed effect for group - we only need one coefficient for it since there are only two groups
- And $(\mathbf{Year} \geq 1994)_t$ is a fixed effect for *time* - one coefficient for two time periods
- You can have more than one set of fixed effects like this! Our interpretation is now within-group *and* within-time
- (i.e. comparing the within-group variation across groups)

Example: Multiple Groups

Multiple Treated and Control Groups

- We can extend DID to having more than two groups, some of which get treated and some of which don't
- And more than two time periods! Multiple before and/or multiple after
- We don't have a full set of interaction terms, we still only need the one, which we can now call $CurrentlyTreated_{it}$
- **If you have more than two groups and/or more than two time periods, then this is what you should be doing**

Multiple treatment groups

- Let's make some quick example data to show this off, with the first treated period being period 7 and the treated groups being 1 and 9, and a true effect of 3

```
did_data ← tibble(group = sort(rep(1:10, 10)),
                  time = rep(1:10, 10)) %>%
  mutate(CurrentlyTreated  = group %in% c(1,9) & time ⩾ 7) %>%
  mutate(Outcome = group + time + 3*CurrentlyTreated + rnorm(100))
did_data
```

```
## # A tibble: 100 × 4
##    group  time CurrentlyTreated Outcome
##    <int> <int> <lgl>              <dbl>
##  1     1     1 FALSE               1.50
##  2     1     2 FALSE               2.43
##  3     1     3 FALSE               3.70
##  4     1     4 FALSE               5.75
##  5     1     5 FALSE               4.57
##  6     1     6 FALSE               6.47
##  7     1     7 TRUE               11.1
##  8     1     8 TRUE               12.6
##  9     1     9 TRUE               12.6
## 10     1    10 TRUE               13.6
## # ℹ 90 more rows
```

Multiple treatment groups

```
# Put group first so the clustering is on group
many_periods_did ← feols(Outcome ~ CurrentlyTreated | group + time, data = did_data)
etable(many_periods_did)
```

```
##                        many_periods_did
## Dependent Var.:                Outcome
##
## CurrentlyTreatedTRUE 3.095*** (0.4991)
## Fixed-Effects:       -----------------
## group                             Yes
## time                              Yes
## ____________________ _________________
## S.E.: Clustered               by: group
## Observations                        100
## R2                              0.96131
## Within R2                       0.34238
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Two-way fixed effects model

- We just ran a two-way fixed effects model
- We have a fixed effect for group and a fixed effect for time
- This is often done when we have a panel dataset with individuals over time
- The generic formula is:

$$y_{it} = \underbrace{\alpha_i}_{\text{Individual FE}} + \underbrace{\alpha_t}_{\text{Time FE}} + \beta CurrentlyTreated_{it} + \varepsilon_{it}$$

- Where did the treatment and pre-post indicator go?

Two-way fixed effects model

- We just ran a two-way fixed effects model
- We have a fixed effect for group and a fixed effect for time
- This is often done when we have a panel dataset with individuals over time
- The generic formula is:

$$y_{it} = \underbrace{\alpha_i}_{\text{Individual FE}} + \underbrace{\alpha_t}_{\text{Time FE}} + \beta CurrentlyTreated_{it} + \varepsilon_{it}$$

- Where did the treatment and pre-post indicator go?
- They're in the fixed effects! We don't need to include them separately

Downers and Assumptions

- So... does this all work?
- That example got pretty close to the truth of 3 but who knows in other cases![1]
- What needs to be true for this to work?

Downers and Assumptions

- So... does this all work?
- That example got pretty close to the truth of 3 but who knows in other cases![1]
- What needs to be true for this to work?
- DID and TWFE give a causal effect as long as *the only reason the gap changed* was the treatment
    - e.g. If Al's earnings were going up $40K *anyway*, then we'd mistakenly attribute $30K of that to college
- For TWFE to have a causal effect with panel data, we assume no endogenous variation across time within unit
    - It gets even messier if we have staggered rollout of treatment
- In DID, we need to assume that there's no endogenous variation *across this particular before/after time change*
- An easier assumption to justify but still an assumption!

[1] We do know: it fails a lot.

Before we finish, a warning!

- DID is so nice and simple that it feels like you can get real flexible with it
- But the stuff we're covering in this class - up to TWFE, *relies very strongly* on the assumptions we made.
- If you break them, the *research design* may hold up, *but the estimator really doesn't* and you may need a different estimator

Context

- TWFE *quickly falls apart* if you have different groups getting treated at different times, called "staggered treatment"

    1. **Forbidden comparison**: Your early treated group becomes a control for your later treated group

    2. If the effect increases/decreases in relative time, your early treated group gives a "bad comparison"

    3. See slides by Andrew Baker for an example explanation of the problem

TWFE with Het. Treatment Timing

Let's create a simple example with three groups and different treatment times:

```r
# Create example data
set.seed(123)
n_periods ← 10; n_groups ← 3; n_obs ← n_periods * n_groups

twfe_data ← data.frame(id = rep(1:n_groups, each = n_periods),
  time = rep(1:n_periods, times = n_groups)) %>%
  mutate( # Group 1 never treated, Group 2 treated at t=5, Group 3 treated at t=8
    treated = (id == 2 & time ⩾ 5) | (id == 3 & time ⩾ 8),
    # True treatment effects: Group 2: effect = 2, Group 3: effect = 4
    true_effect = case_when(
      id == 2 & treated ~ 2,
      id == 3 & treated ~ 4,
      TRUE ~ 0
    ),
    # Generate outcome with: Group fixed effects (id), Time fixed effects (time), Treatment effect
    y = id + time + true_effect + rnorm(n_obs)
  )
```

TWFE with Het. Treatment Timing

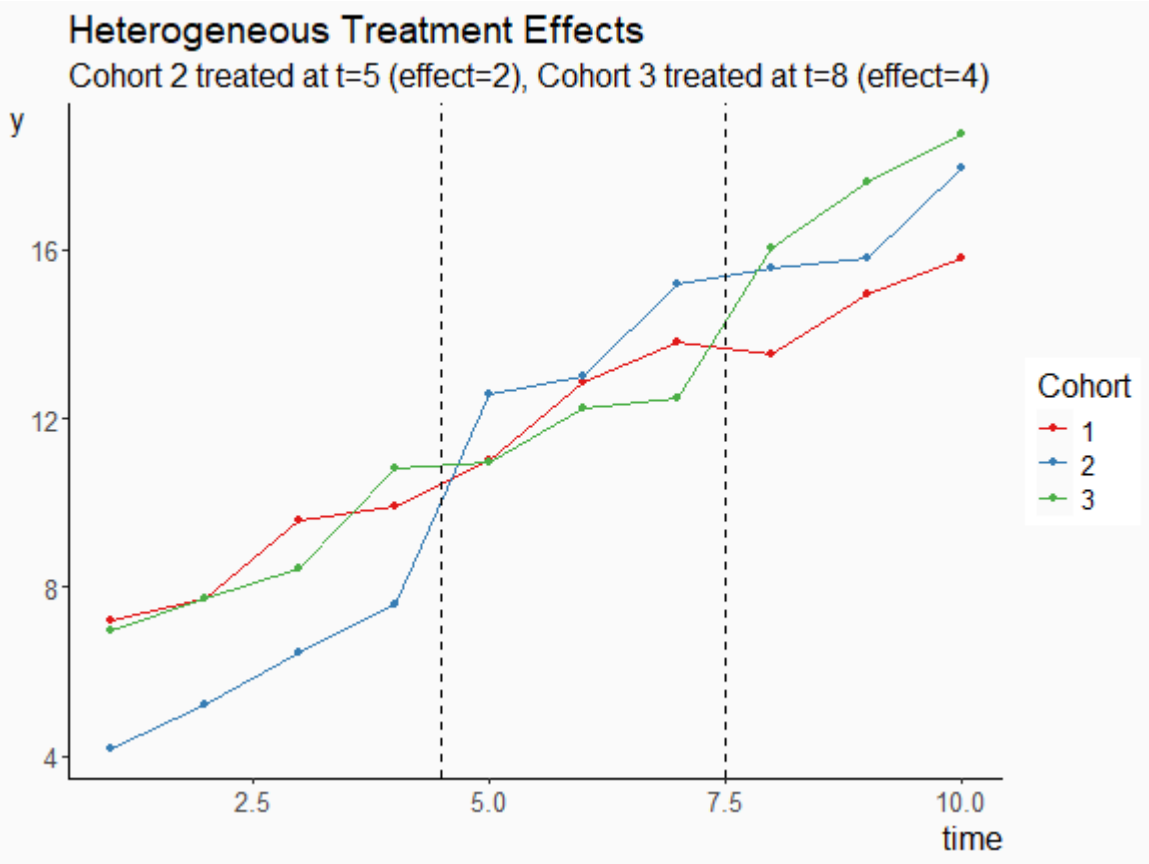
Treatment effects were:

1. Group 2: 2 after t=5
2. Group 3: 4 after t=8

```
# Run TWFE regression
twfe_model ← feols(y ~ treated | id + time, data = twfe_data)
etable(twfe_model) %>% kable(format="markdown")
```

	twfe_model
Dependent Var.:	y
treatedTRUE	3.628 (1.554)
Fixed-Effects:	----------
id	Yes
time	Yes
–	–
S.E.: Clustered	by: id

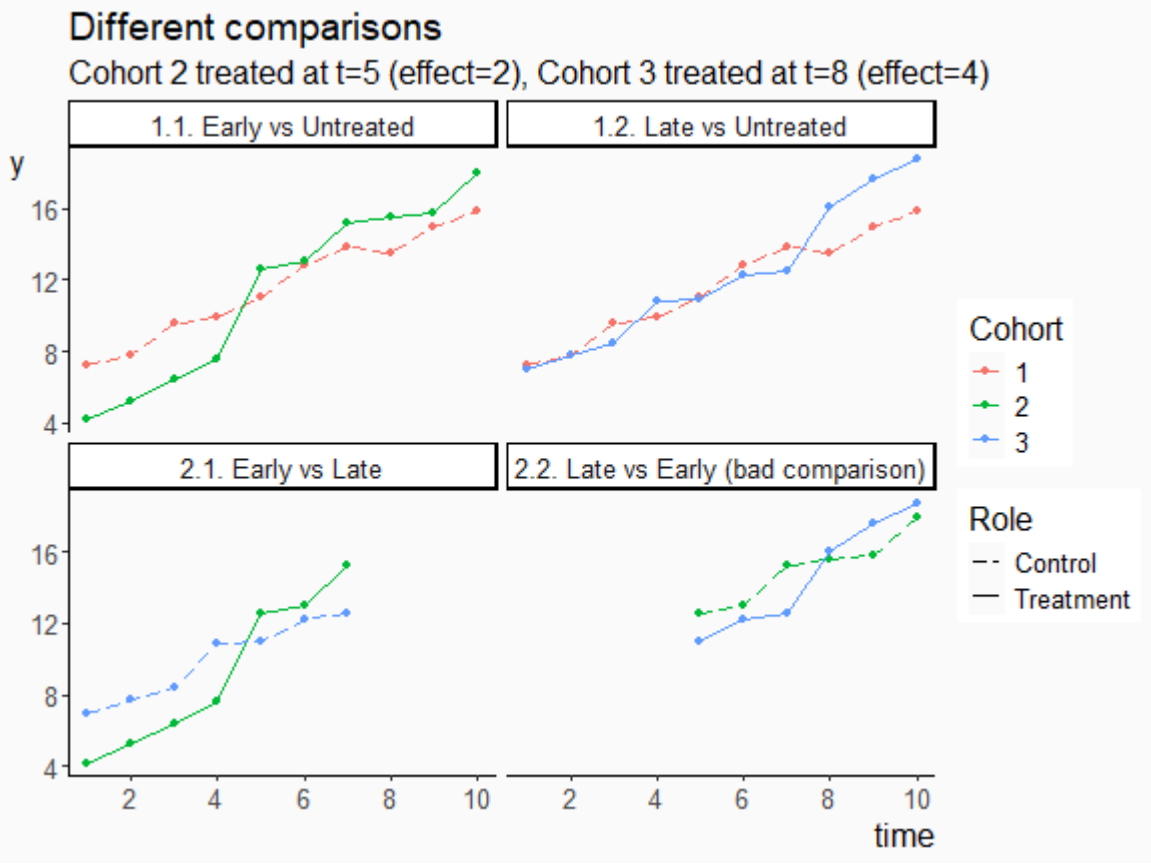
TWFE with Het. Treatment Timing



The TWFE estimate is biased because:

1. Early treated units (Group 2) become controls for later treated units (Group 3)
2. The treatment effects are different across groups (2 vs 4)
3. TWFE weights these comparisons incorrectly

Different comparisons

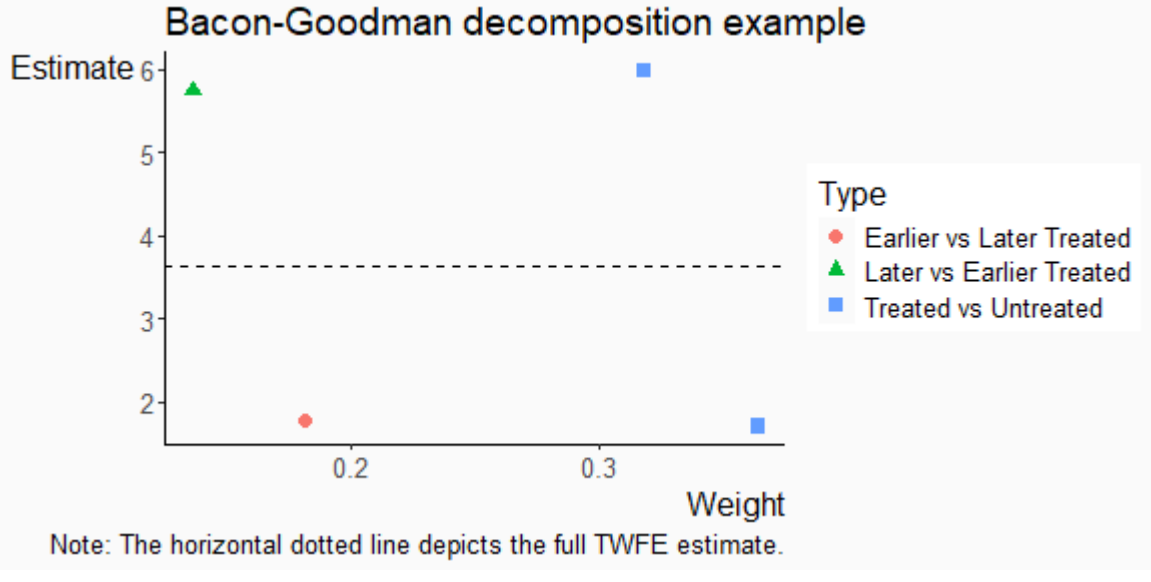


Bacon-Decomposition

```
bgd ← bacon(y ~ treated, data = twfe_data, id = "id", time = "time")
```

```
##                        type  weight avg_est
## 1 Earlier vs Later Treated 0.18182 1.78361
## 2 Later vs Earlier Treated 0.13636 5.73666
## 3     Treated vs Untreated 0.68182 3.69853
```

```
## [1] 3.628289
```



Fixes

- There are many fixes for this problem
- Many go beyond the scope of the class
- But you have the starting tools to understand them
- Also, many are written up to implement using R, Stata, Python, etc.