ECON42720 Causal Inference and Policy Evaluation

7 Fixed Effects and Difference-in-Differences

Ben Elsner (UCD)

Resources for Fixed Effects

Textbook chapter

- Huntington-Klein, The Effect: Ch. 16

Resources for Difference-in-Differences

Textbook chapters

- Cunningham, Causal Inference: The Mixtape, Ch. 9
- Huntington-Klein, The Effect: Ch. 18

YouTube Videos

- Videos 17-21 of my Causal Inference Playlist

Fixed Effects

Start with a regression:

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

If there are unobserved confounders, we have the problem that $E[u_i|X_i] \neq 0$

If we could **observe these confounders**, we could include them in the regression

$$Y_i = \beta_0 + \beta_1 X_i + \boldsymbol{S_i'\delta} + u_i$$

If S_i includes all confounders, $E[u_i|X_i, \boldsymbol{S_i}] = 0$ holds and we have an unbiased and consistent estimator for β_1.

Fixed Effects: Controlling for Unobservables

Problem: We usually can't observe all confounders

Fixed effects allow us to **control for (some) unobserved and observed confounders**

What we need:

- **Panel data**: multiple observations per unit
- or **Grouped data**: multiple units in each group

Fixed Effects with Panel Data

Panel data is data with multiple observations per unit i

$$Y_{it} = \beta_0 + \beta_1 X_{it} + u_{it}$$

Now **add unit fixed effects**:

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \alpha_i + u_{it}$$

The fixed effects α_i can be viewed as **separate dummies for each unit** i

What Fixed Effects Do

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \alpha_i + u_{it}$$

The fixed effects α_i isolate the **within-unit variation in Y_{it} and X_{it}**

Suppose i are countries and t are years. **Interpretation of β_1**:

- If X_{it} goes up in a given country, how does Y_{it} change within the same country?
- So β_1 measures the **average within-country effect** of X_{it} on Y_{it}

The fixed effects α_i **control for all time-invariant observables AND unobservables**

Fixed Effects: Example

We will now go through a simple example: **crime rates and police presence in cities**

- Here, a *group* is a city
- There is *within-city variation* in crime rates and police presence *over time*
- This is the classic use of *fixed effects with panel data*

Data are (to some extent) made up for illustration purposes

Fixed Effects: Example

Causal relationship of interest

$$\text{crime}_i = \alpha + \beta \text{ police presence}_i + u_i$$

City	Year	Murder rate	Police presence
Baltimore	2009	55.4	42
Albuquerque	2009	7.7	28
New York	2009	4.1	30
Pittsburgh	2009	18.6	33
San Francisco	2009	6.1	20
Detroit	2009	43.8	31

Difference-in-Differences: a Quasi-Experimental Design

Some units get treated, some don't... we've heard that before

What's different about difference-in-differences?

- **Treatment assignment** does **NOT** need to be as good as **random**
- The **TREND in outcomes** of the control group is a good **counterfactual** for the trend of the treated group

DiD is arguably one of the **most popular designs in empirical economics**

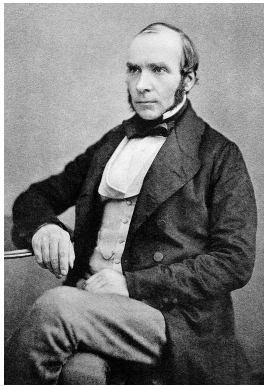
Historical DiD Example: The Cholera Hypothesis

19th century: Cholera was a major disease in Europe

Dominant hypothesis: Cholera is **transmitted through the air**

John Snow in 1854: Cholera is **transmitted through water**

Research design: **Difference-in-differences**



John Snow (1813-1858)
(Source: Wikipedia)

Broad Street Pump in London (Soho)

(Source: Wikipedia)

The Cholera Hypothesis

Snow's theory: **Cholera is transmitted through water**

- People drink contaminated water that contains the cholera bacterium
- The bacterium enters the digestive system and causes cholera
- Through vomiting and diarrhea, the bacterium is excreted and contaminates the water supply further

Some observations:

- Sailors got sick when they went on land but not when staying docked
- Cholera was more prevalent in poor areas with bad hygiene
- Some apartment blocks were affected, other neighbouring ones not

The Cholera Hypothesis

How could Snow test his theory?

- Mind you: experiments were only established in 1935 by Fisher as a means to prove causality
- And you couldn't run an experiment (drink from the Thames if heads, from another source if tails)

Snow's research design

- Some areas in London had their water supply from the Thames
- Others had their water supply from other sources
- Problem: areas were different in many ways

Snow's Research Design

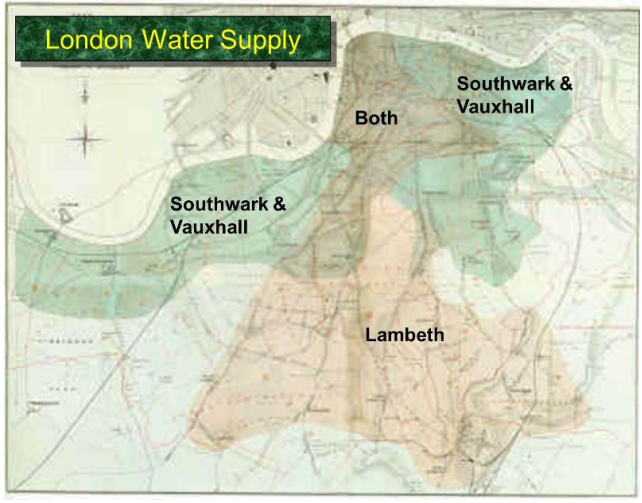
Different boroughs in London had different **water supplies, all from the Thames**

But: in 1849 the Lambeth Water Company switched to a new water source upstream

- This turned out to be cleaner and not contaminated cholera
- The Southwark and Vauxhall Water Company did not switch

Did cholera cases decline in Lambeth after the switch relative to Southwark and Vauxhall?

Lambeth vs. Southwark and Vauxhall Water Supply



(Source: inferentialthinking.com)

John Snow's Data

Much of the data on water suppliers was hand-collected (!) by Snow

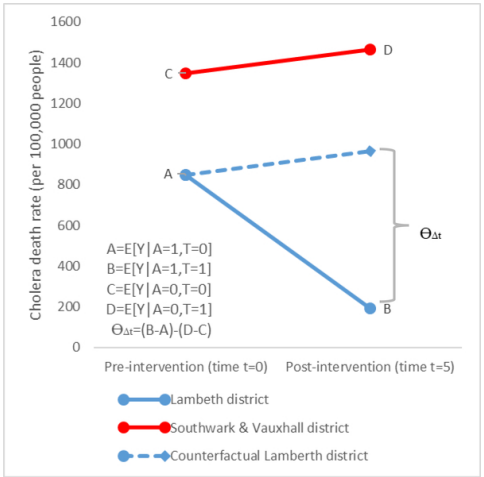
Cholera deaths per 10,000 households in the mid-1850s

Company Name	**1849** Before Switch	**1854** After Switch
Southwark and Vauxhall	135	147
Lambeth	85	19

Things to note

- There were **more deaths in both years in Southwark and Vauxhall**
- **Death rates in Lambeth dropped dramatically** after the switch
- Death rates in Southwark and Vauxhall stayed roughly the same

John Snow Discovered Difference-in-Differences



Source: Caniglia & Murray (2020)

John Snow Discovered Difference-in-Differences

Difference 1: Lambeth vs. Southwark and Vauxhall

- Solid blue vs red line: differences in cholera deaths between the two areas

Difference 2: Before vs. after the switch

- Dotted blue line: projects the trend in Lambeth if the switch had not happened
- This is just the trend of Southwark and Vauxhall

Difference-in-differences: The difference between the solid and dotted blue line

- relative to the counterfactual, the switch reduced cholera deaths by 78 per 10,000 households

John Snow Discovered Difference-in-Differences

Company Name	**1849** Before Switch	**1854** After Switch	**Difference 2**
Southwark and Vauxhall	135	147	+12
Lambeth	85	19	-66
Difference 1	-50	-128	-78

The difference-in-differences is 78 cholera deaths per 10,000 households

References

Caniglia, Ellen C., & Murray, Eleanor J. 2020. Difference-in-Difference in the Time of Cholera: a Gentle Introduction for Epidemiologists. *Current Epidemiology Reports*, **7**, 203–210.

benjamin.elsner@ucd.ie

www.benjaminelsner.com

Sign up for office hours

YouTube Channel

@ben_elsner

LinkedIn

Contact

Prof. Benjamin Elsner
University College Dublin
School of Economics
Newman Building, Office G206
benjamin.elsner@ucd.ie

Office hours: book on Calendly