Difference-in-Differences

Mixtape Session

Roadmap

- Stacking
  - Clean controls
  - Identification and aggregation
  - Heterogeneity analysis

Keep TWFE but avoid differential timing problems

- Problem with TWFE occurred when we used already-treated as controls
- CS and SA used the never-treated (CS and SA), not-yet-treated (CS) or last-treated (SA) as controls
- Each identified group-time ATT ("cohort-specific ATT"), then via weighting based on group shares, aggregate treatment parameters (e.g., ATT)
- Could we still use TWFE, but avoid the use of already-treated as controls? How? Interpretation?
- Cengiz, et al. (2019) call this "clean controls" but now it's called "stacking"

Stacking minimum wages

- Discuss the Cengiz, et al. (2019) as an example of stacking, even though it is a very small part of the paper
- It's a good paper at illustrating an argument built up through tables, figures and various forms of intuitive reasoning that guides estimation
- Data is 1979 to 2016 US state-level panel, 138 "prominent" state-level minimum wage change events
- Main model specification is TWFE, but in a robustness section (Appendix D) they introduce a stacking alternative

Background

- Theory: minimum wages should reduce employment in perfectly competitive labor markets
- Theory: minimum wages will increase employment in monopsony labor markets
- Prior research: “new minimum wage” studies starting with Card and Krueger (1995) found no effect of minimum wages on employment, but others (often authored by David Neumark) found reductions
- Focus had often been on aggregate employment or teens
- Controversial, contentious and unsettled

Data

- Hourly wage data from 1976-2016 NBER out-rotation group of the Current Population Survey broken into wage bins (by $0.25) from $0 to $30

  > *"We use the individual-level NBER Merged Outgoing Rotation Group of the CPS for 1979-2016 to calculate quarterly, state-level distributions of hourly wages. For hourly workers, we use the reported hourly wage, and for other workers, we define the hourly wage to be their usual weekly earnings divided by usual weekly hours. We do not use any observations with imputed wage data to minimize the role of measurement error."*

- Wages are deflated to 2016 dollars to get "real wage"
- There are 117 wage bins that are then collapsed into quarterly, state-level employment counts E_{swt} using person-level sampling weights

Minimum wage data

- Quarterly max of the state-level daily minimum wage series from Vaghul and Zipperer (2016)
- 138 minimum wage events, of which 8.6% of workers were below the minimum wage the year before the event
- Focus will be on “missing” vs “excess” jobs which is a break in the wage frequency just above and just below the minimum wage cutoff

Bite

Constant reference to “bite”. Footnote 2.

> *“When we refer to the ‘bite’ of the minimum wage, or to the extent to which the minimum wage is ‘binding’, we mean how effective the minimum wage is in raising wages at the bottom. Therefore, the bite is a function of (i) how many workers are earning below the minimum wage, (ii) how many of those workers are legally covered by the policy, and (iii) the extent of compliance.”*

You can sometimes hear it referred to as the first stage, using IV language, but as this is a DiD and not IV they tend to use the word “bite” over and over instead.

The idea is that this paper only makes sense if the policy change has “bite”

Rhetorical argument

- Key visual showing excess vs missing jobs helps communicate the idea of “bite” which supports all subsequent analysis
- If no bite, no effects shown can be plausibly causal – so notice, the logic of identification comes from convincingly showing bite, not from the model results themselves
- Carefully construct data into bins so that narrowly employment above and below the minimum wage can be measured
- Do the same for employment per population
- Calculate net changes in employment both for total (main results) and by various slices (heterogenous effects)

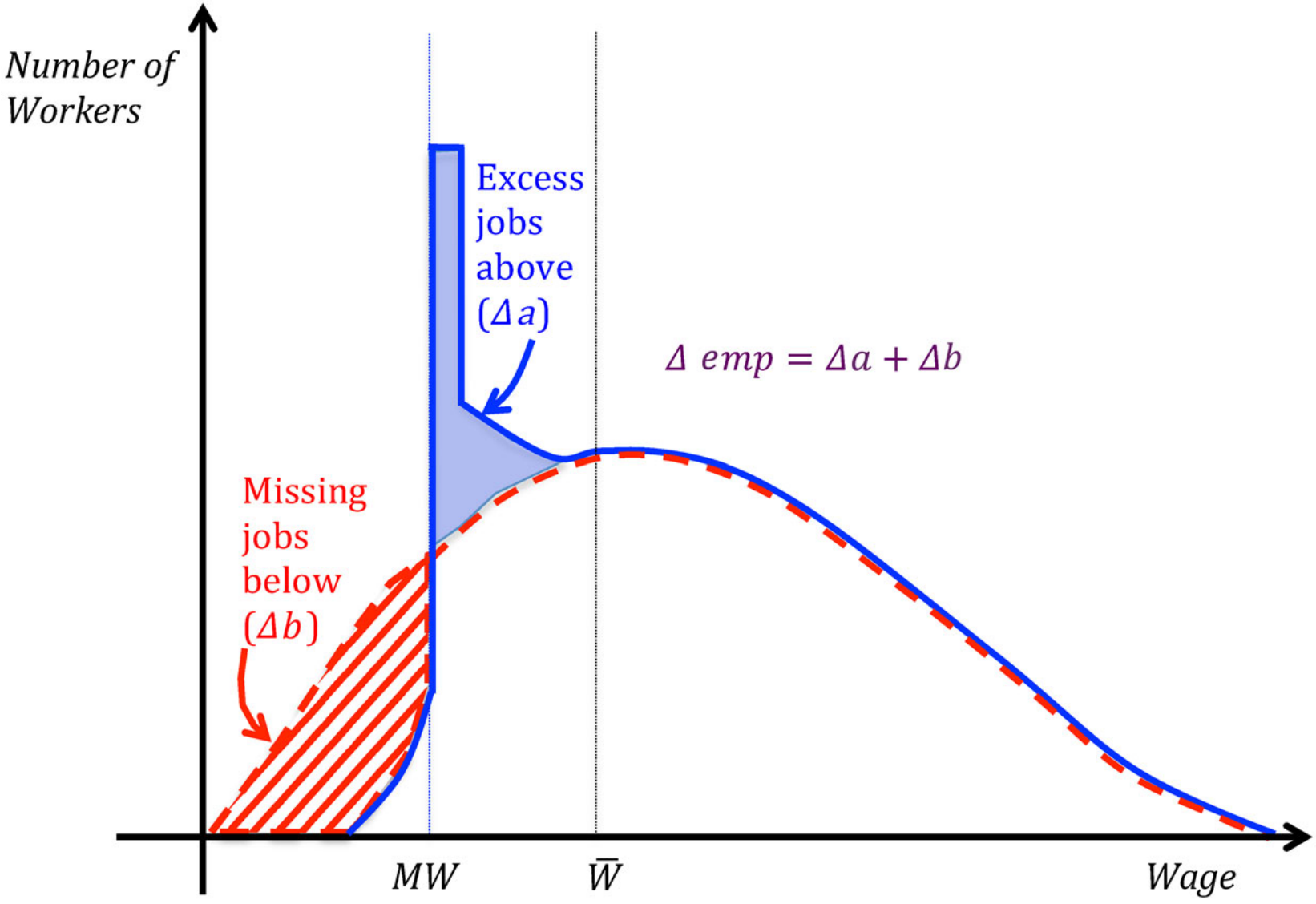


FIGURE I

The Impact of Minimum Wages on the Frequency Distribution of Wages

Null results

- Results we will now see is essentially a lot of zeroes
- Authors fail to find any evidence of an effect on employment other than the “bite” result which I’ll show
- Empirical equation is estimated using TWFE; differential timing problems, τ are event years and k are dollar bins, μ and ρ are state-by-wage-bin and period-by-wage-bin fixed effects, ω are controls for small or federal increases

$$E/N = \sum_{\tau=-3}^{4} \sum_{k=-4}^{17} \alpha_{\tau k} I_{sjt}^{\tau k} + \mu_{sj} + \rho_{jt} + \omega_{sjt} + u_{sjt} \tag{1}$$

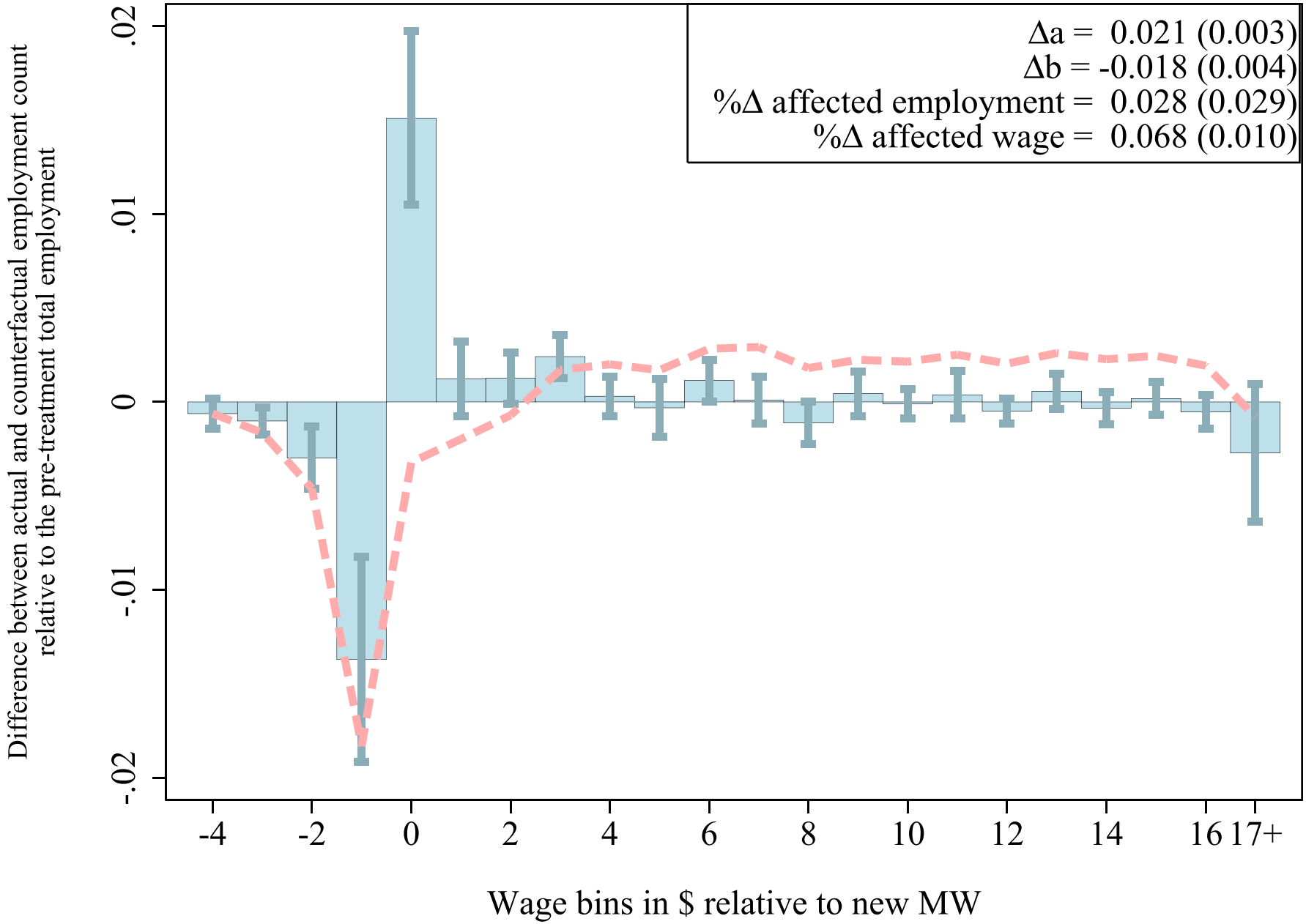


FIGURE II

Impact of Minimum Wages on the Wage Distribution

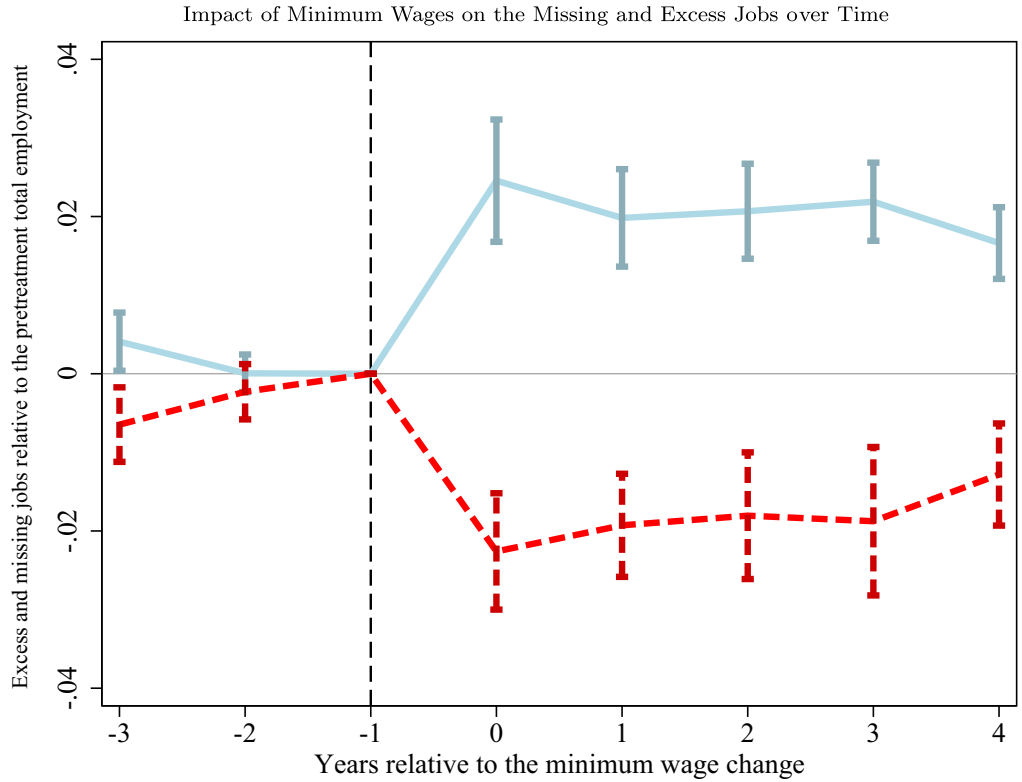

Impact of Minimum Wages on the Missing and Excess Jobs over Time
Excess and missing jobs relative to the pretreatment total employment
.04
.02
0
-.02
-.04
-3
-2
-1
0
1
2
3
4
Years relative to the minimum wage change

FIGURE III

TABLE I

IMPACT OF MINIMUM WAGES ON EMPLOYMENT AND WAGES

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Missing jobs below new MW (Δb)	−0.018***	−0.018***	−0.018***	−0.016***	−0.016***	−0.015***	
	(0.004)	(0.004)	(0.004)	(0.002)	(0.002)	(0.002)	
Excess jobs above new MW (Δa)	0.021***	0.018***	0.020***	0.016***	0.014***	0.015***	
	(0.003)	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)	
%Δ affected wages	0.068***	0.057***	0.068***	0.049***	0.043***	0.050***	0.065***
	(0.010)	(0.010)	(0.012)	(0.010)	(0.010)	(0.011)	(0.010)
%Δ affected employment	0.028	0.000	0.022	−0.002	−0.019	−0.000	0.027
	(0.029)	(0.023)	(0.021)	(0.021)	(0.021)	(0.023)	(0.028)
Employment elasticity w.r.t. MW	0.024	0.000	0.019	−0.001	−0.016	−0.000	0.023
	(0.025)	(0.020)	(0.018)	(0.018)	(0.018)	(0.019)	(0.024)
Emp. elasticity w.r.t. affected wage	0.411	0.006	0.326	−0.032	−0.449	−0.003	0.410
	(0.430)	(0.402)	(0.313)	(0.439)	(0.574)	(0.455)	(0.421)
Jobs below new MW ($\overline{b}_{-1}$)	0.086	0.086	0.086	0.086	0.086	0.086	0.086
%Δ MW	0.101	0.101	0.101	0.101	0.101	0.101	0.101
Number of events	138	138	138	138	138	138	138
Number of observations	847,314	847,314	847,314	847,314	847,314	847,314	14,484
Number of workers in the sample	4,694,104	4,694,104	4,694,104	4,694,104	4,694,104	4,694,104	4,694,104

TABLE II

IMPACT OF MINIMUM WAGES ON EMPLOYMENT AND WAGES BY DEMOGRAPHIC GROUPS

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Missing jobs below new MW (Δb)	-0.065^{***}	-0.032^{***}	-0.114^{***}	-0.023^{***}	-0.028^{***}	-0.094^{***}	-0.020^{***}	-0.004^{***}
	(0.010)	(0.007)	(0.010)	(0.005)	(0.008)	(0.010)	(0.005)	(0.001)
Excess jobs above new MW (Δa)	0.075^{***}	0.038^{***}	0.127^{***}	0.026^{***}	0.028^{***}	0.100^{***}	0.021^{***}	0.004^{***}
	(0.011)	(0.006)	(0.020)	(0.004)	(0.006)	(0.012)	(0.003)	(0.001)
%Δ affected wages	0.080^{***}	0.076^{***}	0.083^{***}	0.072^{***}	0.044^{***}	0.073^{***}	0.051^{***}	0.060^{*}
	(0.014)	(0.014)	(0.018)	(0.011)	(0.012)	(0.011)	(0.013)	(0.032)
%Δ affected employment	0.038	0.043	0.030	0.025	-0.004	0.015	0.015	0.011
	(0.024)	(0.030)	(0.032)	(0.027)	(0.044)	(0.018)	(0.048)	(0.055)
Employment elasticity w.r.t. MW	0.097	0.061	0.125	0.025	-0.005	0.052	0.016	0.003
	(0.061)	(0.042)	(0.134)	(0.027)	(0.058)	(0.062)	(0.049)	(0.014)
Emp. elasticity w.r.t. affected wage	0.475^{*}	0.570	0.356	0.343	-0.086	0.206	0.304	0.184
	(0.268)	(0.386)	(0.317)	(0.362)	(1.005)	(0.233)	(0.904)	(0.841)
Jobs below new MW ($\overline{b}_{-1}$)	0.264	0.145	0.432	0.102	0.133	0.358	0.104	0.027
%Δ MW	0.103	0.103	0.102	0.101	0.100	0.103	0.103	0.103

TABLE II

CONTINUED

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Number of events	138	138	138	138	138	138	138	138
Number of observations	847,314	847,314	847,314	847,314	846,729	847,314	847,314	847,314
Number of workers in the sample	660,771	2,248,711	287,484	2,277,624	781,003	469,226	1,830,393	2,349,485
Sample	Less than high school	High school or less	Teen	Women	Black or Hispanic	High probability	Medium probability	Low probability

Notes. The table reports effects of a minimum wage increase by demographic groups based on the event study analysis (see equation (1)) exploiting 138 state-level minimum wage changes between 1979 and 2016. The table reports five-year averaged post-treatment estimates on missing jobs up to $4 below the new minimum wage, excess jobs at and up to $5 above it, employment, and wages for individuals without a high school degree (column (1)), for individuals with high school degree or less schooling (column (2)), for teens (column (3)), for women (column (4)), for black or Hispanic workers (column (5)). Columns (6)–(8) report the results for groups of workers with differential probability of being exposed to the minimum wage changes. We use the Card and Krueger (1995) demographic predictors to estimate the probability of being exposed (see the text for details). Column 6 shows the results for the workers who have a high probability of being exposed to the minimum wage increase, column (7) for the middle-probability group, and column (8) for the low-probability group. All specifications include wage bin-by-state and wage bin-by-period fixed effects. Regressions are weighted by state-quarter aggregated population of the demographic groups. Robust standard errors in parentheses are clustered by state; significance levels are $^{*}0.10$, $^{**}0.5$, $^{***}0.01$.

The first two rows report the change in number of missing jobs below the new minimum wage (Δb), and excess jobs above the new minimum wage (Δa) relative to the pretreatment total employment. The third row, the percentage change in average wages in the affected bins, ($\%\Delta$W), is calculated using equation (2) in Section 2.2. The fourth row, percentage change in employment in the affected bins, is calculated by dividing change in employment by jobs below the new minimum wage ($\frac{\Delta a+\Delta b}{b_{-1}}$). The fifth row, employment elasticity with respect to the minimum wage, is calculated as $\frac{\Delta a+\Delta b}{\%\Delta MW}$, whereas the sixth row, employment elasticity with respect to the wage, reports $\frac{1}{\%\Delta W}\frac{\Delta a+\Delta b}{b_{-1}}$. The line on the number of observations shows the number of quarter-bin cells used for estimation, while the number of workers refers to the underlying CPS sample used to calculate job counts in these cells.

TABLE III

IMPACT OF MINIMUM WAGES ON EMPLOYMENT AND WAGES BY SECTORS (1992–2016)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Missing jobs below new MW (Δb)	−0.019***	−0.016*	−0.066***	−0.003	−0.011***	−0.101***	−0.033***	−0.017**
	(0.004)	(0.008)	(0.007)	(0.002)	(0.003)	(0.015)	(0.003)	(0.008)
Excess jobs above new MW (Δa)	0.020***	0.011	0.072***	0.005	0.011***	0.101***	0.041***	0.011
	(0.003)	(0.008)	(0.011)	(0.006)	(0.002)	(0.015)	(0.010)	(0.009)
%Δ affected wages	0.058***	0.058	0.056***	0.097	0.056***	0.049***	0.060***	0.073
	(0.011)	(0.073)	(0.014)	(0.086)	(0.013)	(0.012)	(0.021)	(0.078)
%Δ affected employment	0.008	−0.111	0.022	0.051	0.009	−0.001	0.062	−0.101
	(0.031)	(0.136)	(0.037)	(0.163)	(0.044)	(0.026)	(0.080)	(0.145)
Employment elasticity w.r.t. MW	0.007	−0.056	0.060	0.019	0.005	−0.002	0.086	−0.052
	(0.027)	(0.069)	(0.103)	(0.059)	(0.026)	(0.117)	(0.111)	(0.074)
Emp. elasticity w.r.t. affected wage	0.140	−1.910	0.387	0.530	0.166	−0.011	1.040	−1.385
	(0.523)	(3.922)	(0.597)	(1.311)	(0.763)	(0.542)	(1.058)	(2.956)
Jobs below new MW ($\overline{b}_{-1}$)	0.087	0.050	0.270	0.036	0.057	0.434	0.136	0.050
%Δ MW	0.098	0.098	0.098	0.098	0.098	0.098	0.098	0.098

TABLE III

CONTINUED

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Number of events	118	118	118	118	118	118	118	118
Number of observations	554,931	554,931	554,931	554,931	554,931	554,931	554,931	554,931
Number of workers in the sample	2,652,792	358,086	384,498	274,812	1,504,643	156,634	315,397	349,749
Sector	Overall	Tradeable	Nontradeable	Construction	Other	Restaurants	Retail	Manufacturing

Notes. The table reports the effects of a minimum wage increase by industries based on the event study analysis (see equation (1)) exploiting 138 state-level minimum wage changes between 1992 and 2016. The table reports five-year averaged post-treatment estimates on missing jobs up to \$4 below the new minimum wage, excess jobs at and up to \$5 above it, employment, and wages for all sectors (column (1)), tradable sectors (column (2)), nontradeable sectors (column (3)), construction (column (4)), other sectors (column (5)), restaurants (column (6)), retail (column (7)), and manufacturing industries (column (8)). Our classification of tradable, nontradeable, construction, and other sectors follows Mian and Sufi (2014) (see Online Appendix D for the details). Regressions are weighted by state-quarter aggregated population. Robust standard errors in parentheses are clustered by state; significance levels are $^{*}0.10$, $^{**}0.05$, $^{***}0.01$.

The first two rows report the change in number of missing jobs below the new minimum wage (Δb), and excess jobs above the new minimum wage (Δa) relative to the pretreatment total employment. The third row, the percentage change in average wages in the affected bins, ($\%\Delta$W), is calculated using equation (2). The fourth row, percentage change in employment in the affected bins, is calculated by dividing change in employment by jobs below the new minimum wage ($\frac{\Delta a+\Delta b}{\bar{b}_{-1}}$). The fifth row, employment elasticity with respect to the minimum wage, is calculated as $\frac{\Delta a+\Delta b}{\%\Delta MW}$, whereas the sixth row, employment elasticity with respect to the wage, reports $\frac{1}{\%\Delta W}\frac{\Delta a+\Delta b}{\bar{b}_{-1}}$. The line on the number of observations shows the number of quarter-bin cells used for estimation, while the number of workers refers to the underlying CPS sample used to calculate job counts in these cells.

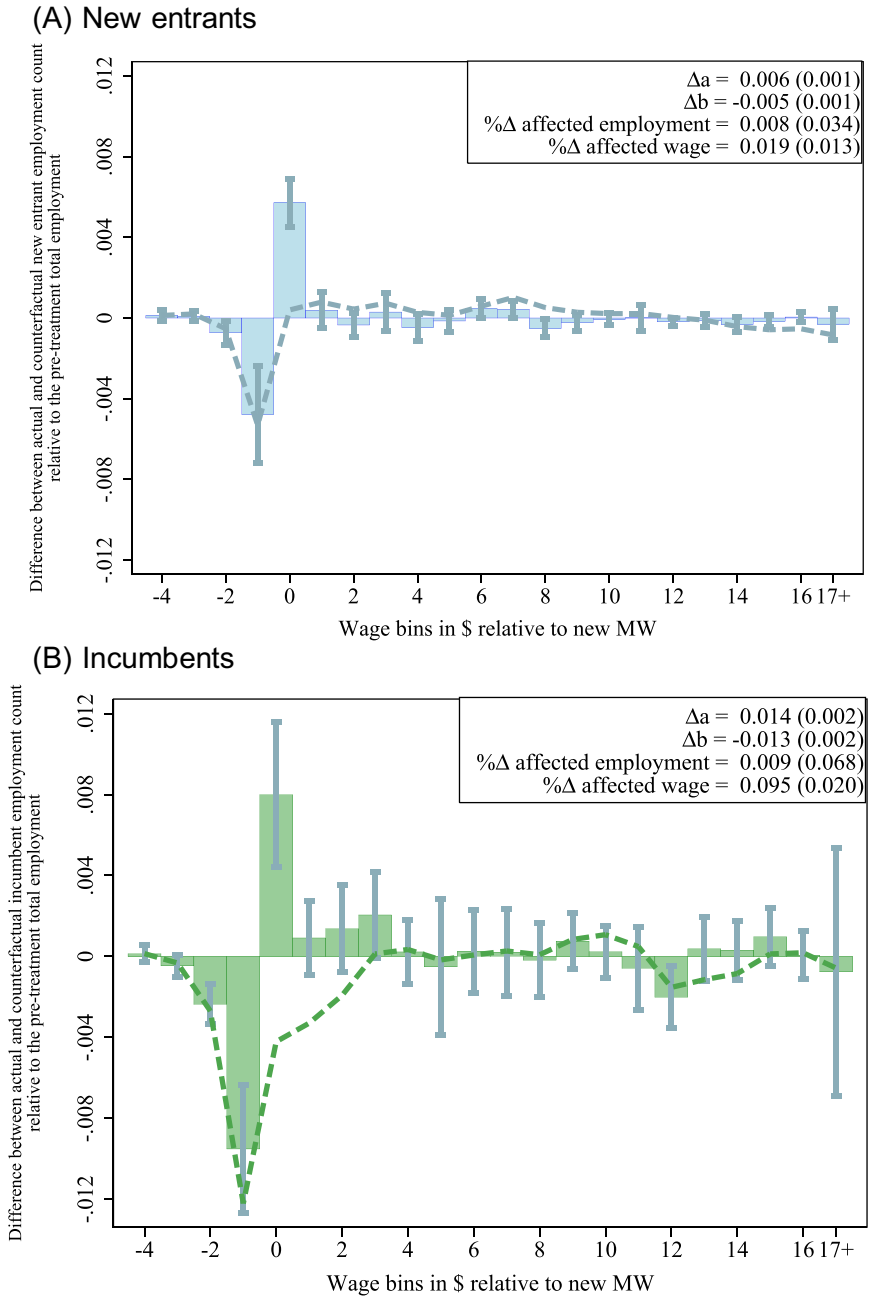


FIGURE IV

Impact of Minimum Wages on the Wage Distribution by Pretreatment Employment Status: New Entrants and Incumbents

Stacking alternative

- TWFE estimation is biased if there are differential timing and heterogenous treatment effects by cohort (SA 2020)
- They propose their own alternative which they call “clean controls” but which is more commonly called “stacking”
- Estimation models are TWFE; weights were unknown at time of writing (but are now known via Gardner 2021)

Dataset construction

Clean controls is done one of two ways:

1. Create 138 datasets, one for each event h where the treatment group is one state and the control are all other states that did not have a minimum wage increase in eight-year panels around event h, balanced in calendar time, inference must adjust for heteroskedasticity with only one treatment date (Ferman and Pinto Restat)
2. "Stack" the 138 datasets, re-centering each treatment date such that data is balanced in "event time" with 3 periods pre-treatment, 4 years post-treatment, and controls are all untreated units from -3 to +4. Several units will appear more than once).

Steps to stacked regression

1. Create separate "event by cohort specific" datasets for each policy cohort (e.g., groups who pass minimum wages in the same year)
   - → Dataset will consist of the relevant policy cohort **plus** controls
   - → Data is structured in "event time" and will be balanced such that panel length is h, perhaps starting point being 3 years prior to treatment and ending point 4 years after or something like that
   - → Each dataset will contain individuals untreated over the h period defined

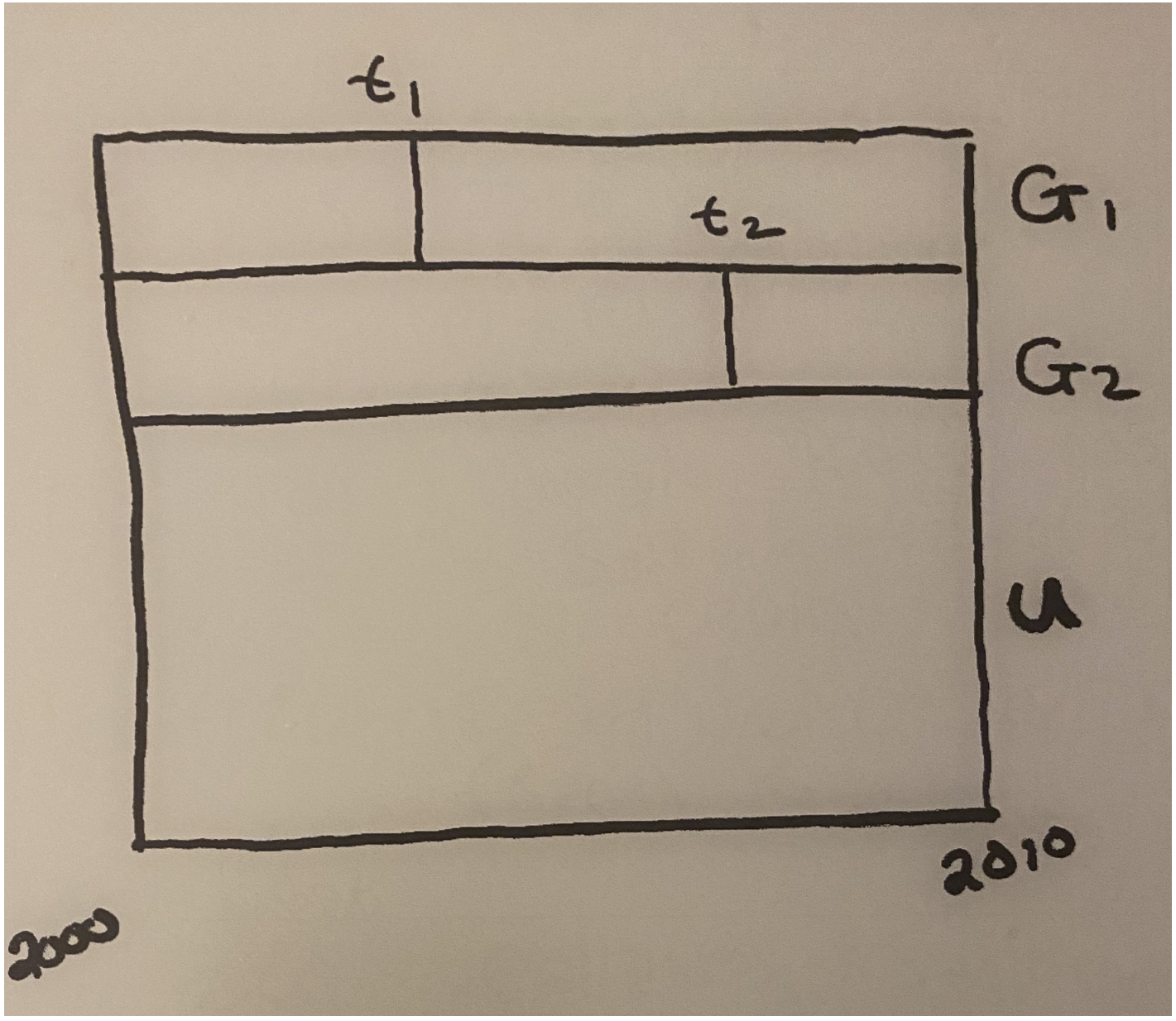
Steps to stacked regression

1. Append each dataset (or what people are now calling “stacking”) to one another
   - → This necessarily replicates control observations though as they are in each datasets
   - → Since the same people are often appearing many many times, you will correct for this in the regression model specification
2. Estimate a simple 2x2 model but include “cohort-by-state” fixed effects so as to account for the multiple appearances of observations from the never-treated control states

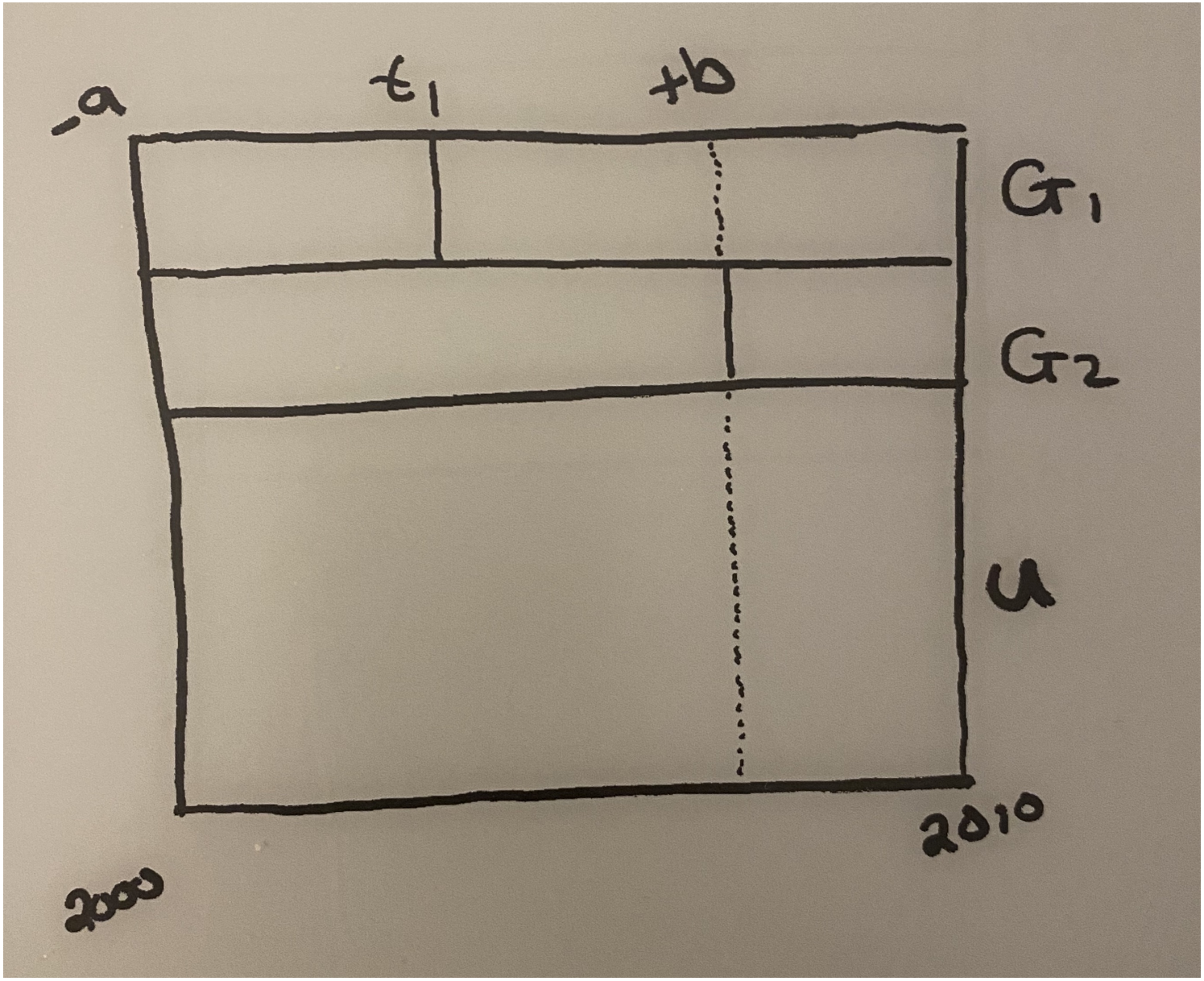
Comments on the approach

- Hard part of this is probably just the careful balancing, saving datasets, then appending, but ultimately not difficult – just watch yourself.
- Because the data is now balanced *in event time*, there is *no differential timing*; it is a simple 2x2
- Recall that the reason TWFE is biased in DiD designs is (1) differential timing and (2) heterogeneity
- Stacked eliminates (1) making (2) irrelevant
- But recall the lessons we learned about including time-varying covariates from Sant'Anna and Zhao (2020) – stacked will suffer from those too

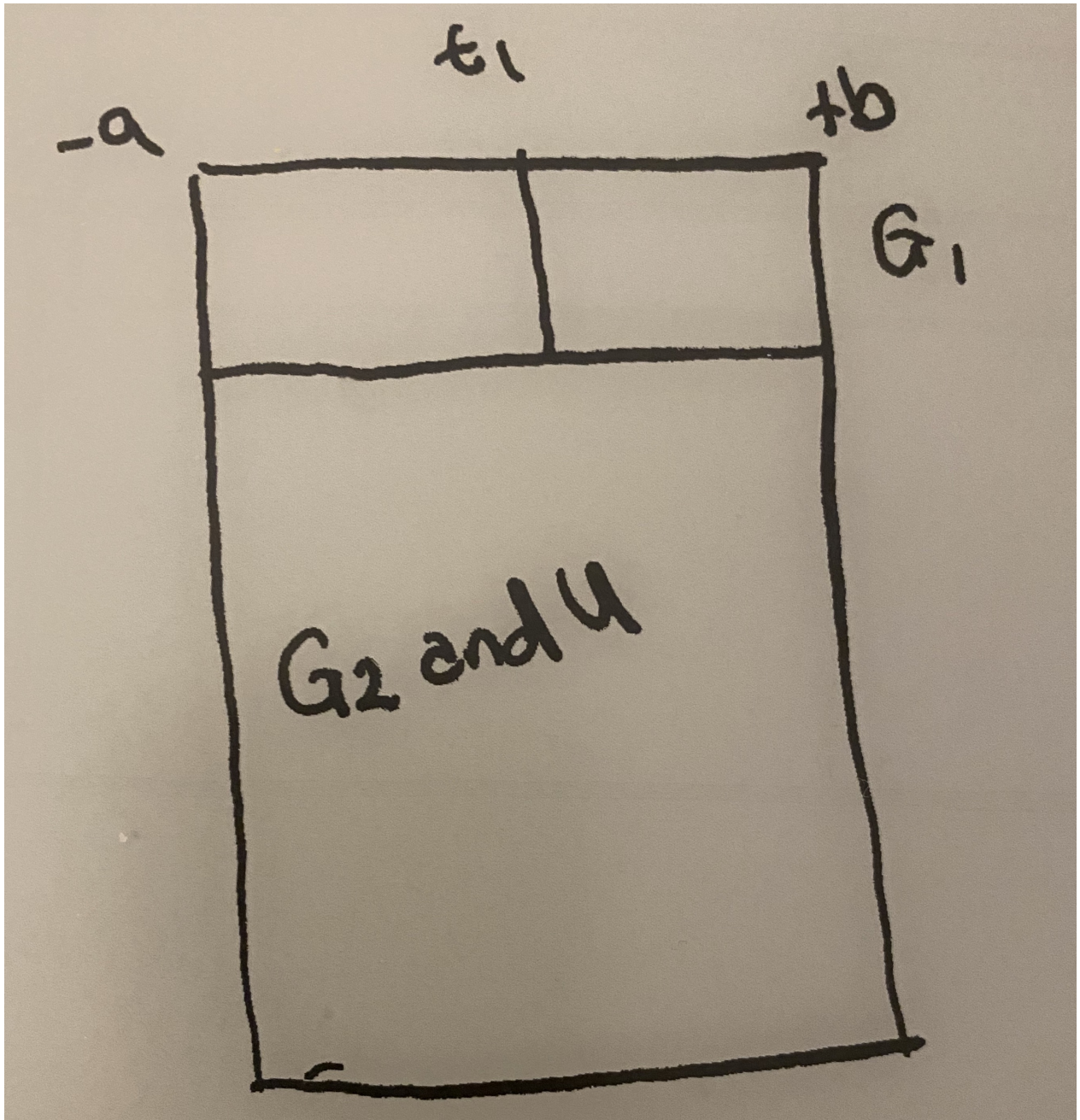
Imbalanced in relative time with differential timing



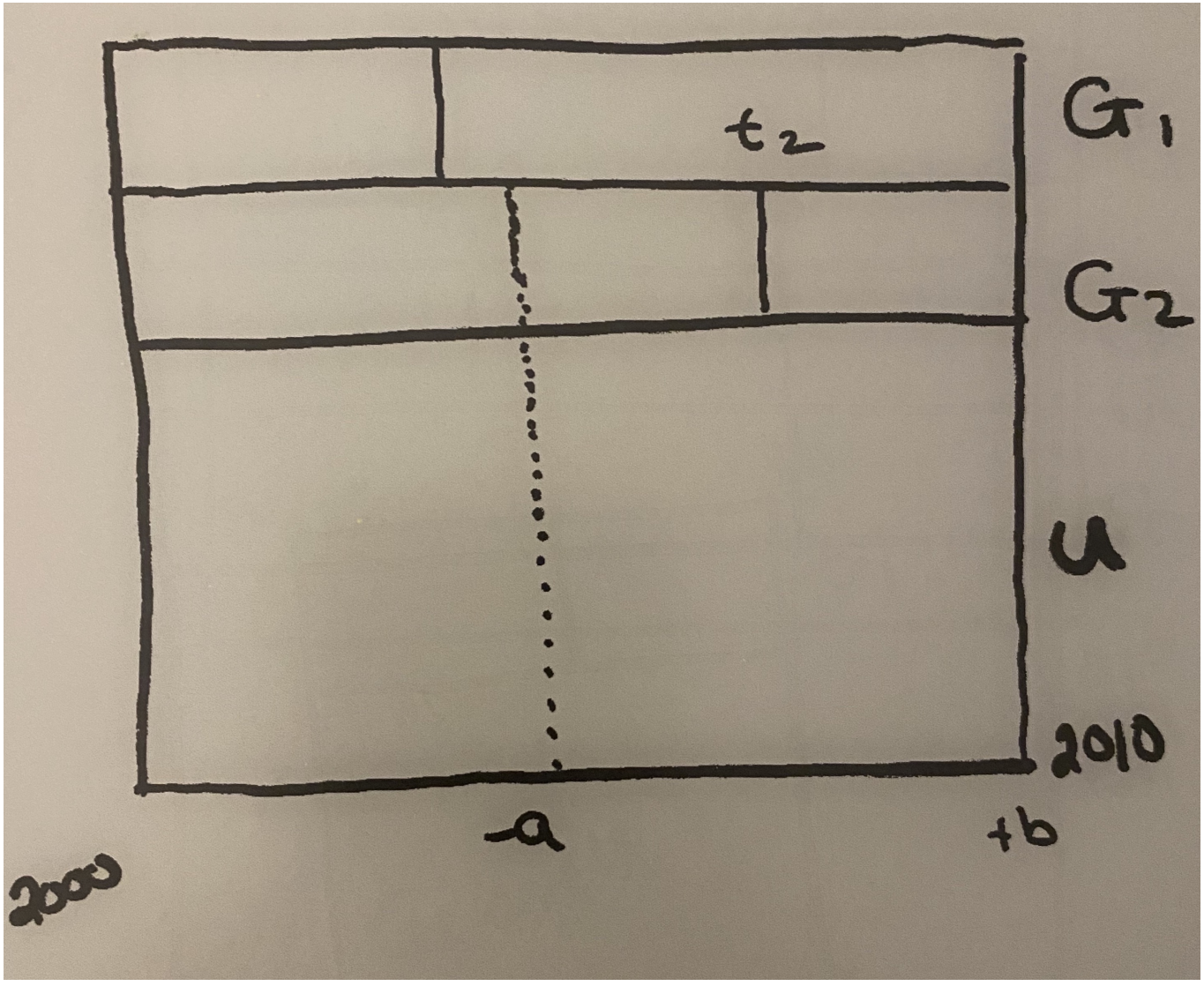
Creating G_1 dataset: choose max pre (-a) and post (+b) periods



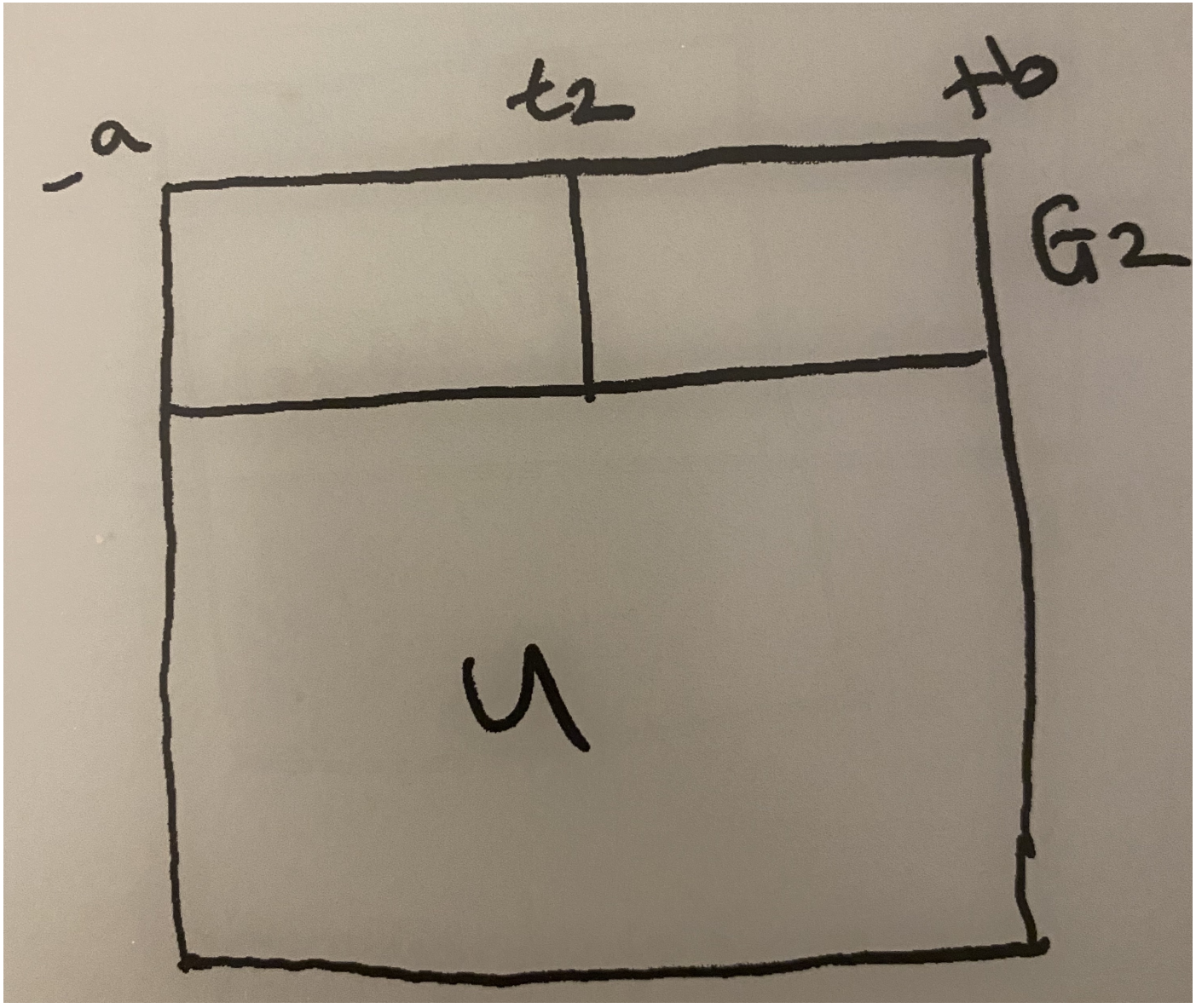
Creating G_1 dataset: keep untreated units on [-a,+b]



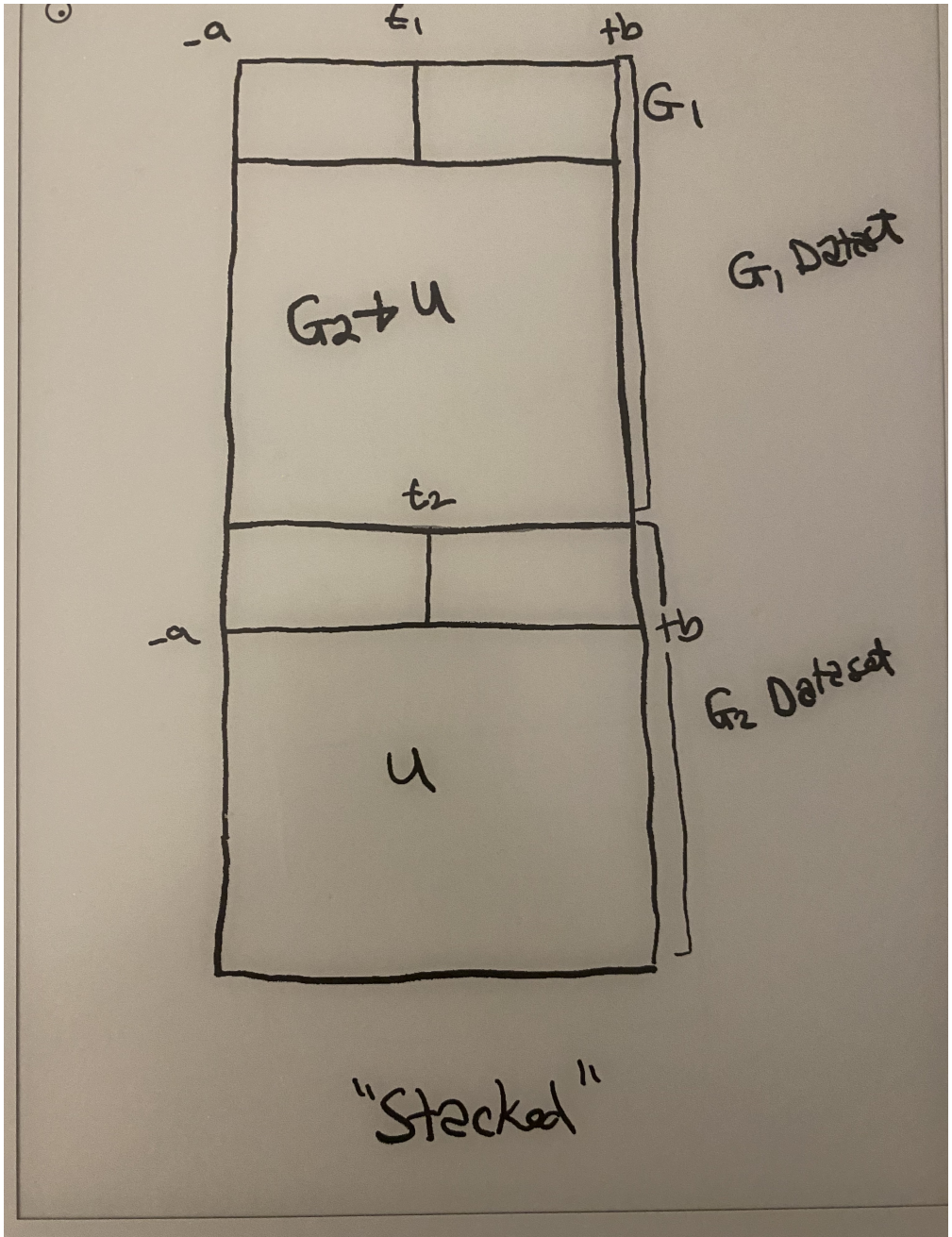
Creating G_2 dataset: keep *only* untreated units on [-a,+b] intervals as controls



Creating G_2 dataset: save the dataset



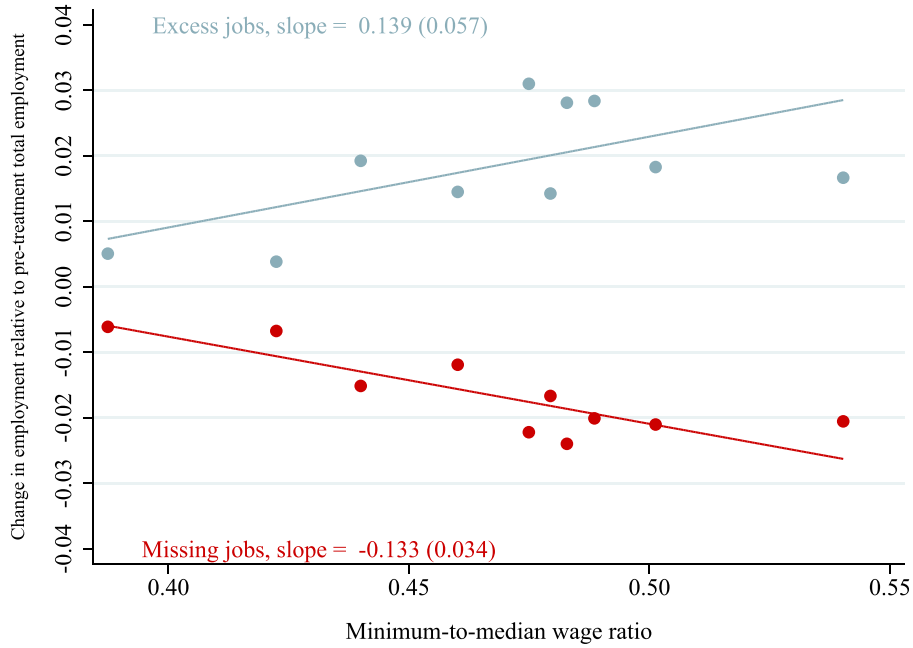
Creating G_2 dataset: stack the datasets



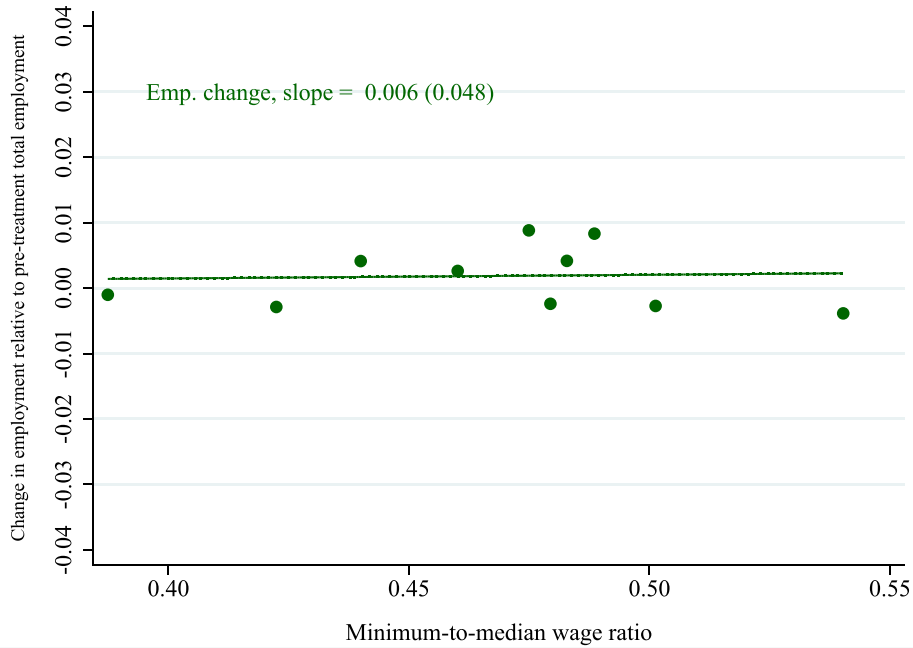
Discussion of the stacked data

- Why doesn't G_1 appear in G_2?
- Notice that U appears in both G_1 and G_2 datasets.
- Unclear to me exactly what Cengiz, et al. (2019) run in their stacked regression, but we probably need to say it now that we want to have a control for "dataset-by-state" fixed effects some units appear more than once (e.g., U).

(A) Missing and excess jobs



(B) Employment change



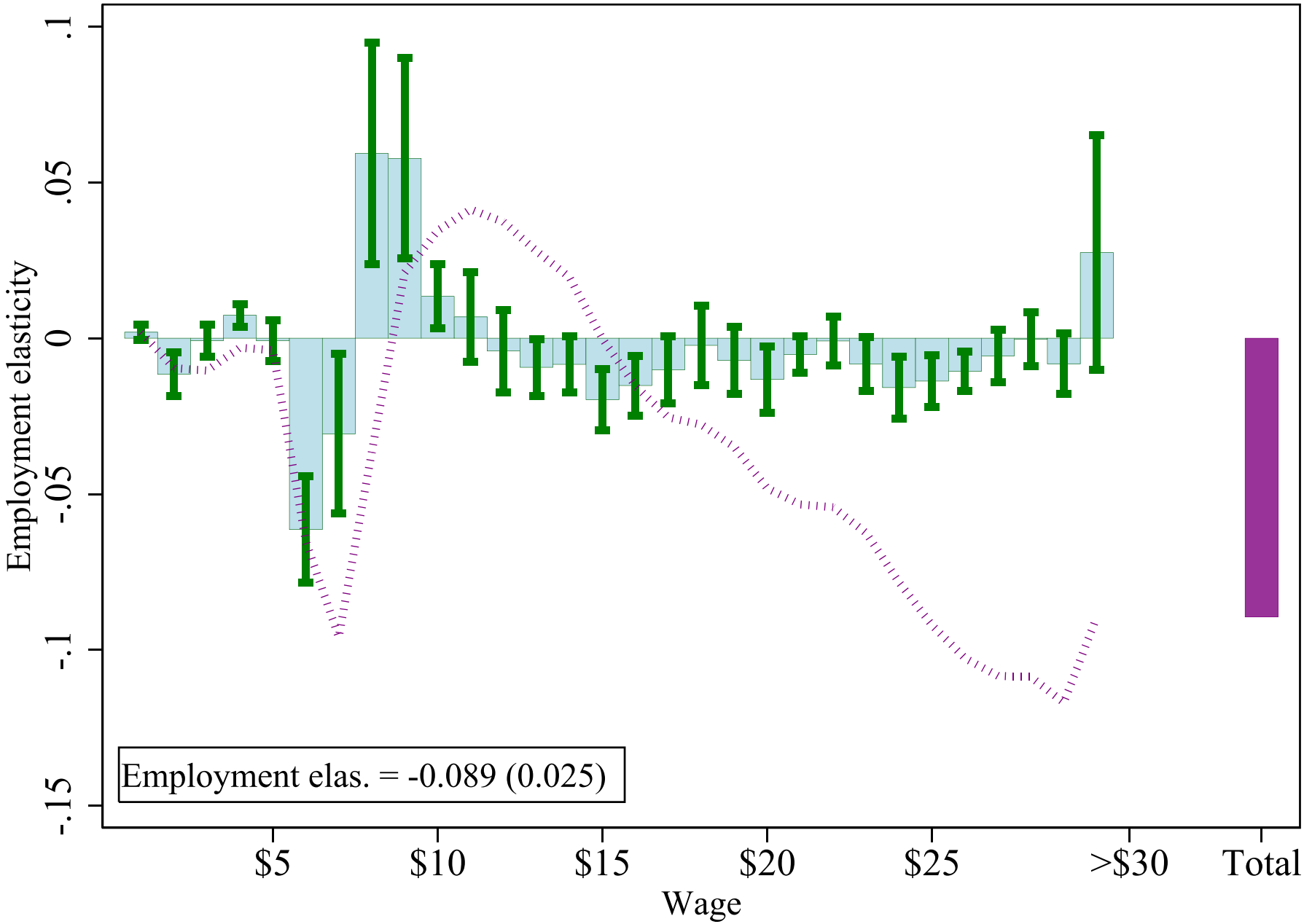
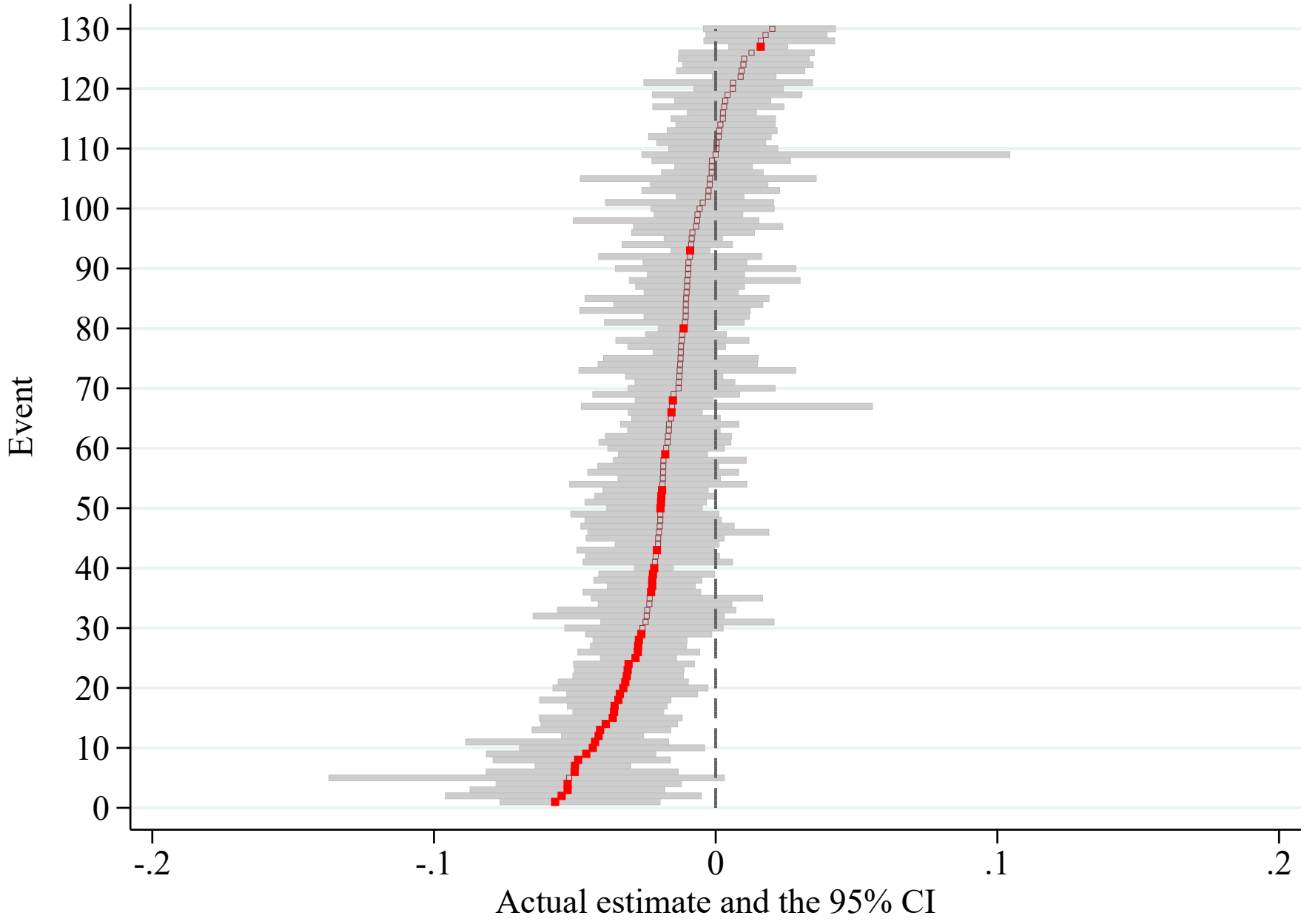
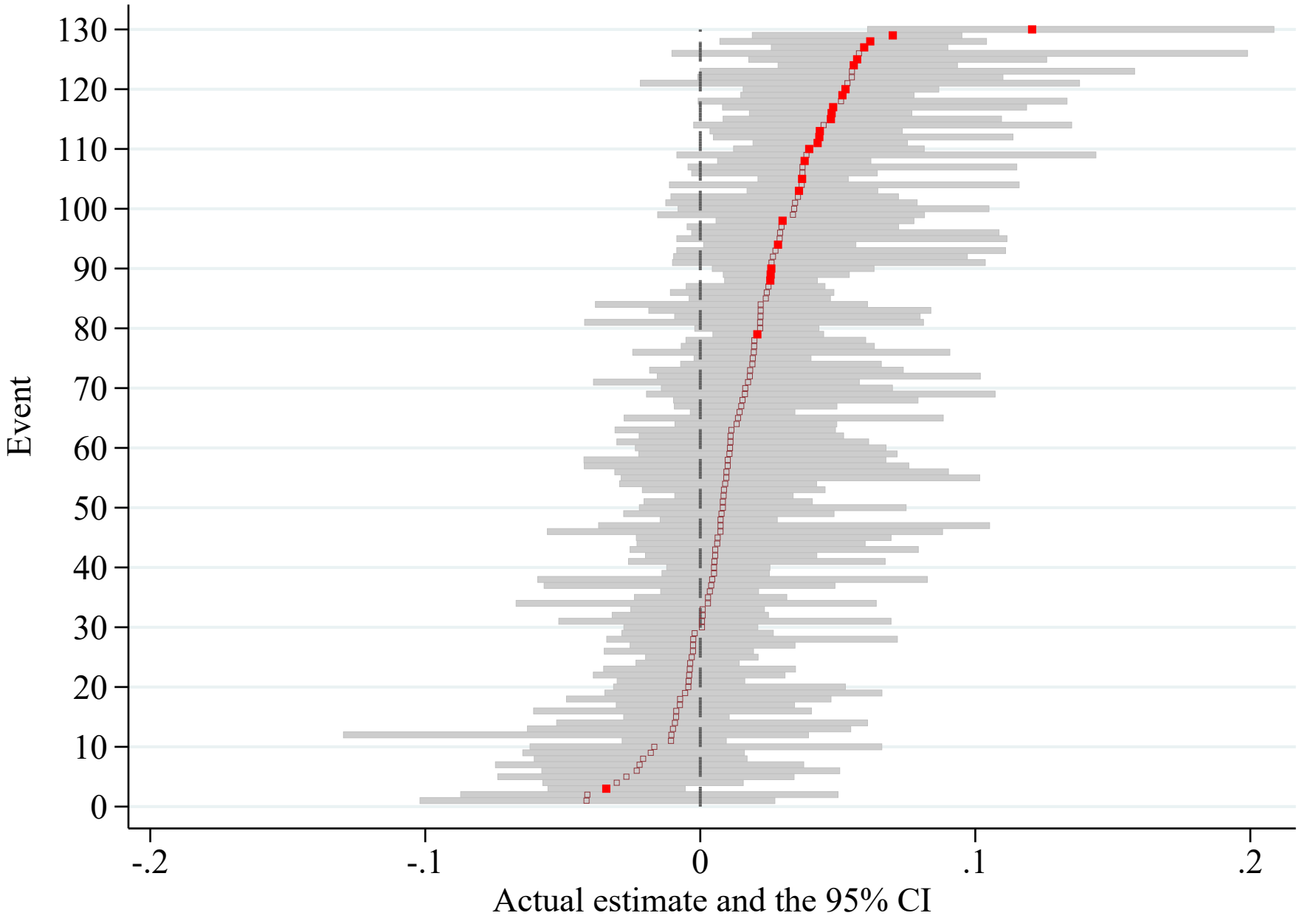

Employment elasticity
.1
.05
0
-.05
-.1
-.15
Employment elas. = -0.089 (0.025)
$5
$10
$15
$20
$25
>$30
Total
Wage

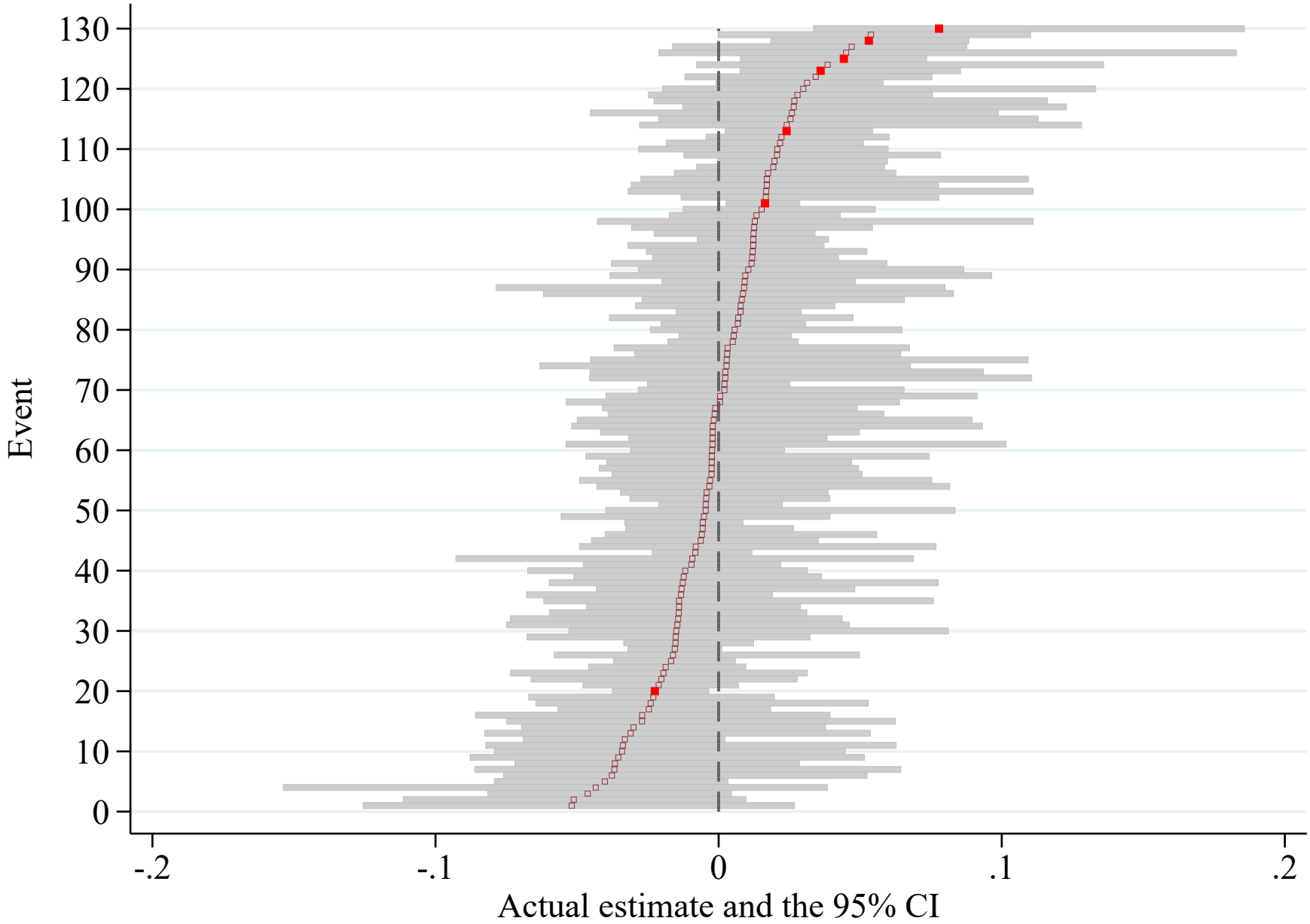
FIGURE VI

Upper part of wage distribution

Logic is used to dismiss effects at upper part of distribution which is only place they find effects


Event
130
120
110
100
90
80
70
60
50
40
30
20
10
0
-.2
-.1
0
.1
.2
Actual estimate and the 95% CI


Event
130
120
110
100
90
80
70
60
50
40
30
20
10
0
-.2
-.1
0
.1
.2
Actual estimate and the 95% CI



(c) Employment change ($\Delta a + \Delta b$)

Unknown parameter

- Recall CS and SA start with parameter (e.g., group-time ATT) then build the estimator that finds it using aggregation
- Stacking goes in reverse: start with TWFE using a restructured dataset with “clean controls”
- So what is stacking identifying? VWATT? What are its assumptions? What are the weights?

Identification

- Stacking is a TWFE estimation method, except that by balancing in relative event time, there is no longer any differential timing
- Thus identification requires a weighted parallel trends assumption, but how exactly?
- The parallel trends is *the same* as what we saw with our group-time representation
- Parallel is assumed to hold first *within* stacked dataset, not in the final regression model, which shapes the fixed effects we must employ to ensure it

Gardner notation

$$Y_{cgpit} = \lambda_{cg} + \lambda_{cp} + \beta D_{cgph} + \varepsilon_{cgpit} \tag{2}$$

c dataset; g group; p period; ith member of group g; tth time period of period p

D_{cgp} is an indicator for whether group g is treated during period p of the group-c dataset; β is the group-period ATT; $\widehat{\beta}$ is estimate from TWFE

Weighted ATT

$$
\begin{aligned}
\widehat{\beta} &= \sum_{g=1}^{G} \sum_{p=g}^{P} w_{gp} \beta_{gp} \\
w_{gp} &= \frac{(1-\pi_c)\pi_c \rho_c}{\overline{P} \sum_{c=1}^{G} (1-\pi_c)\pi_c \rho_c}
\end{aligned}
$$

π_c is the fraction of units treated in period p and ρ_c is the population share of observations for a given group/period

The stacked estimator weights each group's ATT by dataset-specific treatment variance and sample size which slightly overstates the true average bc π and ρ are both fractions.

Like Goodman-Bacon (2021), we see that the weights make $\widehat{\beta}$ slightly biased estimate of ATT.

Heterogeneity is nevertheless important

- When theory predicts that longrun effects of a policy may differ from shortrun effects, then exploring heterogeneity is required
- Effects may also differ for size of policy which creates problems for SUTVA bc of “no hidden variation in treatment”
- But heterogeneity analysis, we said, can introduce p-hacking even subconsiously or due to the “garden of forking paths” (Gelman and Loken 2013)

P-hacking

"Here's the thing: P-values of .05 aren't that hard to find if you sort the data differently or perform a huge number of analyses. In flipping coins, you'd think it would be rare to get 10 heads in a row. You might start to suspect the coin is weighted to favor heads and that the result is statistically significant.

But what if you just got 10 heads in a row by chance (it can happen) and then suddenly decided you were done flipping coins? If you kept going, you'd stop believing the coin is weighted.

Stopping an experiment when a p-value of .05 is achieved is an example of p-hacking. But there are other ways to do it – like collecting data on a large number of outcomes but only reporting the outcomes that achieve statistical significance. By running many analyses, you're bound to find something significant just by chance alone."

Alternatives to p-hacking heterogeneity

1. **Preregistration of study designs**: Scientists publicly commit to an experiment's design before the data collection phase or the analysis phase. Much harder to 'cherry pick' results
2. **Open data sharing**: Journals are increasingly requiring that researchers post their data online, or submit to a data repository
3. **Replication**: Some journals are publishing replications (e.g., Nature's ReScienceX)

Clemens and Strain (2021) use preregistration which is rare in labor or quasi-experimental work but is very common in development economics using RCTs

Researchers must make choices

- Nick Huntington-Klein and several others published a 2021 article in *Economic Inquiry* in which two applied microeconomics papers were assigned to individual researchers with the task of replicating the results from raw data
- Results were concerning:

  > *"We find large differences in data preparation and analysis decisions, many of which would not likely be reported in a publication. No two replicators reported the same sample size. Statistical significance varied across replications, and for one of the studies the effect's sign varied as well. The standard deviation of estimates across replications was 3–4 times the mean reported standard error."*

- Sometimes it isn't the garden of forking paths or p-hacking that is the problem – there is also uncertainty driven by the numerous reasonable choices that researchers must make in order to do any analysis in the first place

Pre-registration and heterogeneity analysis

- Theoretical reasons to suspect effects of the minimum wage may differ depending on whether the bite is large or small
- Because their analysis would be conducting heterogeneity analysis, and heterogeneity analysis was one of the causes of the replication crisis, they would pre-register
- Pre-registered design as extensions of their earlier short-run analysis

Variation in treatment

"Consider an assessment of the causal effect of aspirin on headaches. For the potential outcome with both of us taking aspirin, we obviously need more than one aspirin tablet. Suppose, however, that one of the tablets is old and no longer contains a fully effective dose, whereas the other is new and at full strength. In that case, each of us may have three treatments available: no aspirin, the ineffective tablet, and the effective tablet. There are thus two forms of the active treatment, both nominally labeled "aspirin": aspirin+ and aspirin-." (Imbens and Rubin 2015)

SUTVA

Stable Unit Treatment Value Assumption requires that an individual receiving a specific treatment level cannot receive different forms of that treatment (called the “no hidden variations of treatments” by Imbens and Rubin 2015)

> *“One strategy to make SUTVA more plausible relies on redefining the represented treatment levels to comprise a larger set of treatments, for example, Aspirin-, Aspirin+ and no-aspirin instead of only Aspirin and no-aspirin.” (Imbens and Rubin 2015)*

Great Recession

- Great Recession saw a pause in minimum wage hikes followed by large increases in several states
- Increases in Cengiz, et al. (2019) were around 8 log points on average; minimum wage increases after the Great Recession range from 25 log points to as high as 60 log points (Clemens and Strain 2021)
- “DC, CA and NY had increased their minimum wages by 61, 50 and 53 percent respectively.”

Methodology and Data

- Data: American Community Survey and Current Population Survey:
  - → Years: 2011-2015 with pre-registration commitment to study through 2019
- Methodologies:
  - → TWFE event study
  - → Stacked regression (Cengiz, et al. 2019; Baker, et al. 2020)
  - → Imputation estimator (Borusyak, Jaravel and Spiess 2021)
- Unique characteristics of treatment: large and small increases will be modeled separately

Summary of findings

- Large increases in minimum wages reduced employment rates among individuals with low levels of experience and education by just over 2.5pp
- Relatively small minimum wage increases are variable and centered on zero much like what Cengiz, et al. (2019) found
- Medium-run effects are larger and more negative than short-run effects

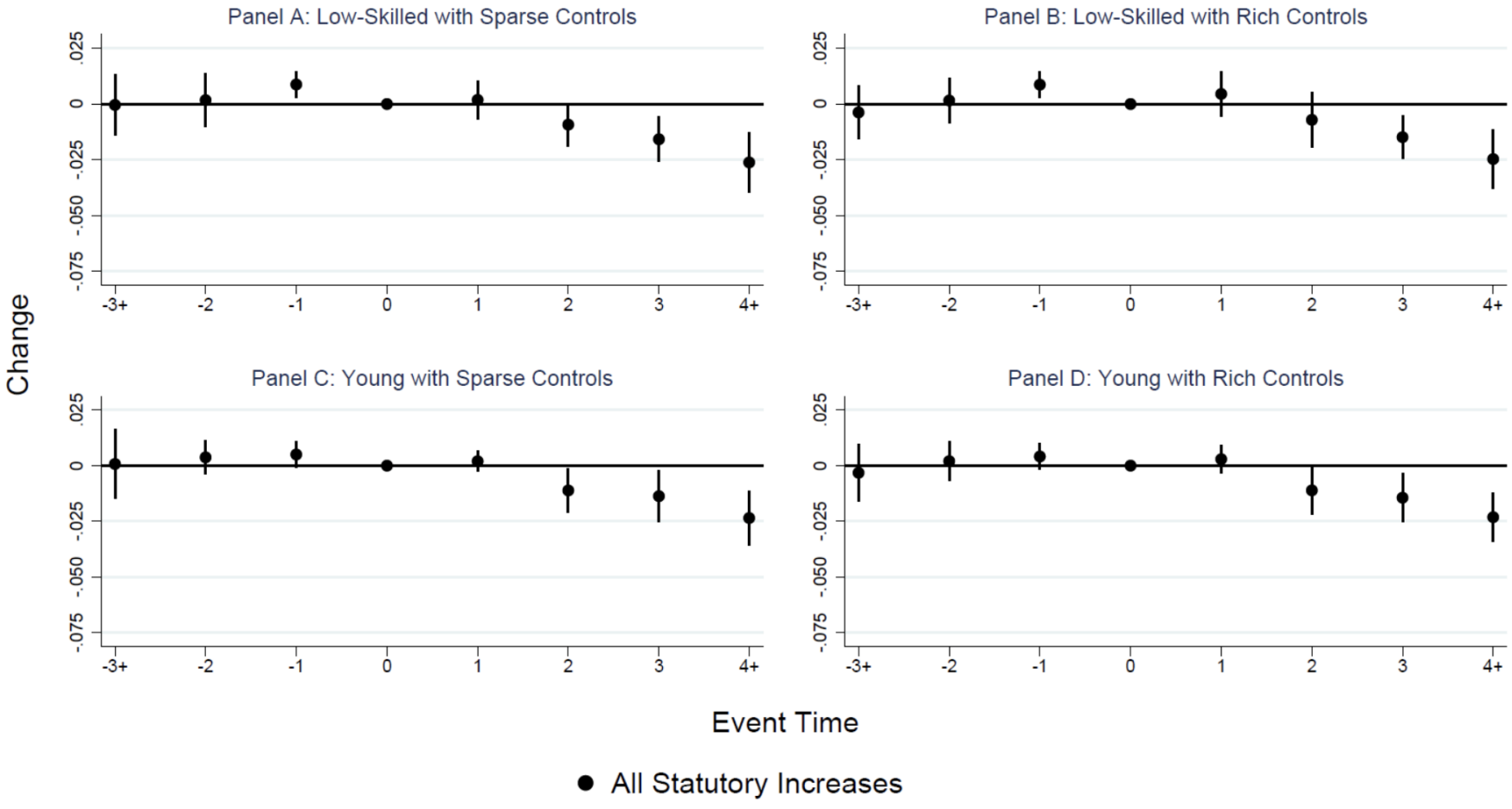
TWFE Event study specification

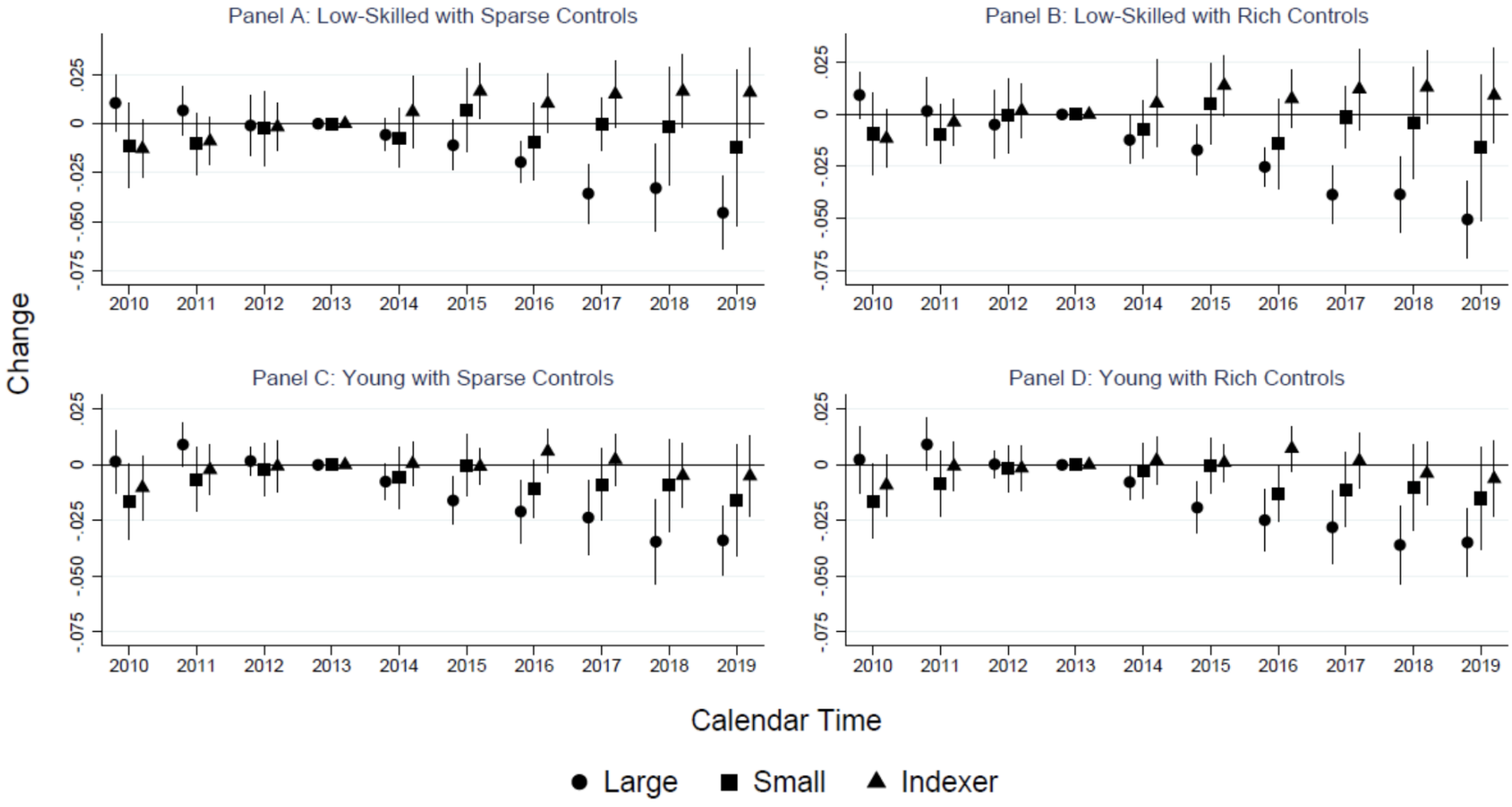
$$Y_{i,s,g(s),t} = \sum_{g(s)\neq 0} \beta_{g(s)} Policy_{g(s)} \times Post_t + \alpha_{1s} State_s + \alpha_{2t} Time_t + X_{i,s,t}\gamma + \varepsilon_{i,s,t}$$

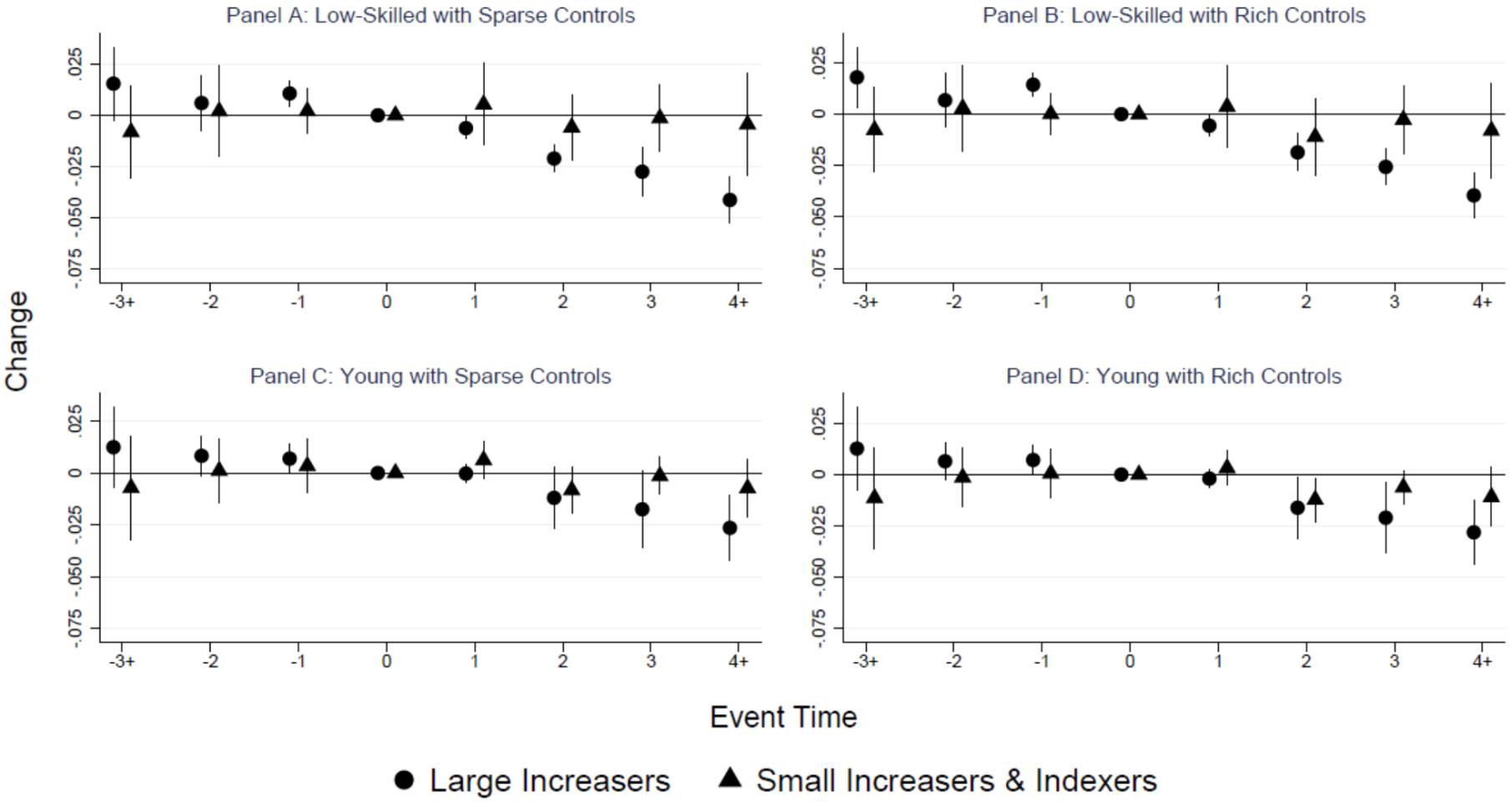
Y is binary for employment for person i in state s in policy category $g(s)$ and time t. Samples are restricted to young (16-21yo) without high school, and young overall (16-25yo). X are the "rich controls" they label later and it includes median house price index, log aggregate personal income per capita, employment rates for different skill levels, and individual-level demographic controls.

Coefficients of interest are the β terms and 2014 will be treated as the transition year. It measures the causal effect of state minimum wage policy changes on employment under standard, albeit nontrivial, assumptions such as homogenous treatment effects over time, no anticipation and parallel trends.

They will also estimate triple difference versions of this model with a within-state control group of untreated workers.


Panel A: Low-Skilled with Sparse Controls
Panel B: Low-Skilled with Rich Controls
Panel C: Young with Sparse Controls
Panel D: Young with Rich Controls
Change
.025
0
-.025
-.050
-.075
-3+
-2
-1
0
1
2
3
4+
Event Time
All Statutory Increases


Panel A: Low-Skilled with Sparse Controls
Panel B: Low-Skilled with Rich Controls
Panel C: Young with Sparse Controls
Panel D: Young with Rich Controls
Change
.025
0
-.025
-.050
-.075
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
Calendar Time
Large
Small
Indexer


Panel A: Low-Skilled with Sparse Controls
Panel B: Low-Skilled with Rich Controls
Panel C: Young with Sparse Controls
Panel D: Young with Rich Controls
Change
.025
0
-.025
-.050
-.075
-3+
-2
-1
0
1
2
3
4+
Event Time
Large Increasers
Small Increasers & Indexers

Extension 1: Stacked regression

- Remember the big idea: create a new, much larger, dataset by appending the same dataset to itself over and over, not in balanced calendar time, but in balanced *event time*.
- Then estimate a simple TWFE model controlling for dataset-by-state fixed effects

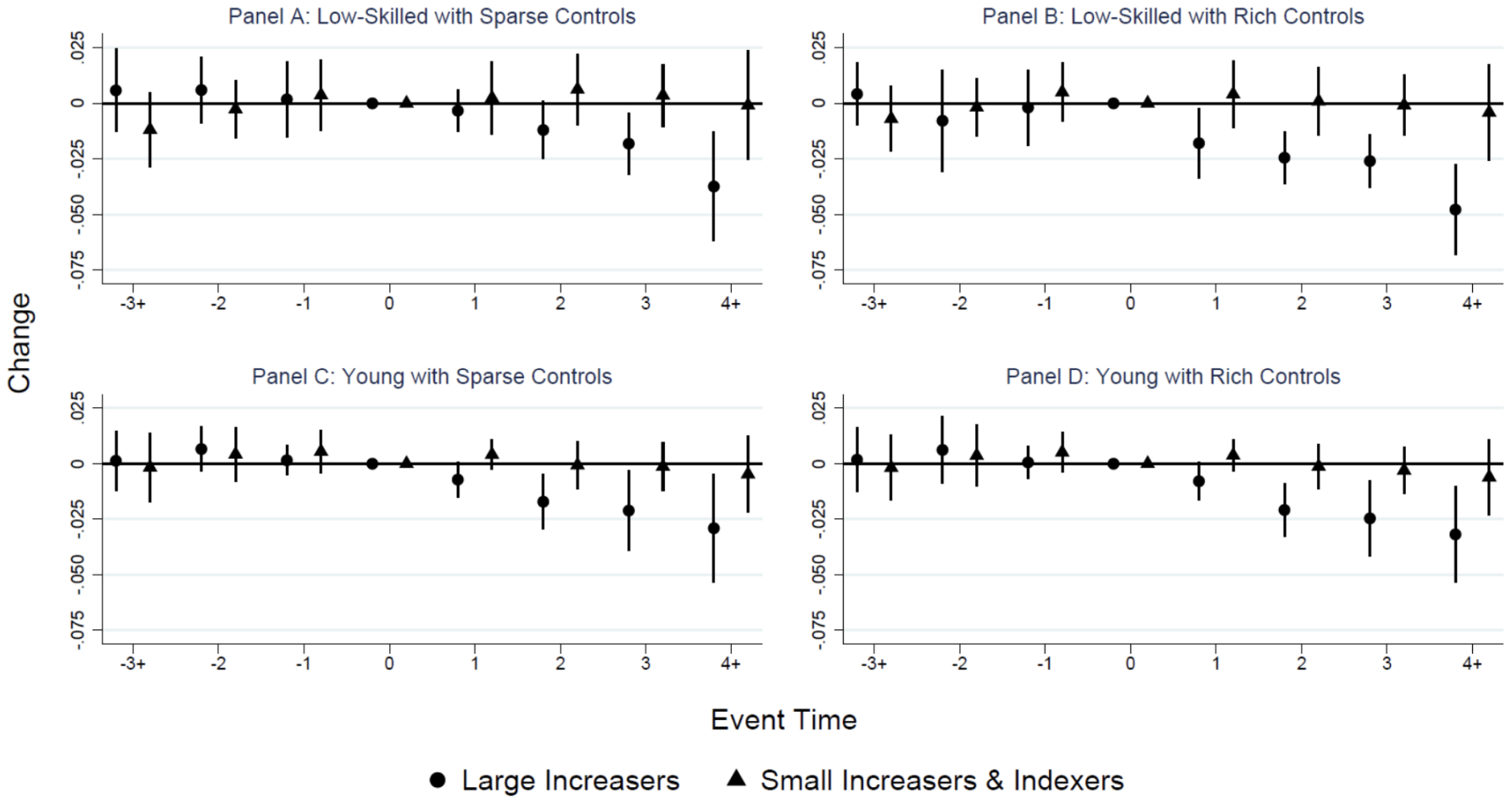


Figure 10. Stacked Event Studies of Changes in Employment Following Large and Small Statutory Minimum Wage Increases: This figure displays coefficients from the "stacked event study" estimator described by equation (5). Event Time is defined such that year "1" corresponds with the year during which a given state enacted its first minimum wage change due to legislation passed during our sample period. We compare estimates for large vs. small increases as defined in the main text. Panels A and B plot coefficients for low-skilled individuals defined as individuals ages 16–25 without a completed high school education. Panels C and D plot coefficients for young individuals defined as all individuals ages 16–21. The samples are from the ACS. Regressions with "sparse controls" include state and year fixed effects, as well as the log of annual *per capita* income and the annual average of the median house price index. Regressions with "rich controls" include all sparse controls plus the three-year lag of both the log of annual *per capita* income and the annual average of the median house price index, as well as a dummy variable for each education group and age. Error bars denote 95 percent confidence intervals around each estimated coefficient. Standard errors are clustered by state.

Briefly concluding remarks

- To understand the rest of their results, we have to first cover the imputation estimator that Borusyak, Jaravel and Speiss (2021) have created
- But to summarize what we found so far, large shocks are indeed credibly causing declines in employment, particular for the affected class of workers (young with or without a high school degree)
- More modest increase are null, though, which is what Cengiz, et al. (2019) also found