



Stockholm
University

Pär Nyman

Methods in Political Science
20 April 2025

Differences-in-differences

Event-focused analysis of panel data

Introduction

Stronger causal claims with temporal data

- Last week: I have a random dataset. Now let us try to model the causal relations in it.
- This week: Something extraordinary happened. Let us tease out its causal effects.
- Focus on DiD or dif-in-dif.
- Briefly visit shift-share / Bartik instruments.

Introduction

How are gig workers affected by a minimum pay per order?

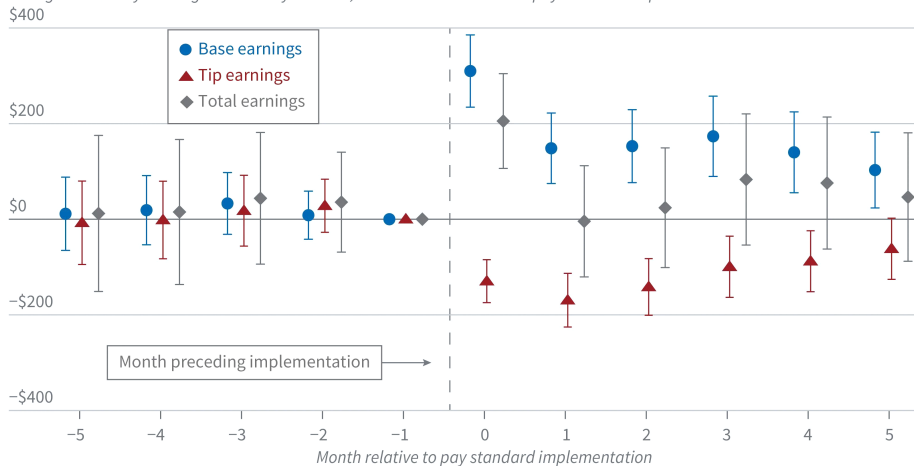
- Gig workers operate outside traditional employment relations and are not always covered by minimum wages or collective agreements.
- In 2024, Seattle established a minimum compensation for delivery tasks of \$0.44 per minute plus \$0.74 per mile, or \$5 per task, whichever is greater.
- How does that affect their earnings?

<https://www.nber.org/papers/w34545>

Introduction

Earnings of Gig Workers in Seattle

Change in monthly earnings for delivery workers, relative to month before pay standard implementation



Introduction

How are gig workers affected by a minimum pay per order?

- Maybe not the response we expected?
 - Base earnings increased immediately.
 - Tips fell (possibly because of app companies).
 - Total earnings returned roughly to baseline, as jobs per worker decreased.
- Authors' conclusion: free entry makes it hard to improve gig worker welfare through minimum payment per job.
- Our conclusion: we need sharp events, preferably with suitable controls, to detect complex responses as these.

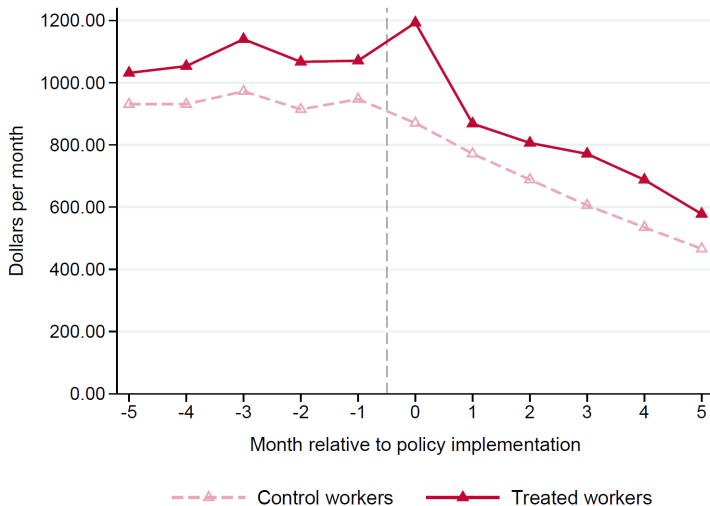
Introduction

How are gig workers affected by a minimum pay per order?

- How did they estimate these effects?
- The shown numbers are not the actual changes, but the difference between the changes for the workers in Seattle and the changes for workers in other parts of Washington.
- The control group's trajectory serves as the counterfactual.
- This is what differences-in-differences is all about.

Introduction

Total earnings

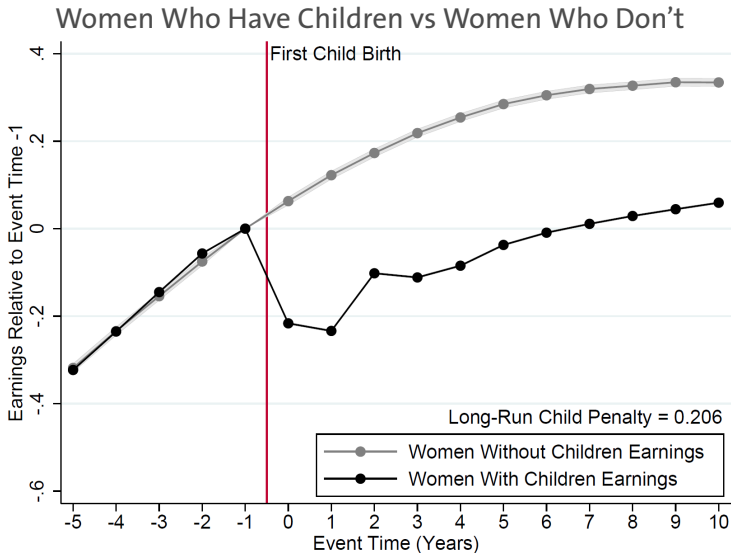


Introduction

Counterfactual

- If we can assume that earnings would have remained constant without the intervention, we can estimate a causal effect by simply comparing the post-treatment to the pre-treatment.
- If that is not the case, we need another counterfactual for how earnings would develop without the treatment.
- Let us look at another example.

Introduction

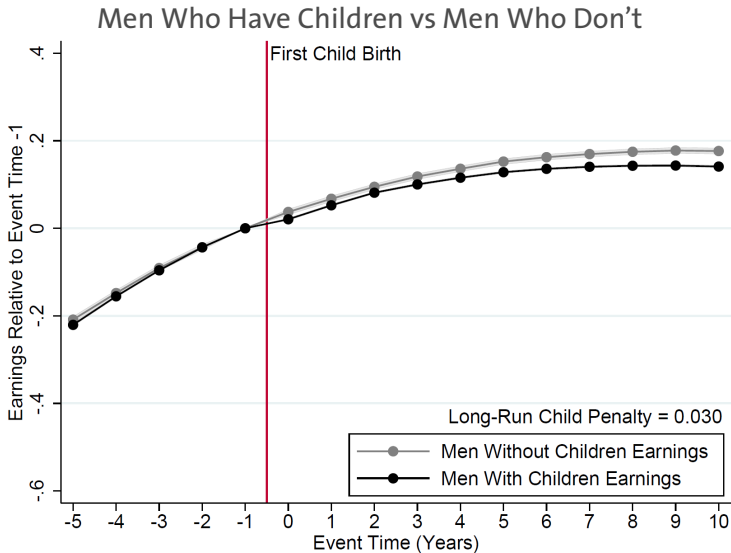


Introduction

Counterfactual

- In this case a control group was selected by finding individuals of the same age who did not have a child (approximation).
- Event-time means it is centered around the births.
- The trajectory is so stable for the treated women that a control group was not necessary (only used in appendix).
- The most interesting (but sad) comparison is to the men.
- They are not a control group, but a relevant comparison.

Introduction



Introduction

Core idea behind differences-in-differences

- One group that received a treatment and one that did not.
- The treatment is often a policy reform or some other sudden event. Should not be understood as a controlled experiment.
- There may be systematic differences between the groups, but they are assumed to follow the same trend.
- The pre- vs post difference for the untreated group serves as the counterfactual.
- The differences-in-differences (change for treated minus change for controls) is our estimate of the treatment effect.

DiD Assumptions

Core assumptions

- Parallel trends in potential outcomes: the change in earnings would have been the same as for the untreated if Seattle had not introduced a minimum pay or the women did not give birth.
- No pre-treatment effects (e.g., no anticipation behaviour)
- Non-interference: the treatment does not affect the control group (we cover it next week).

Today we focus on parallel trends.

DiD Assumptions

Parallel trends

- The DiD version of the exogeneity assumption.
 - In the absence of treatment, the treatment and control groups would have followed parallel trajectories over time.
 - $Y(0)$ follows parallel trends for treated and controls.
- Does not require balance in levels.
- Watch out for time-varying confounders that differentially affect treated and controls around treatment

Not always straightforward. You need to make a judgment.

DiD Assumptions

Parallel trends cannot be tested, but evaluation is central

- Look at the pre-trend (strongest argument?).
- Test for placebo effects (things that should not be affected by the treatment).
- Look at covariate balance over time (can be placebo test).
- The post-trend can be informative too (not common and may be affected by the treatment).
- Theory and intuition.
- Plausibility of treatment effect.

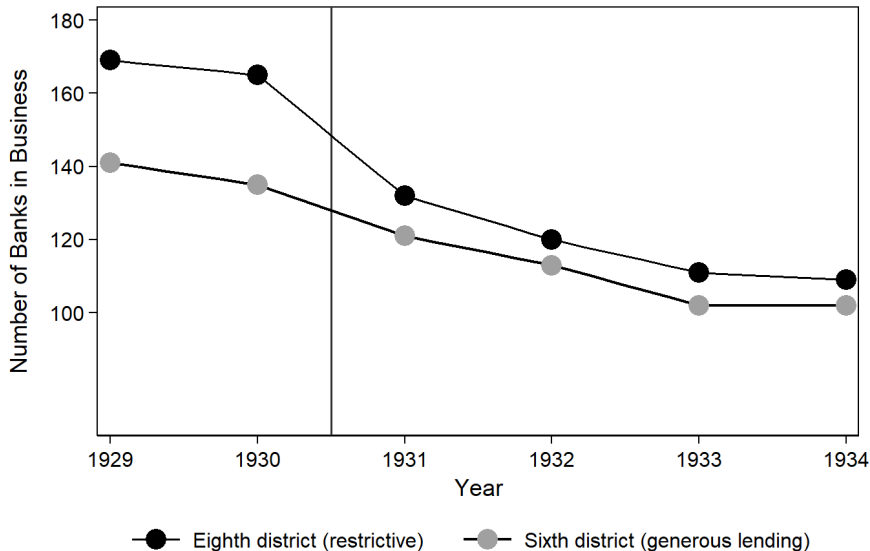
Group discussion: For which of these two studies is the parallel trends assumption the most credible?

DiD Assumptions

Number of banks in eighth vs sixth district.

- Different sides of a FED district border in Mississippi.
- Atlanta Fed (6th) followed a more generous lending policy than the St. Louis Fed (8th).
- Treatment is policy divergence when crisis hit.

DiD Assumptions



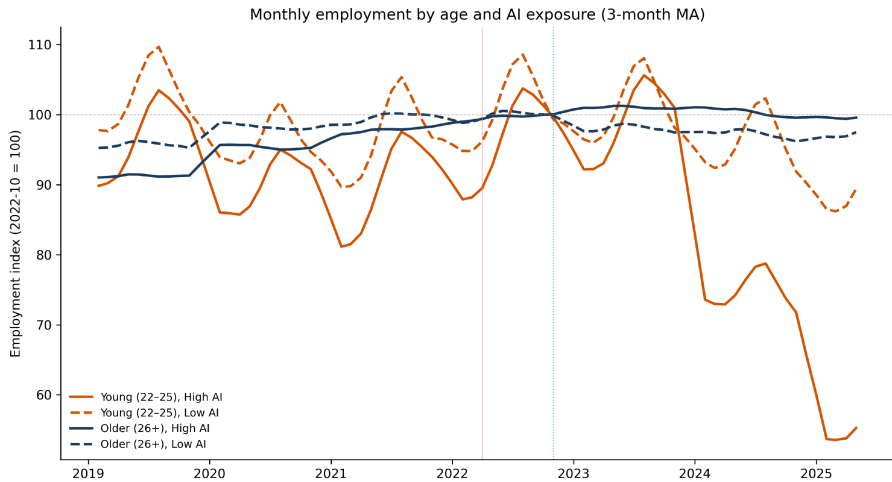
DiD Assumptions

Employment of young people in occupations with high vs low AI exposure

- Employed individuals 22–25 in occupations coded as high AI exposure (Q4) vs occupations coded as low AI exposure (Q1–Q3), three-month averages.
- Vertical lines show possible treatments: first rate hike and the launch of ChatGPT.
- Trigger warning: bad if true.

The data I show are not the exact variation they use in their DiD.

DiD Assumptions



DiD Assumptions

When is the parallel trends assumption more convincing?

- Would banks in the Sixth District have developed like banks in the Eighth District if the Atlanta Fed had followed the more restrictive policy?
- Would youth employment in AI-exposed occupations have developed like in other occupations in the absence of AI?
- Is the DiD the causal effect of lending policy / AI?

DiD Assumptions

Evaluating these two examples

- Looking at trends: similar pre-trend (AI) much more convincing than post-trend (banks).
- Plausibility: Bank-story is more convincing. The AI effect is unrealistically large and why no effect on 26+?
- Other treatments: Diffuse timing of AI treatment, and no immediate response, opens the door for confounders (interest rates, new labour law, etc).
- Trust in authors: How much data mining is behind the graph? Are they open about coding challenges?

DiD Assumptions

Evaluating these two examples

- *My suspicion:* The two declines are what would happen if you miss recent occupational codes on the cohorts which enter the 22–25 group in 2024 and 2025 and smooth that Dec–Jan drop with a 3-month average.
- That is of course not the main point, and not something I expect you to see from a simple graph, but perhaps a relief for your job prospects.

The typical DiD

The canonical dif-in-dif

- Two periods used for estimation: pre- and post-treatment.
- One treatment and one control group.
- Untreated group remains untreated in all periods.
Treatment cannot be reversed.
- Binary treatment.
- Balanced panel with no compositional changes or non-random missingness.

Straightforward to calculate the differences-in-differences.

The typical DiD

Estimating the treatment effect

- A DiD-effect can be calculated from four averages:
 - Change among treated minus change among controls
 - or post-treatment difference minus pre-difference

Generous lending saved 19 banks

	6th	8th	Dif.
1930	135	165	-30
1931	121	132	-11
Change	-14	-33	

$$-14 - -33 = 19$$

$$-30 - -11 = 19$$

The typical DiD

In practice it is estimated with a regression analysis

- First-difference model (can be "long" differences)
- TWFE + treatment indicator (FE on treatment group & time)
- Regress y on treatment group, post-treatment dummy and interaction between the two.

All three techniques provide the same estimate as the manual calculation for the 2x2. We will confirm that at the workshop. More complex settings are often handled with TWFE.

Extensions

So is DiD the same thing as TWFE?

- Sadly, the terms are sometimes used interchangeably.
- Think of DiD as a design/idea and TWFE as a model specification. Not all DiD designs are estimated with TWFE.
- And I don't think every TWFE should be called DiD.
 - Personally I would not use the DiD lingo unless I believe in approximately parallel trends, preferably because I can investigate pre-trends in an event-study graph.
 - A DiD can be more or less typical depending on the deviations from the canonical model.

Extensions

The canonical model is often extended

- The MLDA example in A&P has:
 - Multiple periods.
 - Staggered treatment.
 - Continuous treatment.
- They use TWFE + treatment indicator. Understand it as DiD.
- The modern DiD literature is concerned with all deviations from the canonical 2x2. Results can be misleading, in particular if effects vary across groups or over time.
- IMHO: They are obsessing. But we can learn from them.
- Roth et al provides an accessible summary and a modern contrast to A&P. Read cursorily.

Extensions

Multiple periods

- Better precision with more periods.
- Assumes parallel trends over a longer period.
- An average effect is neither better or worse if there are delayed or time-varying effects.
- Good practice to estimate separate effects for every post-period, allowing lagged and varying effects.

No biggie.

Extensions

Staggered treatment

- What units are treated at different points in time?
- TWFE works well if treatment effects are constant.
- Including comparisons between treated now and treated before improves precision but is biased if treatment effect is delayed or diminishing.
- More weight to units treated in the middle of the time period (more variance in treatment).
- This is an insight we can use for all kinds of panel data.

Extensions

Staggered treatment

	$t = 1$	$t = 2$	$t = 3$
Early treated	Untreated	Treated	Treated
Late treated	Untreated	Untreated	Treated

Clean comparison from $t = 1$ to $t = 2$. Including $t = 3$ provides more information but is problematic if treatment effects evolve over time (because the change for the early treated is still affected by the treatment before $t = 2$).

Extensions

Reversible treatments

- We also need assumptions about dynamics if controls can be previously treated.
- Similar trade-off as with staggered treatments.
- But maybe the parallel trends assumption is more likely when the control group is also treated at some time?
- If we think about this as a trade-off, it is applicable to many other cases when we think about suitable comparisons. Should we go for never treated or treated at another time?

Extensions

Continuous treatments

- If the difference between treated and control is not binary, we need assumptions about linearity.
- Continuous treatments make it harder to create event-study graphs (treatment group and control group over time). And we love those graphs!

Extensions

Take-aways about extensions to the canonical model

- Violations of the canonical model is not a crime. Most likely consequence is some attenuation bias.
- But as always, we like large and abrupt changes that are spaced far between each other. That is the DiD logic.
- For the course: focus on understanding the DiD estimator and evaluating the parallel trends assumption.
- For other purposes: the modern literature may exaggerate the actual problems, but important to be aware of and useful for improving understanding of paneldata in general.

Other approaches

The seminar assignment

- Oral presentations instead of written assignment.
- Never tried before.
- You pick a study and present it for the seminar group.
- No Stata in the assignment = short intro on the workshop.
- I told you to use a dif-in-dif or a shift-share.
- Shift-share useful but not on exam.
- A more formal take on shift-share would require knowledge about IV, which you do not have yet.

Other approaches

The shift-share logic

- The effect of an external shock (the shift) depends on a pre-existing characteristic (the share). The shock-exposure is measured with an interaction ($\text{shift} \times \text{share}$) and is assumed to be exogenous / unrelated to other factors that would have changed the outcome.

Other approaches

Illustration: Bartik (1991)

- Suppose we want to know whether local labour demand raises wages or employment.
- Exogeneity assumption violated: booming places are different (policy, infrastructure, productivity, etc.) and wages and employment affect each other.
- We need exogenous variation in labour demand.

Other approaches

Bartik's classic example

- Start with two cities:
 - City A has a lot of car and steel jobs.
 - City B has more health care and education jobs.
- Then national demand changes:
 - Car and steel decline nationally.
 - Health care grows nationally.
- Then labour demand increases in B relative to A.
- Their different *industry mix at the start* makes them differently exposed to the same national shocks.

This is the Bartik idea: use national industry shifts, weighted by local industry shares, to predict local labour-demand shocks.

Other approaches

The shift-share variation combines two things

$$\text{Shock}_r = \sum_i \text{industry share}_{r,i, \text{ before}} \times \text{national growth shock}_i$$

- We can understand the interaction as a predicted change in local employment, or a proxy for that.
- We do not want to use observed employment because it is affected by possible confounders as well as local wages.
- This interaction is arguably less susceptible to such bias.

Other approaches

Main assumptions in plain language

- The initial industry shares are measured *before* the shocks.
- The national industry shocks are not themselves driven by the specific local outcomes we study.
- The shift-share variable should affect the outcome mainly through the local variable of interest (if it understood as a predictor for that variable rather than something important on its own).

What should make us cautious?

- Places with different industry mixes may already be on different long-run trends (i.e., violation of parallel trends).
- A few industries may dominate the measure.

Other approaches

Another example

- A PhD Candidate I supervise wants to know how party choice is affected by the household's fuel consumption.
- Background: Fuel-prices went up 2022 and was a salient issue in the Swedish election.
 - Fuel consumption is endogenous to many confounders.
 - So much else happened at the same time as prices increased.
- But maybe the interaction between the two is better?
- $\text{Share} = \text{L/km} \times \text{the car's annual mileage.}$
- $\text{Shift} = \text{Fuel prices.}$
- Preliminary results indicate that the incumbent red-green government was punished for increasing fuel prices (they lost more support among those who drive a lot).

Other approaches

Take-away

- DiD uses *timing* plus a comparison group.
- Shift-share uses *differential exposure* to common shocks. Not too different from a DiD with continuous treatment.
- Same goal in both cases: find variation that is closer to exogenous by exploiting sudden shocks.
- Shift-share is not on exam. So why on the seminar?
 - Shift-share designs are relatively common and not always that data-demanding = you can use them.
 - The intuition can be explained in a short presentation.
 - Thinking in these terms helps us understand the rest of the course content better.