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

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

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

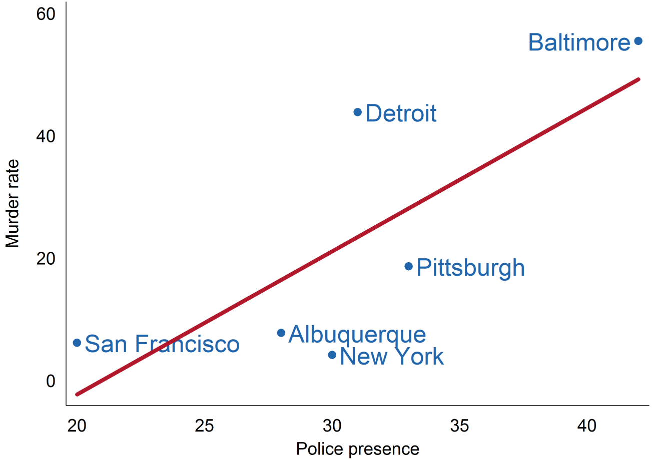
Fixed Effects: Example

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

The Cross-sectional Relationship is Positive. . .



Fixed Effect Regressions

Logic of Fixed Effect Regressions: exploit **variation within subjects over time**

In our case: how does the **murder rate in a city change** when in the same city the police presence increases by 1 unit?

Advantage:

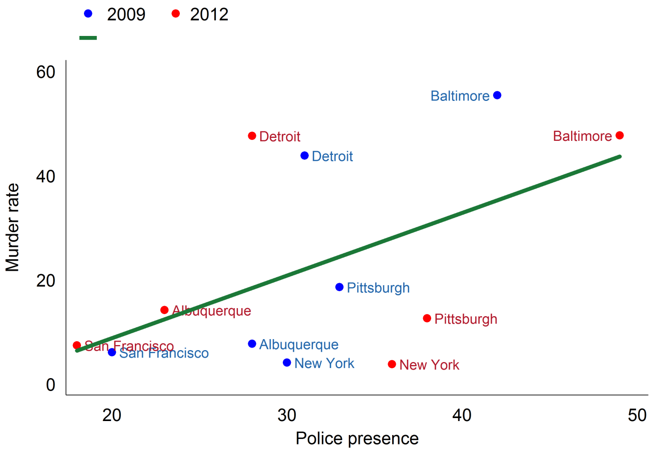
- **fixed city characteristics** are **held constant**
- And as such many determinants why Baltimore has a higher crime rate *and* police presence than San Francisco
- We circumvent an important *selection problem* $\Rightarrow$ eliminates (or reduces) **omitted variable bias**

Now Suppose You Have Panel Data

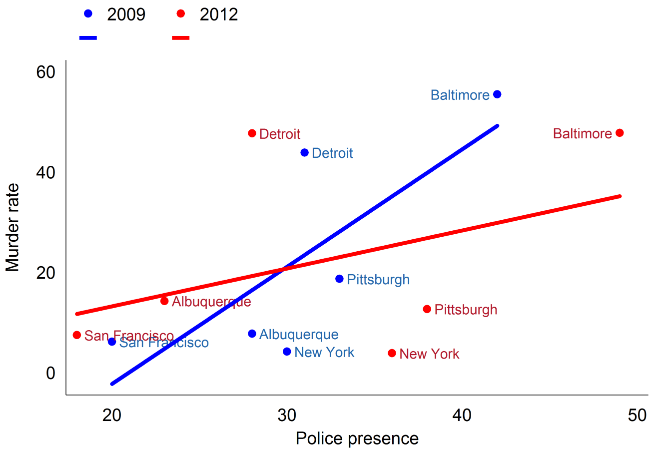
City	Year	Murder rate	Police presence
Baltimore	2009	55.4	42
Baltimore	2012	47.4	49
Albuquerque	2009	7.7	28
Albuquerque	2012	14.2	23
New York	2009	4.1	30
New York	2012	3.8	36
Pittsburgh	2009	18.6	33
Pittsburgh	2012	12.6	38
San Francisco	2009	6.1	20
San Francisco	2012	7.4	28
Detroit	2009	43.8	31
Detroit	2012	47.6	28

{Note: data are fictitious}

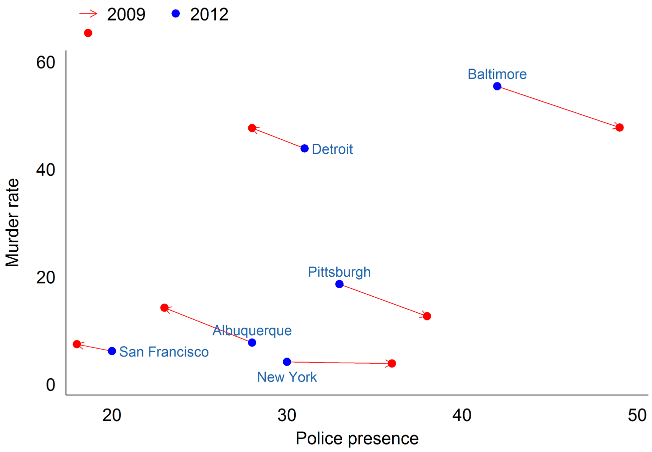
Cross-sectional Relationship in Panel Data: Still Positive



In each year we have a positive association



Now look at within-city changes



Fixed Effect Regressions

A **Fixed Effect Regressions** only relies on the **within-variation**

$$Y_{it} = \beta X_{it} + \alpha_i + \varepsilon_{it}$$

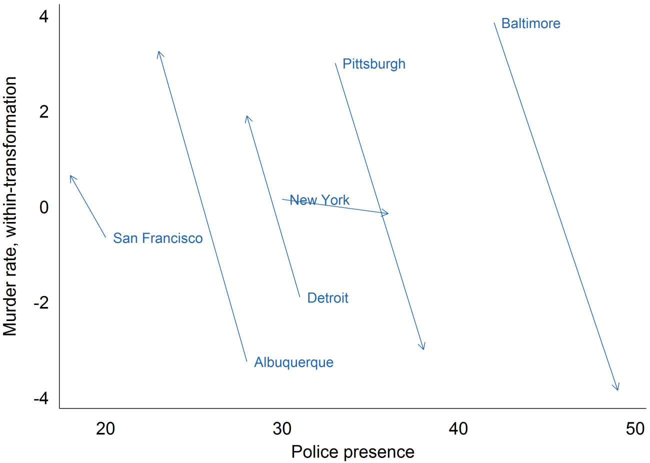
The **between-variation** will be netted out

At the core of the FE regression lies a **within-transformation**

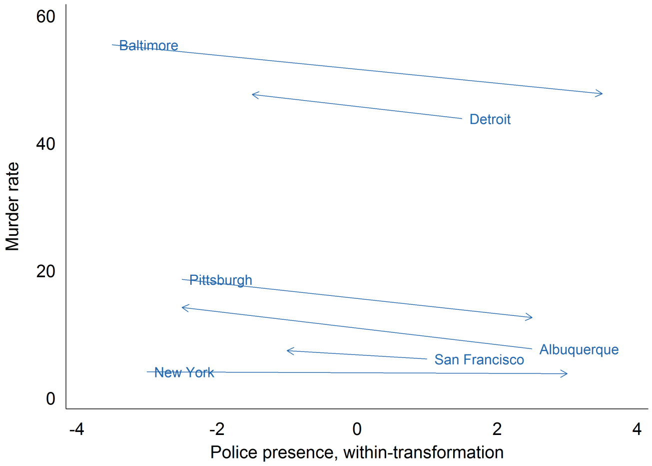
$$Y_{it} - \overline{Y}_i = \beta(X_{it} - \overline{X}_i) + \varepsilon_{it} - \varepsilon_i$$

Takes from each variable the **deviation from the mean**

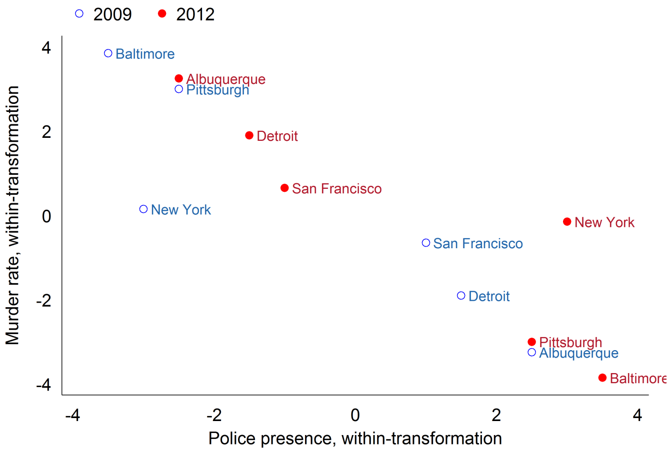
Within-transformation of Y



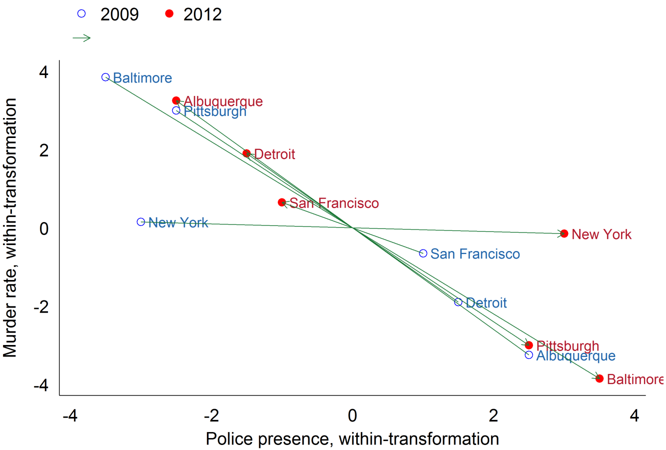
Within-transformation of X



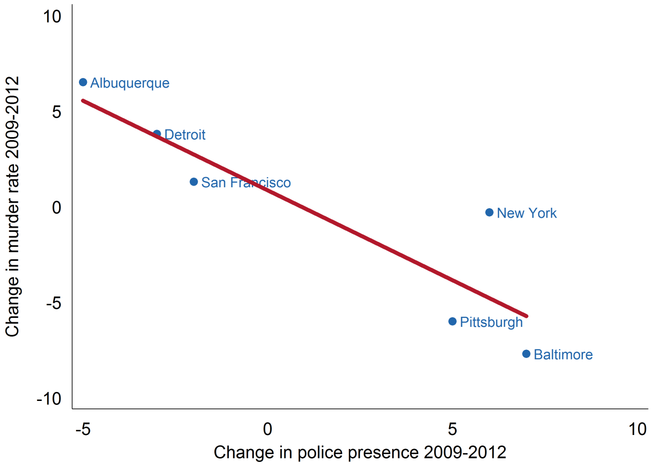
Within-transformation of X and Y



The within-effects in each city



Average within-effect: NEGATIVE!



Summary of the example

The **cross-sectional relationship** between police presence and crime rates is **positive**

- This is **between-city variation**
- It is driven by **differences in city characteristics**
- More crime-prone cities choose to hire more police officers...

We can learn a lot from **within-city variation**

- The city fixed effects **eliminate all time-invariant differences between cities**
- They isolate the **within-city variation** in **all variables**
- The **within-effect** of police presence on crime rates is **negative**

Interpretation: if within a city the police presence goes up by 1 unit, the crime rate goes down by β units

Another way to look at fixed effects

The fixed effects **split the data into many units** – here a unit is a city

A **fixed effect regression performs two tasks** at the same time:

1. it estimates the effect of X on Y within each unit
2. it averages these effects across all units

Fixed Effects and Causality

Fixed effects can **eliminate time-invariant confounders**

$$Y_{it} = \beta X_{it} + \alpha_i + \varepsilon_{it}$$

Causal identification is through **selection on observables**

Conditional independence assumption

$$E[\varepsilon_{it} | X_{it}, \alpha_i] = 0$$

- Conditional on fixed effects, the error term is uncorrelated with X_{it}
- In plain English: **within each unit, X_{it} has to be as good as randomly assigned**

Multiple units and time periods

It is common to have **panel data with many units and many time periods**

- Example: 50 US states over 20 years

We often use **two-way fixed effects**:

- Unit fixed effects (δ_i) absorb all time-invariant differences between units
- Time fixed effects (δ_t) absorb all time trends that are common to all units

The **regression equation is then**

$$Y_{it} = \beta X_{it} + \delta_i + \delta_t + u_{it}$$

Fixed Effects with Grouped Data

Grouped data is data with multiple units $i = 1, \ldots, N$ which belong to distinct groups $g = 1, \ldots, G$.

Example: students in schools, workers in firms, patients in hospitals

Classic case: stratified experiments in within schools

- It is not random who goes to which school
- But within schools, treatment assignment is random

To estimate the treatment effect, we can use **fixed effects for groups**

$$Y_{ig} = \beta X_{ig} + \alpha_g + u_{ig}$$

Example for Identification with Group Fixed Effects: Project STAR

Remember the STAR experiment? Within schools, students were randomly assigned to small classes

- It is not random who goes to which school
- But it is random who gets assigned to small classes within a given school

The **basic regression** run by Krueger (1999) was

$$\text{Test score}_{ig} = \beta \mathbb{1}[\text{Small class}]_{ig} + \alpha_g + \varepsilon_{ig}$$

β is **causally identified because of the random assignment** within schools

Implementation of Fixed Effects Regressions in R

First of all, you need to **have panel data in "long form"**

- Each **row** is an observation for a unit at a certain time

City	Year	Murder rate	Police presence
Baltimore	2009	55.4	42
Baltimore	2012	47.4	49
Albuquerque	2009	7.7	28
Albuquerque	2012	14.2	23
New York	2009	4.1	30
New York	2012	3.8	36
Pittsburgh	2009	18.6	33
Pittsburgh	2012	12.6	38
San Francisco	2009	6.1	20
San Francisco	2012	7.4	28
Detroit	2009	43.8	31
Detroit	2012	47.6	28

Data must not be in wide format!

City	Murder Rate 2009	Police 2009	Murder Rate 2012	Police 2012
Baltimore	55.4	42	47.4	49
Albuquerque	7.7	28	14.2	23
New York	4.1	30	3.8	36
Pittsburgh	18.6	33	12.6	38
San Francisco	6.1	20	7.4	28
Detroit	43.8	31	47.6	28

Can't work with that! If you have such data, use the `pivot` commands from `dplyr` to bring your panel data into long form.

Fixed Effects in R: Preparation

You can use different **R packages to run fixed effect regressions**:

- Use the standard `lm()` and include dummies for units or groups
- Use the `plm` package (`plm()` with the `within` option)
- Use the `fixest` package, which is very efficient, especially when you have many fixed effects

For `plm` and `fixest` and other advanced packages, we need `modelsummary` to display the results

We will showcase these methods with the `gapminder` data

```
library(gapminder)
library(tidyverse)
library(plm)
library(fixest)
library(modelsummary)
```

The Gapminder Data: 142 countries, 12 years

```r
data("gapminder")
head(gapminder)
```

```
## # A tibble: 6 x 6
##   country     continent  year lifeExp      pop gdpPercap
##   <fct>       <fct>     <int>   <dbl>    <int>     <dbl>
## 1 Afghanistan Asia       1952    28.8  8425333      779.
## 2 Afghanistan Asia       1957    30.3  9240934      821.
## 3 Afghanistan Asia       1962    32.0 10267083      853.
## 4 Afghanistan Asia       1967    34.0 11537966      836.
## 5 Afghanistan Asia       1972    36.1 13079460      740.
## 6 Afghanistan Asia       1977    38.4 14880372      786.
```

OLS Regressions with Dummies

Suppose we want to regress life expectancy on GDP per capita

- We want to include 141 country dummies and 11 year dummies
- We can do this easily with `factor()`

```
# Generate log gdp per capita
gapminder$loggdp <- log(gapminder$gdpPercap)

# Plain OLS without dummies
fereg.ols <- lm(lifeExp ~ loggdp, data = gapminder)

# Estimate OLS with country and year dummies
fereg.olsdummies <- lm(lifeExp ~ loggdp + factor(country) + factor(year)
                       , data = gapminder)
```

OLS vs Fixed Effects (dummies)

```
stargazer(fereg.ols, fereg.olsdummies,
          column.labels = c("OLS", "OLS dummies"),
          type = "latex", header = FALSE, digits = 2,
          keep= "loggdp",
          keep.stat = c("n", "adj.rsq"))
```

Table 1:

	Dependent variable:	
	lifeExp	
	OLS	OLS dummies
	(1)	(2)
loggdp	8.41***	1.45***
	(0.15)	(0.27)
Observations	1,704	1,704
Adjusted R^2	0.65	0.93

plm and fixest

```r
# Convert your data frame to a pdata.frame for plm
pdata <- pdata.frame(gapminder, index = c("country", "year"))
pdata$loggdp <- log(pdata$gdpPercap)

# FE estimation with PLM (note: the effect argument is important here)
fereg.plm <- plm(lifeExp ~ loggdp, data = pdata,
                 model = "within", effect="twoways")

# FE estimation with fixest (LOOK HOW SIMPLE!)
fereg.fixest <- feols(lifeExp ~ loggdp |
                        country + year, data = gapminder)
```

Regression results

```
model_list <- list("OLS"=fereg.ols,
                   "OLS dummies"=fereg.olsdummies,
                   "PLM"=fereg.plm,
                   "FIXEST"=fereg.fixest)

# Use modelsummary to create the table
msummary(model_list, output = "latex",
         keep = "loggdp",
         gof_omit = "^(?!.*Num.Obs|.*R2.Adj)")
```

	OLS	OLS dummies	PLM	FIXEST
loggdp	8.405	1.450	1.450	1.450
	(0.149)	(0.268)	(0.268)	(0.679)
Num.Obs.	1704	1704	1704	1704
R2 Adj.	0.652	0.930	−0.078	0.930

Regression results

All regressions that account for fixed effects yield the **same point estimates**

The **standard errors differ**:

- OLS and `plm` do not adjust the standard errors unless we tell them to do so
- `fixest` adjusts the standard errors, in this case for two-way clustering at the country and year level
- Neither is 100% correct! Consensus is to cluster by unit but not time

Why do we **need to adjust the standard errors?**

- Observations within the same unit are likely to be correlated
- Life expectancy today is a function of life expectancy yesterday, and so on

(One reason) Why we need Differences-in-Differences

Consider the fixed effect regression with states i and time periods t

$$Y_{it} = \beta X_{it} + \alpha_i + \varepsilon_{it}$$

Suppose X_{it} is a policy variable: in some period t, a new policy is introduced in some states

- X_{it} could be a dummy that equals one in each period after the policy has been introduced

We can't really argue that the **policy change was as good as random**

- There are probably **good reasons why a policy was introduced** in state i and why at time t

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**

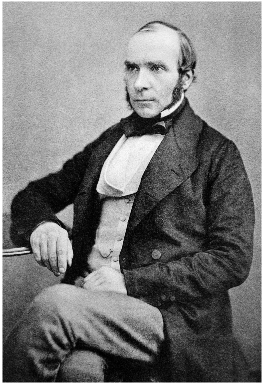
Historical DiD Example: The Cholera Hypothesis

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

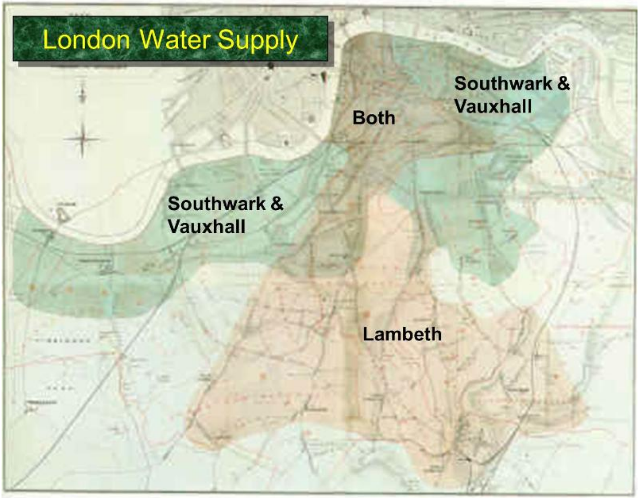
Different boroughs in London had different **water supplies, all from the Thames**

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

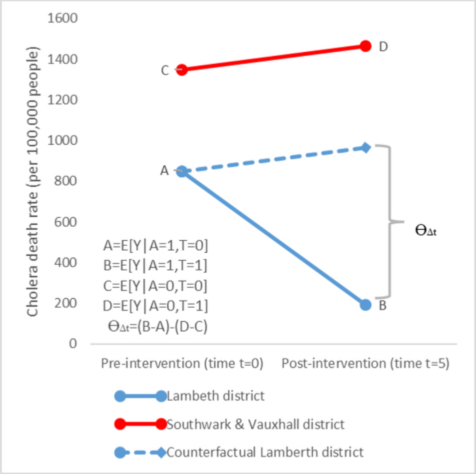
Cholera deaths per 10,000 households in the mid-1850s

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

- Because of the switch, cholera deaths dropped by 78 per 10,000 households in Lambeth

The simple 2×2 DiD

The simple 2×2 DiD is the **canonical difference-in-differences design**

- We have the difference between a **treatment group** k **and an untreated group** U
- ...and the difference before and after k received the treatment $(pre(k), post(k))$

$$\widehat{\delta}_{kU}^{2\times 2} = \underbrace{\left(\overline{y}_k^{\text{post}(k)} - \overline{y}_k^{\text{pre}(k)}\right)}_{\text{Pre-post difference, treated}} - \underbrace{\left(\overline{y}_U^{\text{post}(k)} - \overline{y}_U^{\text{pre}(k)}\right)}_{\text{Pre-post difference, untreated}}$$

$\widehat{\delta}_{kU}^{2\times 2}$ is the estimated ATT for group k

What does the simple 2×2 DiD identify?

Start with **conditional expectations**

$$\widehat{\delta}^{2\times 2}_{kU} = \Big(E\big[Y_k \mid \text{Post}\big] - E\big[Y_k \mid \text{Pre}\big]\Big) - \Big(E\big[Y_U \mid \text{Post}\big] - E\big[Y_U \mid \text{Pre}\big]\Big)$$

Let's use **potential outcomes** and add and **subtract a counterfactual**

$$\widehat{\delta}^{2\times 2}_{kU} = \underbrace{\Big(E\big[Y^1_k \mid \text{Post}\big] - E\big[Y^0_k \mid \text{Pre}\big]\Big) - \Big(E\big[Y^0_U \mid \text{Post}\big] - E\big[Y^0_U \mid \text{Pre}\big]\Big)}_{\text{Switching equation}}$$

$$+ \underbrace{E\big[Y^0_k \mid \text{Post}\big] - E\big[Y^0_k \mid \text{Post}\big]}_{\text{Adding and subtracting the counterfactual}}$$

What does the simple 2×2 DiD identify?

Re-arrange the terms from the previous slide:

$$\widehat{\delta}_{kU}^{2\times 2} = \underbrace{E\left[Y_k^1 \mid \text{Post}\right] - E\left[Y_k^0 \mid \text{Post}\right]}_{\text{ATT}}$$
$$+ \underbrace{\Big[E\left[Y_k^0 \mid \text{Post}\right] - E\left[Y_k^0 \mid \text{Pre}\right]\Big] - \Big[E\left[Y_U^0 \mid \text{Post}\right] - E\left[Y_U^0 \mid \text{Pre}\right]\Big]}_{\text{Non-parallel trends bias in } 2 \times 2 \text{ case}}$$

The simple 2×2 DiD identifies the **average treatment effect on the treated (ATT)**

- but only if the second term is zero...
- that is, only if the **parallel trends assumption** holds

The parallel trends assumption

Notice here: Y^0 appears everywhere $\Rightarrow$ **counterfactual!**

$$\underbrace{\Big[E\big[Y_k^0 \mid \text{Post}\big] - E\big[Y_k^0 \mid \text{Pre}\big]\Big] - \Big[E\big[Y_U^0 \mid \text{Post}\big] - E\big[Y_U^0 \mid \text{Pre}\big]\Big]}_{\text{Non-parallel trends bias in } 2 \times 2 \text{ case}}$$

In plain English: **in the absence of the treatment**, the outcomes of the treated and untreated groups would have **evolved in the same way**

Classic Study: Card & Krueger (1994) on the Effects of Minimum Wages

Economic theory: **higher minimum wages**. . .

- **reduce employment** in **competitive labour markets**
- but it may **increase employment in monopsonistic labour markets**

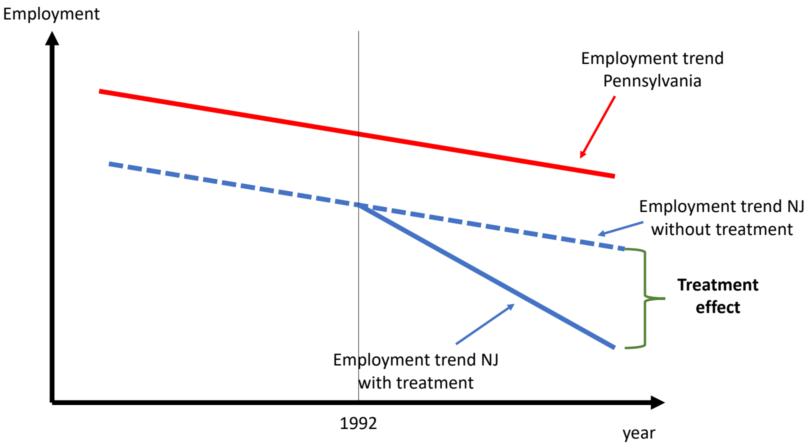
Which model is correct in practice? That's an empirical question

In a controversial study, Card & Krueger (1994) use the minimum wage increase in New Jersey in 1992 to answer this question

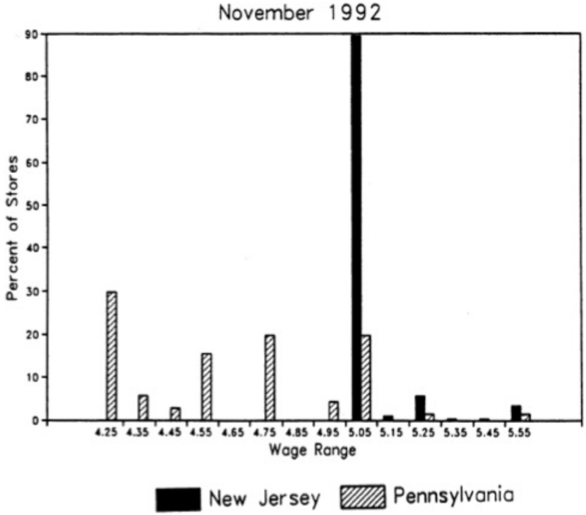
- They use data on workers in fast food restaurants
- They use Pennsylvania as a control state, which did not change its minimum wage

Classic Study: Card & Krueger (1994) on the Effects of Minimum Wages

In 1992, New Jersey increased the minimum wage, while Pennsylvania did not



The minimum wage change did bite



The 2×2 DiD in Card & Krueger (1994)

ATT of interest:

$$\widehat{\delta}^{2\times 2}_{NJ,PA} = \underbrace{E\big[Y^1_{NJ} \mid \text{Post}\big] - E\big[Y^0_{NJ} \mid \text{Post}\big]}_{\text{ATT}}$$

$$+ \underbrace{\Big[E\big[Y^0_{NJ} \mid \text{Post}\big] - E\big[Y^0_{NJ} \mid \text{Pre}\big]\Big] - \Big[E\big[Y^0_{PA} \mid \text{Post}\big] - E\big[Y^0_{PA} \mid \text{Pre}\big]\Big]}_{\text{Non-parallel trends bias}}$$

With **constant state and time effects**, this **maps into the regression**

$$Y_{its} = \alpha + \gamma NJ_s + \lambda D_t + \delta (NJ \times D)_{st} + \varepsilon_{its}$$

The 2×2 DiD in Card & Krueger (1994)

Variable	Stores by state PA (i)	NJ (ii)	Difference, NJ – PA (iii)
1. FTE employment before, all available observations	23.33 (1.35)	20.44 (0.51)	−2.89 (1.44)
2. FTE employment after, all available observations	21.17 (0.94)	21.03 (0.52)	−0.14 (1.07)
3. Change in mean FTE employment	−2.16 (1.25)	0.59 (0.54)	2.76 (1.36)

Does $\widehat{\delta}^{2\times 2}_{NJ,PA} = 2.76$ mean that the minimum wage raised employment?

The 2×2 DiD in Card & Krueger (1994)

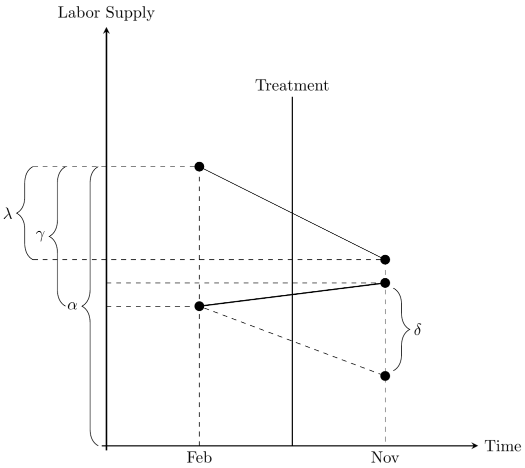
$$Y_{its} = \alpha + \gamma NJ_s + \lambda D_t + \delta(NJ \times D)_{st} + \varepsilon_{its}$$

1. PA pre: α
2. PA post: $\alpha + \lambda$
3. NJ pre: $\alpha + \gamma$
4. NJ post: $\alpha + \gamma + \lambda + \delta$

δ is the **difference-in-differences** estimator!

The ATT in Card & Krueger (1994)

$$Y_{its} = \alpha + \gamma NJ_s + \lambda D_t + \delta(NJ \times D)_{st} + \varepsilon_{its}$$



Providing Evidence for Parallel Trends

We want to **estimate the ATT** but there might be a **non-parallel trends bias**

$$\underbrace{\Big[E[Y_k^0 \mid \text{Post}] - E[Y_k^0 \mid \text{Pre}]\Big] - \Big[E[Y_U^0 \mid \text{Post}] - E[Y_U^0 \mid \text{Pre}]\Big]}_{\text{Non-parallel trends bias in } 2 \times 2 \text{ case}}$$

The Parallel Trends Assumption is an **identification assumption**

- Identification assumptions **cannot be tested!**
- We need to **bring good arguments in favour** of it (difficult)
- And provide **empirical evidence in support** of it (easy?)

Providing Evidence for Parallel Trends: Pre-trends

A common **diagnostics test** is to look at the **pre-trends**

- Suppose treated and control **moved in parallel before** the treatment was given
- ...it is then likely they **would have moved in parallel after**, had the **treatment not been given**

Pre-trends are commonly presented in **event-study graphs**

Event Study Example: Miller *et al.* (2021)

Miller *et al.* (2021) study the impact of the expansion of Medicaid in the U.S. on population mortality

Expansion of Medicaid under the Affordable Care Act (ACA) in 2014

- Health insurance for low-income individuals
- Post-2014: covers everyone with incomes up to 138% of the federal poverty line
- Initially, the ACA was supposed to apply to all states
- But the Supreme Court ruled in 2012 that states could opt out
- 29 states plus DC expanded Medicaid in 2014, 7 later, 14 did not

Data: **Vital statistics data** on deaths linked with individual survey data

Event Study Example: Miller *et al.* (2021)

Difference-in-differences:

- Expansion states vs. non-expansion states
- After vs. before the expansion
- But: “staggered adoption” because there were states that expanded Medicaid later

Event studies consider leads (pre-treatment) and lags (post-treatment) of the treatment date

- The treatment date is the date of the Medicaid expansion
- It is normalised to $t = 0$
- Leads are $\tau = -1, -2, -3, -4, -5, ...$
- Lags are $\tau = 1, 2, 3, 4, 5, ...$

Event Study Example: Miller *et al.* (2021)

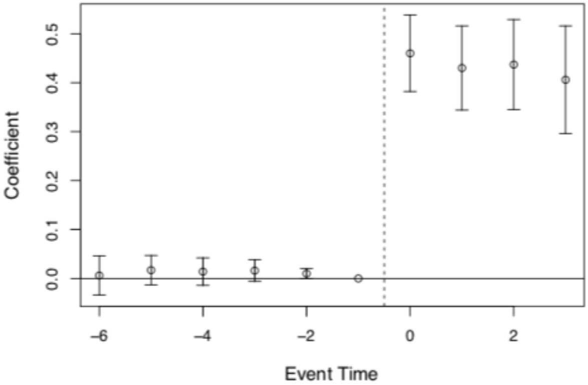
Event study model with q leads and m lags

$$Y_{its} = \gamma_s + \lambda_t + \sum_{\tau=-q}^{-1} \gamma_\tau D_{s\tau} + \sum_{\tau=0}^{m} \delta_\tau D_{s\tau} + x_{ist} + \varepsilon_{ist}$$

- Individuals i, states s, years t
- γ_s state fixed effects, λ_t year fixed effects
- x_{ist} are time-varying controls

Event Study Example: Miller *et al.* (2021)

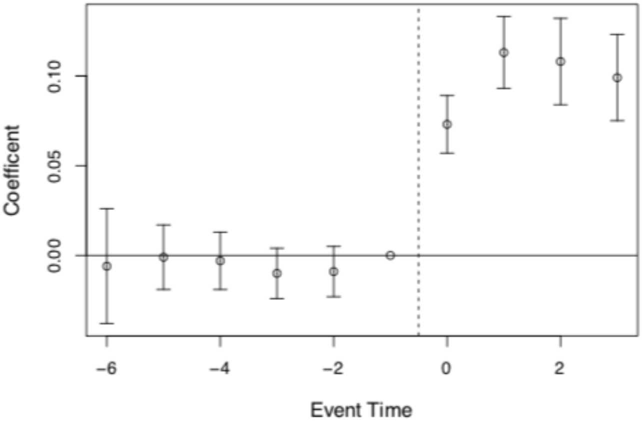
"Zero Stage": Medicaid expansion increased eligibility



(a) Medicaid Eligibility

Event Study Example: Miller *et al.* (2021)

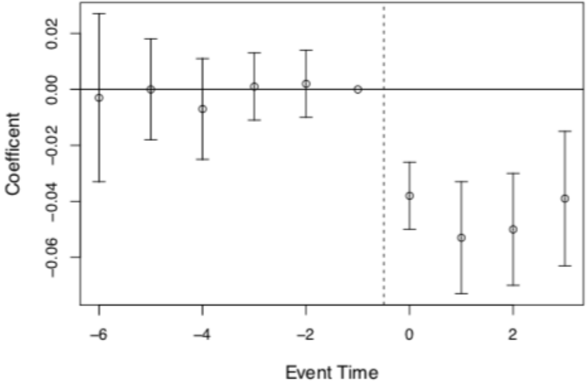
"First Stage I": Medicaid expansion increased coverage



(b) Medicaid Coverage

Event Study Example: Miller *et al.* (2021)

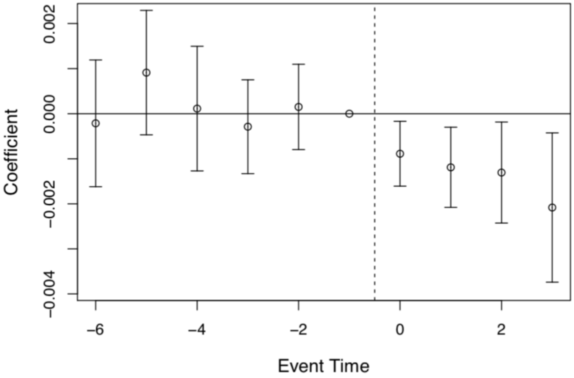
"First Stage II": Medicaid expansion reduced the number of uninsured persons



(c) Uninsured

Event Study Example: Miller *et al.* (2021)

"Reduced Form": Medicaid expansion reduced mortality by about 9%



Lessons from Miller *et al.* (2021)

Do parallel trends hold in the Medicaid expansion case?

- We don't know!

But the **authors provide very compelling evidence**

- The coefficients of the leads are close to zero and statistically insignificant
- The coefficients of the lags are large and statistically significant
- This jump is consistent with the Medicaid expansion and inconsistent with other events

Coefficients of lags are placebo tests

- The coefficients are zero when they should be zero

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.

Card, David, & Krueger, Alan. 1994. Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania. *American Economic Review*, **84**(4), 772–93.

Krueger, Alan B. 1999. Experimental Estimates of Education Production Functions. *The Quarterly Journal of Economics*, **114**(2), 497–532.

Miller, Sarah, Johnson, Norman, & Wherry, Laura R. 2021. Medicaid and Mortality: New Evidence From Linked Survey and Administrative Data*. *The Quarterly Journal of Economics*, **136**(3), 1783–1829.

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