ECON 3510: Political Economy of Development
Lecture 3: Panel Data and Difference in Differences

Instructor: Weizheng Lai

Bowdoin College

Fall 2025

Introduction

- We have focused on cross-sectional data, which include observations from a population at a point in time.
- In this part, we discuss **panel data**, where we have observations for each unit i (say a person or state) across multiple periods t.
- The information at the time dimension can help us deal with certain types of unobserved confounding variables.
- We will discuss how the regression techniques can exploit the advantage of panel data.

Outline

Union Memberships and Wages

- One of the oldest questions in labor economics is the relationship between union membership and wages.
  - Do workers whose wages are set by collective bargaining earn more because of this?
  - Or would they earn more anyway, perhaps because they are more experienced or skilled?

- Suppose we observe a group of workers' union memberships (x_{it}) and wages (y_{it}) over time. $i = 1, 2, \cdots, N$ and $t = 1, 2, \cdots, T$.

- We wish to estimate a population model:
$$y_{it} = \alpha + \beta x_{it} + u_{it}.$$

- We can regress y_{it} on x_{it}. However, we may not be convinced that the regression coefficient captures the causal effect of union membership on wages (β). The union membership may be correlated with unobserved/unmeasurable factors that also determine wages.
$$E(u_{it} \mid x_{it}) \neq 0.$$

Fixed Effects

- It is possible to address this issue if

$$E(u_{it} \mid x_{it}) = \lambda_i \text{ is constant.}$$

  That is, conditional on union membership, the effect of the unobserved determinant on wages is fixed (or time invariant) for an individual. λ_i is thus said to be an **individual fixed effect**.

- The population model is written as

$$y_{it} = \alpha + \beta x_{it} + \lambda_i + v_{it},$$

  where $v_{it} = u_{it} - \lambda_i$ satisfies $E(v_{it} \mid x_{it}, \lambda_i) = 0$. This is a **fixed effects model**.

Fixed Effects Model

- Note that
$$y_{it} = \alpha + \beta x_{it} + \lambda_i + v_{it}$$
has the individual specific intercept $\alpha_i \equiv \alpha + \lambda_i$ and the common slope β.

- The model can be written as
$$y_{it} = \beta x_{it} + \sum_j \alpha_j D_{ij} + v_{it},$$
where D_{ij} is a dummy variable that equals 1 if $i = j$ and 0 otherwise.

- β can be consistently estimated by an OLS regression of y_{it} on x_{it} and $\{D_{ij}\}$. The estimator $\hat{\beta}_{FE}$ is said to be the **fixed effects estimator**.

- Note that the fixed effects model can be estimated due to the panel data structure: an individual i is observed at multiple time points.
  - In consider cross-sectional data ($T = 1$), x_{i1} is perfectly collinear with $\{D_{ij}\}$.

Within Estimator

- Besides directly implementing OLS, there is another way to estimate

$$y_{it} = \alpha_i + \beta x_{it} + v_{it}.$$

- Given i, take the average of the above equation over time $t = 1, 2, \cdots, T$ to get a cross-sectional equation:

$$\bar{y}_i = \alpha_i + \beta \bar{x}_i + \bar{v}_i.$$

  Take the difference between the two equations:

$$y_{it} - \bar{y}_i = \beta(x_{it} - \bar{x}_i) + (v_{it} - \bar{v}_i).$$

  Here, α_i is eliminated. This model relates the deviation of y_{it} from the mean outcome $\bar{y}_i$ to the deviation of x_{it} from the mean treatment $\bar{x}_i$. It limits the comparison to the same individual.

- β can be consistently estimated by an OLS regression of $y_{it} - \bar{y}_i$ on $x_{it} - \bar{x}_i$. Since only *within-individual variation* is used, the estimator is called a **within estimator**, denoted by $\hat{\beta}_{WI}$.
  - Variables without variation over time would be dropped in a fixed effects model.

- In fact, $\hat{\beta}_{WI} = \hat{\beta}_{FE}$. Most software uses this fact to calculate $\hat{\beta}_{FE}$, which avoids creating and adding many dummies.

First Difference Estimator

- We can take the first difference of population models between t and $t-1$:

$$\begin{aligned} y_{it} &= \alpha_i + \beta x_{it} + v_{it} \\ y_{i,t-1} &= \alpha_i + \beta x_{i,t-1} + v_{i,t-1} \\ \Rightarrow y_{it} - y_{i,t-1} &= \beta(x_{it} - x_{i,t-1}) + (v_{it} - v_{i,t-1}) \\ \Delta y_{it} &= \beta \Delta x_{it} + \Delta v_{it}. \end{aligned}$$

  Here, α_i is also eliminated.

- β can be consistently estimated by an OLS regression. $\hat{\beta}_{FD}$ is the first difference estimator.

- When $T = 2$, $\hat{\beta}_{FD} = \hat{\beta}_{WI} = \hat{\beta}_{FE}$.
  - $y_{i2} - \bar{y}_i = y_{i2} - \frac{y_{i1}+y_{i2}}{2} = \frac{1}{2}(y_{i2} - y_{i1})$.

Extension

- **Multi-way fixed effects model:** e.g.,
$$y_{it} = \alpha_i + \delta_t + \beta x_{it} + \varepsilon_{it}.$$
Time fixed effects δ_t control for variables that are constant across units but evolve over time.

- Fixed effects models can be applied even in non-panel data. The idea of fixed effects is to leverage within variation.
    - In Dale and Krueger (2002), they control for dummies of application/application results, essentially fixed effects, so that they can compare students who have the same applications/admissions but choose to attend more or less selective colleges.

- Statistical inference needs to use clustered robust standard errors to control for likely time series correlation in the error term.
    - Clustering by individual assumes independence across individuals.
    - Clustering by group assumes independence across groups of individuals.

Outline

Motivating Example: Card and Krueger (1994)

- What is the effect of minimum wage on employment?
  - Theoretically, in a competitive market, raising the minimum wage moves the equilibrium point up on the downward sloping labor demand curve, thus reducing employment.
- Card and Krueger (1994) study this question, exploiting a dramatic change in the New Jersey state minimum wage.
- On April 1, 1992, NJ raised the state minimum from $4.25 to $5.05 per hour.
- Card and Krueger (1994) collected data on employment at fast food restaurants in New Jersey in February 1992 and again in November 1992.
  - Fast-food restaurants are a leading employer of low-wage workers
- Could we simply compare average employment before and after the minimum wage change? Would the difference tell us the causal effect?

  Maybe not. If there were unobserved employment trends in the fast food industry, then the post-pre difference could in part reflect such unobserved trends.

Controlling for Unobserved Trends

- To control for unobserved trends, Card and Krueger (1994) used data on employment at fast food restaurants in Eastern Pennsylvania. During the period, the minimum wage in Pennsylvania stayed at $4.25.
- The idea is that Eastern PA restaurants can provide a good proxy for employment trends that are not due to the minimum wage change.
- More formally, consider a potential outcome framework.

  Y_{1ist} = employment at restaurant i in state s and time t if the state minimum wage is high.

  Y_{0ist} = employment at restaurant i in state s and time t if the state minimum wage is low.

Controlling for Unobserved Trends (Cont'd)

- The employment change in NJ:

$$\begin{aligned}
\Delta_{\mathrm{NJ}} &\equiv E(Y_{ist} \mid s=\mathrm{NJ}, t=\mathrm{Nov}) - E(Y_{ist} \mid s=\mathrm{NJ}, t=\mathrm{Feb}) \\
&= E(Y_{1ist} \mid s=\mathrm{NJ}, t=\mathrm{Nov}) - E(Y_{0ist} \mid s=\mathrm{NJ}, t=\mathrm{Feb}) \\
&= \underbrace{E(Y_{1ist} \mid s=\mathrm{NJ}, t=\mathrm{Nov}) - E(Y_{0ist} \mid s=\mathrm{NJ}, t=\mathrm{Nov})}_{\text{treatment effect}} \\
&\quad + \underbrace{E(Y_{0ist} \mid s=\mathrm{NJ}, t=\mathrm{Nov}) - E(Y_{0ist} \mid s=\mathrm{NJ}, t=\mathrm{Feb})}_{\text{bias}}.
\end{aligned}$$

  As expected, Δ_{NJ} contains the treatment effect and a term for unobserved trends. The treatment effect is an average treatment effect on the treated (ATT):

$$ATT = E(Y_{1ist} \mid s=\mathrm{NJ}, t=\mathrm{Nov}) - E(Y_{0ist} \mid s=\mathrm{NJ}, t=\mathrm{Nov}).$$

- Similarly, the employment change in PA:

$$\begin{aligned}
\Delta_{\mathrm{PA}} &\equiv E(Y_{ist} \mid s=\mathrm{PA}, t=\mathrm{Nov}) - E(Y_{ist} \mid s=\mathrm{PA}, t=\mathrm{Feb}) \\
&= E(Y_{0ist} \mid s=\mathrm{PA}, t=\mathrm{Nov}) - E(Y_{0ist} \mid s=\mathrm{PA}, t=\mathrm{Feb})
\end{aligned}$$

  It only involves unobserved trends.

Difference-in-Differences

- Assume

$$E(Y_{0ist} \mid s = \text{NJ}, t = \text{Nov}) - E(Y_{0ist} \mid s = \text{NJ}, t = \text{Feb}) = E(Y_{0ist} \mid s = \text{PA}, t = \text{Nov}) - E(Y_{0ist} \mid s = \text{PA}, t = \text{Feb}),$$

  i.e., in the absence of the minimum wage increase, NJ restaurants would have had the same employment trends as PA restaurants. This is often referred to as a **parallel trends assumption.**

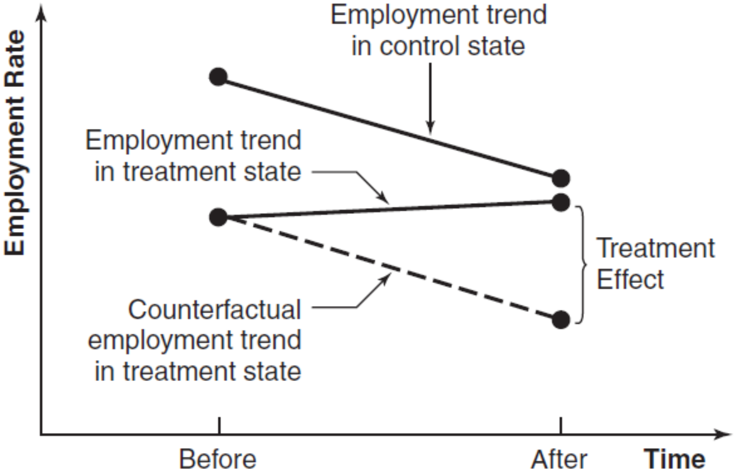
- If the parallel trends assumption holds,

$$\Delta_{\text{NJ}} - \Delta_{\text{PA}} = ATT$$

  identifies the causal effect of the minimum wage increase on employment.

- This approach is called **difference-in-differences (DiD)**. It essentially compares the evolution of the outcome between treatment and control groups (here, NJ vs. PA); the control group helps with modeling the counterfactual trends.

DiD Intuition



Card and Krueger (1994) Results

- No evidence that the minimum wage reduced employment!

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

DiD Regression

- We do not have to compute averages manually. The DiD estimand is a function of conditional expectations. We can use regression to “automate” calculation.
- **Parallel trends assumption:** $E(Y_{0ist} \mid s,t) = \alpha_s + \delta_t$.
- Let D_s be a dummy that equals 1 for the treatment state (NJ). $Post_t$ is a dummy that equals 1 if t is after the policy change (after April 1992).

$$\begin{aligned} Y_{ist} &= (D_s \times Post_t)Y_{1ist} + (1 - D_s \times Post_t)Y_{0ist} \\ &= (Y_{1ist} - Y_{0ist}) \cdot (D_s \times Post_t) + Y_{0ist} \\ &= (Y_{1ist} - Y_{0ist}) \cdot (D_s \times Post_t) + \alpha_s + \delta_t + u_{ist}, \end{aligned}$$

where $u_{ist} = Y_{0ist} - (\alpha_t + \delta_t)$, and $E(u_{ist} \mid s,t) = 0$ under parallel trends.

$$E(Y_{ist} \mid s,t) = E[(Y_{1ist} - Y_{0ist}) \cdot (D_s \times Post_t) \mid s,t] + \alpha_s + \delta_t.$$

DiD Regression (Cont'd)

- We can set up a regression
$$Y_{ist} = \alpha_s + \delta_t + \beta(D_s \times Post_t) + \varepsilon_{ist}.$$
This is a fixed effects model. It can be estimated by OLS.

- **Claim:** Regression coefficient β identifies ATT under parallel trends.

- Why? The regression above models the CEF as
$$\begin{aligned}
E(Y_{ist} \mid s = \text{NJ}, t = \text{Nov}) &= \alpha_{\text{NJ}} + \delta_{\text{Nov}} + \beta \\
E(Y_{ist} \mid s = \text{NJ}, t = \text{Feb}) &= \alpha_{\text{NJ}} + \delta_{\text{Feb}} \\
E(Y_{ist} \mid s = \text{PA}, t = \text{Nov}) &= \alpha_{\text{PA}} + \delta_{\text{Nov}} \\
E(Y_{ist} \mid s = \text{PA}, t = \text{Feb}) &= \alpha_{\text{PA}} + \delta_{\text{Feb}}.
\end{aligned}$$
Thus,
$$\begin{aligned}
\beta = {} & [E(Y_{ist} \mid s = \text{NJ}, t = \text{Nov}) - E(Y_{ist} \mid s = \text{NJ}, t = \text{Feb})] \\
& - [E(Y_{ist} \mid s = \text{PA}, t = \text{Nov}) - E(Y_{ist} \mid s = \text{PA}, t = \text{Feb})] \\
= {} & ATT. \qquad \text{(by parallel trends)}
\end{aligned}$$
Estimator:
$$\hat{\beta} = (\bar{Y}_{\text{NJ,Nov}} - \bar{Y}_{\text{NJ,Feb}}) - (\bar{Y}_{\text{PA,Nov}} - \bar{Y}_{\text{PA,Feb}}).$$

DiD with Multiple Periods

- Card and Krueger (1994) have 2 periods. Often we have more than 2 periods for a DiD analysis.

- We can still estimate
$$Y_{it} = \alpha_i + \delta_t + \beta(D_i \times Post_t) + \varepsilon_{it},$$
and coefficient β gives the ATT for the entire post-treatment period.

- But with more periods, we can do more:
  - We can test whether the parallel trends assumption appears to hold *prior* to treatment.
  - We can analyze how the ATT changes over time.

- How do we do this?

DiD with Multiple Periods (Cont'd)

- Suppose that we have periods $t = -\underline{T}, \cdots, \bar{T}$. Treated units begin getting treatment at period 0.

- For each period $s \neq 0$, we can estimate a 2-period DiD between period s and period -1 (the period just ahead of treatment):

$$\hat{\beta}_s = \underbrace{(\bar{Y}_{1s} - \bar{Y}_{1,-1})}_{\text{Diff for treatment group}} - \underbrace{(\bar{Y}_{0s} - \bar{Y}_{0,-1})}_{\text{Diff for control group}},$$

  where $\bar{Y}_{dt}$ is the average for treatment d in period t.

- Conveniently, $\hat{\beta}_s$'s are equal to the OLS estimates of the regression

$$Y_{it} = \alpha_t + \delta_t + \sum_{s \neq -1} D_i \times \mathbb{1}[t = s] \times \beta_s + \varepsilon_{it}.$$

  This is often called an **event study model** or a **dynamic model**.

  β_s ($s \geq 0$) identifies ATT at post-treatment period s, under parallel trends.

  β_s ($s < 0$) identifies the difference in trends between treatment and control groups in pre-periods. Under parallel trends, we expect them to be 0. **(pre-trends test)**

Example - Medicaid Expansion

- The Affordable Care Act (ACA, aka Obamacare) expanded Medicaid coverage to people with income up to 138% of the federal poverty line.
- Medicaid expansion went into effect in 2014.
- By 2015, 24 states had expanded Medicaid (more have done so since).
- Carey et al. (2020) study the impacts of Medicaid expansion using a DiD design comparing early-adopting states to non-adopters. They use data for 2008–2015.

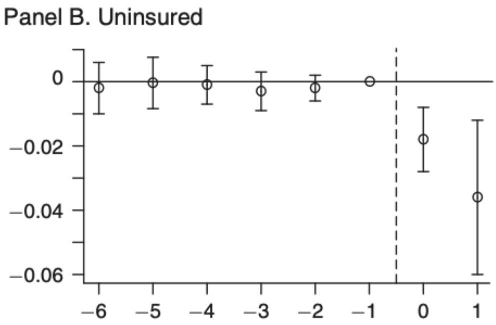
Carey et al. (2020)

- A slightly simplified version of their regression specification is

$$Y_{its} = \phi_t + \lambda_s + \sum_{r \neq -1} D_i \times 1[t = 2014 + r] \times \beta_r + \varepsilon_{ist}$$

where Y_{ist} is outcome for person i in year t in state s, and $D_i = 1$ if in an expansion state.

- Results show similar "pre-trends" but negative effects after treatment.



Some Caution about Parallel Trends

- DiD relies on the parallel trends assumption, which allows for selection bias but requires it to be stable over time. This rules out time-varying confounding factors.

- Often we will be worried about time-varying confounds—e.g., in the Medicaid example, macro-economic factors might be different across states.

- The parallel trends assumption is fundamentally a **functional formal assumption**: it is assumed that

$$E(Y_{it} \mid i,t) = \alpha_i + \delta_t.$$

  To make parallel trends more plausible, we can extend it to allow for more complex functional forms:

  - Time-varying variables: $E(Y_{it} \mid i,t) = \alpha_i + \delta_t + \mathbf{X}_{it}'\gamma$;
  - Group-specific linear trends: $E(Y_{it} \mid i,t) = \alpha_i + \delta_t + \phi_{region(i)} \times t + \mathbf{X}_{it}'\gamma$;
  - Heterogeneous dynamics: $E(Y_{it} \mid i,t) = \alpha_i + \phi_{region(i)} \times \delta_t + \mathbf{X}_{it}'\gamma$.

Some Caution about Parallel Trends (Cont'd)

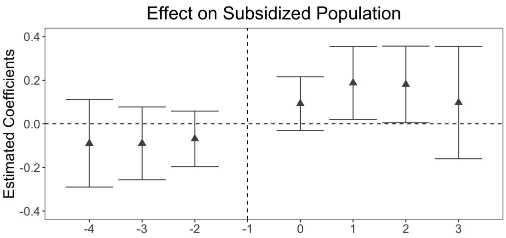
- Testing for pre-treatment differences ("pre-trends") can help increase our confidence in the research design. But they're not perfect. Why?

1. Just because trends were parallel *beforehand* doesn't mean that they would continue to be afterwards.
   - E.g., there is another policy change that has exactly the same timing as the treatment policy of interest. The DiD estimand won't be able to separate the effects of the two policies.
   - We may use institutional knowledge to rule out confounding policies or determine if such policies could affect the outcome of interest.

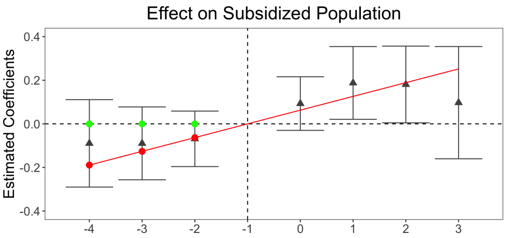
2. Often our estimates of pre-trends are noisy so we're not sure whether they're actually zero or not.
   - Recall when we fail to reject $H_0 : \beta_{\text{pre}} = 0$, it doesn't mean we "accept" that β_{pre} is a true zero.

Testing Pre-Trends

- In addition to looking at the point estimates of pre-trends, it's important to consider **what the CIs rule out**.
- A good rule of thumb for whether a plot is convincing is whether you can draw a smooth line through all the confidence intervals.



(a) Are you convinced there's an effect here?



(b) Maybe not!

DiD Extension I: Non-Binary Treatment

- We have focused on binary treatment D_i.
- The DiD analysis can be applied when treatment is measured with a continuous variable T_i, which captures the intensity/dosage of a treatment.

Example: Finkelstein (2007)

- Medicare is a US government program introduced in 1965 to provide health insurance to *all* the elderly. However, the impact of Medicare can vary due to differences in the uninsured rate.
- A stylized version of Finkelstein (2007)'s regression:

$$\log(y_{ijt}) = \alpha_j + \delta_t + \sum_{k=1948}^{1975} T_{z(j)} \times \mathbb{1}\{t = k\} + \varepsilon_{ijt}.$$

y_{ijt} is the outcome of hospital i in county j and year t. $T_{z(j)}$ is the uninsured rate among the elderly in 1963—a higher value should relate to a greater influence of Medicare.

DiD Extension II: Staggered Timing

- We have focused on the treatment that is implemented at the same time for all treated units. But the timing may vary, e.g., states pass a policy in different years.

- The DiD analysis can be applied in this setting by running OLS regressions like:

$$Y_{it} = \alpha_i + \delta_t + \beta D_{it} + \varepsilon_{it}$$

  where $D_{it} = 1$ if unit i is treated in period t.
  - This incorporates the classical case that we have discussed (define $D_{it} = D_i \times Post_t$).
  - Under parallel trends, $\beta \approx ATT$.

- Event study model:

$$Y_{it} = \alpha_i + \delta_t + \sum_{s \neq -1} \beta_s D_{it}^s + \varepsilon_{it},$$

  where $D_{it}^s = 1$ if i is s periods relative to the start period of its treatment; $= 0$ if never treated.

- *(Not required.)* Staggered DiD involves two types of comparisons: treated vs. control; newly treated vs. not yet treated; and newly treated vs. already treated.
  - In recent years, a growing econometric literature studies how to use reasonable comparisons in staggered DiD (e.g., see a review by Roth et al., 2023).

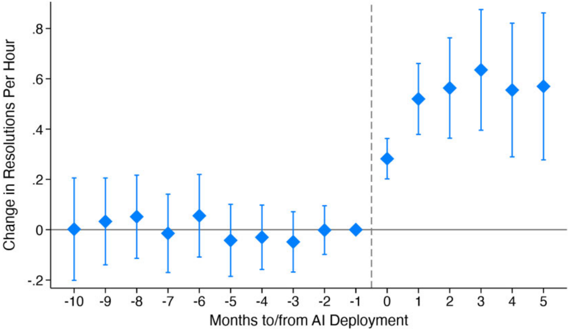
Example: Brynjolfsson et al. (2025)

- Brynjolfsson et al. (2025) studies how the generative AI–based conversational assistant affects customer-support agents' productivity (measured by # resolutions per hour).

- They estimate:

$$y_{it} = \alpha_i + \delta_t + \beta AI_{it} + \mathbf{X}'_{it}\gamma + \varepsilon_{it},$$

where $AI_{it} = 1$ if AI has been activated for agent i at time t.



DiD Extension III: Non-Panel Data

- A DiD analysis doesn't require panel data.
- The key is that we need to observe outcomes for treated and control groups both before and after the treatment, so that we can compare changes over time between the two groups.
  - Repeated cross-sectional data can suffice, or data with information relevant to the timing of the treatment.

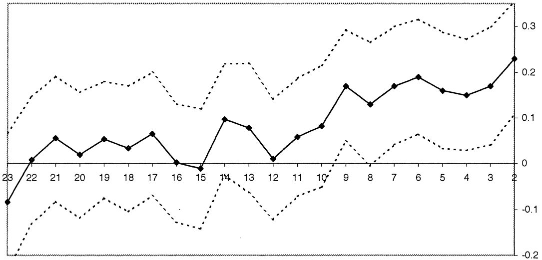
Example: Duflo (2001)

- Duflo (2001) studies the impact of Indonesia's large primary school construction program in 1974.
  - The program constructed more schools in some regions but less in others.
  - The program should benefit students younger than 12 in 1974, who haven't completed primary school.

- She uses the 1995 intercensal survey of Indonesia (a cross section). But she can observe an individual's birth year and birth region. She estimates

$$S_{ijt} = \alpha_j + \delta_t + \sum_{k \neq 24} \beta_k (P_j \times \mathbb{1}\{t - 1974 = j\}) + \text{controls} + \varepsilon_{ijt},$$

where S_{ijt} is education of individual i born in region j and in year t. P_j is the number of schools constructed in region j.



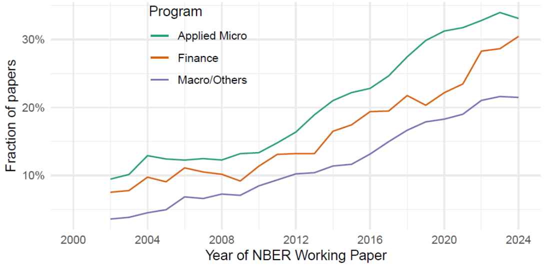
Age in 1974

FIGURE 1. COEFFICIENTS OF THE INTERACTIONS AGE IN 1974* PROGRAM INTENSITY IN THE REGION OF BIRTH IN THE EDUCATION EQUATION

DiD Summary

- DiD checklist:
  - Does your dataset allow you to run a DiD?
  - What assumption do you need for a causal interpretation?
  - How do you justify your assumption?

- DiD has become one of the most popular methods in modern economics (and perhaps even social sciences). Use it with caution!



(a) Difference-in-differences

Source: Goldsmith-Pinkham (2024)

References I

Brynjolfsson, Erik, Danielle Li, and Lindsey Raymond (2025). “Generative AI at work”. *The Quarterly Journal of Economics*, qjae044.

Card, David and Alan B Krueger (1994). “Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania”. *The American Economic Review* 84.4, pp. 772–793.

Carey, Colleen M, Sarah Miller, and Laura R Wherry (2020). “The impact of insurance expansions on the already insured: the Affordable Care Act and Medicare”. *American Economic Journal: Applied Economics* 12.4, pp. 288–318.

Dale, Stacy Berg and Alan B Krueger (2002). “Estimating the payoff to attending a more selective college: An application of selection on observables and unobservables”. *The Quarterly Journal of Economics* 117.4, pp. 1491–1527.

Duflo, Esther (2001). “Schooling and labor market consequences of school construction in Indonesia: Evidence from an unusual policy experiment”. *American Economic Review* 91.4, pp. 795–813.

Finkelstein, Amy (2007). “The aggregate effects of health insurance: Evidence from the introduction of Medicare”. *The Quarterly Journal of Economics* 122.1, pp. 1–37.

Goldsmith-Pinkham, Paul (2024). “Tracking the credibility revolution across fields”. *arXiv preprint arXiv:2405.20604*.

Roth, Jonathan et al. (2023). “What’s trending in difference-in-differences? A synthesis of the recent econometrics literature”. *Journal of Econometrics* 235.2, pp. 2218–2244.