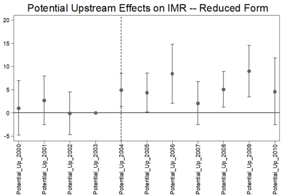
Session IV
Introduction to difference-in-differences

Evaluating public policies

Arthur Heim (PSE & Cnaf)



February 20, 2023*

Outline

1. Introduction
   - In the previous sessions:
   - Leaving the experimental ideal
   - What's a difference of difference
   - What's the plan
2. DID: the 2x2 case
3. Case study: Minimum wage by Card and Krueger (AER 1994)
4. Multiple groups, multiple periods
5. Modern DiD: Application to the minimum wage debate

Introduction

In the previous sessions:

* Source: Dias, Rocha, and Soares (2023) ; Difference-in-differences estimates of the effect of glyphosate on infant birth outcomes

In the previous sessions:

- We have seen how randomized control trials can retrieve causal effects
- We discussed how to use linear regressions to estimate these parameters and, more broadly, what regressions can and cannot estimate.
- We discussed the role of the conditional independence assumption and its implementation in multivariate regressions.
- We have seen two important issues for inference: heteroskedasticity and clustering

Now, we leave the experimental ideal for the ”natural experiments” world

Introduction

Leaving the experimental ideal

- Most of the time we do not directly manipulate treatment assignment, even less often randomly.
- How can we deal with selection bias if we do not randomize ? How do we estimate causal effects when treatment is **endogenous** ?

Introduction

Leaving the experimental ideal

- Most of the time we do not directly manipulate treatment assignment, even less often randomly.
- How can we deal with selection bias if we do not randomize ? How do we estimate causal effects when treatment is **endogenous** ?
- Econometricians have come a long way to define settings and hypotheses that get you close to a randomized experiment. That's why these methods are sometimes refereed to as "quasi-experiments" or "natural experiments"
- This lecture is about **differences in differences** (DiD or Diff-in-Diff), one of the (if not the) most popular method used in empirical work to estimate the impact of a policy.
- Identification comes from a non-refutable assumption called **parallel trend**: Treated and control outcomes would have followed the same path (parallel) had the policy not been implemented.
- The idea is simple (so, easy to communicate) and there are many settings where it is a very well fitted design.

What's a difference of difference

When simple differences are not enough

1. You can always compare participants (treated) and non-participants (untreated), even controlling for some characteristics using e.g. OLS.
   - **Problem:** There may be unobserved differences associated with both treatment status and outcomes. In that case, estimates suffer from omitted variable bias.
2. If you have panel data or repeated cross sections, you can compare before and after treatment.
   - **Problem:** "Spontaneous" evolution or trend that's hard to distinguish from treatment effect.

The fundamental problem of causal inference is that the potential outcomes in the **counterfactual cannot be observed**. Can we estimate functions of this counterfactual using other observations ?

What's a difference of difference

Sometimes two differences are better than one

- What if treated and untreated had a similar evolution in the past ?
- Said differently, what if the (observed and unobserved) differences between the two groups remain constant ?
- Then if we take the difference *after - before* for both groups, we remove the common spontaneous evolution. If we take the difference between the groups over both period, we remove the constant difference between the two groups. If we take the difference between these two differences, we remove the spontaneous evolution in time and the permanent differences between the two groups. What remains is the treatment effect.
- **Conditions:** Having data before and after treatment for both groups and plausible parallel trends.

Introduction

Warnings

- In the simplest case: you have two groups, two periods. One group is treated at the second period.
- But many settings involve comparing multiple groups, multiple periods, treatment starting at different time for different people.
- Until recently, economists used their favourite regression tools in these more generalized settings with multiple periods, multiple groups or multiple groups affected at different dates, thinking the generalization was straightforward.
- This proved to be wrong unless one makes strong restrictions on the treatment effect heterogeneity and dynamics. Furthermore, the usual estimation methods may be very biased.
- **NB** The second part of this class is a bit more advanced but it would have been wrong not to give you an idea of what's trending in the econometrics of difference in differences.

Introduction

What's the plan

1. The 2x2 case: theory and illustration.
2. Case study on minimum wage
3. Multiple groups, multiple period - an introduction
4. New case study on minimum wage

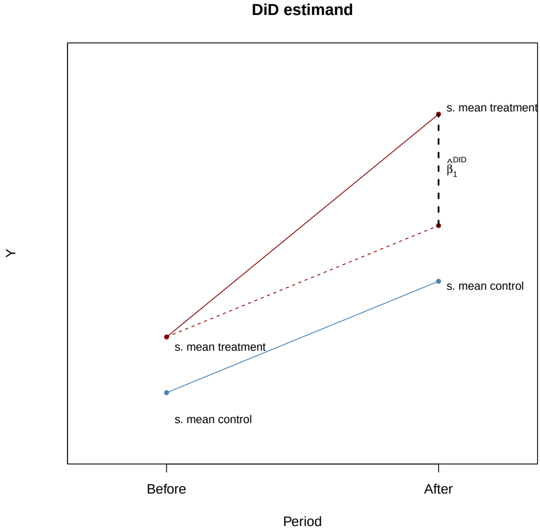
Outline

DID: the 2x2 case

Intuition

- We consider two groups, one of which is affected by a policy. We observe both groups before the policy is implemented and after.
- We can recover the average treatment effect on the treated if we assume parallel trend which is equivalent to one of this two definition:
  1. Suppose the difference between treated and control individuals are constant on average
  2. Suppose groups are on the same path
- Then the **before/after difference** remove the common evolution between groups, the **difference between groups** remove the constant difference between the groups.
- the double difference identify an average treatment effect.
- Condition 1 or 2 are called the parallel trend assumption.

Intuition


DiD estimand
s. mean treatment
$\hat{\beta}_1^{DID}$
s. mean control
s. mean treatment
s. mean control
Y
Before
After
Period

Simulation using R

```r
# set sample size
n <- 400
# define treatment effect
TEffect <- 3
# generate treatment dummy
TDummy <- c(rep(0, n/2), rep(1, n/2))
set.seed(666)
# simulate pre- and post-treatment values of the dependent variable
y_pre <- 7 + rnorm(n, 0, 2)
y_pre[1:n/2] <- y_pre[1:n/2] - 1
y_post <- 7 + 2 + TEffect * TDummy + rnorm(n, 0, 2)
y_post[1:n/2] <- y_post[1:n/2] - 1

dfDiD <- as.data.frame(cbind(y_post, y_pre, TDummy))

dflong <- dfDiD %>%
    pivot_longer(cols = c(y_post, y_pre), names_to = c("period"), values_to = "Y") %>%
    mutate(time = ifelse(period == "y_post", 1, 0), period = factor(period, levels = c("y_pre",
        "y_post")))

averages <- dflong %>%
    group_by(period, TDummy) %>%
    summarise(Ybar = mean(Y), sd = sd(Y), n = n(), se = sd/sqrt(n))

control_increase = averages$Ybar[averages$period == "y_post" & averages$TDummy ==
    0] - averages$Ybar[averages$period == "y_pre" & averages$TDummy == 0]
```

Graphical representations

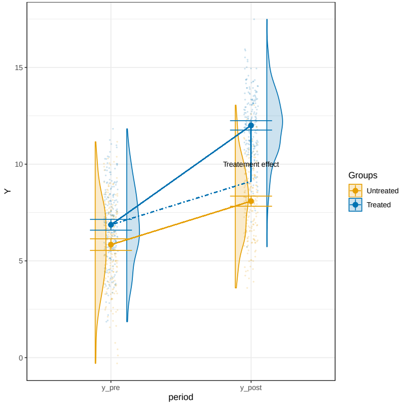


Figure 2: Simulated data

Identification in the 2x2 case

Setting and notations

- There is a lot more to know, I strongly advise to read the difference-in-differences chapter in Cunningham (2018).
- As usual, let Y be our outcome but let's be more flexible on the treatment status notation and define treatment groups k and an untreated group U.
- There is a pre-period for the treatment group, $pre(k)$; a post-period for the treatment group, $post(k)$; a pre-treatment period for the untreated group $pre(U)$; and a post period for the untreated group $post(U)$.
- The DiD estimator is defined as:

$$\hat{\delta}_{kU}^{2\times 2} = \left(\bar{Y_k}^{post(k)} - \bar{Y_k}^{pre(k)}\right) - \left(\bar{Y_U}^{post(U)} - \bar{Y_U}^{pre(U)}\right)$$

Identification in the 2x2 case

Retrieving causal effect of interest

- The DiD estimator is defined as:

$$\hat{\delta}_{kU}^{2\times 2} = \left(\bar{Y_k}^{post(k)} - \bar{Y_k}^{pre(k)}\right) - \left(\bar{Y_U}^{post(U)} - \bar{Y_U}^{pre(U)}\right)$$

- Assuming random sampling from a large population, these empirical averages can be rewritten in population expectation:

$$\hat{\delta}_{kU}^{2\times 2} = \Big(\mathbb{E}[Y_k|post(k)] - \mathbb{E}[Y_k|pre(k)]\Big) - \Big(\mathbb{E}[Y_U|post(U)] - \mathbb{E}[Y_U|pre(U)]\Big)$$

- Assuming SUTVA, observed quantities reveal potential outcomes

$$\begin{aligned}\hat{\delta}_{kU}^{2\times 2} =& \Big(\mathbb{E}[Y_k(k)|post(k)] - \mathbb{E}[Y_k(0)|pre(k)]\Big) \\ & - \Big(\mathbb{E}[Y_U(0)|post(U)] - \mathbb{E}[Y_U(0)|pre(U)]\Big)\end{aligned}$$

Identification in the 2x2 case

Retrieving causal effect of interest

- Now, the usual trick: add and substract counterfactual values and re-arrange:

$$
\begin{aligned}
\hat{\delta}_{kU}^{2\times 2} &= \mathbb{E}[Y_k(k)|post(k)] - \mathbb{E}[Y_k(0)|pre(k)] \\
&\quad - (\mathbb{E}[Y_U(0)|post(U)] - \mathbb{E}[Y_U(0)|pre(U)]) \\
&\quad + \mathbb{E}[Y_k(0)|post(k)] - \mathbb{E}[Y_k(0)|post(k)] \\
&= \underbrace{\mathbb{E}[Y_k(k)|post(k)] - \mathbb{E}[Y_k(0)|post(k)]}_{ATT} + \\
&\quad (\mathbb{E}[Y_k(0)|post(k)] - \mathbb{E}[Y_k(0)|pre(k)]) \\
&\quad - (\mathbb{E}[Y_U(0)|post(U)] - \mathbb{E}[Y_U(0)|pre(U)]) \\
&= ATT + \text{Non parallel-trend bias}
\end{aligned}
\tag{1}
$$

- The DiD estimator retrieves the **average treatment effect on the treated** if and only if the second term zeros out, although it's based on a pure theoretical construct $\mathbb{E}[Y_k(0)|post(k)]$, the situation of the treated after the treatment occurred had they not been treated.

Identification in the 2x2 case

Retrieving causal effect of interest

- So, assuming parallel trend means assuming:

$$\mathbb{E}[Y_k(0)|post(k)] - \mathbb{E}[Y_k(0)|pre(k)] = \mathbb{E}[Y_U(0)|post(U)] - \mathbb{E}[Y_U(0)|pre(U)] \quad (2)$$

- The parallel trends assumption can be rationalized by imposing a particular generative model for the untreated potential outcomes:

$$Y_{it}(0) = \alpha_i + \phi_t + \varepsilon_{it} \quad (3)$$

- If we assume the previous **data generating process** where ε_{it} is **mean-independent** of D_i, then parallel trend holds.
- This model allows treatment to be assigned **non-randomly** based on characteristics that affect the **level** of the outcome α_i, but requires the treatment assignment to be **mean-independent** of variables that affect the **trend** in the outcome (ε_{it}).
- In other words, parallel trends **allows for the presence of selection bias**, but the bias from selecting into treatment must be the same across periods.

Identification in the 2x2 case

Retrieving causal effect of interest

- Actually, parallel trends come with an "implicit" assumption of **no anticipation** which states that the treatment has no causal effect prior to its implementation.
- This is important for identification of the ATT, since otherwise the changes in the outcome for the treated group between period 1 and 2 could reflect not just the causal effect in period $t = 2$ but **also the anticipatory effect** in period $t = 1$

$$Y_{i1}(1) = Y_{i1}(0) \; \forall \; i \; with \; D_i = 1 \tag{4}$$

- This assumption also relates to the problem of "Ashenfelter dips" in program evaluation where we typically observe a negative shock before entering a program (Heckman and Smith 1999).
- If Y is earnings, and t1 is measured at the time of a transitory earnings dip, and if non-participants do not experience the dip, then the previous equation will be violated, because the time path of no-program earnings between t1 and t2 will be different between participants and non-participants.
- The Ashenfelter dip can also be a form of **collider bias** if eligibility is conditioned on outcome (e.g. being unemployed)

Estimation

From theory to practice

- the DID estimator can be non-parametrically estimated by computing :

$$\hat{\delta}_{kU}^{2\times 2} = \underbrace{\left(\bar{Y_k}^{post(k)} - \bar{Y_k}^{pre(k)}\right)}_{\Delta_k} - \underbrace{\left(\bar{Y_U}^{post(U)} - \bar{Y_U}^{pre(U)}\right)}_{\Delta_U}$$

- In a large population framework and an $i.i.d.$ sample, the associates standard errors are:

$$SE_{\hat{\delta}} = \sqrt{\frac{S\left(\Delta_k\right)}{n_k} + \frac{S\left(\Delta_U\right)}{n_u}} \quad (5)$$

- Or, we can estimate the model using OLS and only dummies:

$$Y_{it} = \alpha + \beta D_i + \gamma post_t + \delta D_i \times post_t + \varepsilon_{it} \quad (6)$$

- Or equivalently:

$$Y_i(post) - Y_i(pre) = \alpha + \delta D_i + \varepsilon_i$$

Estimation

Regression in the 2x2 case

$$Y_{it} = \alpha + \beta D_i + \gamma post_t + \delta D_i \times post_t + \varepsilon_{it} \tag{7}$$

- **Note**: D_i in this specification code for individuals or groups who are "ever treated" and the interaction captures the treatment "switching".
- The regression estimates the conditional expectation function based on two dummies and their interaction

Estimation

Regression in the 2x2 case

$$Y_{it} = \alpha + \beta D_i + \gamma post_t + \delta D_i \times post_t + \varepsilon_{it} \tag{7}$$

- **Note**: D_i in this specification code for individuals or groups who are "ever treated" and the interaction captures the treatment "switching".
- The regression estimates the conditional expectation function based on two dummies and their interaction
- **fully saturated regression**: What do we get ?

$$\mathbb{E}[Y_{it}|D_i, post] = \alpha + \beta D_i + \gamma post_t + \delta D_i \times post_t$$

- Hence $\mathbb{E}[Y_{it}|D_i = 0, post = 0] = \alpha$, the average of the untreated before,
- $\mathbb{E}[Y_{it}|D_i = 0, post = 1] = \alpha + \gamma$, the average of the untreated after
- $\mathbb{E}[Y_{it}|D_i = 1, post = 0] = \alpha + \beta$, the average of the treated before
- $\mathbb{E}[Y_{it}|D_i = 1, post = 1] = \alpha + \beta + \gamma + \delta$, the average of treated after

Estimation

Regression in the 2x2 case

- Isolate δ in the last equation by substituting the other parameters by their expectation equivalent and voilà:
- $\delta = \mathbb{E}[Y_{it}|D_i = 1, post = 1] - \mathbb{E}[Y_{it}|D_i = 1, post = 0] - (\mathbb{E}[Y_{it}|D_i = 0, post = 1] - \mathbb{E}[Y_{it}|D_i = 0, post = 0])$
- The coefficient of the interaction between group and time correspond to the difference-in-differences estimand
- It has a **causal interpretation if and only if** the parallel trend hold.
- **Note**: estimating this regression works great with RCT because randomisation imply parallel trend. In that case, the coefficient on D_i should be 0 as randomisation should remove baseline differences but if it didn't, the DID correct this baseline imbalance and remove it from the post-exposure difference between treated and control.

Estimation

Estimating the effect on the data generated previously

```
didhand <- summary(y_post[TDummy == 1] - y_pre[TDummy == 1] - (y_post[TDummy == 0] -
    y_pre[TDummy == 0]))
sehand <- sqrt((var(y_post[TDummy == 1] - y_pre[TDummy == 1]))/length(y_post[TDummy ==
    1] - y_pre[TDummy == 1]) + (var(y_post[TDummy == 0] - y_pre[TDummy == 0]))/length(y_post[TDummy ==
    0] - y_pre[TDummy == 0]))
# Or the regression models
didreg <- lm(Y ~ period * TDummy, dflong)
didreg2 <- lm(I(y_post - y_pre) ~ TDummy)
print(paste("DID by hand rounded:", round(didhand[4], 2), " And it's homoskedastic SE:",
    round(sehand, 3)))

[1] "DID by hand rounded: 2.89  And it's homoskedastic SE: 0.263"
```

Estimation

Results

	(1)	(2)
(Intercept)	5.841***	2.248***
	(0.138)	(0.186)
periody_post	2.248***	
	(0.195)	
TDummy	1.027***	2.889***
	(0.195)	(0.263)
periody_post × TDummy	2.889***	
	(0.276)	
Num.Obs.	800	400
R2	0.590	0.233
R2 Adj.	0.588	0.231
RMSE	1.95	2.62

By "hand", we obtained 2.8891185, the exact same coefficient as the regressions. Standard errors use the formula in equation (5) which is equivalent to those obtained with the second regression.

Estimation

Inference

- OLS estimates of equation (7) provide consistent estimates and asymptotically valid confidence intervals of the ATT when the parallel trend and no anticipation assumptions are combined with the assumption of independent sampling.
- The asymptotics is based on large population and fixed number of period. With a **balanced panel** or **repeated cross section** of $i.i.d.$ observations, The variance of the error is consistently estimable using standard clustering methods that allow for arbitrary serial correlation at the unit level (Liang and Zeger 1986; Bertrand, Duflo, and Mullainathan 2004).
- The same logic easily extends to cases where the observations are individual units who are members of independently-sampled clusters (e.g. states), and the standard errors are clustered at the appropriate level, provided that the number of treated and untreated clusters both grow large.

Conditional parallel trends

A slightly different identification strategy

- Sometimes, parallel trend is not a plausible assumption but it may be the case that conditional on some characteristics you would get parallel trend
- E.g.: your outcome of interest is wages, the treatment and control groups have different education levels, and the trends affecting the wages of high/low education workers differ. Then, the following assumption may be more plausible than the standard common trends assumption:
- Let $\boldsymbol{X_i}$ be a vector of time invariant covariates for unit i. Conditional parallel trends means assuming:

$$(\mathbb{E}[Y_k(0)|post(k), \boldsymbol{X_i}] - \mathbb{E}[Y_k(0)|pre(k), \boldsymbol{X_i}]) \tag{8}$$

$$= (\mathbb{E}[Y_U(0)|post(U), \boldsymbol{X_i}] - \mathbb{E}[Y_U(0)|pre(U), \boldsymbol{X_i}]) \tag{9}$$

- This assumption is **neither stronger nor weaker** than the unconditional parallel trend.
- Conditional parallel trend **does not imply** unconditional parallel trend and unconditional parallel trend **does not imply** conditional parallel trend ;

Conditional parallel trends

Estimating DiD under conditional parallel trend

- Applied researchers who want to account for covariates in their DID specification often just include covariates in their regression.
- Specifically, they estimate

$$Y_{it} = \beta_0 + \beta_1 1\{G_i = k\} + \beta_2 1\{T = t\} + \beta_3 1\{T = t\} 1\{G_i = k\} + X_i'\theta + u_{it} \tag{10}$$

- Now **this is a parametric assumption** ; we impose **structure**.
- That regression identifies a causal effect if it corresponds to the true model generating the potential outcomes, i.e. if

$$Y_{it}(0) = \beta_0 + \beta_1 1\{G_i = k\} + \beta_2 1\{T = t\} + X_i'\theta + u_{it}$$

- That means that the treatment effect is **constant & additive**.
- One obvious problematic restriction is that this model does not allow the effect of time on the outcome to depend on X_i, i.e. no different trajectories for different groups $\left(\mathbb{E}[Y_{ipost}(0) - Y_{ipre}(0) \mid G_i, X_i] = \beta_2\right)$, while this was the reason why we wanted to account for covariates in the first place.

Conditional parallel trends

Estimating DiD under conditional parallel trend

- An alternative approach is to allow for covariate-specific trends and treatment effect in DiD settings is the regression adjustment procedure.
- this would be similar to a modification of (10) that interacts $\mathbf{X_i}$ with both treatment group and time dummies.
- However, the parameters obtained from this regression is usually not the average treatment effect on the treated because of the treatment-variance weighting of the OLS.
- It works if both treated and control units have roughly the same covariate distribution (strong overlap) and treatment effect is homogeneous.
- Another way of estimating DID under conditional parallel trend is proposed by Heckman et al. (1998):
  1. Estimate the conditional expectation of the outcome **among untreated units**,
  2. and then average these "predictions" using the empirical distribution of $\mathbf{X_i}$ among treated units.
- We need not restrict ourselves to linear models for the CEF and can use more flexible semi-/non-parametric methods instead.

Conditional parallel trends

Estimating DiD under conditional parallel trend

- Abadie (2005) proposes a **propensity score estimator** that requires performing at most one non-parametric estimation. His estimator relies on the following result:

Theorem

If conditional parallel trend holds and if $0 < P(G_i = k \mid X_i) < 1$ *almost surely, then*

$$\mathbb{E}[Y_{i,post} - Y_{i,pre} \mid G_i = k] - \mathbb{E}\Big[(Y_{i,post} - Y_{i,pre}) \frac{\frac{P(G_i=k|X_i)}{P(G_i=k)}}{\frac{P(G_i=U|X_i)}{P(G_i=U)}} \mid G_i = U \Big]$$
$$= \underbrace{\mathbb{E}[Y_{i,post}(1) - Y_{i,post}(0) \mid G_i = k]}_{ATT} \tag{11}$$

- Where $P(G_i = k \mid X_i)$ is the probability of being treated conditional on covariate. This is called The propensity score and is usually estimated using a Logit/Probit regression.
- the conditional DID estimator of Abadie (2005) weight the control group units so that the distribution of covariates $\boldsymbol{X}$ is more balanced. In words, we give more weights to observations in the control groups that "looks" more like the treated units.

Treatment effect dynamics with 2 groups

One coefficient is not enough

- When we have access to repeated cross sections or panel data around the window of treatment, it's frustrating to only show one DiD coefficient.
- Treatment effect may evolve with time and we may want to see that.
- In situations like that, we have testable implications of our identification assumption. Before treatment, we shouldn't see any difference between those who will be treated and the untreated.
- To test common trends assumption, and to estimate dynamic treatment effects, it turns out one just needs to estimate the following regression (Wing, Simon, and Bello-Gomez 2018):

$$Y_i = \alpha + \sum_{t=1}^{\bar{t}} \beta_t \mathbb{1}(T_i = t) + \theta_1 \mathbb{1}(G_i = 1) + \sum_{t=1}^{\bar{t}} \gamma_{1t} \mathbb{1}(T_i = t) \mathbb{1}(G_i = 1) + u_i$$

- One can show that for any t,

$$\begin{aligned}\gamma_{1t} = & E\left(Y_i \mid G_i = 1, T_i = t\right) - E\left(Y_i \mid G_i = 1, T_i = 0\right) \\ & - \left(E\left(Y_i \mid G_i = 0, T_i = t\right) - E\left(Y_i \mid G_i = 0, T_i = 0\right)\right)\end{aligned}$$

- the DID comparing the evolution of the mean outcome from period 0 to t in groups 0 and 1 .

Outline

Case study: Minimum wage by Card and Krueger (AER 1994)

Context

- **Economic theory prediction:** In a competitive labor market, increases in the minimum wage would decrease the employment level of minimum wage workers
- David Card and Alan Krueger. 1994. "Minimum Wages and Employment: A Case Study of the Fast Food Industry in New Jersey and Pennsylvania." *American Economic Review* 84, no. 4 (September): 772–793
- Analyze the effect of a minimum wage increase in New Jersey using a differences in differences methodology
- In February 1992 NJ increased the state minimum wage from $4.25 to $5.05
- Pennsylvania's minimum wage stayed at $4.25
- They surveyed about 400 fast food stores both in NJ and in PA both before and after the minimum wage increase in NJ

Case study: Minimum wage by Card and Krueger (AER 1994)

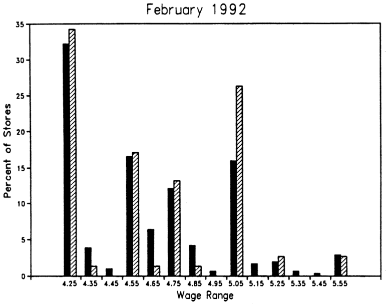
Context

Figure 3: Map of New-Jersey

Case study: Minimum wage by Card and Krueger (AER 1994)

Context

Figure 4: Distribution of hourly wage before the minimum wage increase in New-Jersey (Black)



Case study: Minimum wage by Card and Krueger (AER 1994)

Context

Figure 5: Distribution of hourly wage after the minimum wage increase in New-Jersey (Black)

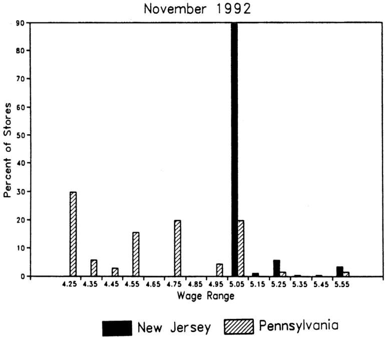


FIGURE 1. DISTRIBUTION OF STARTING WAGE RATES

Case study: Minimum wage by Card and Krueger (AER 1994)

Minimum wage: the great debate

Minimum Wage: A "Game of Thrones" but characters are economic theories

- In competitive labor markets at equilibrium, wage rate equals the marginal productivity so minimum wage should reduce total employment... **in a static equilibrium**

Case study: Minimum wage by Card and Krueger (AER 1994)

Minimum wage: the great debate

Minimum Wage: A "Game of Thrones" but characters are economic theories

- In competitive labor markets at equilibrium, wage rate equals the marginal productivity so minimum wage should reduce total employment... **in a static equilibrium**
- But workers' productivity **increase with experience** and if a higher wage **reduces turnover**, or if higher wage **incentivize workers** to be more productive (also maybe more job satisfaction), there may be some **human capital** counter-balancing effects

Case study: Minimum wage by Card and Krueger (AER 1994)

Minimum wage: the great debate

Minimum Wage: A "Game of Thrones" but characters are economic theories

- In competitive labor markets at equilibrium, wage rate equals the marginal productivity so minimum wage should reduce total employment... **in a static equilibrium**
- But workers' productivity **increase with experience** and if a higher wage **reduces turnover**, or if higher wage **incentivize workers** to be more productive (also maybe more job satisfaction), there may be some **human capital** counter-balancing effects
- **Macroeconomics:** Higher minimum wage may increase the **aggregated demand** which may lead to more job creations

Case study: Minimum wage by Card and Krueger (AER 1994)

Minimum wage: the great debate

Minimum Wage: A "Game of Thrones", but the characters are economic theories

- **Firm bargaining power**: In some cases, firms may have bargaining power that enables them to pass on some or all of the costs of a minimum wage increase to consumers in the form of higher prices. If consumers are willing to pay higher prices, firms may not need to reduce employment to maintain their profitability.

Case study: Minimum wage by Card and Krueger (AER 1994)

Minimum wage: the great debate

Minimum Wage: A "Game of Thrones", but the characters are economic theories

- **Firm bargaining power**: In some cases, firms may have bargaining power that enables them to pass on some or all of the costs of a minimum wage increase to consumers in the form of higher prices. If consumers are willing to pay higher prices, firms may not need to reduce employment to maintain their profitability.
- **Monopsony power**: In some labor markets, employers may have monopsony power, which means they are the only or dominant buyer of labor in the market. In such cases, a minimum wage increase may not lead to reduced employment, as employers may have been paying wages below the efficient wage level due to their market power.

Case study: Minimum wage by Card and Krueger (AER 1994)

Minimum wage: the great debate

Minimum Wage: A "Game of Thrones", but the characters are economic theories

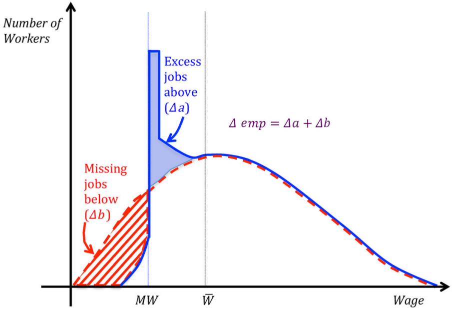
- **Firm bargaining power**: In some cases, firms may have bargaining power that enables them to pass on some or all of the costs of a minimum wage increase to consumers in the form of higher prices. If consumers are willing to pay higher prices, firms may not need to reduce employment to maintain their profitability.
- **Monopsony power**: In some labor markets, employers may have monopsony power, which means they are the only or dominant buyer of labor in the market. In such cases, a minimum wage increase may not lead to reduced employment, as employers may have been paying wages below the efficient wage level due to their market power.
- **Social norms:** Minimum wage increases may signal to workers that their labor is valued and respected, which can increase their motivation and commitment to their job. Additionally, minimum wage increases may signal to employers that they should value their workers and invest in their training and development, which can lead to higher productivity and increased employment.

Case study: Minimum wage by Card and Krueger (AER 1994)

Minimum wage: the great debate

Minimum wage: the great debate

Figure 6: Theoretical minimum wage effect on the wage distribution from (Cengiz et al. 2019)



Case study: Minimum wage by Card and Krueger (AER 1994)

Estimations

Figure 7: Estimation strategy in (Card and Krueger 1994)

(1a) $\Delta E_i = a + \mathbf{b}\mathbf{X}_i + c\mathrm{NJ}_i + \varepsilon_i$

or

(1b) $\Delta E_i = a' + \mathbf{b}'\mathbf{X}_i + c'\mathrm{GAP}_i + \varepsilon'_i$

where ΔE_i is the change in employment from wave 1 to wave 2 at store i, $\mathbf{X}_i$ is a set of characteristics of store i, and NJ_i is a dummy variable that equals 1 for stores in New Jersey. GAP_i is an alternative measure of the impact of the minimum wage at store i based on the initial wage at that store (W_{1i}):

$\mathrm{GAP}_i = 0$ for stores in Pennsylvania

$= 0$ for stores in New Jersey with $W_{1i} \geq \$5.05$

$= (5.05 - W_{1i}) / W_{1i}$ for other stores in New Jersey.

GAP_i is the proportional increase in wages at store i necessary to meet the new minimum rate. Variation in GAP_i reflects both

Case study: Minimum wage by Card and Krueger (AER 1994)

Estimations

Figure 8: Mean differences between New-Jersey and Pennsilvania

TABLE 3—AVERAGE EMPLOYMENT PER STORE BEFORE AND AFTER THE RISE IN NEW JERSEY MINIMUM WAGE

	Stores by state			Stores in New Jersey[a]			Differences within NJ[b]	
Variable	PA (i)	NJ (ii)	Difference, NJ − PA (iii)	Wage = \$4.25 (iv)	Wage = \$4.26–\$4.99 (v)	Wage ≥ \$5.00 (vi)	Low–high (vii)	Midrange–high (viii)
1. FTE employment before, all available observations	23.33 (1.35)	20.44 (0.51)	−2.89 (1.44)	19.56 (0.77)	20.08 (0.84)	22.25 (1.14)	−2.69 (1.37)	−2.17 (1.41)
2. FTE employment after, all available observations	21.17 (0.94)	21.03 (0.52)	−0.14 (1.07)	20.88 (1.01)	20.96 (0.76)	20.21 (1.03)	0.67 (1.44)	0.75 (1.27)
3. Change in mean FTE employment	−2.16 (1.25)	0.59 (0.54)	2.76 (1.36)	1.32 (0.95)	0.87 (0.84)	−2.04 (1.14)	3.36 (1.48)	2.91 (1.41)
4. Change in mean FTE employment, balanced sample of stores[c]	−2.28 (1.25)	0.47 (0.48)	2.75 (1.34)	1.21 (0.82)	0.71 (0.69)	−2.16 (1.01)	3.36 (1.30)	2.87 (1.22)
5. Change in mean FTE employment, setting FTE at temporarily closed stores to 0[d]	−2.28 (1.25)	0.23 (0.49)	2.51 (1.35)	0.90 (0.87)	0.49 (0.69)	−2.39 (1.02)	3.29 (1.34)	2.88 (1.23)

Notes: Standard errors are shown in parentheses. The sample consists of all stores with available data on employment. FTE (full-time-equivalent) employment counts each part-time worker as half a full-time worker. Employment at six closed stores is set to zero. Employment at four temporarily closed stores is treated as missing.

[a]Stores in New Jersey were classified by whether starting wage in wave 1 equals \$4.25 per hour ($N = 101$), is between \$4.26 and \$4.99 per hour ($N = 140$), or is \$5.00 per hour or higher ($N = 73$).

[b]Difference in employment between low-wage (\$4.25 per hour) and high-wage (≥ \$5.00 per hour) stores; and difference in employment between midrange (\$4.26–\$4.99 per hour) and high-wage stores.

[c]Subset of stores with available employment data in wave 1 and wave 2.

[d]In this row only, wave-2 employment at four temporarily closed stores is set to 0. Employment changes are based on the

Case study: Minimum wage by Card and Krueger (AER 1994)

Estimations

Figure 9: Mean differences between New-Jersey and Pennsilvania

TABLE 4—REDUCED-FORM MODELS FOR CHANGE IN EMPLOYMENT

	Model				
Independent variable	(i)	(ii)	(iii)	(iv)	(v)
1. New Jersey dummy	2.33 (1.19)	2.30 (1.20)	—	—	—
2. Initial wage gap[a]	—	—	15.65 (6.08)	14.92 (6.21)	11.91 (7.39)
3. Controls for chain and ownership[b]	no	yes	no	yes	yes
4. Controls for region[c]	no	no	no	no	yes
5. Standard error of regression	8.79	8.78	8.76	8.76	8.75
6. Probability value for controls[d]	—	0.34	—	0.44	0.40

Notes: Standard errors are given in parentheses. The sample consists of 357 stores with available data on employment and starting wages in waves 1 and 2. The dependent variable in all models is change in FTE employment. The mean and standard deviation of the dependent variable are -0.237 and 8.825, respectively. All models include an unrestricted constant (not reported).

[a] Proportional increase in starting wage necessary to raise starting wage to new minimum rate. For stores in Pennsylvania the wage gap is 0.

[b] Three dummy variables for chain type and whether or not the store is company-owned are included.

[c] Dummy variables for two regions of New Jersey and two regions of eastern Pennsylvania are included.

[d] Probability value of joint F test for exclusion of all control variables.

Case study: Minimum wage by Card and Krueger (AER 1994)

Estimations

Figure 10: Mean differences between New-Jersey and Pennsilvania

TABLE 5—SPECIFICATION TESTS OF REDUCED-FORM EMPLOYMENT MODELS

	Change in employment		Proportional change in employment	
Specification	NJ dummy (i)	Gap measure (ii)	NJ dummy (iii)	Gap measure (iv)
1. Base specification	2.30 (1.19)	14.92 (6.21)	0.05 (0.05)	0.34 (0.26)
2. Treat four temporarily closed stores as permanently closed[a]	2.20 (1.21)	14.42 (6.31)	0.04 (0.05)	0.34 (0.27)
3. Exclude managers in employment count[b]	2.34 (1.17)	14.69 (6.05)	0.05 (0.07)	0.28 (0.34)
4. Weight part-time as 0.4 × full-time[c]	2.34 (1.20)	15.23 (6.23)	0.06 (0.06)	0.30 (0.33)
5. Weight part-time as 0.6 × full-time[d]	2.27 (1.21)	14.60 (6.26)	0.04 (0.06)	0.17 (0.29)
6. Exclude stores in NJ shore area[e]	2.58 (1.19)	16.88 (6.36)	0.06 (0.05)	0.42 (0.27)
7. Add controls for wave-2 interview date[f]	2.27 (1.20)	15.79 (6.24)	0.05 (0.05)	0.40 (0.26)
8. Exclude stores called more than twice in wave 1[g]	2.41 (1.28)	14.08 (7.11)	0.05 (0.05)	0.31 (0.29)
9. Weight by initial employment[h]	—	—	0.13 (0.05)	0.81 (0.26)
10. Stores in towns around Newark[i]	—	33.75 (16.75)	—	0.90 (0.74)
11. Stores in towns around Camden[j]	—	10.91 (14.09)	—	0.21 (0.70)
12. Pennsylvania stores only[k]	—	−0.30 (22.00)	—	−0.33 (0.74)

Notes: Standard errors are given in parentheses. Entries represent estimated coefficient of New Jersey dummy [columns (i) and (iii)] or initial wage gap [columns (ii) and (iv)] in regression models for the change in employment or the percentage change in employment. All models also include chain dummies and an indicator for company-owned stores.

Case study: Minimum wage by Card and Krueger (AER 1994)

Discussion

Main Results of Card and Krueger (1994)

- Minimum wage increase **did not lead to job losses** in fast-food restaurants in New Jersey compared to those in eastern Pennsylvania.
- Employment actually increased in New Jersey relative to Pennsylvania after the minimum wage increase.
- Wages increased for low-wage workers in New Jersey relative to Pennsylvania.
- There was no evidence of significant price increases at fast-food restaurants in New Jersey.
- The study triggered a long academic debate that's still very active because it challenged the conventional wisdom that raising the minimum wage leads to job losses.

Case study: Minimum wage by Card and Krueger (AER 1994)

Discussion

Critiques of Card and Krueger (1994)

- Small sample size: Only 410 fast-food restaurants in New Jersey and eastern Pennsylvania.
- Unreliable data: Payroll records not designed for research purposes, may contain errors.
- Lack of statistical significance: Differences in employment levels between states not statistically significant.
- Unrepresentative comparison group: Comparison group may not be a good representation of control group.
- Theoretical limitations: Study did not account for potential long-term effects of minimum wage increases.

Answer to the critics: Card and Krueger (2000)

More data

Figure 11: Adding more counties and observations to the initial sample with additional data sources

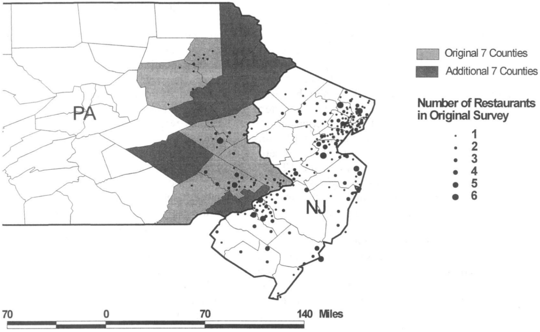


FIGURE 1. AREAS OF NEW JERSEY AND PENNSYLVANIA COVERED BY ORIGINAL SURVEY AND BLS DATA

Answer to the critics: Card and Krueger (2000)

More data

Figure 12: New estimate of the average effect on employment are close to 0 and insignificant

TABLE 2—BASIC REGRESSION RESULTS; BLS ES-202 FAST-FOOD DATA AND CARD-KRUEGER SURVEY DATA

	Dependent variable:			
	Change in levels		Proportionate change	
Explanatory variables	(1)	(2)	(3)	(4)
A. *All of New Jersey and 7 Pennsylvania Counties, BLS Data*				
New Jersey indicator	0.536	0.225	0.007	0.009
	(1.017)	(1.029)	(0.029)	(0.029)
Chain dummies and subunit dummy variable	No	Yes	No	Yes
Standard error of regression	10.09	9.99	0.286	0.281
R^2	0.001	0.029	0.000	0.046
B. *All of New Jersey and 14 Pennsylvania Counties, BLS Data*				
New Jersey indicator	0.946	0.272	0.045	0.032
	(0.856)	(0.859)	(0.024)	(0.024)
Chain dummies and subunit dummy variable	No	Yes	No	Yes
Standard error of regression	10.80	10.63	0.303	0.294
R^2	0.002	0.042	0.005	0.071
C. *Original Card-Krueger Survey Data*				
New Jersey indicator	2.411	2.488	0.029	0.030
	(1.323)	(1.323)	(0.050)	(0.049)
Chain and company-ownership dummies	No	Yes	No	Yes
Standard error of regression	10.28	10.25	0.385	0.382
R^2	0.009	0.025	0.001	0.024

Notes: Each regression also includes a constant. Sample size is 564 for panel A, 687 for panel B, and 384 for panel C. Subunit dummy variable equals one if the reporting unit is a subunit of a multiunit employer. For comparability with the BLS data, employment in the CK sample is measured by the total number of full- and part-time employees. Standard errors are in

Answer to the critics: Card and Krueger (2000)

Long term effects

Figure 13: Comparing the effect with more period before and after with repeated cross sections

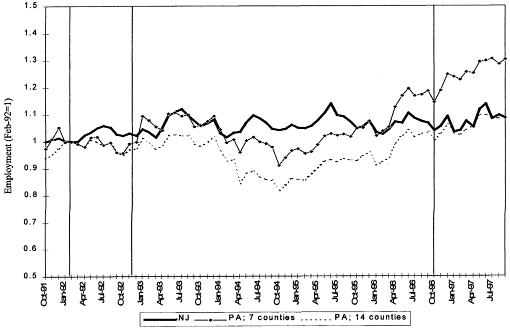


FIGURE 2. EMPLOYMENT IN NEW JERSEY AND PENNSYLVANIA FAST-FOOD RESTAURANTS, OCTOBER 1991 TO SEPTEMBER 1997

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.

Answer to the critics: Card and Krueger (2000)

Long term effects

Figure 14: Adding controls in the regression remove the difference in employment

TABLE 5—ESTIMATED REGRESSION MODELS FOR CHANGE IN AVERAGE PAYROLL HOURS/35, BNW DATA

	Specification: (1)	(2)	(3)	(4)	(5)	(6)	(7)
New Jersey	−0.85 (0.49)	—	—	—	−0.36 (0.44)	−0.66 (0.41)	−0.09 (0.42)
NW subsample (1 = yes)	—	−3.49 (0.42)	—	—	−3.44 (0.43)	—	—
Chain dummies:							
Roy Rogers	—	—	−3.56 (0.81)	—	—	−3.14 (0.85)	−1.98 (0.89)
Wendy's	—	—	−0.85 (0.67)	—	—	−0.71 (0.67)	−1.35 (0.70)
KFC	—	—	−6.51 (0.90)	—	—	−6.30 (0.90)	−6.56 (0.89)
Company-owned	—	—	−0.89 (0.76)	—	—	−1.31 (0.81)	−0.72 (0.95)
Payroll data type:							
Biweekly	—	—	—	1.73 (0.52)	—	—	1.65 (0.52)
Monthly	—	—	—	−2.60 (0.48)	—	—	−1.06 (0.89)
R^2	0.01	0.23	0.41	0.30	0.23	0.10	0.45
Standard error of regression	3.47	3.07	2.70	2.95	3.08	3.32	2.62

Notes: Standard errors are in parentheses. Sample consists of 235 stores. Dependent variable in all models is the change in average weekly payroll hours divided by 35 between wave 1 and wave 2.

Answer to the critics: Card and Krueger (2000)

Main Results

- Re-analysis of original 1994 study with updated data and improved methodology.
- Confirmed previous findings that minimum wage increases do not lead to job losses in the fast-food industry.
- Expanded analysis to include more states and industries, finding no evidence of job losses due to minimum wage increases.
- The study challenged the view that minimum wage increases lead to job losses across all industries and regions.

Outline

1. Introduction
2. DID: the 2x2 case
3. Case study: Minimum wage by Card and Krueger (AER 1994)
4. Multiple groups, multiple periods
   - Mostly harmless, really ?
   - What's in the 2WFE:(Goodman-Bacon 2021) intuition and results
   - It get worse with heterogenous treatment effects
   - Let's simulate data
   - Callaway and Sant'Anna (2020) solve these issues
   - Using Callaway and Sant'Anna (2020) on the previous data

Multiple groups, multiple periods

Mostly harmless, really ?

- In many settings, individuals do not receive the treatment at the same "calendar" time but we are interested in using this differential timing as a source of comparison.
- If you follow Angrist and Pischke (2008), a seemingly *mostly harmless* nat-
ural extension to the Dif-in-Dif model is the two-way fixed effect regression:

  > *"It's also easy to add additional (units) or periods to the regression setup... [and] it's easy to add additional covariates."*

$$Y_{it} = \gamma_i + \lambda_t + \delta^{DD} D_{it} + \varepsilon_{it} \tag{12}$$

- where γ_i and λ_i are individual and time fixed effects and D_{it} the indicator for treatment that indicate when people get treated.
- It's easy to modify this regression equation to add controls, specific trends etc.

Mostly harmless, really ?

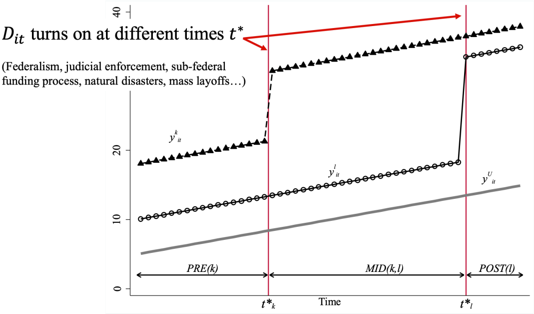
It turns out it wasn't mostly harmless econometrics

- Developed literature now on the issues with TWFE DiD with "staggered treatment timing" (Sun and Abraham 2020; Borusyak, Jaravel, and Spiess 2021; Callaway and Sant'Anna 2020; Goodman-Bacon 2021; de Chaisemartin and D'Haultfoeuille 2021; Chaisemartin and D'Haultfœ uille, Aout 2020; Athey and Imbens 2018a), probably more that I don't know about.
- Two recent survey if you want to go deeper:
  1. Clément de Chaisemartin and Xavier D'Haultfœ uille. 2021. *Two-Way Fixed Effects and Differences-in-Differences with Heterogeneous Treatment Effects: A Survey.* SSRN Scholarly Paper ID 3980758. Rochester, NY: Social Science Research Network, December 8, 2021
  2. Jonathan Roth et al. 2021. "What's Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature," 54
- Two sources of problem: heterogeneous treatment effect one one side, "hiden" weightings and wrong comparison in the regression on the other.
- I provide intuition and some ways to solve the problem but know there is more to this.

What's in the 2WFE:(Goodman-Bacon 2021) intuition and results

2 groups are treated at different dates, one group is never treated

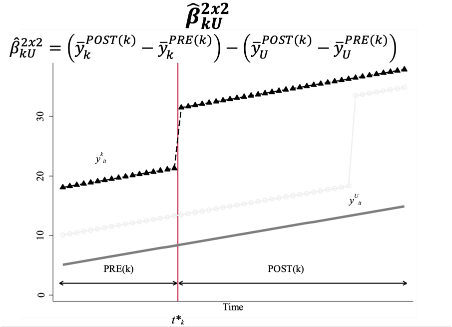
Figure 15: Comparison of outcomes over time



What's in the 2WFE:(Goodman-Bacon 2021) intuition and results

2 groups are treated at different dates, one group is never treated

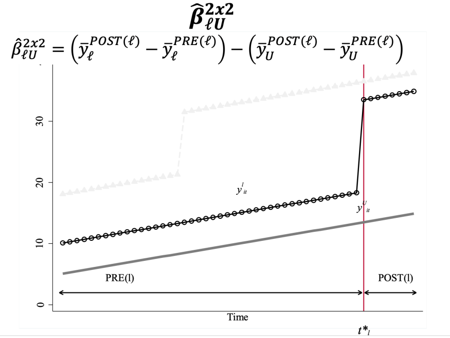
Figure 16: First Difference in differences: Early-treated with never treated



What's in the 2WFE:(Goodman-Bacon 2021) intuition and results

2 groups are treated at different dates, one group is never treated

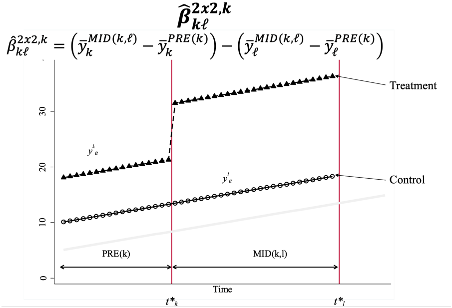
Figure 17: Second Difference in differences: Late-treated with never treated



What's in the 2WFE:(Goodman-Bacon 2021) intuition and results

2 groups are treated at different dates, one group is never treated

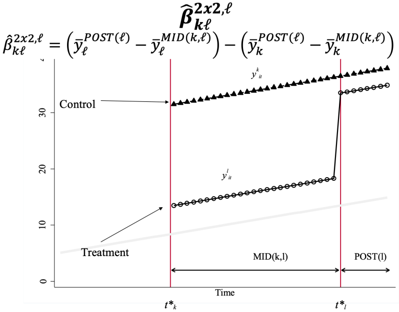
Figure 18: Third Difference in differences: Early-treated with Late-treated



What's in the 2WFE:(Goodman-Bacon 2021) intuition and results

2 groups are treated at different dates, one group is never treated

Figure 19: Fourth Difference in differences: Early-treated with Late-treated



What's in the 2WFE:(Goodman-Bacon 2021) intuition and results

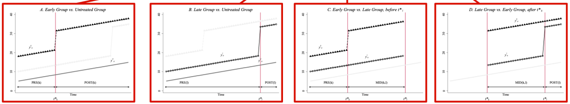
2 groups are treated at different dates, one group is never treated

Figure 20: The DID coefficient is a weighted average of all DIDs, some we don't want

$$y_{it} = \alpha_i + \alpha_t + \hat{\beta}^{DD} D_{it} + u_{it}$$

For three groups:

$$\hat{\beta}^{DD} = s_{kU}\hat{\beta}^{2x2}_{kU} + s_{\ell U}\hat{\beta}^{2x2}_{\ell U} + s^{k}_{k\ell}\hat{\beta}^{2x2,k}_{k\ell} + s^{k}_{k\ell}\hat{\beta}^{2x2,\ell}_{k\ell}$$



2x2 DDs: subsamples with two groups (treat/control) and two periods (pre/post)

Multiple groups, multiple periods

What's in the 2WFE:(Goodman-Bacon 2021) intuition and results

- δ^{DD} is just the weighted average of the four 2x2 treatment effects. The weights are a function of the size of the subsample, relative size of treatment and control units, and the timing of treatment in the sub sample.
- Already-treated units act as controls even though they are treated.
- Given the weighting function, panel length alone can change the DiD estimates substantially, even when each δ^{DD} does not change.
- Groups treated closer to middle of panel receive higher weights than those treated earlier or later.

Overall TWFE don't do what people thought they did.

Multiple groups, multiple periods

It get worse with heterogenous treatment effects

- There are two types of heterogeneous treatment effects:
  1. **Heterogeneous effects across groups**
     - The difference in potential outcomes differs across groups
     - In other words, the same treatment would lead to different responses in different groups/units
  2. **Heterogeneous effects within groups over time**
     - Need to see this relative to a counterfactual time path
     - The difference between the actual path and the counterfactual changes over time
     - Example: treatment pushes units onto a different time trend

It get worse with heterogenous treatment effects

Figure 21: Meme to wrap it up

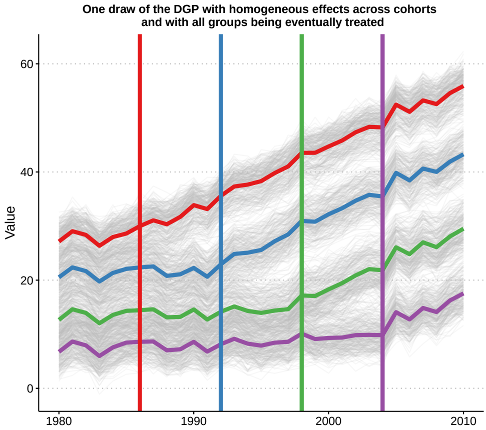
Let's simulate data

Scenario for a DGP

- Imagine e.g. we want to estimate the impact of new metro stations in the neighbourhood on rent prices.
- We consider 3 new metro stations opening at different time and 40 neighbourhoods where we randomly sample 250 rents by square meters each year for instance (repeated cross section)
- We generate a model with homogeneous treatment effect, but dynamic, then a second model where late adopters have small treatment effects.
- We estimate the model using TWFE and compare with the true effect.

Let's simulate data

DGP from the DID package documentation by Callaway and Sant'Anna (2020)



Let's simulate data

Estimating dynamic treatment effects via TWFE event-study regressions

- Given that we are interested in treatment effect dynamics, we then proceed to consider a classical two-way fixed-effects (TWFE) event study specification

$$Y_{i,t} = \alpha_i + \alpha_t + \gamma_k^{-K} D_{i,t}^{<-K} + \sum_{k=-K}^{-2} \gamma_k^{\text{lead}} D_{i,t}^{k} + \sum_{k=0}^{L} \gamma_k^{\text{lag}} D_{i,t}^{k} + \gamma_k^{L+} D_{i,t}^{>L} + \varepsilon_{i,t}$$

- where $D_{i,t}^{k} = 1\{t - G_i = k\}$ is an "event-study" dummy variable that takes value one if a unit i is k periods away from initial treatment at time t and zero otherwise, $D_{i,t}^{<-K} = 1\{t - G_i < -K\}$ and $D_{i,t}^{>L} = 1\{t - G_i > L\}$ are defined analogously. For instance, $D_{i,t}^{0}$ is equal to one if the unit i is first treated at time $t, D_{i,t}^{1}$ is equal to one if a one period has passed since treatment started (treatment lags), etc. Alternatively we have that $D_{i,t}^{-2}$ is equal to one if a unit i will be treated in two periods from t (treatment leads). In this exercise we set K and L to be equal to 5 .
- Up to today, it is customary to interpret estimates of γ_k^{lags} as "good" measures of the average treatment effect for being exposed to treatment for k periods, and estimates of γ_k^{leads} as measures of pre-trends. Our first exercise here is to assess if this is OK-ish.

Let's simulate data

TWFE regression with leads and lags using R

- So far we used $lm\_robust(\cdot)$ from estimater but it is not optimal for panel data and fixed effect regressions. Thus we use $lfe :: felm(\cdot)$ to estimate the model.
- make dummy columns and generate pre-post dummies

```
data <- data %>%
  mutate(rel_year = year - cohort_year) %>%
  dummy_cols(select_columns = "rel_year") %>%
  \# generate pre and post dummies
  mutate(Pre = ifelse(rel_year < -5, 1, 0),
         Post = ifelse(rel_year > 5, 1, 0))
```

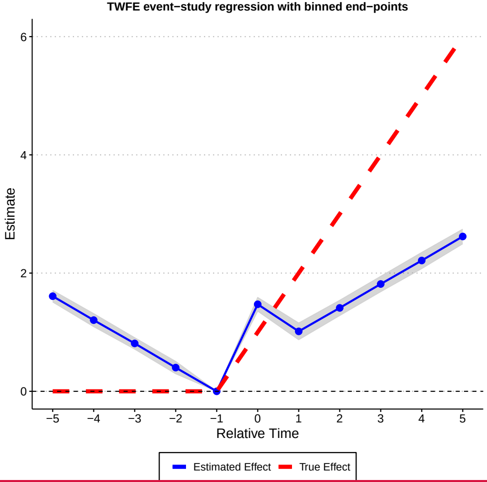
- Then we estimate the model:

```
mod <- lfe::felm(dep_var ~ Pre + `rel_year_-5` + `rel_year_-4` + `rel_year_-3` +
    `rel_year_-2` + rel_year_0 + rel_year_1 + rel_year_2 + rel_year_3 + rel_year_4 +
    rel_year_5 + Post | unit + year | 0 | state, data = data, exactDOF = TRUE)
```

- We then compare with the true effect we generated which is an increase of 1 unit each for the treated.

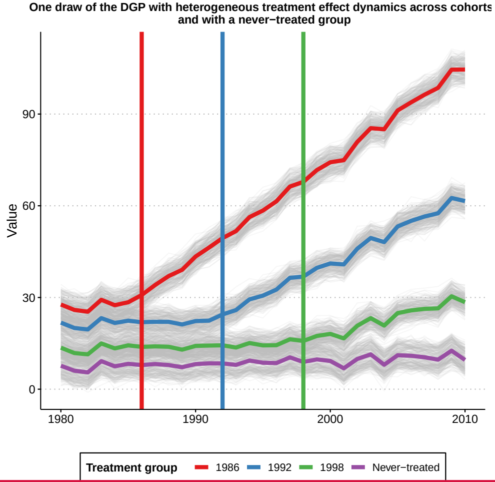
Let's simulate data

Estimation of the event-study using TWFE



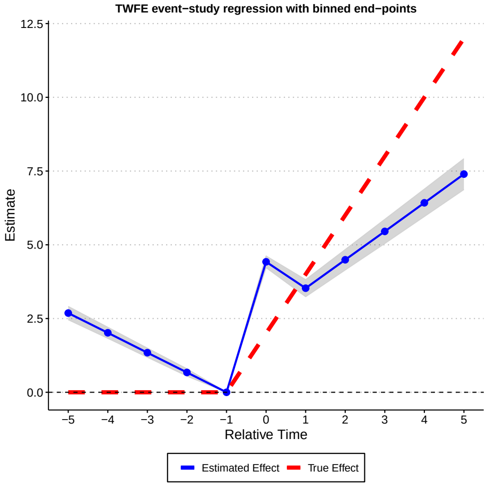
Let's simulate data

Setting with heterogenous treatment effect



Let's simulate data

Estimation with the same model as before



Let's simulate data

What do we make of that ?

- The results above show that these TWFE event-study type estimates are severely biased for the true treatment effects.
- Furthermore, using the estimates of coefficient of treatment leads as a way to find evidence of "pre-trends" is very problematic, as illustrated above.
- The reason for that is well described by Sun and Abraham (2020)
- Now, putting it simply, the results above highlight that such TWFE linear regression **should not be used to highlight treatment effect dynamics**!

Let's simulate data

What do we make of that ?

- The results above show that these TWFE event-study type estimates are severely biased for the true treatment effects.
- Furthermore, using the estimates of coefficient of treatment leads as a way to find evidence of "pre-trends" is very problematic, as illustrated above.
- The reason for that is well described by Sun and Abraham (2020)
- Now, putting it simply, the results above highlight that such TWFE linear regression **should not be used to highlight treatment effect dynamics**!

Then, what do we do ?

Callaway and Sant'Anna (2020) solve these issues

Two papers and a solution for *almost* every problems

- In a first paper Sant'Anna and Zhao (2020) clarify hypotheses to estimate ATT using DID for any time difference, and new estimator close to Abadie (2005) that is non-parametric and "doubly robust".
- In a second paper Callaway and Sant'Anna (2020) propose to estimate each group-time average treatment effect using the doubly-robust estimator of their companion paper and come-up with ways to aggregate relevent treatment effects with appropriate weights to obtain meaningfull parameters.
- In words, they compute every ATTs for each group at each date and turn them into a weighted ATT.
- They discuss different situations whether there is never-treated groups or only not-yet-treated groups.

Callaway and Sant'Anna (2020) solve these issues

The $ATT(g,t)$ parameter of Sant'Anna and Zhao (2020)

$$ATT(g,t) = E\left[Y_t^1 - Y_t^0 \mid G_g = 1\right]$$

$$\mathrm{ATT}(g,t) = E\left[\left(\frac{G_g}{E\left[G_g\right]} - \frac{\frac{p_g(X)C}{1-p_g(X)}}{E\left[\frac{p_g(X)C}{1-p_g(X)}\right]}\right)\left(Y_t - Y_{g-1}\right)\right]$$

C Indicator for never-treated group G_g Indicators for groups treated at different times Propensity score $p_g(X) = P\left(G_g = 1 \mid X, G_g + C = 1\right)$

Callaway and Sant'Anna (2020) solve these issues

The $ATT(g,t)$ parameter of Sant'Anna and Zhao (2020)

$(Y_t - Y_{g-1})$: Long differences between outcomes in period t and the period before group g was treated

$$\left(\frac{G_g}{E\left[G_g\right]} - \frac{\frac{p_g(X)C}{1-p_g(X)}}{E\left[\frac{p_g(X)C}{1-p_g(X)}\right]} \right)$$

- The expression in parentheses is a weighting function to balance the treated and control group on covariates
- Control units with similar characteristics to the treated groups are getting more weight

Callaway and Sant'Anna (2020) solve these issues

What to do with these ATT(g,t)

- Can aggregate the ATT(g; t) across time and groups
- This will allow for the estimation of more interesting parameters
- One can also use this estimator to look at pre-trends
- Inference is done through bootstrapping
- Read carefully the paper and documentation to make sure your setting fit their hypotheses ; it's very clear in the paper.

Using Callaway and Sant'Anna (2020) on the previous data

The *did* Package

- Along with the paper, the authors provide a very well documented package called did
- It allows to estimate did models for 2xt periods or any ATT(g,t) when you have many groups many period
- Then you can agreggate these ATT(g,t) to get a weighted average of the ATT for the event study.
- To estimate ATT(g,t), we run

```
mod <- did::att_gt(yname = "dep_var", tname = "year", idname = "unit", gname = "cohort_year",
    control_group = "notyettreated", bstrap = FALSE, data = data, print_details = FALSE)
```

- To aggregate to an event study we run

```
event_std <- did::aggte(mod, type = "dynamic")
# get the basic SE
att.egt <- event_std$att.egt
```

- We then compare with the true effect we generated which is an increase of 1 unit each for the treated.

Using Callaway and Sant'Anna (2020) on the previous data

Estimations on the homogenous model

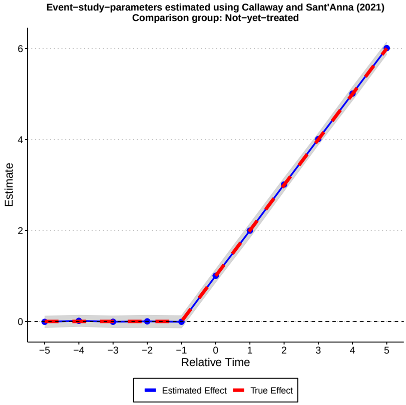


Figure 26: Estimation on the homogenous model using Callaway Sant'Anna (2020)

Multiple groups, multiple periods

Other estimators

- Callaway and Sant'Anna (2020) is a very effective estimator but other scholars proposed alternative estimators that are more or less fitted to different situations.
- For instance, Borusyak, Jaravel, and Spiess (2022) define the "imputed estimator" for event studies, Sun and Abraham (2020) propose another using a saturated specification in 2WFE ; Gardner (2022) cleverly use the Frisch-Waugh-Lovel theorem and use 2 partial regressions to correct bias,...
- Next session (with Denis Fougère), you will see two more related estimators:
  1. Triple differences
  2. Synthetic controls

Multiple groups, multiple periods

Some recent papers with methodological advances

- Testing for parallel pre-trends (Freyaldenhoven, Hansen, and Shapiro 2019; Rambachan and Roth 2020)
- Estimating dynamic treatment effects (Borusyak, Jaravel, and Spiess 2021; Sun and Abraham 2020)
- Re-weighting to recover relevant parameters (Callaway and Sant'Anna 2021; de Chaisemartin and D'Haultfoeuille 2021)
- Adjusting inference for (failed) pre-trend tests (Roth 2022)
- Machine learning meets DiD (Athey and Imbens 2018b)
- Fuzzy designs: instrumental dif-in-dif (Chaisemartin and D'Haultfoeuille 2017)

Outline

Modern DiD: Application to the minimum wage debate

Setting

- Special report prepared for Her Majesty's treasury in 2019 by Arindrajit Dube. 2019b. *Impacts of Minimum Wages: Review of the International Evidence.* Independent report. London, UK: HM Treasury and Department for Business, Energy & Industrial Strategy
- **Main goal** Comparing staggered adoption of minimum wage increases across the US states and estimate the effect on wage and employment for those likely affected
- Use stacked regression (yet another DiD estimator) to estimate the event-study effect of minimum wage increases.
- Derived from another work published in the Quarterly Journal of Economics (Cengiz et al. 2019)

Modern DiD: Application to the minimum wage debate

Setting

Figure 27: Data used (from Dube (2019a))

The analysis in this report closely follows the methodology used by Cengiz et al. Our primary source of data is the Consumer Population Survey Outgoing Rotation Group (CPS-ORG) for 2011-2018. The CPS-ORG provides individual level data that is used to estimate the quarterly distribution of hourly wages and employment for each state. We estimate this distribution using only observations with non-imputed earnings. The CPS-ORG provides a direct measurement of the hourly wage for hourly workers. For non-hourly workers, we estimate the hourly earnings as the respondent's usual weekly earnings divided by their usual hours worked per week. We next deflate the hourly wages to 2018 dollars using the monthly

Modern DiD: Application to the minimum wage debate

Setting

Figure 28: Emirical strategy (from Dube (2019a))

The regression specification used here is a stacked difference-in-difference as follows:

$$\frac{E_{hsjt}}{N_{hst}} = \sum_{\tau=-3}^{2} \sum_{k=-4}^{17} \alpha_{\tau k} I_{hsjt}^{\tau k} + \mu_{hsj} + \Omega_{hsjt} + u_{hsjt} \quad (1)$$

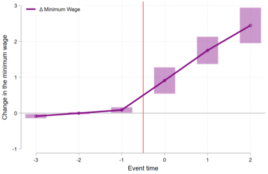
where E_{hsjt} is the employment count in \$1 wage bin j in state s for event h and during quarter t and N_{hst} is the population of state s during quarter t for event h. The wage bins are constructed relative to the new minimum wage for each event. $I_{hsjt}^{\tau k}$ is a treatment dummy variable taking value one if the minimum wage was raised in state s, τ time periods from date t for each dollar group k. Here τ represents event time in years relative to the minimum wage change for $\tau < 1$. For example, $\tau = 0$ represents the first full year following the first minimum wage increase. The $\tau = 1$ category includes all intermediate periods between the first and the penultimate year of the post-treatment period, while $\tau = 2$ represents the last full year of the post treatment period (i.e., 2018). This slightly non-standard way of delineating event time allows us to look at the effect in the most recent period in calendar time (2018), which is of particular interest given the phased-in nature of the minimum wage increases we are studying (more on this below). Indeed, the key estimate of interest is the most recent period effect, where the minimum wage is the highest.

Modern DiD: Application to the minimum wage debate

Setting

Figure 29: Effect of minimum wage increase on New job and lost jobs (from Dube (2019a))

Figure A1— Evolution of the Minimum Wage in Treated Versus Control States

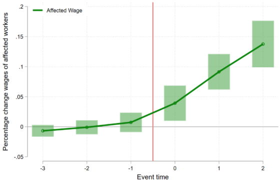


Notes: This figure shows the evolution of the minimum wage in treated versus control states. The estimates are found from a stacked difference-in-difference model that is similar to Equation (1) but doesn't estimate the change across the wage distribution. Specifically, we regress the quarterly, state minimum wage on treatment indicators I^{τ}_{hst} that value one if the minimum wage was raised in state s τ time periods from date t of event h. Here τ represents event-time in years relative to the minimum wage change for $\tau < 1$. The event-time $\tau = 1$ includes all time periods after one year of the minimum wage change but before the last year of the minimum wage change, while $\tau = 2$ represents the last full year of post treatment period (i.e., 2018). The purple line depicts the average change in the minimum wage in the treated group relative to the control group. The shaded area is the 95% confidence interval based on standard errors that are clustered by state.

Modern DiD: Application to the minimum wage debate

Setting

Figure 30: Effect of minimum wage increase on wages of affected workers (from Dube (2019a))



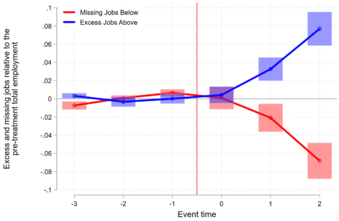
(b) Weighted by population size

Notes: These figures show the effect of high minimum increases on affected wages with and without population weights. The green line depicts the average percentage change in wages for affected workers relative to the average wage of affected workers in states that increased their minimum wage above \$10.50, over the three years prior to the initial raise. Workers are "affected" if their wage is less than five dollars above the new minimum wage. The shaded area is the 95% confidence interval based on standard errors that are clustered by the state, calculated using the delta method. Here τ represents event-time in years relative to the minimum wage change for $\tau < 1$. The event-time $\tau = 1$ includes all time periods after one year of the minimum wage change but before the last year of the minimum wage change, while $\tau = 2$ represents the last full year of post treatment period (i.e., 2018).

Modern DiD: Application to the minimum wage debate

Setting

Figure 31: Effect of minimum wage increase on New job and lost jobs (from Dube (2019a))



(b) Weighted by population

Modern DiD: Application to the minimum wage debate

Setting

Figure 32: Notes on the previous graph (from Dube (2019a))

Notes: These figures show the effect of high minimum increases on missing and excess jobs with and without population weights. The regression specification is given in equation (1) and includes state-event-dollar group fixed effects and event-time-dollar group fixed effects. The specification estimates the employment effect for every dollar group.

The red line depicts the average change in missing jobs. Jobs are "missing" if their wage is less than the new minimum wage. The missing jobs below the new minimum wage are estimated as the averaged effects for the dollar-groups below the new minimum wage. The red shaded area is the 95% confidence interval for the missing jobs below the new minimum wage based on standard errors that are clustered by state.

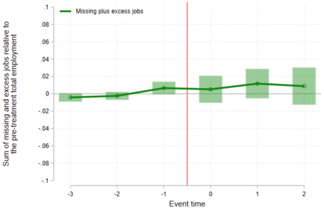
The blue line depicts the average change in excess jobs. Jobs are "excess" if their wage is at least the new minimum wage but less than five dollars above the new minimum wage. The excess jobs above the new minimum wage are estimated as the averaged effects for the dollar-groups between the new minimum wage and five dollars above the new minimum wage. The blue shaded area is the 95% confidence interval for the excess jobs above the new minimum wage based on standard errors that are clustered by state. Here τ represents event-time in years relative to the minimum wage change for $\tau < 1$. The event-time $\tau = 1$ includes all time periods after one year of the minimum wage change but before the last year of the minimum wage change, while $\tau = 2$ represents the last full year of post treatment period (i.e., 2018).

Modern DiD: Application to the minimum wage debate

Setting

Figure 33: Net Effect of minimum wage on employment (from Dube (2019a))

(a) Unweighted



(b) Weighted by Population

Notes: These figures show the effect of high minimum wage increases on affected employment with and without population weights. The red line depicts the average percentage change in wages for affected workers relative to the average employment of affected workers in states that increased their minimum wage above $10.50 over the three years prior to the initial raise. Workers are "affected" if their wage is less than five dollars above the new minimum wage. The shaded area is the 95% confidence interval based on standard errors that are clustered by the state. The regression specification is given in equation (1) and includes state-event-dollar group fixed effects and event-time-dollar group fixed effects. The specification estimates the

Outline

Wrap-up

A rather extensive introduction to difference in differences

- In the 2x2 case, under (conditional) parallel trend, no anticipation and random sampling, the DiD estimator correspond to the Average treatment effect on the treated
- With 2 groups or two (or more period), we can estimate the event-study or aggregated DID with regressions or related estimations methods.
- Discussion on a famous application on minimum wage
- Multiple groups and multiple periods create problems for estimation (not identification)
- Very active literature

Next sessions: Triple differences and synthetic controls

- Enjoy your winter break, no mandatory readings.

Bibliography I

- Abadie, Alberto. 2005. "Semiparametric Difference-in-Differences Estimators." *The Review of Economic Studies* 72, no. 1 (January): 1–19.
- Angrist, Joshua D., and Jörn-Steffen Pischke. 2008. *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton University Press.
- Athey, Susan, and Guido Imbens. 2018a. "Design-Based Analysis in Difference-In-Differences Settings with Staggered Adoption." September 1, 2018.
- ———. 2018b. "Design-Based Analysis in Difference-In-Differences Settings with Staggered Adoption." September 1, 2018.
- Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. 2004. "How Much Should We Trust Differences-in-Differences Estimates?" *The Quarterly Journal of Economics* 119 (1): 249–275.
- Borusyak, Kirill, Xavier Jaravel, and Jann Spiess. 2021. "Revisiting Event Study Designs: Robust and Efficient Estimation." August 27, 2021.
- ———. 2022. *Revisiting Event Study Designs: Robust and Efficient Estimation,* arXiv:2108.12419, April 19, 2022.
- Callaway, Brantly, and Pedro H. C. Sant'Anna. 2021. "Difference-in-Differences with Multiple Time Periods." *Journal of Econometrics,* Themed Issue: Treatment Effect 1, 225, no. 2 (December 1, 2021): 200–230.
- Callaway, Brantly, and Pedro H.C. Sant'Anna. 2020. "Difference-in-Differences with Multiple Time Periods." *Journal of Econometrics* (December): S0304407620303948.

Bibliography II

- Card, David, and Alan Krueger. 1994. "Minimum Wages and Employment: A Case Study of the Fast Food Industry in New Jersey and Pennsylvania." *American Economic Review* 84, no. 4 (September): 772–793.
- Card, David, and Alan B Krueger. 2000. "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania: Reply." *American Economic Review* 90, no. 5 (December 1, 2000): 1397–1420.
- Cengiz, Doruk, Arindrajit Dube, Attila Lindner, and Ben Zipperer. 2019. "The Effect of Minimum Wages on Low-Wage Jobs*." *The Quarterly Journal of Economics* 134, no. 3 (August 1, 2019): 1405–1454.
- Chaisemartin, C De, and Xavier D'Haultfoeuille. 2017. "Fuzzy Differences-in-Differences." *Review of Economic Studies,* no. 1, 1–30.
- Chaisemartin, Clément De, and Xavier D'Haultfœ uille. Aout 2020. *Differences-in-Differences Estimators of Intertemporal Treatment Effects.* Working Paper.
- Cunningham, Scott. 2018. *Causal Inference: The Mixtape.*
- De Chaisemartin, Clément, and Xavier D'Haultfœ uille. 2021. *Two-Way Fixed Effects and Differences-in-Differences with Heterogeneous Treatment Effects: A Survey.* SSRN Scholarly Paper ID 3980758. Rochester, NY: Social Science Research Network, December 8, 2021.
- De Chaisemartin, Clément, and Xavier D'Haultfoeuille. 2021. "Difference-in-Differences Estimators of Intertemporal Treatment Effects." May 29, 2021.
- Dias, Mateus, Rudi Rocha, and Rodrigo R Soares. 2023. "Down the River: Glyphosate Use in Agriculture and Birth Outcomes of Surrounding Populations." *The Review of Economic Studies* (February 6, 2023): rdad011.

Bibliography III

- Dube, Arindrajit. 2019a. *Evidence from Seven US States on the Impact of High Minimum Wages.* Technical Annex A. London: Her Majesty Treasury.
- ———. 2019b. *Impacts of Minimum Wages: Review of the International Evidence.* Independent report. London, UK: HM Treasury and Department for Business, Energy & Industrial Strategy.
- Freyaldenhoven, Simon, Christian Hansen, and Jesse M. Shapiro. 2019. "Pre-Event Trends in the Panel Event-Study Design." *American Economic Review* 109, no. 9 (September 1, 2019): 3307–3338.
- Gardner, John. 2022. *Two-Stage Differences in Differences,* arXiv:2207.05943, July 12, 2022.
- Goodman-Bacon, Andrew. 2021. "Difference-in-Differences with Variation in Treatment Timing." *Journal of Econometrics* (June): S0304407621001445.
- Heckman, James, Hidehiko Ichimura, Jeffrey Smith, and Petra Todd. 1998. "Characterizing Selection Bias Using Experimental Data." *Econometrica* 66 (5): 1017–1098.
- Heckman, James J., and Jeffrey A. Smith. 1999. "The Pre-Programme Earnings Dip and the Determinants of Participation in a Social Programme. Implications for Simple Programme Evaluation Strategies." *The Economic Journal* 109 (457): 313–348.
- Liang, Kung-Yee, and Scott L. Zeger. 1986. "Longitudinal Data Analysis Using Generalized Linear Models." *Biometrika* 73 (1): 13–22.
- Rambachan, Ashesh, and Jonathan Roth. 2020. "Design-Based Uncertainty for Quasi-Experiments." August 4, 2020.

Bibliography IV

- Roth, Jonathan. 2022. "Pre-Test with Caution: Event-Study Estimates after Testing for Parallel Trends." *American Economic Review: Insights.*
- Roth, Jonathan, Pedro H C Sant'Anna, Alyssa Bilinski, and John Poe. 2021. "What's Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature," 54.
- Sant'Anna, Pedro H.C., and Jun Zhao. 2020. "Doubly Robust Difference-in-Differences Estimators." *Journal of Econometrics* 219, no. 1 (November): 101–122.
- Sun, Liyang, and Sarah Abraham. 2020. "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects." *Journal of Econometrics* (December): S030440762030378X.
- Wing, Coady, Kosali Simon, and Ricardo A. Bello-Gomez. 2018. "Designing Difference in Difference Studies: Best Practices for Public Health Policy Research." *Annual Review of Public Health* 39, no. 1 (April): 453–469.

Outline

1. Introduction
2. DID: the 2x2 case
3. Case study: Minimum wage by Card and Krueger (AER 1994)
4. Multiple groups, multiple periods
5. Modern DiD: Application to the minimum wage debate
6. Wrap-up
7. Appendix