

Precept 12: Causality with Repeated Measurements

Soc 400: Applied Social Statistics

Alex Kindel¹

Princeton University

December 13, 2018

¹Thanks to Ian Lundberg for providing examples.

Today's Agenda

- A stratification walkthrough
- Repeated measurements
 - Diff-in-diff
 - Fixed effects
- Some coding advice
- General review

Stratification

Two weighting strategies:

- ATE = sum of CATEs weighted by probability of being in each stratum
- ATT = sum of CATEs weighted by probability of being in each stratum *conditional on receiving treatment*

ATE

$$\sum_x E[Y(1) - Y(0)|X = x]P(X = x)$$

ATT

$$\sum_x E[Y(1) - Y(0)|X = x]P(X = x|D = 1)$$

Stratification in dplyr

To RStudio!

Repeated observations: Big ideas

- Difference in difference
- Fixed effects

Difference in difference

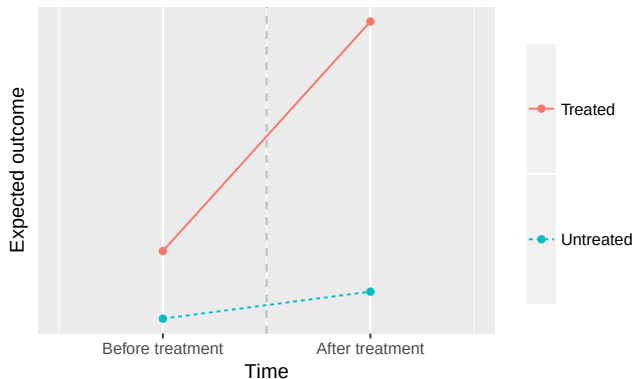
The difference in difference (DID) estimator is:

$$Y_{i,t=1} - Y_{i,t=0} = \beta_0 + \beta_1(D_{i,t=1} - D_{i,t=0}) + u_i$$

Difference in difference

The difference in difference (DID) estimator is:

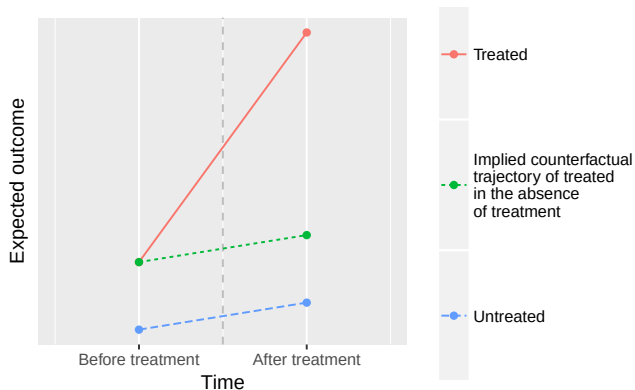
$$Y_{i,t=1} - Y_{i,t=0} = \beta_0 + \beta_1(D_{i,t=1} - D_{i,t=0}) + u_i$$



Difference in difference

The difference in difference (DID) estimator is:

$$Y_{i,t=1} - Y_{i,t=0} = \beta_0 + \beta_1(D_{i,t=1} - D_{i,t=0}) + u_i$$



Things to know about the DID estimator:

Things to know about the DID estimator:

- It has only two time points

Things to know about the DID estimator:

- It has only two time points
- Some people get treatment between the two time points

Things to know about the DID estimator:

- It has only two time points
- Some people get treatment between the two time points
- It estimates
 - 1 Difference 1: The change in the outcome between $t = 0$ and $t = 1$ among the control group

Things to know about the DID estimator:

- It has only two time points
- Some people get treatment between the two time points
- It estimates
 - ① Difference 1: The change in the outcome between $t = 0$ and $t = 1$ among the control group
 - ② Difference 2: The change in the outcome between $t = 0$ and $t = 1$ among the treated group

Things to know about the DID estimator:

- It has only two time points
- Some people get treatment between the two time points
- It estimates
 - ① Difference 1: The change in the outcome between $t = 0$ and $t = 1$ among the control group
 - ② Difference 2: The change in the outcome between $t = 0$ and $t = 1$ among the treated group
 - ③ The difference in the differences: The causal estimate is $(2) - (1)$

More things to know about DID

- It assumes that the change in Y that would have happened in the treated group in the absence of treatment is the same as the observed change in Y in the control group

More things to know about DID

- It assumes that the change in Y that would have happened in the treated group in the absence of treatment is the same as the observed change in Y in the control group
 - This assumption is untestable.

More things to know about DID

- It assumes that the change in Y that would have happened in the treated group in the absence of treatment is the same as the observed change in Y in the control group
 - This assumption is untestable.
 - This assumption is often dubious.

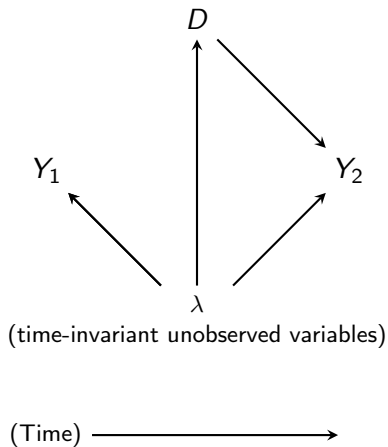
More things to know about DID

- It assumes that the change in Y that would have happened in the treated group in the absence of treatment is the same as the observed change in Y in the control group
 - This assumption is untestable.
 - This assumption is often dubious.
 - If the pre-treatment outcomes are different, it is hard to believe that the slopes would be the same in the absence of treatment.
- It is robust to unobserved time-invariant confounders

More things to know about DID

- It assumes that the change in Y that would have happened in the treated group in the absence of treatment is the same as the observed change in Y in the control group
 - This assumption is untestable.
 - This assumption is often dubious.
 - If the pre-treatment outcomes are different, it is hard to believe that the slopes would be the same in the absence of treatment.
- It is robust to unobserved time-invariant confounders
- It is sensitive to time-varying confounders

DID is good if
(subscripts index time)



Fixed effects

Fixed effects is an **extension of the DID idea to many time points.**

Fixed effects

Fixed effects is an **extension of the DID idea to many time points**.

The person fixed effects estimator with individuals indexed by i and survey years indexed by t estimates a unique intercept for each respondent in the data.

$$Y_{it} = \alpha_i + \beta D_{it} + \epsilon_{it}$$

Things to know about the fixed effects estimator

- It is often called a within-person estimator.
- It estimates how changes in T are associated with changes in Y , for individual i .

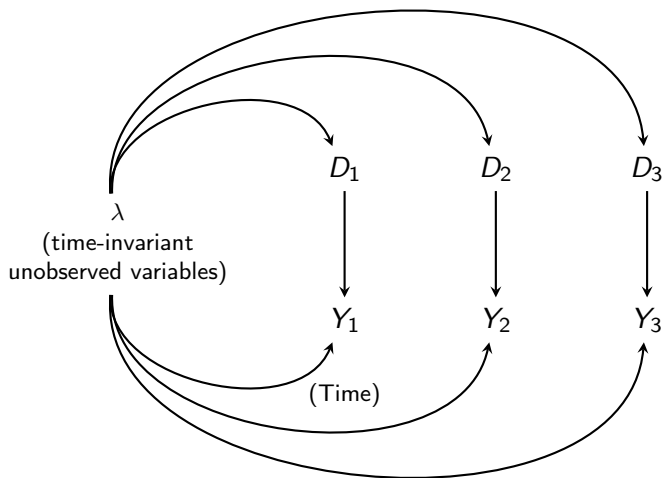
Things to know about the fixed effects estimator

- It is often called a within-person estimator.
- It estimates how changes in T are associated with changes in Y , for individual i .
- It accounts for all confounders that are constant within person (do not change over time)
 - This includes all unobserved time-invariant confounders

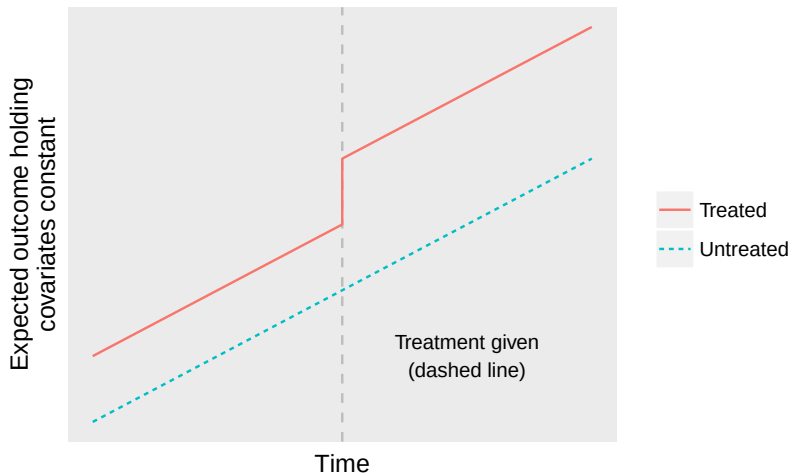
Things to know about the fixed effects estimator

- It is often called a within-person estimator.
- It estimates how changes in T are associated with changes in Y , for individual i .
- It accounts for all confounders that are constant within person (do not change over time)
 - This includes all unobserved time-invariant confounders
- But it is sensitive to time-varying confounders

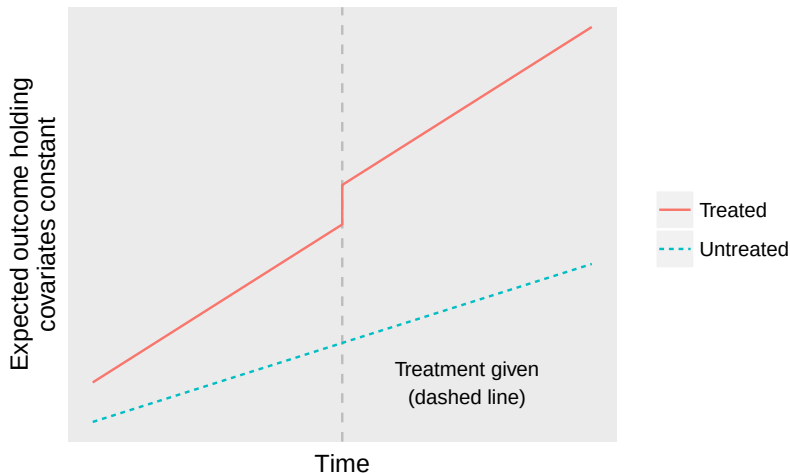
Fixed effects is good if
(subscripts index time)



Fixed effects is good if



Fixed effects is bad if



Repeated observations example: Marriage and men's wages

- Suppose you want to estimate the causal effect of marriage on men's wages.
- You have repeated wage observations on individuals before and after marriage.

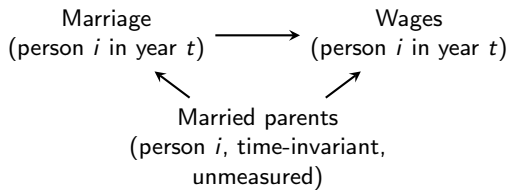
Repeated observations example: Marriage and men's wages

- Suppose you want to estimate the causal effect of marriage on men's wages.
- You have repeated wage observations on individuals before and after marriage.
- In the following scenarios, what estimation strategy would you use?

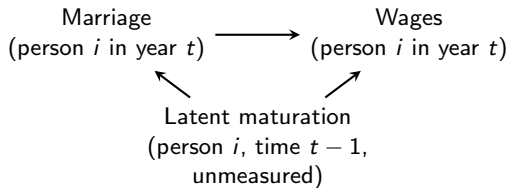
Scenarios

- ① Having married parents causes men to marry and causes higher wages.
- ② People “mature” at different ages. Something changes, and you get your act together to become an adult. When you hit this “latent maturation,” it causes you to marry and to earn more in the next year.

Scenario
1



Scenario
2



Repeated observations example: Answer

Repeated observations example: Answer

- **Scenario 1:** Fixed effects adjusts for time-invariant confounders, so would be preferred here.

Repeated observations example: Answer

- **Scenario 1:** Fixed effects adjusts for time-invariant confounders, so would be preferred here.
- **Scenario 2:** This is a case of time-varying unobserved confounding, so neither approach will yield a consistent causal estimate!

Coding advice

- This isn't a software engineering course—we don't grade your code

Coding advice

- This isn't a software engineering course—we don't grade your code
- That said, learning to write good code will save you pain and heartache for years to come

Coding advice

- This isn't a software engineering course—we don't grade your code
- That said, learning to write good code will save you pain and heartache for years to come
- Here are some common issues I've seen and how to fix them

Coding advice

- ① Magic numbers
- ② Abstraction: DRY code and WET code
- ③ Style
- ④ Commenting
- ⑤ Portability
- ⑥ Organization

Any more questions?

Come to the review session!

January 14th in the afternoon