Difference-in-Differences & two-way fixed effects

Nils Droste

2022 ClimBEco course

Causal Inference from observational data

Synopsis: Today, we will be looking into *the* classical research design for inferring causal effects from observational data (i.e. when experiments are unethical or infeasible), and its recent developments

Causal Inference from observational data

Synopsis: Today, we will be looking into *the* classical research design for inferring causal effects from observational data (i.e. when experiments are unethical or infeasible), and its recent developments

In particular, we will develop an understanding of

Causal Inference from observational data

Synopsis: Today, we will be looking into *the* classical research design for inferring causal effects from observational data (i.e. when experiments are unethical or infeasible), and its recent developments

In particular, we will develop an understanding of

- quasi / natural experiments

Causal Inference from observational data

Synopsis: Today, we will be looking into *the* classical research design for inferring causal effects from observational data (i.e. when experiments are unethical or infeasible), and its recent developments

In particular, we will develop an understanding of

- quasi / natural experiments
  - Difference-in-Differences

Causal Inference from observational data

Synopsis: Today, we will be looking into *the* classical research design for inferring causal effects from observational data (i.e. when experiments are unethical or infeasible), and its recent developments

In particular, we will develop an understanding of

- quasi / natural experiments
  - Difference-in-Differences
  - (two-way) fixed-effects regressions

Causal Inference from observational data

Synopsis: Today, we will be looking into *the* classical research design for inferring causal effects from observational data (i.e. when experiments are unethical or infeasible), and its recent developments

In particular, we will develop an understanding of

- quasi / natural experiments
  - Difference-in-Differences
  - (two-way) fixed-effects regressions
  - staggered treatment

Definition

Quasi- / natural experiment:

A setting where a subpopulation is treated with an intervention of sorts that occurs due to non-random assignment processes (outside of the researchers influence if its called *natural*).

Recall potential outcome approximation

We may choose to infer an average treatment effect (ATE) I: i.e. $I = \{A, B...\}$ by comparing the average outcomes of treated individuals a from A with the one of untreated individuals $b \in B$:

$$E\{\Delta Y_i\} \approx E\{Y_a(1)\} - E\{Y_b(0)\} \quad (1)$$

For such a case we can exploit *random chance* within sufficiently large samples to make these groups comparable.

Recall potential outcome approximation

We may choose to infer an average treatment effect (ATE) I: i.e. $I = \{A, B...\}$ by comparing the average outcomes of treated individuals a from A with the one of untreated individuals $b \in B$:

$$E\{\Delta Y_i\} \approx E\{Y_a(1)\} - E\{Y_b(0)\} \tag{1}$$

For such a case we can exploit *random chance* within sufficiently large samples to make these groups comparable.

But what if we do not have a random assignment (and there may be a selection-bias and / or substantial differences between groups)?

a famous case

By comparing a treated with an untreated group over 2+ periods, we can control for (time-constant) differences between groups.

a famous case

By comparing a treated with an untreated group over 2+ periods, we can control for (time-constant) differences between groups.

A classic example is the Card and Krueger (AER, 1994), comparing fast-food worker employment in Pennsylvania (PA) and New Jersey (NJ) before and after a minimum wage raise in NJ in 1992.

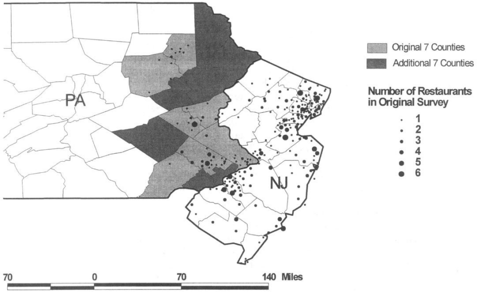


FIGURE 1. AREAS OF NEW JERSEY AND PENNSYLVANIA COVERED BY ORIGINAL SURVEY AND BLS DATA

Image source: Card and Krueger 2000

results

Here is a (tweaked) version of Card and Krueger 1994, Figure 1.

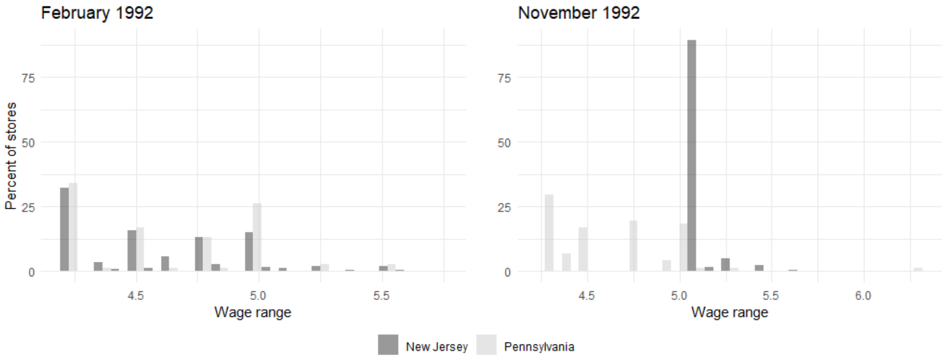


Image source: Card and Krueger 1994

potential outcome notation

Here, we are interested in the average treatment effect on the treated (ATT)

$$ATT = E[Y_i(1) - Y_i(0)|D = 1] \tag{2}$$

potential outcome notation

Here, we are interested in the average treatment effect on the treated (ATT)

$$ATT = E[Y_i(1) - Y_i(0)|D = 1] \tag{2}$$

For this to be a consistent estimator, we will need a set of conditions to hold, some of which we can test, others we will need to assume.

potential outcome notation

Here, we are interested in the average treatment effect on the treated (ATT)

$$ATT = E[Y_i(1) - Y_i(0)|D = 1] \quad (2)$$

For this to be a consistent estimator, we will need a set of conditions to hold, some of which we can test, others we will need to assume.

In particular, the parallel-trends assumption:

$$E[Y_{it}(0) - Y_{it-1}(0)|D = 1] = E[Y_{it}(0) - Y_{it-1}(0)|D = 0] \quad (3)$$

potential outcome notation

Here, we are interested in the average treatment effect on the treated (ATT)

$$ATT = E[Y_i(1) - Y_i(0)|D = 1] \tag{2}$$

For this to be a consistent estimator, we will need a set of conditions to hold, some of which we can test, others we will need to assume.

In particular, the parallel-trends assumption:

$$E[Y_{it}(0) - Y_{it-1}(0)|D = 1] = E[Y_{it}(0) - Y_{it-1}(0)|D = 0] \tag{3}$$

because then, we can assume

$$ATT = E[Y_t - Y_{t-1}|D = 1] - E[Y_t - Y_{t-1}|D = 0] \tag{4}$$

Directed acyclic graphs

Difference-in-Differences (DID)

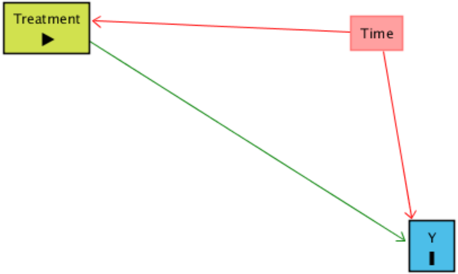
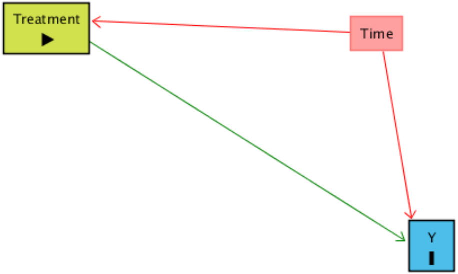


Image source: Huntington-Klein 2018

Directed acyclic graphs

Difference-in-Differences (DID)



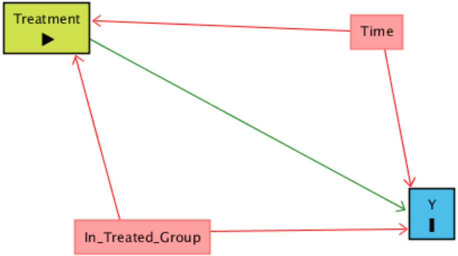


Image source: Huntington-Klein 2018

Directed acyclic graphs

Difference-in-Differences (DID)

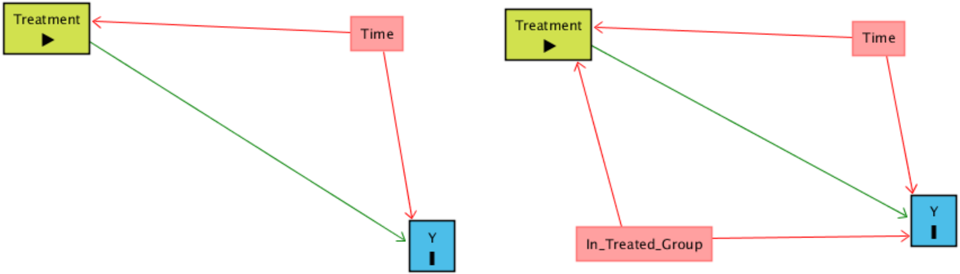
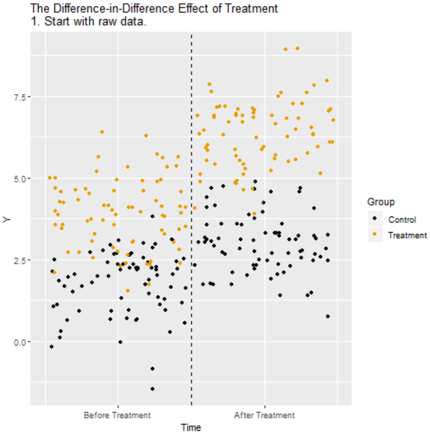


Image source: Huntington-Klein 2018

→ Accounting for differences between groups over time enables not just the estimation of time and group effects but also the differences-in-differences (i.e. the interaction of time and group effects).

Estimation



DID Estimator

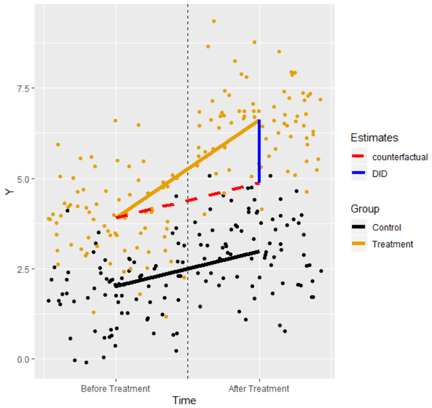
assume n individual units i, and $t = 2$ time periods, we can estimate the effect of a treatment ocurring at $P_{t=1}$, affecting the treated subpopulation $D_i = 1$

$$\begin{aligned} Y_i = \beta_0 + \beta_1 D_i + \beta_2 P_t \\ + \beta_3 D_i \times P_t + \varepsilon_i \end{aligned} \tag{5}$$

with D_i = Treatment, P_t = Period Dummy.

β_3 gives us an estimate of the diff-in-diff treatment effect.

Estimation



DID Estimator

assume n individual units i, and $t = 2$ time periods, we can estimate the effect of a treatment ocurring at $P_{t=1}$, affecting the treated subpopulation $D_i = 1$

$$\begin{aligned} Y_i = \beta_0 + \beta_1 D_i + \beta_2 P_t \\ + \beta_3 D_i \times P_t + \varepsilon_i \end{aligned} \tag{5}$$

with D_i = Treatment,
P_t = Period Dummy.

β_3 gives us an estimate of the diff-in-diff treatment effect.

regression

To estimate an example ATT, we can use the Card and Krueger (1994) data

```
> did_model <- lm(emptot ~ time + treated + time:treated,
                  data = card_krueger_1994_mod)
> summary(did_model)

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    23.331      1.072  21.767   <2e-16 ***
time           -2.166      1.516  -1.429   0.1535
treated        -2.892      1.194  -2.423   0.0156 *
time:treated    2.754      1.688   1.631   0.1033
```

Note: heteroskedasticity and autocorrelation robust standard errors should be computed

intermediate summary

A difference-in-differences approach allows us to

- compare a treatment group with an untreated quasi-counterfactual
- even under conditions of a non-random assignment
- assuming that the groups behave comparably enough

Panel Data

Let us suppose a situation with repeated measurements for multiple individual(s) (units), i.e. cross-sectional time-series, or panel data.

An example panel data structure

individual (i)	**time (t)**	$\mathbf{Y}_{it}$	$\mathbf{X}_{it}$	$\mathbf{D}_{it}$
A	1	0.8	0.3	0
A	2	0.7	0.2	0
A	3	0.5	0.2	1
B	1	1.2	0.4	0
B	2	1.1	0.5	0
B	3	0.9	0.6	1
...	...	...	...	...

Fixed-Effects

An exemplary study

ELSEVIER
Available online at www.sciencedirect.com
SCIENCE DIRECT®
Ecological Economics 55 (2005) 527–538
ECOLOGICAL ECONOMICS
www.elsevier.com/locate/ecolecon

ANALYSIS

Environmental pressure group strength and air pollution: An empirical analysis

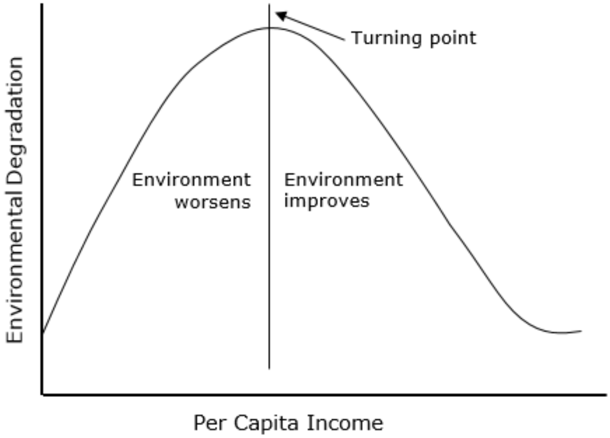
Seth Binder, Eric Neumayer*

Department of Geography and Environment and Center for Environmental Policy and Governance (CEPG), London School of Economics and Political Science, Houghton Street, London WC2A 2AE, UK

Received 7 December 2003; received in revised form 22 October 2004; accepted 14 December 2004
Available online 24 February 2005

Fixed-Effects

An exemplary study



Theoretical underpinning for Binder and Neumayer 2005: Environmental Kuznets Curve. Image source: Wikipedia

Fixed-Effects

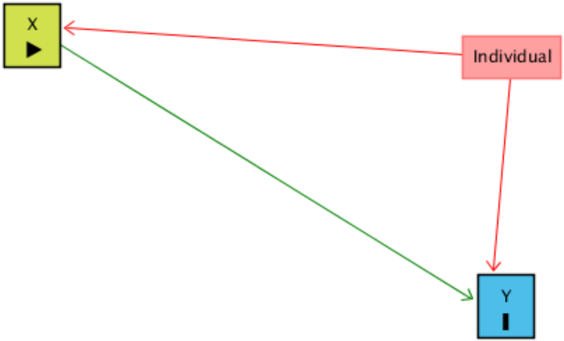
Replicated results from table 2, model 1 Binder and Neumayer 2005

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  -57.55012   16.60671  -3.465 0.000548 ***
lnengopc      -0.51121    0.11878  -4.304 1.82e-05 ***
lnenergy       1.00887    0.60455   1.669 0.095425 .
lngdp         13.81819    4.17975   3.306 0.000975 ***
lngdpsq       -0.88657    0.27384  -3.238 0.001239 **
polity        -0.05079    0.03023  -1.680 0.093135 .
```

Note: This is not a fixed effects regression but an OLS

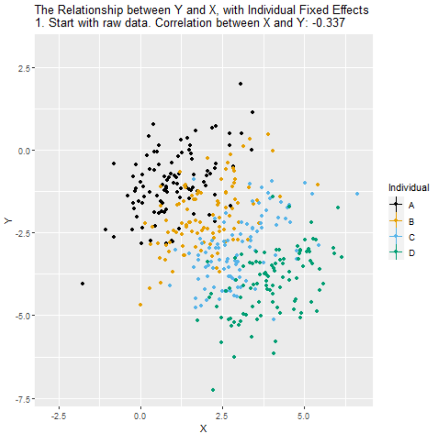
Panel Data

One way fixed-Effects (FE) structure

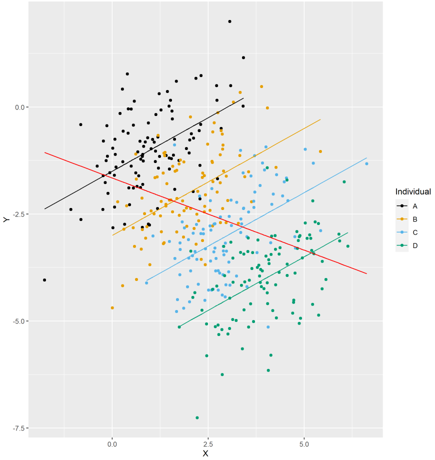


Accounting for individual fixed effects. Image source: Huntington-Klein 2018

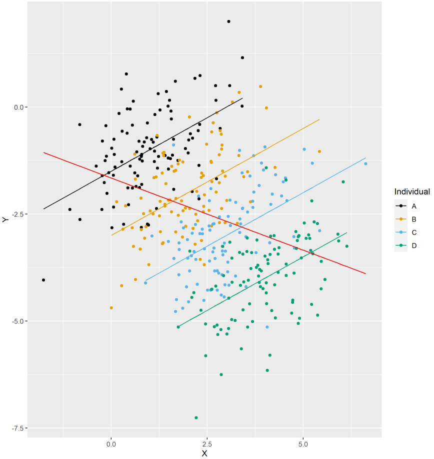
Fixed-Effects



Fixed-Effects



Fixed-Effects



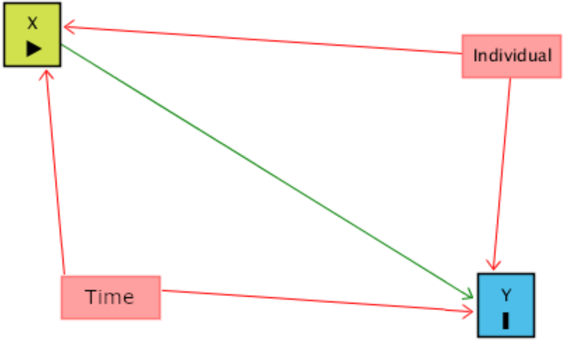
FE Estimator

$$Y_i = \alpha_i + \beta X_{it} + \varepsilon_{it} \quad (6)$$

with α_i = individual, time-invariant effect, and X_i = a variable of interest. The so called *within* transformation accounts for α_i through demeaning such that we can consistently estimate $\partial Y / \partial X$

Fixed-Effects

Two way fixed-effects



Accounting for individual and time (two-way) fixed effects. Image source: Huntington-Klein 2018

Estimator

The two-way fixed effects model can be formulated as

$$Y_{it} = \alpha_i + \theta_t + \tau D_{it} + \beta X_{it} + \varepsilon_i \tag{7}$$

where α_i individual and θ_t time fixed effects, respectively
Note that D here is a binary treatment indicator ($D_{it} = D_i \times P_t$),
possibly a vector of control variables X_i,
and error term ε_i, assumed to be normally distributed and centered around 0,
indepdendent of everything else.

time-trends

This allows to take both individual differences and shocks in time into account

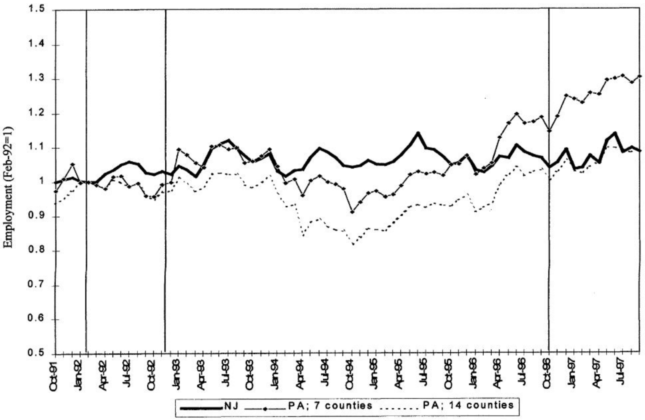


FIGURE 2. EMPLOYMENT IN NEW JERSEY AND PENNSYLVANIA FAST-FOOD RESTAURANTS, OCTOBER 1991 TO SEPTEMBER 1997

Note: Vertical lines indicate dates of original Card-Krueger survey and the October 1996 federal minimum-wage increase.
Source: Authors' calculations based on BLS ES-202 data.

Image source: Card and Krueger 2000

intermediate summary

A two-way fixed effects approach allows us to

- control for time-constant, unobserved heterogeneity between individuals
- control for common time shocks, that affect all individuals
- a multi-period 2WFE approach resembles DiD when
  - treatment is simultaneous
  - effects are homogeneous

Staggered treatment and other current developments

Over the last 6 or so years, DiD methdology has seen quite some development.

I will briefly introduce

Staggered treatment and other current developments

Over the last 6 or so years, DiD methdology has seen quite some development.

I will briefly introduce

- staggered treatment / event-time studies

Staggered treatment and other current developments

Over the last 6 or so years, DiD methdology has seen quite some development.

I will briefly introduce

- staggered treatment / event-time studies
- non-parallel trend corrections

Staggered treatment and other current developments

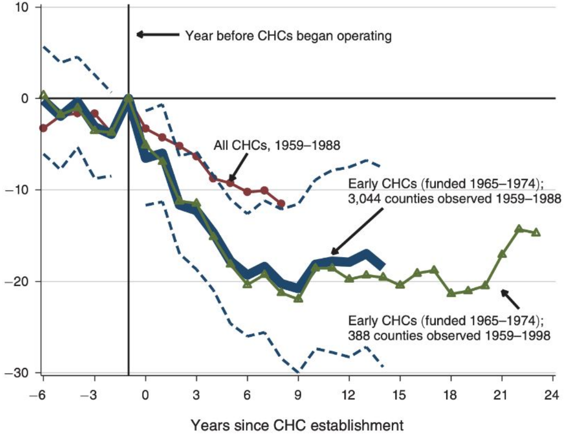
Over the last 6 or so years, DiD methdology has seen quite some development.

I will briefly introduce

- staggered treatment / event-time studies
- non-parallel trend corrections
- heterogeneous treatment corrections

intuition

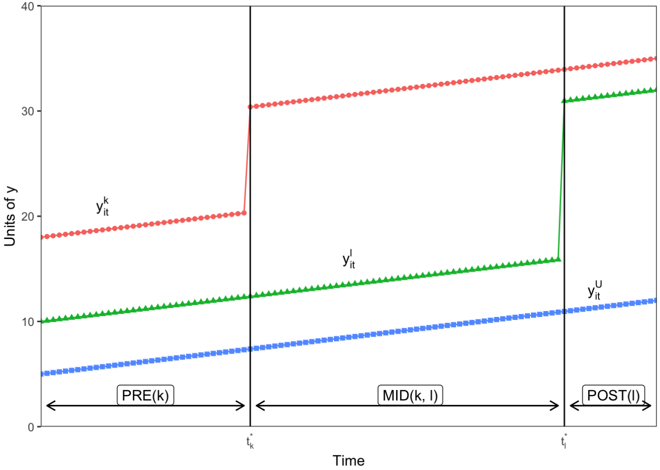
What if treatment happened at different times?



Reduction in mortality by Community Health Centers (CHC). Image source: Bailey and Goodman-Bacon 2015

intuition

We may run into time series length related weighting problems



DiD with variations in treatment timing. Image source: Goodman-Bacon 2018, as reproduced by A.C. Baker

intuition

Recall,

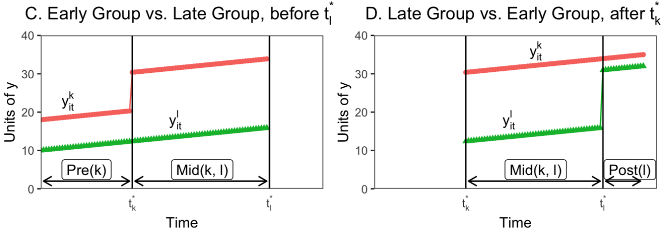
$$ATT = E[Y_t - Y_{t-1}|D = 1] - E[Y_t - Y_{t-1}|D = 0]$$

which is a comparison of mean differences between groups over time. Straightforward for 2 periods.

Now, we have multiple periods.

intuition

Picture a case with multiple periods, three groups, and thus four comparisons



DiD with variations in treatment timing. Image source: Goodman-Bacon 2018, as reproduced by A.C. Baker

notation

Turns out: 2WFE DiD estimator is a weighted average of these comparisons

Theorem 1. Difference-in-Differences Decomposition Theorem
Assume that the data contain $k = 1, \dots, K$ groups of units ordered by the time when they receive a binary treatment, $t_k^* \in (1, T]$. There may be one group, U, that never receives treatment. The OLS estimate, $\hat{\beta}^{DD}$, in a two-way fixed-effects model (2) is a weighted average of all possible two-by-two DD estimators.

$$\hat{\beta}^{DD} = \sum_{k \neq U} s_{kU} \hat{\beta}_{kU}^{2x2} + \sum_{k \neq U} \sum_{\ell > k} s_{k\ell} \left[\mu_{k\ell} \hat{\beta}_{k\ell}^{2x2,k} + (1 - \mu_{k\ell}) \hat{\beta}_{k\ell}^{2x2,\ell} \right] \quad (7)$$

Where the two-by-two DD estimators are:

$$\hat{\beta}_{kU}^{2x2} \equiv \left(\bar{y}_k^{POST(k)} - \bar{y}_k^{PRE(k)} \right) - \left(\bar{y}_U^{POST(j)} - \bar{y}_U^{PRE(j)} \right)$$

$$\hat{\beta}_{k\ell}^{2x2,k} \equiv \left(\bar{y}_k^{MID(k,\ell)} - \bar{y}_k^{PRE(k)} \right) - \left(\bar{y}_\ell^{MID(k,\ell)} - \bar{y}_\ell^{PRE(k)} \right)$$

$$\hat{\beta}_{k\ell}^{2x2,\ell} \equiv \left(\bar{y}_\ell^{POST(\ell)} - \bar{y}_\ell^{MID(k,\ell)} \right) - \left(\bar{y}_k^{POST(\ell)} - \bar{y}_k^{MID(k,\ell)} \right)$$

the weights are:

$$s_{kU} = \frac{n_k n_U \bar{D}_k (1 - \bar{D}_k)}{\widehat{var}(\tilde{D}_{it})}$$

$$s_{k\ell} = \frac{n_k n_\ell (\bar{D}_k - \bar{D}_\ell)(1 - (\bar{D}_k - \bar{D}_\ell))}{\widehat{var}(\tilde{D}_{it})}$$

$$\mu_{k\ell} = \frac{1 - \bar{D}_k}{1 - (\bar{D}_k - \bar{D}_\ell)}$$

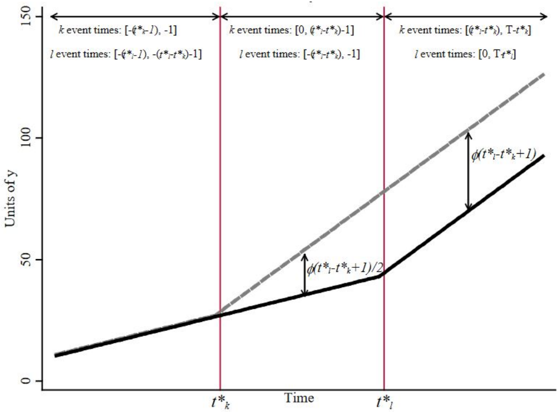
and $\sum_{k \neq U} s_{kU} + \sum_{k \neq U} \sum_{\ell > k} s_{k\ell} = 1$.

Proof: See appendix A.

Bacon decomposition theorem. Image source: Goodman-Bacon 2018, see also his tweet-thread

notation

Even worse, a positive but staggered slope change, can even change the sign of the DD estimate



Staggered treatment with linear slope change. Image source: Goodman-Bacon 2018, see also his tweet-thread

Proposed solutions I

Several works on the issue:

Proposed solutions I

Several works on the issue:

- Athey and Imbens 2020
  - under random adoption dates, a weighted DID is unbiased
  - comparing treated with not-yet treated

Proposed solutions I

Several works on the issue:

- Athey and Imbens 2020
  - under random adoption dates, a weighted DID is unbiased
  - comparing treated with not-yet treated
- Goodman-Bacon 2018
  - a time-in-treatment weighted DiD, fixing the weights to gain balance
  - R package bacondecomp

Proposed solutions II

Several works on the issue:

Proposed solutions II

Several works on the issue:

- Callaway and Sant'Anna 2020
  - bootstrapped inference with pre-intervention conditioning on co-variates
  - R package did

Proposed solutions II

Several works on the issue:

- Callaway and Sant'Anna 2020
  - bootstrapped inference with pre-intervention conditioning on co-variates
  - R package did
- Sun and Abraham 2020
  - time-to-treatment (cohort) weighted approach, comparing to never-treated
  - implemented in fixest R package

Heterogeneous Treatment

Further works on DiD design / application issues:

- Chaisemartin and D'Haultfœuille 2020
  - allowing for treatment effects "heterogeneous across groups and over time periods"
  - R package DIDmultiplegt
- Imai and Kim 2020
  - identify some problematic negative weigthing in 2WFE
  - they propose weighting or matching Imai 2020

Non-parallel trends

However, all these approaches, assume (to some extent) parallel trends

- Jonathan Roth 2021
  - test for parallel-trends, R package pretends
  - robust inference strategy under non-parralel trends Rambachan and Roth 2020, R package HonestDiD

Summary

Summary of DID

Instantaneous (Static) Effects	**Dynamic Effects**
Potential Outcomes	
$Y_{it}(d)$ for $d \in \{0,1\}$	$Y_{it}(g)$ for $g \in \{1, \ldots, T, \infty\}$
Homogeneous Effects	
Target Parameter	
$\tau = \mathbb{E}\left[Y_{it}(1) - Y_{it}(0) \mid D_{it} = 1\right]$	$\tau_\ell = \mathbb{E}\left[Y_{ig+\ell}(g) - Y_{ig+\ell}(\infty) \mid G_i = g\right]$
Estimators	
$\widehat{\tau}_{\mathrm{MM}} = \left(\bar{Y}_{1,\mathrm{post}} - \bar{Y}_{1,\mathrm{pre}}\right) - \left(\bar{Y}_{0,\mathrm{post}} - \bar{Y}_{0,\mathrm{pre}}\right)$	$\widehat{\tau}_{\ell,\mathrm{DDID}} : Y_{it} = \alpha_i + \gamma_t + \sum_{\ell \in \mathcal{L}} D_i P_t^\ell \tau_\ell + \varepsilon_{it}$
$\widehat{\tau}_{\mathrm{OLS}} : Y_{it} = \mu + \alpha D_i + \gamma P_t + \tau D_i P_t + \varepsilon_{it}$	$\widehat{\tau}_{\ell,\mathrm{ES}} : Y_{it} = \alpha_i + \gamma_t + \sum_{\ell \in \mathcal{L}} D_{it}^\ell \tau_\ell + \varepsilon_{it}$
$\widehat{\tau}_{\mathrm{TWFE}} : Y_{it} = \alpha_i + \gamma_t + \tau D_{it} + \varepsilon_{it}$	
Key references	
Card and Krueger (1994)	Ashenfelter and Card (1984)
Bertrand, Duflo, and Mullainathan (2004)	Jacobson, LaLonde, and Sullivan (1993)
Heterogeneous Effects	
Target Parameter	
$\tau_{gt} = \frac{1}{N_{gt}} \sum_{i=1}^{N_{gt}} \left[Y_{it}(1) - Y_{it}(0)\right]$	$\tau_t(g) = \mathbb{E}\left[Y_{it}(g) - Y_{it}(\infty) \mid G_i = g\right]$
Estimators	
$\mathrm{DID}_{c,t} = \sum_{i,t} \left[\rho_{it}^{h_1(c)} \left(Y_{it} - Y_{it-1}\right) - \rho_{it}^{h_2(c)} \left(Y_{it} - Y_{it-1}\right)\right]$	$\widehat{\tau}_t(g) = \left(\bar{Y}_{g,t} - \bar{Y}_{\infty,t}\right) - \left(\bar{Y}_{g,g-1} - \bar{Y}_{\infty,g-1}\right)$
$\widehat{\tau}_{\mathrm{dCdH}} = \sum_{t=2}^{T} \left(\omega_{+,t} \mathrm{DID}_{+,t} + \omega_{-,t} \mathrm{DID}_{-,t}\right)$	$\widehat{\tau}_{\mathrm{CS}} = \frac{1}{\sum_t \sum_{g:g \le t} N_g} \sum_t \sum_{g:g \le t} \tau_t(g)$
Key references	
de Chaisemartin and d'Haultfoeuille (2020)	Callaway and Sant'Anna (forthcoming)
Goodman-Bacon (2021)	Sun and Abraham (2021)

Schönholzer 2022, cf. Roth and Sant'Anna 2021

Summary

Re: DiD and (adjusted) 2WFE estimators

- compare treated individual (units) with a reference group
- who we compare to whom deserves special attention
- software solutions available

References I

Athey, Susan and Guido W Imbens (2020). 'Design-based analysis in Difference-In-Differences settings with staggered adoption'. In: *Journal of Econometrics*. DOI: 10.1016/j.jeconom.2020.10.012.

Bailey, Martha J and Andrew Goodman-Bacon (2015). 'The war on poverty's experiment in public medicine: Community health centers and the mortality of older Americans'. In: *American Economic Review* 105.3, pp. 1067–1104. DOI: 10.1257/aer.20120070.

Binder, Seth and Eric Neumayer (2005). 'Environmental pressure group strength and air pollution: An empirical analysis'. In: *Ecological Economics* 55.4, pp. 527–538. DOI: 10.1016/j.ecolecon.2004.12.009.

Callaway, Brantly and Pedro H.C. Sant'Anna (2020). 'Difference-in-Differences with multiple time periods'. In: *Journal of Econometrics* xxxx, pp. 1–31. DOI: 10.1016/j.jeconom.2020.12.001.

Card, David and Alan B Krueger (1994). 'Minimum wages and employment: a case study of the fast-food industry in New Jersey and Pennsylvania'. In: *American Economic Review* 84.4, pp. 772–793. DOI: 10.3386/w4509.

— (2000). 'Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania: Reply'. In: *American Economic Review* 90.5, pp. 1397–1420. DOI: 10.1257/aer.90.5.1397.

Chaisemartin, Clément de and Xavier D'Haultfœuille (2020). 'Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects'. In: *American Economic Review* 110.9, pp. 2964–2996. DOI: 10.1257/aer.20181169.

References II

Goodman-Bacon, Andrew (2018). 'Difference-in-Differences With Variation in Treatment Timing'. In: *National Bureau of Economic Research* 17.5, pp. 684–694. DOI: 10.3386/w25018.

Imai, Kosuke and In Song Kim (2020). 'On the Use of Two-Way Fixed Effects Regression Models for Causal Inference with Panel Data'. In: *Political Analysis*, pp. 1–11. DOI: 10.1017/pan.2020.33.

Rambachan, Ashesh and Jonathan Roth (2020). 'An Honest Approach to Parallel Trends'.

Roth, Jonathan (2021). 'Pre-test with Caution: Event-study Estimates After Testing for Parallel Trends'.

Roth, Jonathan and Pedro H.C. Sant'Anna (2021). 'Efficient Estimation for Staggered Rollout Designs'. In: pp. 1–52. arXiv: 2102.01291. URL: http://arxiv.org/abs/2102.01291.

Sun, Liyang and Sarah Abraham (2020). 'Estimating dynamic treatment effects in event studies with heterogeneous treatment effects'. In: *Journal of Econometrics* xxxx, pp. 1–25. DOI: 10.1016/j.jeconom.2020.09.006.