Part D: Panel Data Methods

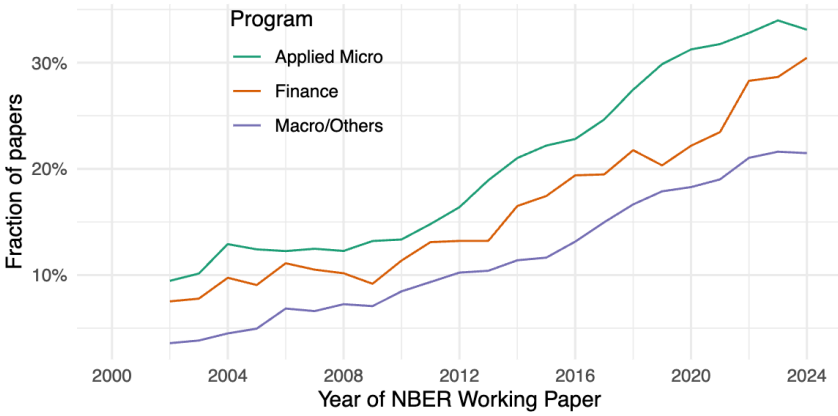
D2: Canonical Difference-in-Differences

Kirill Borusyak

ARE 213 Applied Econometrics

UC Berkeley, Fall 2025

It's everywhere



(a) Difference-in-differences

Fraction of NBER working papers mentioning DiD (Goldsmith-Pinkham (2024), Fig.5a)

D2 outline

1. $2x2$ DiD
2. DiD and event studies with multiple units and periods
3. Discussion of conventional event studies

Difference-in-differences: Idea

- We are interested in causal effects of some binary treatment
- But treatment is not randomly assigned across units
- "Quasi-experimental" contrasts are easier to find in panel data
- Card and Krueger (1994): effect of minimum wages on employment
  - On April 1, 1992, NJ raised the min.wage from \$4.25 to \$5.05
  - Min.wage in PA stayed at \$4.25
  - Measure average employment at fast food restaurants after (Nov 1992) minus before (Feb 1992), in NJ minus PA

Card and Krueger (1994)

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)

(Card-Krueger Table 3, reproduced from MHE Table 5.2.1)

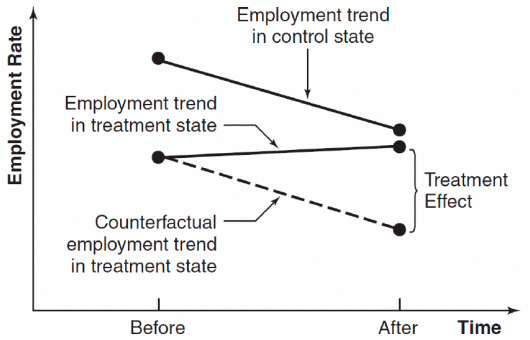
Is this causal?

- Define potential outcomes $Y_{it}(d)$ for $d = 0, 1$ (low/high min.wage), with $\tau_{it} = Y_{it}(1) - Y_{it}(0)$ and $Y_{it} = Y_{it}(D_{it})$

D_{it}	$i = PA$	$i = NJ$
$t = 1$	0	0
$t = 2$	0	1

- Estimand: $ATT = \tau_{NJ,2}$
- Assumptions for this potential outcomes formulation:
  - No spillovers
  - No anticipation effects: Y_{i1} does not depend on D_{i2}
  - Parallel trends...

Parallel trends (PTA)



(MHE Figure 5.2.1, corrected by Peter Hull)

Parallel trends: $\underbrace{\mathbb{E}\left[Y_{NJ,2}(0) - Y_{NJ,1}(0)\right]}_{\text{unobserved}} = \underbrace{\mathbb{E}\left[Y_{PA,2}(0) - Y_{PA,1}(0)\right]}_{\text{observed}}$

Note: notation for a fixed sample, $\mathbb{E}\left[Y_{it}(0)\right]$ depends on i, t. Expectations taken with respect to other determinants of the outcome

PTA = TWFE for $Y(0)$

$$\mathbb{E}\left[Y_{NJ,2}(0) - Y_{NJ,1}(0)\right] = \mathbb{E}\left[Y_{PA,2}(0) - Y_{PA,1}(0)\right]$$

is equivalent to

$$\mathbb{E}\left[Y_{it}(0)\right] = \alpha_i + \beta_t \qquad \text{for } i = NJ, PA;\ t = 1, 2$$

$\Leftarrow$: Obvious

$\Rightarrow$: Let $\alpha_i = \mathbb{E}\left[Y_{i,1}(0)\right]$, $\beta_1 = 0$,
$\beta_2 = \mathbb{E}\left[Y_{NJ,2}(0) - Y_{NJ,1}(0)\right] = \mathbb{E}\left[Y_{PA,2}(0) - Y_{PA,1}(0)\right]$

2x2 DiD estimator

- $\hat{\tau} = \left(Y_{NJ,2} - Y_{NJ,1} \right) - \left(Y_{PA,2} - Y_{PA,1} \right)$ is unbiased for ATT
- $\hat{\tau}$ can be obtained as OLS from the TWFE specification

$$Y_{it} = \tilde{\alpha}_i + \tilde{\beta}_t + \tau D_{it} + \varepsilon_{it}$$

  - Proof: with 2 periods, the first-differenced equation gives the same $\hat{\tau}$,

$$Y_{i2} - Y_{i1} = \tilde{\beta} + \tau \underbrace{\left(D_{i2} - D_{i1} \right)}_{=\mathbf{1}[i=NJ]} + \left(\varepsilon_{i2} - \varepsilon_{i1} \right)$$

Problems with 2x2 designs (1)

1. Effectively 4 observations: can't separate τ from other shocks in NJ relative to PA

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)

These SE are not clustered, and cannot be!

Problems with 2x2 designs (2)

Clustering is necessary: there are state-level fluctuations all the time

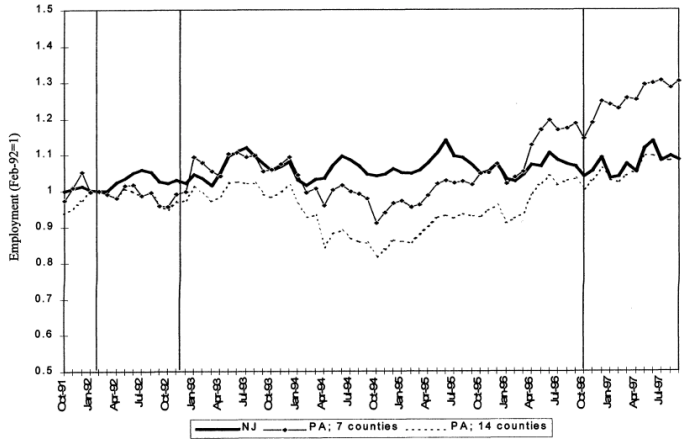


FIGURE 2. EMPLOYMENT IN NEW JERSEY AND PENNSYLVANIA FAST-FOOD RESTAURANTS, OCTOBER 1991 TO SEPTEMBER 1997

(Card and Krueger (2000) for an extended time period)

Problems with 2x2 designs (2)

1. Effectively 4 observations: can't separate τ from other shocks in NJ relative to PA
2. Can't falsify parallel trends

Solution: have more (pre-)periods and more units

	$i = A$	$i = B$	$i = C$	$i = D$
$t = 1$				
$t = 2$				
$t = 3$				
$t = 4$				
$t = E = 5$				
$t = 6$				

Outline

Causal structure with multiple periods

We continue to write $Y_{it} = Y_{it}(D_{it})$ but have in mind three distinct scenarios:

1. Policy switches on at $t = 5$ and stays on; has no lagged effects
2. Same but with lagged effects: $Y_{i6} = Y_{i6}(D_{i5}, D_{i6})$
   - But in the data, $D_{i5} \equiv D_{i6}$
   - $\Rightarrow$ Can't separate contemporaneous effects of D_{i6} from delayed effects of D_{i5}
   - $\Rightarrow$ Simplify notation to $Y_{i6}(D_{i6})$, with $Y_{i6}(1) - Y_{i6}(0) \equiv Y_{i6}(1,1) - Y_{i6}(0,0)$
3. A single event happens at $t = 5$ but can have lagged effects: $Y_{i6} = Y_{i6}(D_{i5})$
   - Since $D_{i5} \equiv D_{i6}$, again simplify notation to $Y_{i6}(D_{i6})$
   - D_{it} = having ever been exposed to the event

PTA with multiple units and periods

- PTA at the unit level (in a fixed sample):

$$\begin{aligned} &\mathbb{E}\left[Y_{it}(0) - Y_{i,t-1}(0)\right] = \mathbb{E}\left[Y_{jt}(0) - Y_{j,t-1}(0)\right] \qquad \forall i, j, t \\ \iff &\mathbb{E}\left[Y_{it}(0)\right] = \alpha_i + \beta_t \qquad \forall i, t \end{aligned}$$

- Or impose PTA at the treatment/control group level (in a random sample):

$$\mathbb{E}\left[Y_{it}(0) - Y_{i,t-1}(0) \mid G_i = 1\right] = \mathbb{E}\left[Y_{it}(0) - Y_{i,t-1}(0) \mid G_i = 0\right], \qquad \forall t$$

where $G_i = \mathbf{1}\left[\text{treatment group}_i\right]$

Static TWFE specification

Assume a balanced panel. **Static TWFE** specification:

$$Y_{it} = \tilde{\alpha}_i + \tilde{\beta}_t + \tau D_{it} + \varepsilon_{it}, \qquad D_{it} = G_i \times \underbrace{Post_t}_{=\mathbf{1}[t \geq E]}$$

- $\hat{\tau} = (\bar{Y}_{\text{treated,post}} - \bar{Y}_{\text{treated,pre}}) - (\bar{Y}_{\text{control,post}} - \bar{Y}_{\text{control,pre}})$
- Under PTA,

$$\mathbb{E}[\hat{\tau}] = \left(\bar{\alpha}_{\text{treated}} + \bar{\beta}_{\text{post}} + ATT - \bar{\alpha}_{\text{treated}} - \bar{\beta}_{\text{pre}}\right) - \left(\bar{\alpha}_{\text{control}} + \bar{\beta}_{\text{post}} - \bar{\alpha}_{\text{controls}} - \bar{\beta}_{\text{pre}}\right) = ATT$$

  which is the avg τ_{it} across all treated units and "post" periods

- This result doesn't require constant or static effects but it is not very useful with strong dynamics

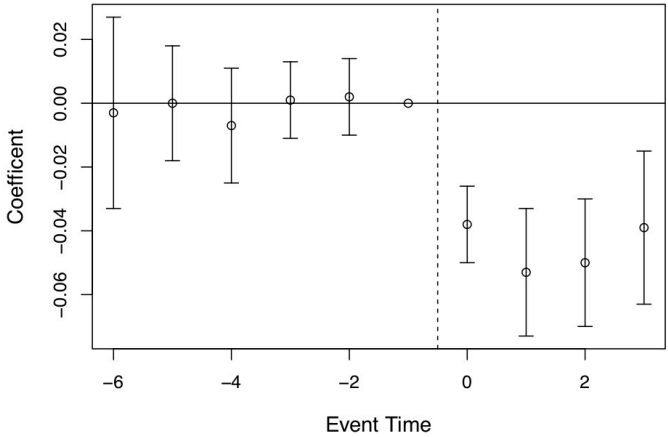
Dynamic event study specification

"**Event study**" dynamic specification:

$$Y_{it} = \tilde{\alpha}_i + \tilde{\beta}_t + \sum_{h=-(E-1)}^{T-E} \tau_h \mathbf{1}\left[t - E = h\right] \times G_i + \varepsilon_{it}, \qquad \tau_{-1} = 0$$

- $\hat{\tau}_h = (\bar{Y}_{\text{treated},E+h} - \bar{Y}_{\text{treated},E-1}) - (\bar{Y}_{\text{control},E+h} - \bar{Y}_{\text{control},E-1})$
- For $h \geq 0$, unbiased for ATT h periods after the event
- For $h < -1$, unbiased for zero under PTA: yields a "pre-trends" test

Example event study plot



(Mixtape Fig. 9.6 based on Miller et al. (2021, Fig. I.E))

Outline

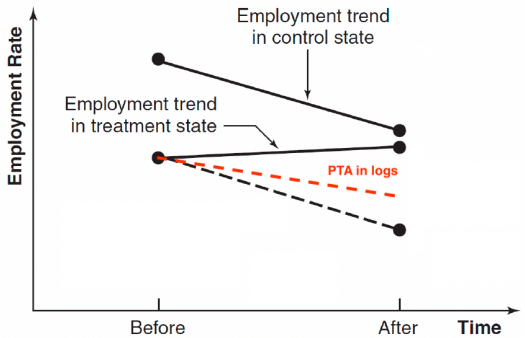
#1: Pre-trend tests are at best suggestive

- We imposed parallel trends in $Y_{it}(0)$ both pre- and post-treatment
  - Parallel trends after $t = E - 1$ are key and untestable
  - Parallel trends before $t = E - 1$ are not necessary but make pre-trend tests useful
- Never say:
  - "My identification assumption is parallel pre-trends"
  - "The pre-trend test passes, which implies that my diff-in-diff strategy is solid"

#2: Justifying PTA *ex ante*

- **Design-based approach**: assignment into treatment vs control group is random
  - Balance on levels in addition to trends $\Longrightarrow$ have better tools than DiD! (Roth and Sant'Anna (2023a))
  - No need for DiD in RCTs!
- **Model-based approach**: stability of the environment, i.e. unobserved factors
  - No balance on levels, only on trends
  - Better with large shocks and in the short-run
  - Levels vs logs matters (Roth and Sant'Anna (2023b))

PTA in levels vs. logs



Justifying PTA is important

Common bad practice: don't justify PTA, just check for pre-trends

From David McKenzie's 2022 blog post (link):

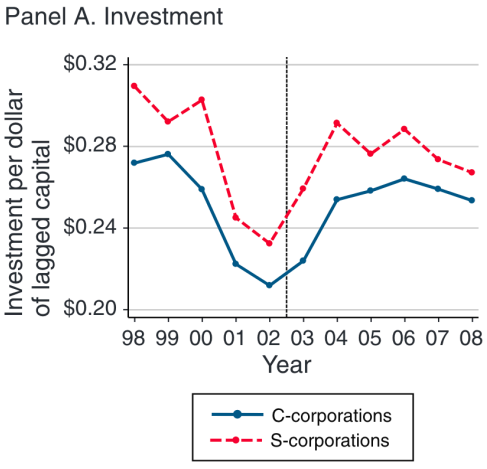
- [Three recent QJE] papers all use a variety of statistical methods to assess and assert the validity of their DiD estimates [...] But I found all three lacking in providing a discussion of why parallel trends should hold

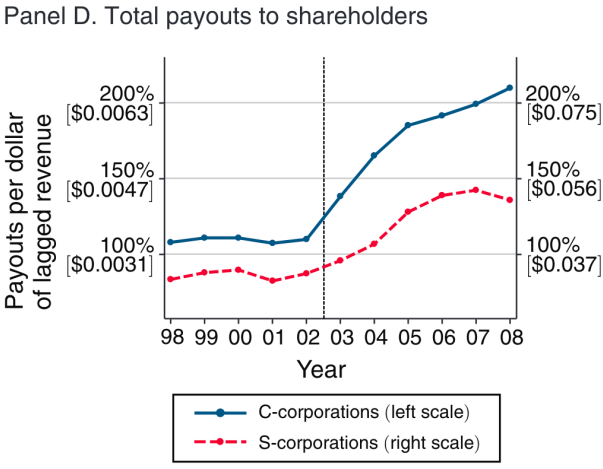
- It is commonplace in papers that use IV to spend a lot of time arguing for the plausibility of the ultimately untestable exclusion restriction, but I think DiD papers have not done the same for the untestable assumption about future trends

Example: Yagan (2015)

- Studies the effects of the 2003 dividend tax cut from 38.6% to 15%
- Treatment group: C-corporations; control group: S-corporations
- S-corporations don't pay dividend tax; otherwise similar taxation
- *"The identifying assumption is not random assignment of C- versus S-status; it is that C- and S-corporation outcomes would have trended similarly in the absence of the tax cut"*
- Justifications:
  1. *"C- and S-corporations of the same ages operate in the same narrow industries and at the same scale"*
  2. *"Contemporaneous stimulative tax provisions applied almost identically"*
  3. *"Key outcomes empirically trended similarly for C- and S-corporations before 2003"*

Yagan (2015): Plotting raw data





What would you ask the author?

#3: Pre-testing problems (Roth, 2022)

Estimators of causal effects and pre-trends are correlated, e.g. by using the same reference period

$$(\bar{Y}_{\text{treated},E+h} - \bar{Y}_{\text{control},E+h}) - (\bar{Y}_{\text{treated},E-1} - \bar{Y}_{\text{control},E-1})$$
$$\text{vs.} \quad (\bar{Y}_{\text{treated},E-\ell} - \bar{Y}_{\text{control},E-\ell}) - (\bar{Y}_{\text{treated},E-1} - \bar{Y}_{\text{control},E-1})$$

- Suppose you only report the results if some pre-trend test doesn't reject ("pre-testing")
- If PTA holds, you are distorting inference for causal effects
- If PTA doesn't hold, you are changing the bias; under some conditions increase it

#4: Event studies conflate estimation and testing

Borusyak et al. (2024):

- If PTA holds, why not use all pre-periods for more efficient estimation?

$$Y_{it} = \tilde{\alpha}_i + \tilde{\beta}_t + \sum_{h \geq 0} \tau_h \mathbf{1}\left[t - E = h\right] \times G_i + \varepsilon_{it}$$

  with all $t < E$ as reference periods ($\hat{\tau}_h = (\bar{Y}_{\text{tr},E+h} - \bar{Y}_{\text{tr,pre}}) - (\bar{Y}_{\text{co},E+h} - \bar{Y}_{\text{co,pre}})$)

- Testing PTA is a separate exercise: using untreated data only,

$$Y_{it} = \tilde{\alpha}_i + \tilde{\beta}_t + \sum_{h=-L}^{-1} \tau_h \mathbf{1}\left[t - E = h\right] \times G_i + \varepsilon_{it}$$

  with $t < E - L$ reference periods ($\hat{\tau}_h = (\bar{Y}_{\text{tr},E+h} - \bar{Y}_{\text{tr},t<E-L}) - (\bar{Y}_{\text{co},E+h} - \bar{Y}_{\text{co},t<E-L})$)

#5: Pre-trend tests are not always conceptually informative

Mann and Pozzoli (2024) study the effects of low-skill immigration into Danish municipalities on industrial robot adoption by firms in 1995-2019

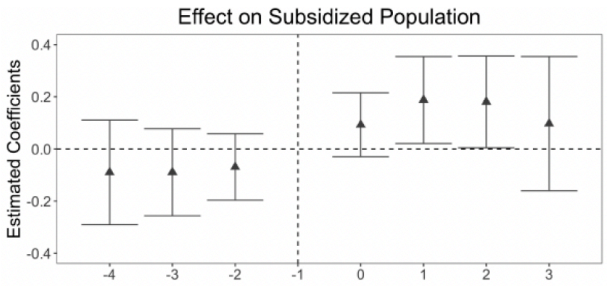
- How informative are their regressions of pre-period (1993-95) changes in robot adoption and other variables on the 1995-2019 change in low-skill immigration?

Table 3
Pre-sample trends and long-run changes in robotization.

	Δ robot users (broad) (1993–1995) (1)	Δ imports (1993–1995) (2)	Δ exports (1993–1995) (3)
Δ Non-West Img IV (1995–2019)	0.476 (0.583)	0.300 (0.262)	−0.272 (0.482)
N	94	94	94
R-sq	0.355	0.166	0.067

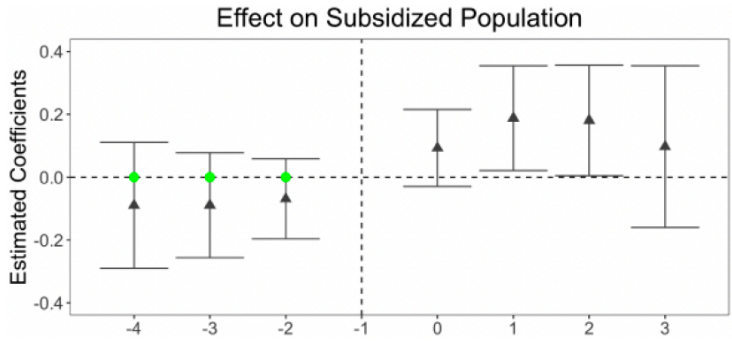
#6: Pre-trend tests are not always powerful (Roth, 2022)

E.g. He and Wang (2017) study the effects of college-graduated bureaucrats placed to Chinese villages on a measure of poverty



(From Jonathan Roth's Mixtape Session slides)

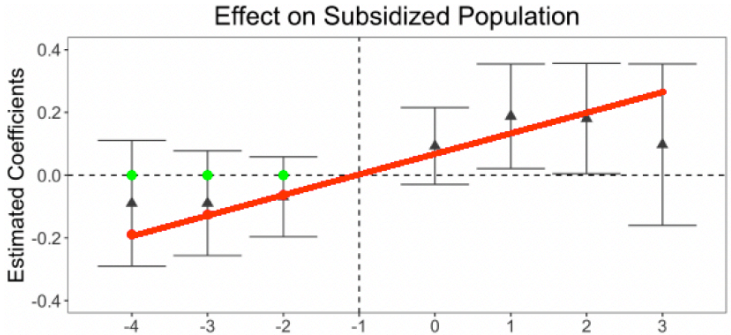
#6: Pre-trend tests are now always powerful (Roth, 2022)



(From Jonathan Roth's Mixtape Session slides)

- *"The estimated coefficients on the leads of treatment ... are statistically indifferent from 0. ... We conclude that the pretreatment trends in the outcomes in both groups of villages are similar, and villages without [treatment] can serve as a suitable control group for villages with [treatment] in the treatment period."*

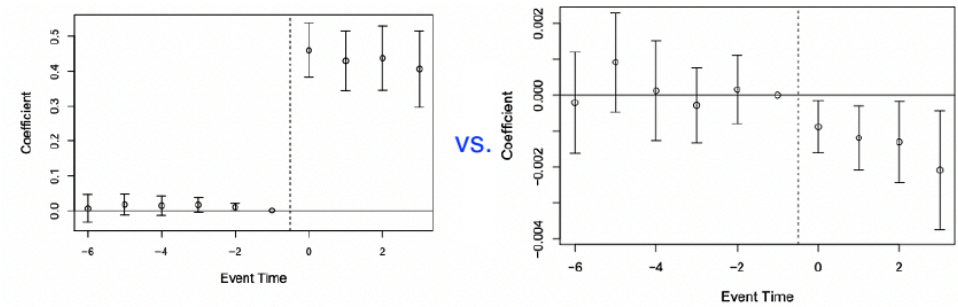
#6: Pre-trend tests are now always powerful (Roth, 2022)



(From Jonathan Roth's Mixtape Session slides)

The "Straight line test"

- The more nonlinearity is needed to kill the effects, the more robust the results are



- See Stata and R package *pretrends* accompanying Roth (2022)
- Having many pre-periods is also important

References I

Borusyak, K., X. Jaravel, and J. Spiess (2024): "Revisiting Event Study Designs: Robust and Efficient Estimation," *Review of Economic Studies*, 91, 3253–3285.

Card, D. and A. Krueger (1994): "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersry and Pennsylvania," *The American Economic Review*, 84, 772–793.

——— (2000): "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania: Reply," *American Economic Review*, 90, 1397–1420.

Goldsmith-Pinkham, P. (2024): "Tracking the Credibility Revolution across Fields," *Working Paper*.

He, G. and S. Wang (2017): "Do College Graduates Serving as Village Officials Help Rural China?" *American Economic Journal: Applied Economics*, 9, 186–215.

Mann, K. and D. Pozzoli (2024): "Robots and immigration," *Journal of Economic Behavior & Organization*, 227, 106708.

Miller, S., N. Johnson, and L. R. Wherry (2021): "Medicaid and Mortality: New Evidence From Linked Survey and Administrative Data," *The Quarterly Journal of Economics*, 136, 1783–1829.

References II

Roth, J. (2022): "Pre-test with Caution: Event-study Estimates After Testing for Parallel Trends," *American Economic Review: Insights*, 4, 305–322.

Roth, J. and P. H. Sant'Anna (2023a): "Efficient Estimation for Staggered Rollout Designs," *JPE: Micro*, arXiv: 2102.01291.

——— (2023b): "When Is Parallel Trends Sensitive to Functional Form?" *Econometrica*, 91.

Yagan, D. (2015): "Capital Tax Reform and the Real Economy: The Effects of the 2003 Dividend Tax Cut," *American Economic Review*, 105, 3531–3563.