Difference-in-differences

Christoph Hanck

Summer 2023

UDE

Across within variation

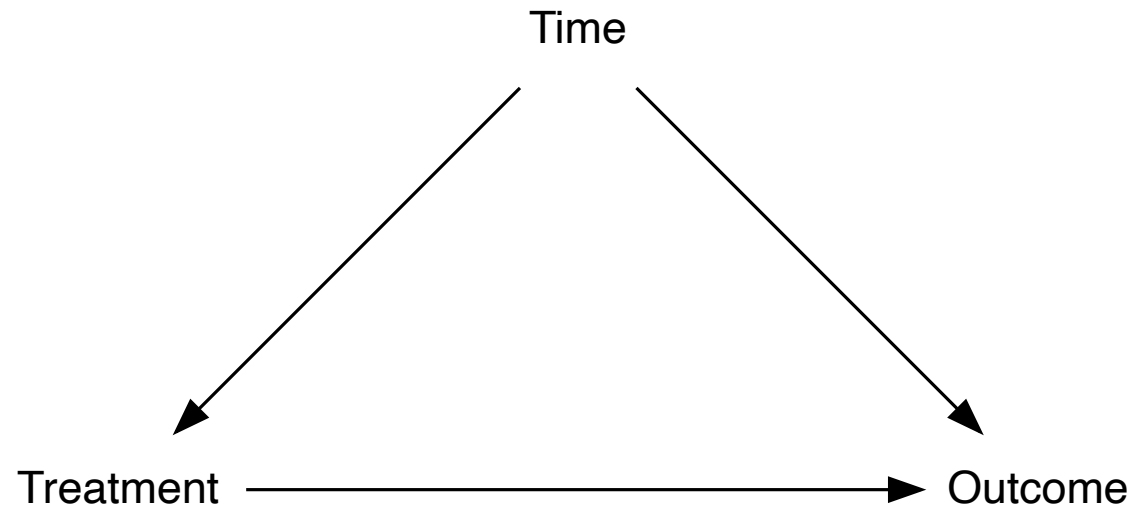**Time can be a back door**

Time

Treatment ⟶ Outcome

Figure 2: Basic time backdoor

Across within variation

Closing a back door with another back door

- We often ask how much of the change in the world is due to a treatment that occurred at a particular *time* which is a back door

- Identifying the treatment effect (closing the back door) is difficult: *all* variation in $Treatment$ is explained by $Time$:

  Individuals are either in a before-treatment time and untreated, or in an after-treatment time and treated.

- Event studies solve this problem by using before-treatment information to *construct* a counterfactual

- Differences-in-differences (DiD) considers a second group that is *never* treated (another back door) to introduce within variation to compare to that group which may be different!
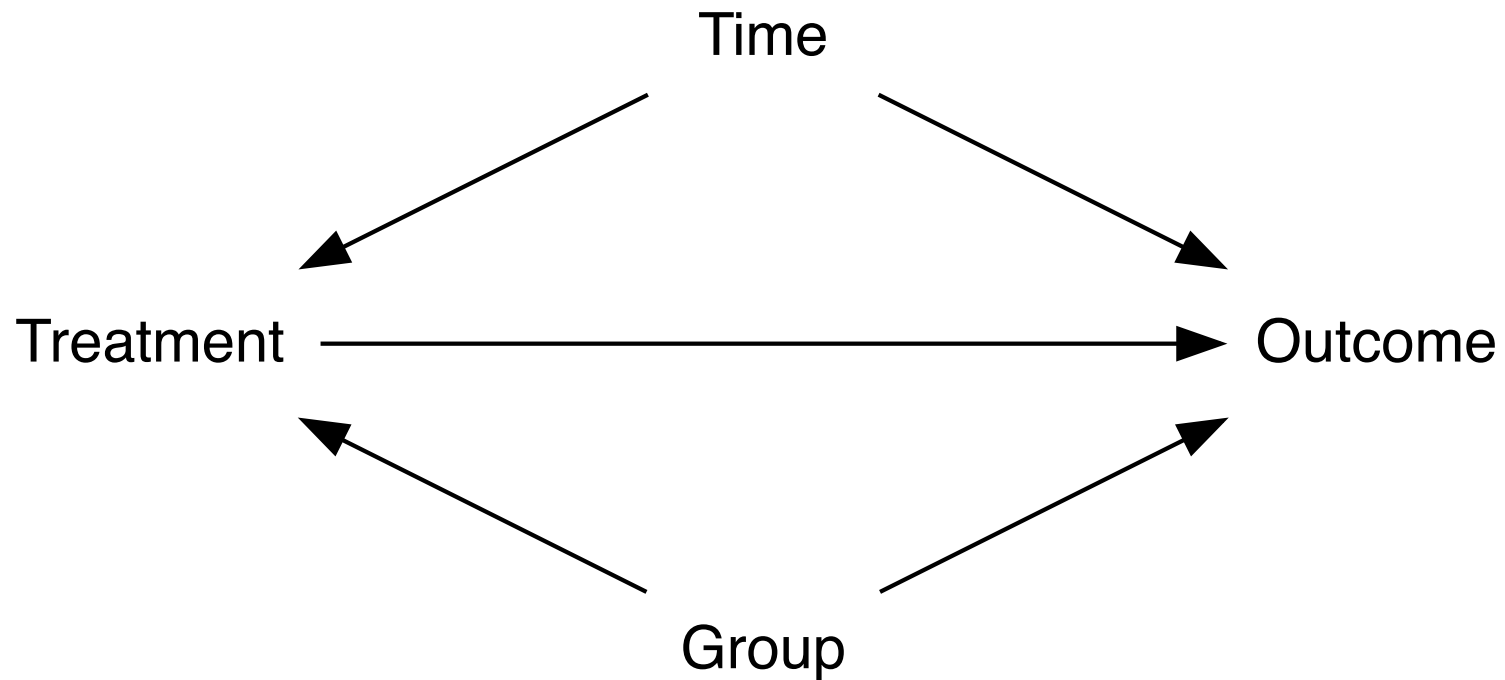
Across within variation

Figure 4: Two back doors that can be closed by DiD

Across within variation

DiD closes a back door with another back door

Steps

1. **Obtain *within* differences**:

   Control for group differences by isolating the *within* variation for both the treated group and untreated group. This **closes back doors through *Group***.

2. **Obtain differences in the differences**:

   Compare the within variation in the treated group to the within variation in the untreated group. Because the within variation in the untreated group is affected by time, this controls for time differences and **closes the back door through *Time***.

Difference-in-differences

Treatment effects in DiD

- DiD compares what is seen for the treated group after treatment against the best guess at what the treatment group would have been without treatment

- The difference between being treated and not being treated for the group that actually gets treated is being isolated

  → **average treatment on the treated**

- The DiD estimate is all about how effective the treatment was for the groups that actually got it

Difference-in-differences

> ## Example: Difference-in-differences and dirty water
>
> - Snow's (1855)[1] study on cholera spreading by dirty drinking water is very similar to a modern-day difference-in-differences research design, and can be easily discussed in those terms (Coleman 2019).
>
> - Water taken in from the parts of the Thames that were downstream of London (by the Lambeth Company) contained everything that Londoners dumped in the river, including plenty of fecal matter from people infected with cholera

[1] Snow, John. 1855. *On the Mode of Communication of Cholera*. John Churchill.

🏠

Difference-in-differences

Example: Difference-in-differences and dirty water

- Between 1849 and 1854 (before and after periods), a policy required the Lambeth Company to move their water intake upstream of London

- Lambeth moving their intake source gives the **treated group:** people in areas where the water came from the Lambeth company, and an **untreated group:** anyone in an area without Lambeth's water supply. (After all, maybe Cholera simply went away for some other reasons)

Research Question: Did areas getting water from Lambeth see their Cholera numbers go down from 1849 to 1854 relative to areas getting no water from Lambeth?

Difference-in-differences

> ## Example: Effect of Lambeth on cholera
>
> | Region supplier | Death rates (1849) | Death rates (1854) |
> |---|:---:|:---:|
> | Non-Lambeth only (dirty) | 134.9 | 146.6 |
> | Lambeth + others (mix dirty and clean) | 130.1 | 84.9 |
>
> Note: Death rates are deaths per 10,000 of the 1851 population.
>
> This yields a treatment effect of $(85 - 130) - (147 - 135) \approx -57$.

Mechanics of DiD

> ## Example: Effect of active choice on organ donor
>
> - People are assumed not to be organ donors in most states of the US. They *may opt-in* to be a potential organ donor when signing up for a driver's license.
>
> - In some states there is an *active choice* rule: when one signs up for a driver's license, one is *asked to choose whether or not* to be a donor.
>
> - In July 2011, California switched from an opt-in to an active choice rule (a treatment!)

[2] Kessler, Judd B., and Alvin E. Roth. 2014. *Don't Take 'No' for an Answer: An Experiment with Actual Organ Donor Registrations*. National Bureau of Economic Research.

Mechanics of DiD

Example: Effect of active choice on organ donor

- Kessler and Roth (2014)[2] compare California against the twenty-five states that either have opt-in or a verbally given question with no fixed response (a difference).

- Specifically, they compared the states on the basis of how their organ donation rates changed from before July 2011 to after (a difference in differences).

![home icon]

Mechanics of DiD

Example: Effect of active choice on organ donor
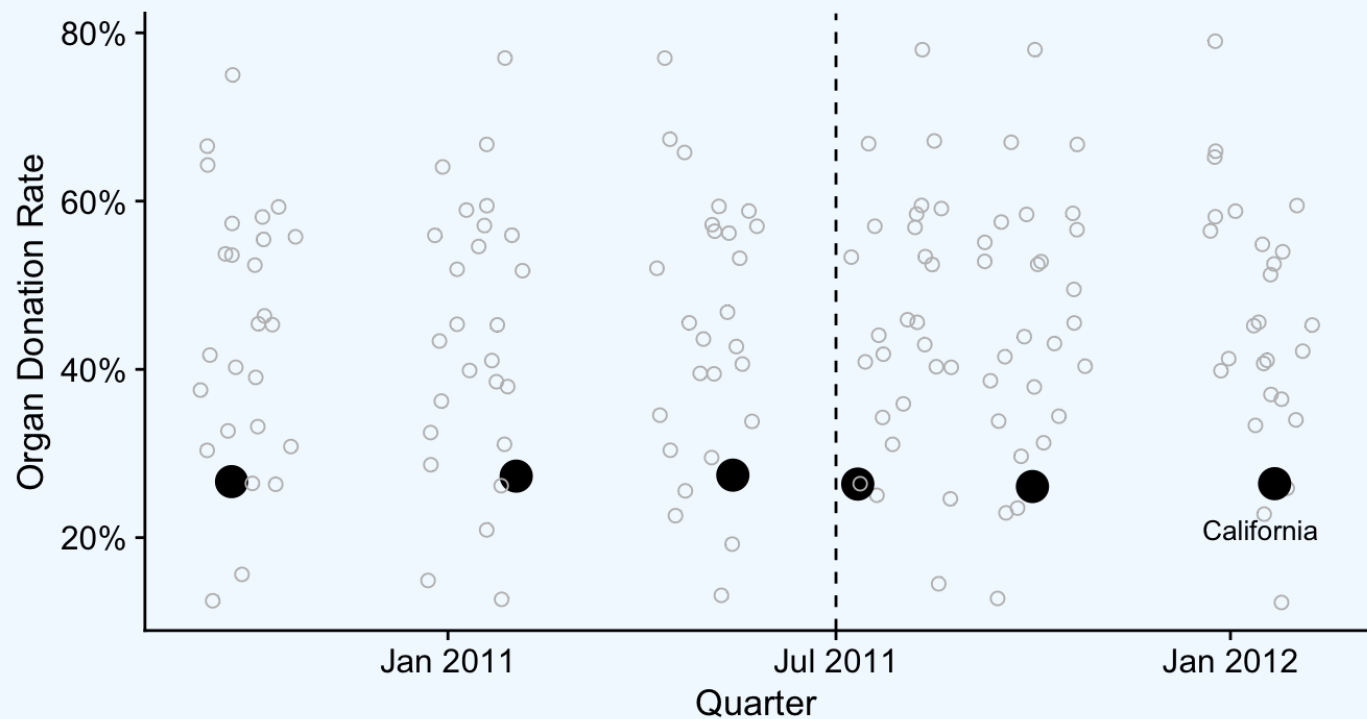
Jitter has been added to the x-axis to make points easier to see, since data is quarterly.

Figure 5: Organ donation rates in California and other states
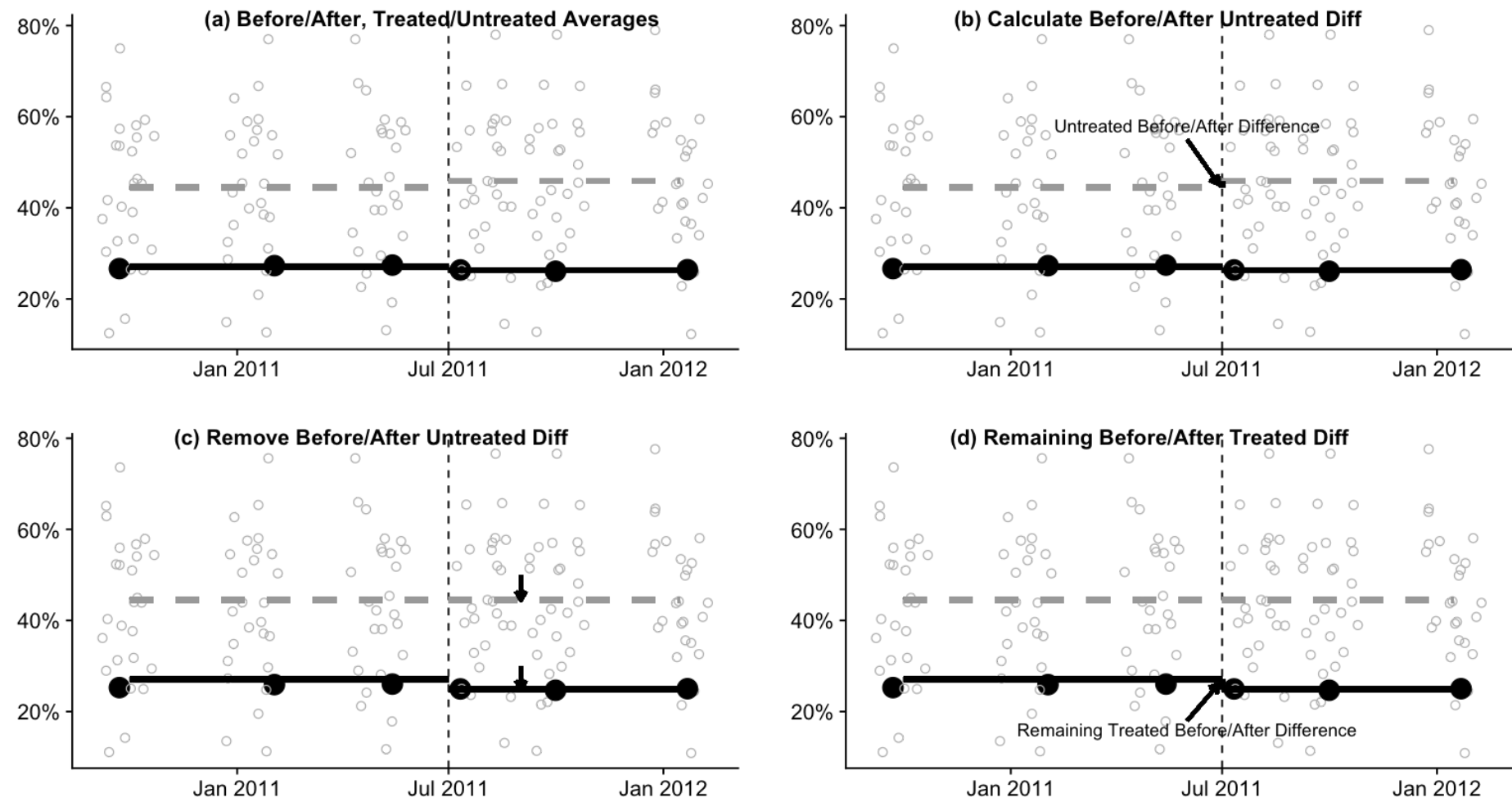
Example: Effect of active choice on organ doner

Figure 6: Organ donation rates in California and other states

Untreated groups and parallel trends

Parallel trends assumption

- If no treatment had occurred, the difference between the treated group and the untreated group would have stayed the same in the post-treatment period as it was in the pre-treatment period

- Parallel trends is *inherently unobservable*: it is about the counterfactual of what would have happened if treatment had not occurred.

Untreated groups and parallel trends

Parallel trends assumption

- The DiD design targets the change in the untreated group to represent all non-treatment changes in the treated group.

- That way, once the untreated group's change is subtracted, only the treated group's change is left.

- DiD cannot work without parallel trends:

  If the gap between the two groups would have changed from the pre-period to the post-period without treatment, then this non-treatment-related change will get mixed up with the treatment-related change.

Untreated groups and parallel trends

Parallel trends assumption

- The difference between pre-treatment and post-treatment in the treated group is

$$EffectOfTreatment + OtherTreatedGroupChanges$$

- $OtherUntreatedGroupChanges$ is the difference between pre-treatment and post-treatment in the untreated group

- Difference-in-difference subtracts one from the other, giving us

$$EffectOfTreatment + OtherTreatedGroupChanges - OtherUntreatedGroupChanges$$

Untreated groups and parallel trends

How to pick an untreated comparison group if a parallel trend is required?

Outcome of the untreated group must change by the same amount as the treated group (if treatment had occurred)

There are a few good signs that can be looked for:

- The treated group and untreated groups are generally similar

- The untreated group is unlikely to suddenly change around the time of treatment

- The treated group and untreated groups had similar trajectories for the dependent variable before treatment

    → Check prior trends, placebo test

Untreated groups and parallel trends

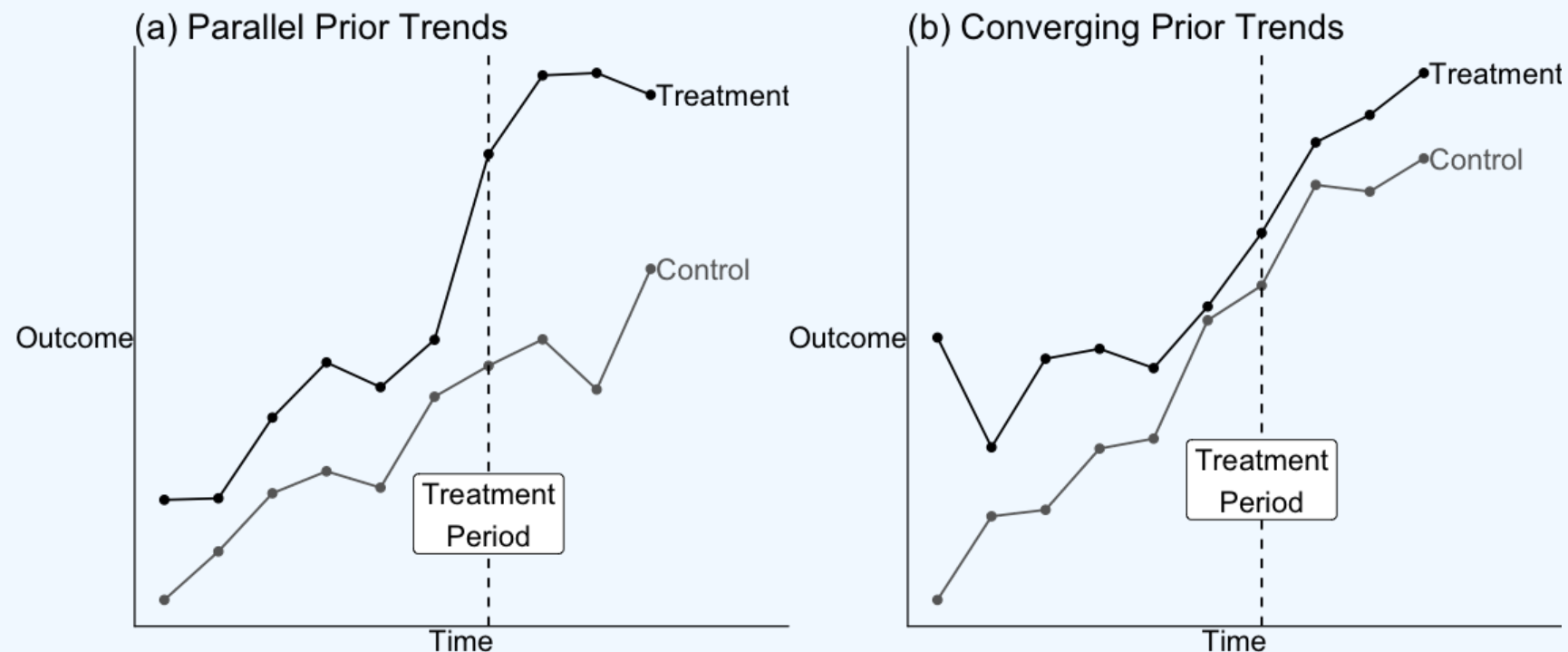## Example: Plausible vs. implausible prior trends

Figure 7: A Graph Where the Prior Trends Test Looks Good for DID, and a Graph Where It Does not

Untreated groups and parallel trends

Placebo test for DiD

In a placebo test, a situation is taken where a treatment was applied.

1. Ignore all the data from the periods where treatment was actually applied

2. Use pre-treatment data, pick a few different periods and pretend that a treatment was applied at that time.

3. Obtain DiD estimates at the pretended treatment dates.

If a DID "effect" is found consistently at those pretended treatment dates, that indicates that something may be awry about the parallel trends assumption.

How to perform DiD?

Two-way fixed effects

The goal here is to control for group differences, and also control for time differences. The regression is

$$Y = \alpha_g + \alpha_t + \beta_1 Treated + \epsilon.$$

- α_g is a set of fixed effects for the group — in the simplest form, just "treated" or "untreated"
- α_t is a set of fixed effects for the time period just "before treatment" and "after treatment"
- $Treated$ is a binary variable indicating that one is being treated right now
- β_1 is the difference-in-differences effect

Control variables that change over time can be incorporated.

How to perform DiD?

Two-way fixed effects

Another way to write the same model equation if only two groups and two time periods is present is

$$Y = \beta_0 + \beta_1 \text{TreatedGroup} + \beta_2 \text{AfterTreatment} + \beta_3 \text{TreatedGroup} \times \text{AfterTreatment} + \epsilon.$$

- TreatedGroup is an indicator of the group being treated

- AfterTreatment is an indicator of the "post"-treatment period

- The third term is an interaction term, in effect an indicator for being in the treated group and in the post-treatment period

🏠

How to perform DiD?

Two-way fixed effects

- This third term is equivalent to $\mathrm{Treated}$ in the last equation, and $\widehat{\beta}_3$ is the DiD estimate

- By standard interaction-term interpretation, β_3 tells us how much bigger the $\mathrm{TreatedGroup}$ effect is in the $\mathrm{AfterTreatment}$ than in the before-period

- Whichever way the equation is written, this approach is called the *two-way fixed effects DiD estimator* since it has two sets of fixed effects, one for group and one for time period

- This model is generally estimated using standard errors that are clustered at the group level

Two-way fixed effects

Advantages

- Highly intuitive, straightforward to use with software that implements fixed effects estimation

- Allows to account for *multi-group designs* where some groups are treated and some are not, rather than just one treated and untreated group.

Disadvantage

Two-way fixed effects does not work very well for **rollout designs** also known as **staggered treatment timing**, where the treatment is applied at different times to different groups.

Two-way fixed effects

The test of prior trends

The simplest form uses the regression model

$$Y = \alpha_g + \beta_1 Time + \beta_2 Time \times Group + \epsilon$$

to test $H_0 : \beta_2 = 0$, i.e., the prior trends do not differ across group.

- α_g are group-specific intercepts
- $\beta_2 Time \times Group$ allows the time trend to be different for each group
- More complex specifications can be made by adding polynomial terms or other nonlinearities to the model

Long-term effects

Dynamic DiD

- DiD can be modified to allow the effect to differ in each time period. In other words, there is a possibility to have **dynamic treatment effects.**

- A common way of doing this is to first generate a centered time variable, which is just the original time variable minus the treatment period:

  Time in the last period before treatment is $t = 0$, the first period with treatment implemented is $t = 1$ the second-to-last period before treatment is $t = -1$ and so on.

- Estimate

$$Y_k = \alpha_g + \alpha_t + \beta_{-T_1}\text{Treated} + \beta_{-(T_1-1)}\text{Treated} + \cdots + \beta_{-1}\text{Treated} + \beta_1\text{Treated} + \cdots + \beta_{T_2}\text{Treated} + \epsilon,$$

  where the Treated regressors are interactions of Treatment given and each of the time periods (for T_1 periods before the treatment period and T_2 periods afterwards).

Long-term effects

Advantages

- The before-treatment coefficients $\beta_{T_1}, \beta_{-(T_1-1)}, \ldots, \beta_{-1}$ should be (close to) zero

- The after-treatment coefficients $\beta_1, \ldots, \beta_{T_2}$ show the DiD estimated effect in the relevant period:

  The effect one period *after* treatment is β_1 and so on.

Disadvantage

- Regular DiD takes advantage of all the data in the entire after period to estimate the effect.

- Each period's effect estimate in the dynamic treatment effects approach relies mostly on data from that one period.

  $\rightarrow$ Shortage of data and hence less precise estimation!

Long-term effects

Dynamic DiD

- When interpreting the results, everything is relative to the (omitted) $t = 0$ effect (as always when there is a categorical variable, everything is relative to the omitted group)

- There should be no actual effect in period 0. But if there was, results will be wrong!

- It's common to represent dynamic estimates graphically with time at the X axis and the DiD estimates and (usually) confidence intervals on the Y axis

🏠

Long-term effects
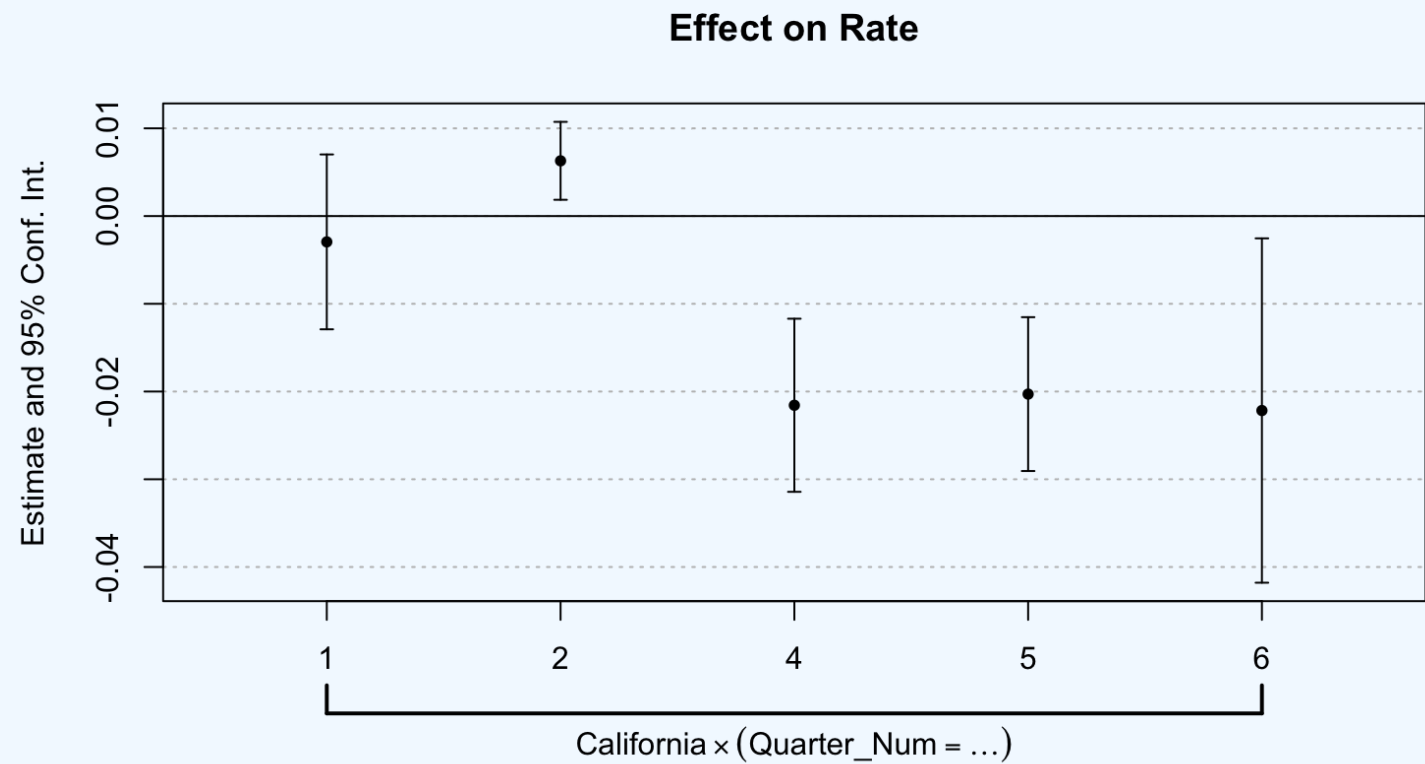
Example: Effect of active choice on organ donor − ctd.

Figure 8: Dynamic effect of active-choice phrasing on organ donation rates

Rollout designs and multiple treatment periods

> **Definition: Rollout design**
>
> A rollout design is when the groups get treated at different times.

Rollout designs and multiple treatment periods

Example: High Speed internet on new businesses

- We are interested in the impact of having access to high-speed internet on the formation of new businesses

- King County got broadband in 2001, Pierce County got it in 2002, and Snohomish County got it in 2003

- They each have a before and after period, but those treatment times are not all the same

Rollout designs and multiple treatment periods

Rollout designs are tricky

From a statistical perspective, tossing a bunch of valid DiD designs together makes the two-way fixed effects regression invalid:

The estimator does not work because this setup leads already-treated groups to get used as untreated groups.

→ We get biased DiD estimates. Estimates of the treatment effect may be negative, even if the true effect is positive for everyone in the sample!

〽️ Doing multiple treatment periods right

Rollout designs are tricky

- Handling multiple treatment periods in DiD where some groups are treated at different times than others is an active area of research

- There are two ways of addressing this problem:

  1. Using models for dynamic treatment effects that are modified to fix the staggered rollout problem (cf. Sun and Abraham 2020[3])

  2. The method described in Callaway and Sant'Anna (2020)[4]

[3] Sun, Liyang, and Sarah Abraham. 2020. *Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects*. Journal of Econometrics.
[4] Callaway, Brantly, and Pedro HC Sant'Anna. 2020. *Difference-in-Differences with Multiple Time Periods*. Journal of Econometrics.

ᴽ Doing multiple treatment periods right

1. Dynamic treatment effects approach (Sun and Abraham 2020)

Dynamic treatment effect models modified for staggered rollout can help in the case of staggered DiD

- They separate out the time periods when the effects take place

- They allow to check model violations by prior trends checks

- They give insight in how the effect evolves (evolving treatment effects are one of the problems with two-way fixed effect!)

    → Opportunity to separate things out and fix them.

1. Dynamic treatment effects approach (Sun and Abraham 2020)[5]

- Sun and Abraham (2020) propose to interact time-centered-on-treatment-time dummies with group membership, i.e., each group and time period has its own coefficient.

  → Avoids comparisons we do not want to make, since now the regression model is barely comparing anything.

- Comparisons are then up to us. We can average the corresponding coefficient estimates together in a way that yields a time-varying treatment effect.

2. Callaway and Sant'Anna (2020)

- Estimate *group-time treatment effects*: average treatment effects on the group treated in a particular time period. This gives an effect estimate for each time period where the treatment was new to *someone*

- Compare Y between each treatment group and the untreated group, and use propensity score matching to improve estimates

  → Group-time treatment effects reflect the post-treatment outcomes of the groups treated in that period against the never-treated groups that are most similar to those treated groups

- 'Summarising' treatment effects can than be obtained as a weighted averages of the group-time treatment effects

⌁ Picking an untreated group with matching

Good control groups are important

- DiD only works if the comparison group is good (parallel trends)

- Since parallel trends cannot be checked directly, an untreated group needs to be picked (or a set of untreated groups) good enough that the assumption is plausible

- When having potential untreated groups, one can choose between them (or aggregate them together) by matching untreated and treated groups (as in Callaway and Sant'Anna 2020)

Idea

- Match each treated group with an untreated group, or produce a set of weights for the untreated groups based on similarity to the treated groups using a set of pre-treatment predictor variables

- Run the the usual DiD with the matching groups/weights applied

⌂ 🪜 Picking an untreated group with matching

A step further: Synthetic control

> ## Definition: Synthetic control
>
> In synthetic control, one matches the treated group to a bunch of untreated groups based not just on prior covariates but also on *prior outcomes*.

Successful synthetic control matching forces prior trends are to be the same because weights have been specifically chosen for the untreated groups that have the same average outcomes as your treated group in each prior period.

⌂ ⌐ Picking an untreated group with matching

Combined matching and DiD

- DiD controls for any differences between treated and untreated groups that is *constant* over time using group fixed effects

- However, DiD is inconclusive about why certain groups come to be treated and others do not

  $\rightarrow$ back doors may be present:

  If there is some back door between 'becomes a treated group' and 'evolution of the outcome in the post-treatment period', identification fails

- Matching can close the back doors between *which groups become treated and when* and the outcome, thus getting parallel trends back!

⌁ Picking an untreated group with matching

A general problem: Regression to the mean

The basic idea:

- if a variable is far above its typical average of this period, then it's likely to go down next period, i.e., it *regresses back* towards the mean

- A problem arises if the pre-period outcome levels are related to the probability of treatment:

  DiD cannot adjust for the researcher assigning treatment to subjects with a (random) extreme outcome

  → The effect estimate will be biased!

⛷ Picking an untreated group with matching

A general problem: Regression to the mean

> ## Example: Effect of policy on unemployment
>
> - Two cities A and B have been matched based on very similar covariates
>
> - Policymakers are planning to put a job training program in place and want to know the effects of the program on unemployment
>
> - They choose City A for the program since unemployment is currently bad in City A. B as the control group.

🪜 Picking an untreated group with matching

A general problem: Regression to the mean

> ## Example: Effect of policy on unemployment
>
> After the policy goes into effect, unemployment might get better in City A for two reasons:
>
> 1. The effect of the policy
>
> 2. Regression to the mean — A may just have had an unusually bad period when policymakers were choosing where to put the training program
>
> DiD cannot tell the two apart!

A general problem: Regression to the mean

Example: Effect of policy on unemployment

The matching emphasizes comparisons that are especially subject to regression to the mean:

- This is only a problem because A and B are matched: If a bunch of untreated cities were used, or a random city from a set of potential comparisons, the bias would not be there

- That is because B was selected as a good match for an unusually bad time in A's history

🏠 ⌁ The unfurling logic of DiD

Example: Training program on educational income disparities

- Consider a teacher training program to help ease educational income disparities that is introduced in some districts but not others

- The relationship between parental income and student test scores should be weaker with the introduction of the training program.

- We consider

$$TestScore = \beta_0 + \beta_1 Income + \epsilon$$

🏠 📶 The unfurling logic of DiD

Example: Training program on educational income disparities – ctd.

- Interest is to perform DiD on β_1 instead of Y:

$$(\beta_1^{Treated,\,After} - \beta_1^{Treated,\,Before}) - (\beta_1^{Untreated,\,After} - \beta_1^{Unreated,\,Before})$$

- In order to know the within variation in the effect we may use the model

$$TestScore = \beta_0 + \beta_1 Income + \beta_2 After + \beta_3 Income \times After + \epsilon$$

- For the treated group, β_3 in the above model gives us

$$\beta_1^{Treated,\,After} - \beta_1^{Treated,\,Before}$$

Example: Training program on educational income disparities – ctd.

- For the untreated group, β_3 gives us

$$\beta_1^{Untreated,After} - \beta_1^{Unreated,Before}$$

- Everything can be combined into one regression with a triple-interaction term:

$$
\begin{aligned}
TestScore = \beta_0 &+ \beta_1 Income \\
&+ \beta_2 After + \beta_3 Income \times After \\
&+ \beta_4 Treated + \beta_5 Treated \times Income + \beta_6 Treated \times After \\
&+ \beta_7 Treated \times Income \times After + \epsilon
\end{aligned}
$$

- This is difference-in-differences but *on a relationship* rather than the average of an outcome!

📈 The unfurling logic of DiD

Applications

- Aside from applying DiD to effects themselves (relationships), the logic can also be applied to other kinds of summary descriptions of a single variable rather than the mean (like DiD would do)

- An example is in using DiD with quantile regression: a form of regression that looks at how predictors affect the *distribution* of a variable

- DiD can also be applied to DiD itself and get the difference-in-difference-in-differences model, also known as triple-differences (DiDiD)

🪜 The unfurling logic of DiD

Applications

- DiDiD could be used to see how a newly implemented policy *changes* a DID-estimated effect

- However, DiDiD is also used to help strengthen the parallel trends assumption by finding a treated group that should not be affected at all, and subtracting out their effect