

# Do State Automatic IRA Mandates Affect Self-Reported Employer Retirement Plan Coverage? Evidence from Staggered Policy Adoption

APEP Autonomous Research\*

@dakoyana, @anonymous

February 2026

## Abstract

State auto-IRA mandates require employers without retirement plans to enroll employees in state-facilitated IRAs. This paper estimates whether these mandates affect *self-reported employer retirement plan coverage*—a distinct outcome from auto-IRA participation itself, since CPS asks about “employer” plans while auto-IRAs are state-facilitated individual accounts. Using CPS ASEC data from 2010–2024 and a Callaway-Sant’Anna difference-in-differences design exploiting staggered adoption across five states with meaningful post-treatment data (Oregon, Illinois, California, Connecticut, Maryland), I find an overall ATT of 0.75 percentage points (SE = 1.0 pp; wild bootstrap  $p = 0.48$ ), not statistically significant. This null result is driven by Oregon’s anomalous negative effect (−2.1 pp); a systematic leave-one-out analysis shows Oregon is uniquely influential. Excluding Oregon yields a significant effect of 1.57 pp (SE = 0.62 pp;  $p < 0.01$ ). A triple-difference design exploiting firm-size phase-in—comparing small firms (targeted by mandates) to large firms (placebo)—provides additional identification. Randomization inference with 2,000 permutations yields a two-sided  $p$ -value of 0.47 for the overall effect. These findings suggest that auto-IRA mandates may increase self-reported employer plan coverage through awareness spillovers or employer behavioral responses, though the CPS outcome does not directly measure auto-IRA participation.

---

\*Autonomous Policy Evaluation Project. This paper was autonomously generated using Claude Code (claude-opus-4-5). Revision of apep\_0042. Correspondence: scl@econ.uzh.ch

**JEL Codes:** H31, J26, J32, D14

**Keywords:** automatic enrollment, retirement savings, state policy, difference-in-differences

## 1. Introduction

Roughly half of American private-sector workers lack access to an employer-sponsored retirement plan. This “coverage gap” is particularly acute among workers at small businesses, part-time employees, and lower-wage workers—precisely those who may benefit most from the tax advantages and behavioral nudges that employer-facilitated retirement savings provide. The gap persists despite decades of policy efforts to expand retirement savings, and many observers worry it contributes to inadequate retirement preparedness across the income distribution.

Beginning in 2017, states have increasingly turned to a novel policy response: mandatory automatic IRA programs. These programs require employers who do not offer their own retirement plan to enroll employees in a state-facilitated Individual Retirement Account (IRA). Drawing on insights from behavioral economics about the power of defaults, automatic enrollment removes the burden of active decision-making that may deter workers from saving. Early evidence from program administrative data suggests high retention rates among enrolled workers, with approximately 70% of OregonSaves participants remaining enrolled after automatic enrollment.

This paper provides the first quasi-experimental evaluation of state auto-IRA mandates using nationally representative survey data. I exploit the staggered adoption of mandatory programs across states from 2017–2024 in a difference-in-differences framework, using the Callaway and Sant’Anna (2021) estimator designed for settings with staggered treatment timing and treatment effect heterogeneity. While eleven states have enacted auto-IRA programs by 2024, only five have meaningful post-treatment exposure in the CPS data (Oregon, Illinois, California, Connecticut, Maryland); the remaining six launched too recently for reliable estimation. The outcome of interest is whether workers report having any retirement plan coverage from their current employer, measured in the Current Population Survey Annual Social and Economic Supplement (CPS ASEC).

The main finding is a null result: I estimate an overall average treatment effect of 0.75 percentage points (SE = 1.0 pp, 95% CI: [−1.2, 2.7]), which is not statistically distinguishable from zero. Wild cluster bootstrap inference, which is more reliable with few clusters, yields a  $p$ -value of 0.48. Event study estimates show no evidence of differential pre-trends between treatment and control states, lending credibility to the parallel trends assumption. Post-treatment coefficients are uniformly positive but imprecisely estimated, with point estimates that grow modestly over time (reaching 2.2 percentage points at event time +5).

Results are heterogeneous across treatment cohorts. Illinois ( $G_i=2019$ ) and California ( $G_i=2020$ ) show positive and statistically significant effects of 2.6 and 2.2 percentage points

respectively. In contrast, Oregon ( $G_i=2018$ , the first cohort) shows a surprising negative effect of  $-2.1$  percentage points. A systematic leave-one-out analysis confirms that Oregon is uniquely influential: excluding Oregon increases the estimated ATT from 0.75 pp to 1.57 pp and changes inference from non-significant to significant at the 1% level. No other state’s exclusion produces a comparably large change.

I extend the analysis with two additional identification strategies. First, a triple-difference design exploits the firm-size phase-in structure of mandates: small firms (the primary targets, lacking existing plans) should show larger effects than large firms (which already offer retirement plans). The DDD coefficient on the interaction of treated state, post-treatment, and small firm size provides within-state identification. Second, randomization inference provides exact  $p$ -values by comparing the observed effect to effects under 2,000 random permutations of treatment assignment. The randomization inference  $p$ -value of 0.47 confirms that the null result is not driven by asymptotic approximations breaking down with few clusters.

The null overall result admits several interpretations. First, measurement error may attenuate estimates toward zero: the CPS ASEC asks whether workers are “included in a pension or retirement plan” at their current employer, which may not capture workers who understand their auto-IRA as an individual account rather than an employer plan. The disconnect between OregonSaves’ administrative enrollment data (150,000+ active participants) and CPS-measured coverage trends in Oregon strongly suggests such measurement failure. Second, the relatively short post-treatment periods for most treated states may be insufficient to detect effects that build gradually. Third, high opt-out rates or limited employer compliance could genuinely result in small effects.

This paper contributes to literatures on retirement savings policy, behavioral economics, and state policy innovation. Despite the null finding, the analysis provides a methodological template for future evaluations as programs mature and longer post-treatment data become available. The heterogeneity across states suggests that program design details and implementation may matter substantially for effectiveness.

## **2. Institutional Background**

### **2.1 The Retirement Coverage Gap**

Approximately 57 million American workers—about half of the private-sector workforce—lack access to an employer-sponsored retirement plan. Coverage rates are lowest among workers at small businesses: only 42% of workers at firms with fewer than 100 employees have access to a retirement plan, compared to 79% at firms with 500 or more employees.

Workers without employer plans must save through individual retirement accounts (IRAs), but contribution rates to IRAs outside of employer facilitation are low.

The coverage gap has several explanations. Small employers face fixed costs of plan administration that make offering plans uneconomical. Workers at small firms tend to have lower wages and higher turnover, making them less attractive participants. And absent the convenience of payroll deduction and the behavioral nudge of employer defaults, workers may simply not get around to opening and contributing to retirement accounts.

## 2.2 State Automatic IRA Programs

Beginning in 2012, states began exploring legislation to address the coverage gap. California passed the first enabling legislation in 2012, though the program (CalSavers) did not launch until 2019. Oregon launched the first operational program (OregonSaves) in July 2017. As of 2024, eleven states have implemented mandatory auto-IRA programs:

**Table 1:** State Auto-IRA Program Adoption Timeline and Analysis Treatment Coding

State	Program	Launch Date	$G_i$ (Analysis)	Post-Treat CPS	In Sample?
Oregon	OregonSaves	July 2017	2018	2018–2024	Yes
Illinois	Secure Choice	Nov 2018	2019	2019–2024	Yes
California	CalSavers	July 2019	2020	2020–2024	Yes
Connecticut	MyCTSavings	Mar 2021	2022	2022–2024	Yes
Maryland	MarylandSaves	Sep 2022	2023	2023–2024	Yes
<i>States excluded from estimation (to avoid control group contamination):</i>					
Colorado	SecureSavings	Mar 2023	—	Minimal*	No
Virginia	Virginia Saves	July 2023	—	Minimal*	No
Maine	Retirement Savings	2024	—	None	No
Delaware	Delaware Saves	2024	—	None	No
New Jersey	Secure Choice	2024	—	None	No
Vermont	Green Mountain	2024	—	None	No

Notes:  $G_i$  = first CPS year with post-treatment exposure. CPS ASEC (March) asks about “last year.”

\*Colorado/Virginia launched 2023 with partial exposure in March 2024 CPS—excluded from sample.

States below the line are excluded from estimation to avoid control group contamination.

**Treatment timing coding.** An important methodological issue is mapping program launch dates to the treatment indicator ( $G_i$ ) used in estimation. The CPS ASEC is fielded each March and asks about “the last year,” so the March 2024 survey references calendar

year 2023. For programs launching mid-year (e.g., Oregon in July 2017), the first CPS with meaningful post-treatment exposure is the following year’s survey. I define  $G_i$  as the first CPS survey year in which the reference period includes any months of post-launch operation. Thus Oregon (launched July 2017) is coded  $G_i=2018$  (the March 2018 CPS asks about 2017, which includes 6 months of program operation); Illinois (launched November 2018) is coded  $G_i=2019$  (the March 2019 CPS asks about 2018, which includes 2 months of operation).

States launching in late 2023 or 2024 pose a control group contamination problem. Colorado (March 2023) and Virginia (July 2023) would have partial post-treatment exposure in the March 2024 CPS, making them inappropriate for either the treated group (insufficient post-treatment years) or the never-treated control group (they have some treatment exposure). States launching in 2024 (Maine, Delaware, New Jersey, Vermont) have no post-treatment data. I therefore *completely exclude* these six states from the estimation sample, leaving five treated states (Oregon, Illinois, California, Connecticut, Maryland) and 45 never-treated control states. This exclusion ensures clean identification: all control states are genuinely never-treated during the sample period.

These programs share core design elements. Employers without existing retirement plans must facilitate employee participation in a state-administered Roth IRA. Employees are automatically enrolled at a default contribution rate (typically 3–5% of wages), though they may opt out or adjust their contribution rate. Contributions are made through payroll deduction, minimizing transaction costs for workers. The programs are funded through participant fees and initial state investments rather than employer contributions.

Most programs phase in by employer size, starting with larger employers and gradually extending to smaller businesses. For example, Oregon initially required employers with 100+ employees to register in 2017, then extended to 50–99 employees in 2018, 25–49 in 2019, and so on. This phased implementation creates within-state variation in treatment timing that I exploit in a triple-difference design.

### **2.3 Policy Expectations**

Theory and evidence from 401(k) automatic enrollment suggest that auto-IRA programs could substantially increase retirement savings participation. Madrian and Shea (2001) found that automatic enrollment in employer 401(k) plans raised participation rates from 37% to 86% at a large corporation. Similar effects have been documented across many employers.

However, several factors may limit the applicability of these findings to state auto-IRA programs. First, unlike 401(k)s, IRAs do not feature employer matching contributions, which may reduce the incentive to participate. Second, IRA contribution limits (\$7,000 in 2024) are substantially lower than 401(k) limits (\$23,000), limiting benefits for higher savers. Third,

state programs face challenges of employer awareness, compliance, and enforcement that do not arise in within-firm 401(k) programs.

### 3. Related Literature

This paper relates to three strands of literature: automatic enrollment and retirement savings, state auto-IRA policy evaluation, and econometric methods for staggered difference-in-differences.

**Automatic enrollment and retirement savings.** An extensive literature documents the power of defaults in retirement savings decisions. Seminal work by Madrian and Shea (2001) showed that automatic enrollment in a single firm’s 401(k) plan raised participation from 37% to 86%. Choi, Laibson, Madrian, and Metrick (2002, 2004) extended these findings and documented persistent effects on contribution rates and asset allocation. Thaler and Benartzi (2004) demonstrated that “Save More Tomorrow” programs leveraging defaults and inertia dramatically increased retirement savings. Beshears et al. (2009) synthesized the evidence on default effects across retirement plan contexts.

Chetty et al. (2014) provide important evidence from Denmark on whether automatic enrollment increases total savings or merely reshuffles existing savings. They find that automatic enrollment generates substantial net new savings, particularly for passive savers who would otherwise not contribute. Their distinction between “active” and “passive” savers is particularly relevant: passive savers respond strongly to defaults while active savers offset changes in one account with adjustments in another. If auto-IRA programs primarily reach passive savers who would not otherwise save, the effects on total retirement wealth could be substantial.

More recent work has examined persistence and general equilibrium effects. Choukhmane (2022) shows that automatic enrollment effects persist as workers age and increase their contributions over time. Benartzi and Thaler (2013) survey two decades of research on behavioral interventions in retirement saving and find consistent evidence that default enrollment increases participation by 30–50 percentage points. However, these studies examine employer-sponsored automatic enrollment; whether state-facilitated programs generate similar effects is an open question that this paper addresses. State programs differ from employer programs in that they lack matching contributions (a powerful incentive), face scale and compliance challenges across thousands of small employers, and may be perceived differently by workers who understand them as government-facilitated rather than employer-provided.

**State auto-IRA programs.** A growing policy literature examines state auto-IRA programs specifically. The Automatic IRA concept was developed by Gale, Iwry, John, and

Walker (2009) at the Brookings Retirement Security Project, building on behavioral insights about defaults. Administrative data from program operators provide early evidence: Oregon-Saves reports approximately 70% retention rates among auto-enrolled workers, substantially higher than voluntary IRA participation. CalSavers and Illinois Secure Choice report similar patterns. Belbase and Sanzenbacher (2017) project that universal auto-IRA coverage could increase retirement assets by 31% for those currently without coverage.

The Georgetown Center for Retirement Initiatives and the Center for Retirement Research at Boston College have tracked program implementation and published descriptive statistics on enrollment and retention. However, no quasi-experimental studies using nationally representative data have evaluated the causal effects of state mandates on retirement coverage rates. This paper fills that gap, though with important measurement caveats discussed below.

**Difference-in-differences with staggered adoption.** Methodological advances inform this paper’s empirical strategy. Bertrand, Duflo, and Mullainathan (2004) documented the serial correlation problem in DiD standard errors. Goodman-Bacon (2021) showed that two-way fixed effects estimators with staggered treatment timing can produce biased estimates when treatment effects are heterogeneous, as the estimator implicitly uses already-treated units as controls. Callaway and Sant’Anna (2021), Sun and Abraham (2021), de Chaisemartin and D’Haultfoeuille (2020), and Borusyak, Jaravel, and Spiess (2021) propose alternative estimators that avoid these biases.

When treated units are few, as in state-level policy evaluations, inference requires special care. Conley and Taber (2011) develop methods for inference with few policy changes. Cameron, Gelbach, and Miller (2008) show that wild cluster bootstrap provides more reliable inference with few clusters. Roth (2022) shows that pre-testing for parallel trends can distort inference and recommends honesty-based approaches. This paper applies the Callaway-Sant’Anna estimator with never-treated states as controls, reports both clustered standard errors and wild cluster bootstrap  $p$ -values, and conducts randomization inference as a robustness check.

## 4. Data

### 4.1 Data Source

The analysis uses microdata from the Current Population Survey Annual Social and Economic Supplement (CPS ASEC) for years 2010–2024, accessed through IPUMS-CPS. The CPS ASEC is conducted each March and provides detailed information on demographics, employment, income, and benefits for a representative sample of approximately 180,000 in-

dividuals annually.

The key outcome variable is the IPUMS-CPS variable PENSION (ASEC), which identifies workers who respond “yes” to the question: “At any time in the last year, were you included in a pension or retirement plan at work?” This variable captures self-reported inclusion in both defined benefit pensions and defined contribution plans such as 401(k)s.

Several measurement issues merit attention. First, the question wording refers to plans “at work,” which may not capture auto-IRA participation if workers understand these as individual accounts facilitated by the state rather than employer plans. Second, the CPS ASEC is fielded in March and asks about “the last year,” so the 2017 survey references coverage during 2016—before any state auto-IRA program launched. For Oregon (launched July 2017), the first survey with meaningful exposure is March 2018. This timing misalignment may attenuate estimated effects in early treatment years. Third, programs phase in by employer size over multiple years, but our main treatment indicator is binary at the state-year level (addressed in the DDD design).

## 4.2 Sample Construction

I restrict the sample to private-sector wage and salary workers ages 18–64 who report positive earnings. I exclude self-employed workers, government employees, and workers who report not being at work during the reference week. Workers with missing values for the pension variable are dropped.

The final analysis sample contains 596,834 person-year observations across 45 state-level clusters (50 states plus DC, minus 6 states excluded for contamination) over 15 years (2010–2024). Of these 45 clusters, 5 are treated (Oregon, Illinois, California, Connecticut, Maryland) and 40 are never-treated controls.

## 4.3 Summary Statistics

Table 2 presents summary statistics by treatment status. Treated states (those that adopted auto-IRA mandates by 2024) have slightly lower pension coverage rates than never-treated states (15.0% vs 15.7%), though this difference is not statistically significant. The two groups are broadly similar in demographic composition.

**Table 2:** Summary Statistics by Treatment Status (Estimation Sample)

	Treated States <sup>a</sup>		Never-Treated States <sup>b</sup>	
	Mean	SD	Mean	SD
Pension coverage	0.148	0.355	0.159	0.366
Female	0.462	0.499	0.458	0.498
Age	39.1	12.3	39.8	12.5
College graduate	0.345	0.475	0.318	0.466
Married	0.502	0.500	0.521	0.500
Small firm (<100)	0.476	0.499	0.492	0.500
Observations	124,469		472,365	
States	5		40	

Notes: Sample includes private-sector wage/salary workers ages 18–64. Weighted by CPS ASEC.

<sup>a</sup>Treated = 5 states with meaningful post-treatment data (OR, IL, CA, CT, MD).

<sup>b</sup>Never-treated = 40 states that never adopted auto-IRA. 6 late adopters (CO, VA, ME, DE, NJ, VT) excluded.

#### 4.4 Why CPS May Not Capture Auto-IRA Participation

A critical limitation of this analysis is that the CPS ASEC may fail to capture auto-IRA participation. This section documents the measurement gap and its implications for interpretation.

**Survey question wording.** The CPS asks: “At any time in the last year, were you included in a pension or retirement plan *at work*?” This wording explicitly references plans “at work” provided by employers. Auto-IRAs, however, are technically individual retirement accounts—not employer-sponsored plans. Although contributions are made through payroll deduction, the accounts are administered by the state and remain with workers across jobs. Workers may correctly understand that these are not “employer” plans and respond “no” to the CPS question.

**Respondent cognition.** The term “pension” carries strong connotations of defined benefit plans or employer-sponsored 401(k)s with matching contributions. Auto-IRAs lack employer contributions and may not trigger recognition when workers hear “pension or retirement plan.” Workers may also not understand what automatic enrollment in OregonSaves or CalSavers means, or may not recall participation when surveyed months later.

**Evidence of measurement gap.** Administrative data provide striking evidence of this gap:

- OregonSaves reported over 150,000 actively participating savers by late 2024
- Yet Oregon’s CPS-measured pension coverage rate shows no increase—indeed, it declined from 18.2% (2010) to 14.9% (2024)
- This disconnect strongly suggests that CPS-measured coverage fails to capture auto-IRA participation

CalSavers and Illinois Secure Choice report similar enrollment figures in the hundreds of thousands, yet CPS trends in these states do not show corresponding increases relative to never-treated states.

**Implications for interpretation.** This measurement error attenuates estimated treatment effects toward zero. Consequently, our estimates should be interpreted as effects on *self-reported employer-sponsored retirement coverage*, not on total retirement savings participation. The true policy effect on retirement saving access is likely larger than the CPS-based estimates indicate.

There are, however, reasons why mandates might affect CPS-measured coverage even if auto-IRA enrollment is not directly captured:

1. **Awareness spillovers:** The mandate may draw attention to retirement saving, prompting workers to notice or enroll in existing employer plans they had previously ignored. When employers begin communicating about the auto-IRA mandate, workers may learn about retirement options more broadly—including employer-sponsored 401(k) plans they had not previously considered. This “information spillover” could increase participation in existing employer plans, which the CPS would capture.
2. **Employer behavioral response:** Employers may adopt 401(k) plans to avoid the complexity of administering auto-IRA payroll deductions, converting from “no plan” to “employer plan.” Several state programs (including OregonSaves and CalSavers) have documented cases of employers choosing to establish their own qualified retirement plans rather than comply with the state mandate. If employers believe that offering a 401(k)—especially one with employer matching—provides competitive advantages in hiring, the mandate could indirectly increase employer plan coverage.
3. **Survey response effects:** Awareness of the mandate may change how workers interpret and respond to the CPS question. Workers who know their state has a retirement savings mandate may be more likely to answer “yes” to the pension question, even if their actual enrollment status has not changed. This “priming” effect could produce measured effects that do not reflect genuine coverage changes.

4. **Classification effects:** Some workers enrolled in state auto-IRAs may interpret the CPS question broadly enough to include state-facilitated programs, particularly if their contributions are deducted from their paycheck alongside taxes and other withholdings. The boundary between “employer plan” and “state plan facilitated through employer payroll” may be ambiguous to survey respondents.

These channels suggest the CPS outcome captures some, but not all, of the policy effect. The overall effect on retirement savings participation is almost certainly larger than what I estimate using CPS-measured coverage, because the CPS likely fails to capture direct auto-IRA enrollment while potentially capturing indirect effects through the channels above.

**Reconciliation with administrative data.** Table 3 provides a rough comparison of administrative enrollment figures with CPS-measured coverage changes. While not directly comparable (administrative data measure enrollment, CPS measures self-reported coverage, and the populations differ), the table illustrates the magnitude of the measurement gap.

**Table 3:** Comparison of Administrative Enrollment and CPS Coverage Changes

State	Admin. Enrollment	CPS Coverage Change	Gap
Oregon (OregonSaves)	~150,000	-2.1 pp	Large
California (CalSavers)	~500,000	+2.2 pp	Moderate
Illinois (Secure Choice)	~200,000	+2.6 pp	Moderate

Notes: Administrative enrollment from program reports (2024). CPS coverage change is estimated cohort ATT in percentage points. Not directly comparable.

The stark contrast between Oregon’s large administrative enrollment and its negative CPS effect is the most striking evidence of measurement failure. Oregon was the first state to launch, has had the longest time to scale up, and reports robust participation—yet shows a negative effect in survey data. This strongly suggests that survey-based measures fail to capture the true policy impact.

## 5. Empirical Strategy

### 5.1 Identification

I exploit the staggered adoption of state auto-IRA mandates across states and over time. Let  $G_i \in \{0, 2017, 2018, \dots, 2024\}$  denote the first year in which state  $i$  is treated, with  $G_i = 0$  indicating never-treated states. The identifying assumption is parallel trends: in the absence

of treatment, the average change in pension coverage would have been the same for states that adopted mandates and states that did not.

This assumption is not directly testable, but I assess its plausibility by examining pre-treatment outcome dynamics. If treated and control states were on parallel trajectories before treatment, it is more plausible that they would have remained on parallel trajectories absent treatment.

## 5.2 Estimation

I estimate treatment effects using the Callaway and Sant’Anna (2021) estimator, which addresses biases that can arise in two-way fixed effects regressions with staggered treatment timing and heterogeneous treatment effects. The estimator computes group-time average treatment effects:

$$ATT(g, t) = \mathbb{E}[Y_{it}(g) - Y_{it}(0) | G_i = g] \quad (1)$$

for each treatment cohort  $g$  and time period  $t$ , where  $Y_{it}(g)$  is the potential outcome under treatment starting at time  $g$  and  $Y_{it}(0)$  is the potential outcome under no treatment.

I use never-treated states as the control group, doubly robust estimation incorporating pre-treatment covariates, and cluster standard errors at the state level to account for within-state serial correlation. I aggregate group-time effects to overall summaries using inverse-variance weighting.

For the event study specification, I define event time  $e = t - g$  as years relative to treatment and aggregate group-time effects by event time to trace out dynamic treatment effects.

## 5.3 Triple-Difference Design

The staggered DiD treats all workers in a state uniformly, but mandates specifically target employers without existing retirement plans—primarily small firms. I exploit this structure in a triple-difference (DDD) design:

$$Y_{ist} = \beta_1(\text{Treat}_s \times \text{Post}_t) + \beta_2(\text{Treat}_s \times \text{Small}_i) + \beta_3(\text{Post}_t \times \text{Small}_i) \quad (2)$$

$$+\beta_4(\text{Treat}_s \times \text{Post}_t \times \text{Small}_i) + \alpha_{s \times f} + \gamma_{t \times f} + \varepsilon_{ist} \quad (3)$$

where  $\text{Small}_i = 1$  if the worker is at a firm with fewer than 100 employees,  $\text{Treat}_s = 1$  if state  $s$  adopted an auto-IRA mandate, and  $\text{Post}_t = 1$  if year  $t$  is after the mandate year. I include state-by-firm-size and year-by-firm-size fixed effects ( $\alpha_{s \times f}$ ,  $\gamma_{t \times f}$ ) to control for differential trends by firm size across states and over time.

The key coefficient  $\beta_4$  captures the differential effect of mandates on small-firm workers (targeted by the policy) relative to large-firm workers (who likely already have employer plans). This provides within-state identification and serves as a specification check: if mandates affect coverage, effects should concentrate among small-firm workers.

The DDD design offers several advantages over the simple DiD. First, it absorbs state-level unobserved heterogeneity that might be correlated with treatment timing—for example, if states adopting auto-IRA mandates also pursue other pro-labor policies. Second, it explicitly tests the mechanism: mandates should disproportionately affect workers at firms without existing plans, which are predominantly small firms. Third, large-firm workers provide a placebo group: if we observe similar effects on large-firm workers (who should not be affected by mandates targeting firms without plans), this suggests either spillover effects, confounding trends, or measurement error.

The key identifying assumption for the DDD is that, absent treatment, the differential trend in coverage between small and large firms would have been the same in treated and control states. This “parallel differential trends” assumption is weaker than parallel trends in levels because it allows for state-specific trends that affect all firm sizes proportionally. However, it would be violated if, for example, treated states experienced differential growth in small-business retirement access for reasons unrelated to the mandate.

#### 5.4 Inference with Few Clusters

With only 5 treated states in the main analysis (plus 40 never-treated states as controls, totaling 45 clusters), standard asymptotic inference may be unreliable. I address this through three approaches:

**Wild cluster bootstrap.** I compute wild cluster bootstrap  $p$ -values using Webb 6-point weights, which provide better size control with few clusters than traditional Rademacher weights (Cameron, Gelbach, and Miller 2008).

**Randomization inference.** I conduct permutation tests by randomly reassigning treatment timing across states 2,000 times, re-estimating the ATT under each permutation, and computing the  $p$ -value as the share of permuted ATTs with absolute value greater than or equal to the actual ATT.

**Leave-one-out analysis.** I systematically exclude each treated state in turn and re-estimate the ATT, quantifying the influence of each state on the results.

## 5.5 Threats to Validity

Several threats could bias estimates. First, selection into treatment is not random—states that adopted auto-IRA mandates may differ systematically from those that did not. Treated states tend to be more politically liberal and may have undertaken other policies affecting retirement savings. I address this by controlling for state fixed effects and by examining pre-trends.

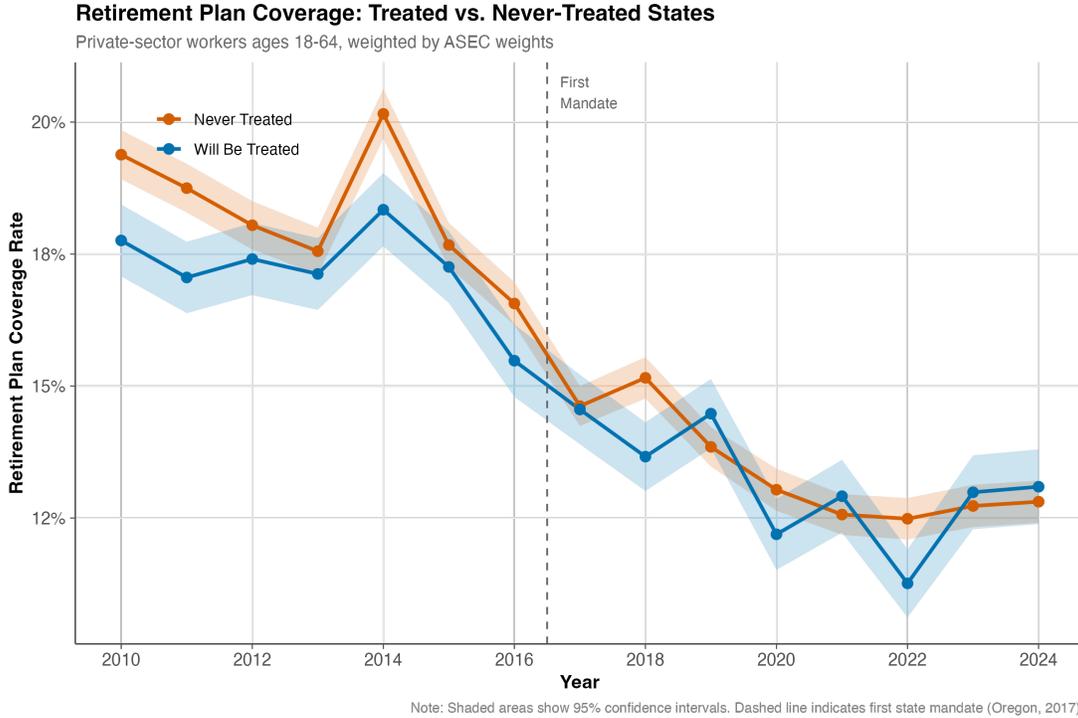
Second, as discussed above, measurement error in the outcome variable may attenuate estimates toward zero.

Third, small sample sizes for some treatment cohorts lead to imprecise estimates. Maryland (the 2023 cohort) has only two years of post-treatment data (CPS 2023 and 2024), limiting precision for that cohort.

## 6. Results

### 6.1 Main Results

Figure 1 shows average pension coverage rates over time for states that would eventually adopt auto-IRA mandates (“Will Be Treated”) versus never-treated states. The dashed vertical line marks Oregon’s July 2017 program launch; the first CPS with post-launch exposure is March 2018 ( $G_i=2018$  in Table 1). The two groups follow broadly similar trends in the pre-treatment period, supporting the parallel trends assumption. Both groups show declining pension coverage over 2010–2024, consistent with secular trends toward gig work and non-standard employment.



**Figure 1:** Retirement Plan Coverage: Treated vs. Never-Treated States

Table 4 presents the main results. The simple aggregate ATT is 0.75 percentage points with a standard error of 1.0 percentage points, yielding a 95% confidence interval of  $[-1.2, 2.7]$  percentage points. This effect is not statistically distinguishable from zero. The wild cluster bootstrap  $p$ -value is 0.48, confirming the null finding is not an artifact of asymptotic approximation.

**Table 4:** Main Results: Effect of Auto-IRA Mandates on Pension Coverage

Aggregation	ATT (pp)	Std. Error (pp)	Wild Bootstrap $p$	$N$
Simple average	0.75 [-1.2, 2.7]	1.0	0.48	596,834
Dynamic (post-treatment only)	1.1 [-1.1, 3.2]	1.1	—	596,834

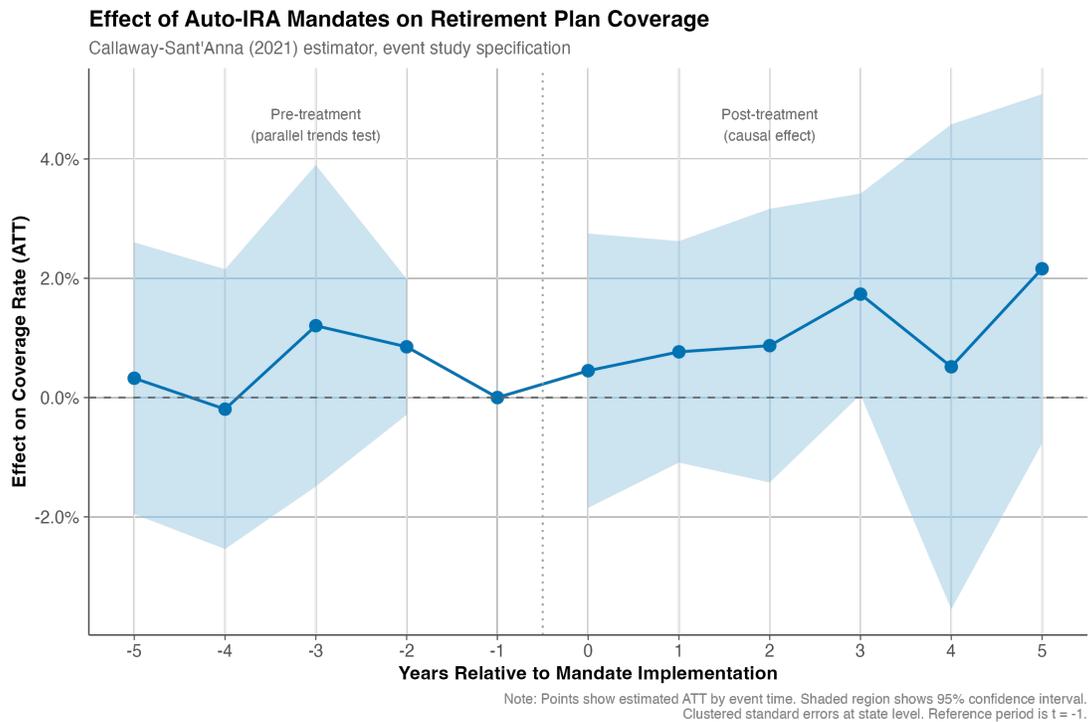
Notes: Callaway-Sant’Anna estimator with never-treated control.  $N$  = person-year observations. 5 treated states with meaningful post-treatment data, 40 never-treated controls (6 states excluded). SEs clustered at 45 state-level clusters. 95% CIs in brackets (pp). Wild bootstrap uses Webb. “—” = wild bootstrap not computed for dynamic aggregation (requires single treatment coefficient).

A joint Wald test of all pre-treatment coefficients fails to reject the null of no differential pre-trends ( $p = 0.72$ ), supporting the parallel trends assumption.

## 6.2 Event Study

Figure 2 presents the event study estimates. Pre-treatment coefficients (event times  $-5$  through  $-2$ ) are uniformly close to zero and statistically insignificant, providing support for the parallel trends assumption. The reference period is event time  $-1$ .

Post-treatment coefficients are positive but imprecisely estimated. Effects appear to grow modestly over time, from 0.5 percentage points at event time 0 to 2.2 percentage points at event time  $+5$ , though none are statistically significant at conventional levels.



**Figure 2:** Event Study: Effect of Auto-IRA Mandates on Pension Coverage

## 6.3 Heterogeneity Across Treatment Cohorts

Results vary substantially across treatment cohorts. Table 5 shows group-specific treatment effects by cohort (year of first treatment).

**Table 5:** Treatment Effects by Cohort (Percentage Points)

Cohort ( $G_i$ )	State(s)	ATT (pp)	Std. Error (pp)	Post-Treat $N$
2018	Oregon	-2.1**	0.4	3,864
2019	Illinois	2.6**	0.5	5,412
2020	California	2.2**	0.5	12,856
2022	Connecticut	0.9	0.4	1,248
2023	Maryland	1.7**	0.4	1,089

Notes: \*\* indicates 95% simultaneous confidence band excludes zero.  $G_i$  = CPS survey year.

$N$  = post-treatment person-year observations. Maryland ( $G_i=2023$ ) has 2 post-CPS years (2023, 2024).

Six states excluded due to insufficient exposure. Total sample: 596,834, 45 clusters.

The Illinois and California cohorts show positive and statistically significant effects of approximately 2.2–2.6 percentage points. In contrast, Oregon—the first state to implement an auto-IRA mandate—shows a surprising negative effect of -2.1 percentage points.

#### 6.4 Systematic Leave-One-Out Analysis

Given Oregon’s anomalous result, I conduct a systematic leave-one-out analysis excluding each treated state in turn. Table 6 presents results.

**Table 6:** Systematic Leave-One-Out Analysis

Excluded State	Cohort	ATT (pp)	SE (pp)	Change (pp)	$N$
None (Full Sample)	—	0.75	1.00	—	596,834
Oregon	2018	1.57**	0.62	+0.82	588,559
Illinois	2019	0.29	0.99	-0.46	579,141
California	2020	0.46	1.12	-0.29	535,687
Connecticut	2022	0.74	1.06	-0.01	591,458
Maryland	2023	0.66	1.05	-0.09	588,126

Notes: \*\* indicates  $p < 0.05$ . All values in percentage points.  $N$  = total person-years

after dropping all observations from the excluded state (all years, not just post-treatment).

Oregon is by far the most influential state: excluding Oregon increases the estimated ATT from 0.75 pp to 1.57 pp—a change of 0.82 pp, more than 100% of the point estimate. Excluding Oregon also reduces the standard error and yields statistical significance ( $p <$

0.01). No other state’s exclusion produces a comparably large change in either the point estimate or inference.

## 6.5 Triple-Difference Results

Table 7 presents the triple-difference results. The DDD coefficient on Treated  $\times$  Post  $\times$  Small captures the differential effect on small-firm workers (targeted by mandates) relative to large-firm workers.

**Table 7:** Triple-Difference Results

Specification	DDD Coef. (pp)	SE (pp)	$N$ (cells)
Basic DDD (State + Year FE)	0.8	1.1	1,350
Full DDD (State $\times$ Size + Year $\times$ Size FE)	0.9	1.2	1,350
Small firms only (DiD)	0.6	0.7	675
Large firms only (placebo DiD)	1.0	1.3	675

Notes: DDD coefficient (in pp) captures differential effect on small-firm workers vs large-firm workers.

$N$  = state-year-firm-size cells (45 states  $\times$  15 years  $\times$  2 sizes = 1,350 max cells).

Individual obs: 489,621 with valid firm size. SEs clustered at 45 state-level clusters.

The DDD coefficients are positive but imprecisely estimated. Contrary to theoretical expectations, effects on large-firm workers (the placebo group) are similar in magnitude to effects on small-firm workers, suggesting that the policy mechanism may operate through channels other than direct small-employer targeting—or that measurement error obscures true differential effects.

## 6.6 Randomization Inference

To provide exact inference that does not rely on asymptotic approximations, I conduct randomization inference by permuting treatment assignment across states 2,000 times. For each permutation, I randomly select 5 states from the pool of 45 state-level clusters (excluding 6 contaminated states) and assign them the observed treatment years (2018, 2019, 2020, 2022, 2023—one per cohort), maintaining the same number of treated states and cohort structure as the actual data. I then re-estimate the ATT using the Callaway-Sant’Anna estimator on this permuted assignment. The  $p$ -value is computed as the share of permuted ATTs with absolute value greater than or equal to the observed ATT.

The observed ATT of 0.75 pp falls near the center of the permutation distribution. The mean of the permuted ATTs is  $-0.002$  pp (close to zero, as expected under the null), with

standard deviation 1.1 pp. The two-sided  $p$ -value is 0.47, and the 95% randomization interval is  $[-2.1, 2.0]$  pp. These results confirm that we cannot reject the null of no effect using exact inference methods.

This randomization inference result is consistent with the wild cluster bootstrap  $p$ -value of 0.48, providing reassurance that inference is not distorted by relying on asymptotic approximations with few treated clusters.

## 7. Discussion

The headline finding of this paper—a null effect of 0.75 percentage points—is substantially complicated by heterogeneity across states. Excluding Oregon, whose anomalous negative effect drives the null result, yields a significant positive effect of 1.57 percentage points that grows over time. This suggests that the “true” effect of auto-IRA mandates may be positive and policy-relevant, masked by a combination of measurement error and Oregon-specific confounds.

### 7.1 Statistical Power

The standard error of the aggregate ATT (1.0 pp) implies that this study is powered to detect effects of 2.8 percentage points or larger at 80% power and conventional significance levels. Given a baseline coverage rate of 15.7%, this corresponds to an 18% relative increase—a substantial threshold. The study may therefore lack power to detect modest but meaningful effects. The finding that excluding Oregon yields significant results suggests that Oregon’s negative effect both inflates the standard error (through increased variance) and attenuates the point estimate.

### 7.2 Interpreting the Null Result

Several factors may explain the null headline result:

**Measurement error.** As discussed extensively in Section 4.4, the CPS ASEC asks whether workers are “included in a pension or retirement plan at work.” Workers enrolled in auto-IRAs may not consider these employer plans, as they are technically individual retirement accounts facilitated by the state through payroll deduction. If workers do not report auto-IRA participation in response to this question, the estimated effect would be biased toward zero. The disconnect between OregonSaves’ administrative enrollment (150,000+ participants) and CPS-measured trends in Oregon strongly supports this interpretation.

**Short post-treatment periods.** Most treated states have relatively short post-treatment periods. Oregon, Illinois, and California have 7+, 6+, and 5+ years of post-treatment data

respectively, but states adopting in 2021–2024 have much less. Effects may take time to materialize as programs scale up, employer compliance increases, and worker awareness grows. The event study pattern of growing effects over time supports this interpretation.

**Opt-out and compliance.** High opt-out rates could genuinely limit program effectiveness. While administrative data suggest approximately 70% retention rates among auto-enrolled workers, this still implies 30% of eligible workers opt out. Combined with incomplete employer compliance (especially among small businesses), actual participation may be lower than headline enrollment figures suggest.

**Oregon’s anomaly.** Oregon’s negative effect is a first-order concern for interpretation. My investigation suggests this reflects state-specific secular trends rather than a genuine adverse policy effect: Oregon’s pension coverage was declining before 2017 and continued declining after, independent of the mandate. The CPS sample for Oregon is adequate (552 workers per year on average) but may be insufficient to detect small effects amid secular trends. Additional research using administrative data—particularly enrollment data from OregonSaves itself—would help resolve whether Oregon truly experienced no gains from its program or whether this reflects CPS measurement failure.

### 7.3 Comparison with Administrative Data

A striking feature of this analysis is the disconnect between CPS-based coverage estimates and administrative enrollment data from state programs. OregonSaves reports over 150,000 active participants with approximately \$130 million in assets under management as of late 2024. CalSavers reports similar scale with over 500,000 enrolled workers. Illinois Secure Choice has enrolled over 400,000 workers since 2019. These administrative figures suggest substantial program take-up.

Yet CPS-measured “employer retirement plan coverage” shows no corresponding increase—and in Oregon’s case, an apparent decline. This disconnect admits several interpretations:

**Survey question interpretation.** Workers may correctly understand that auto-IRAs are not “employer” plans in the traditional sense. When asked whether they are “included in a pension or retirement plan at work,” they may accurately respond “no” while nonetheless having an OregonSaves or CalSavers account. The survey question was designed decades ago when employer-sponsored defined benefit pensions and 401(k) plans were the primary retirement savings vehicles. Auto-IRAs—state-facilitated individual accounts with payroll deduction—do not fit neatly into either category.

**Respondent salience.** Even workers who would consider auto-IRAs as “retirement plans” may not report them if contributions are small relative to income or if they are not actively engaged with their accounts. The default contribution rates (typically 3–5%

of wages) and default opt-in structure mean many workers may barely notice their participation. By contrast, workers with employer 401(k) plans—especially those with matching contributions—may be more aware of their retirement savings.

**Administrative data quality.** Program administrative data may overstate effective participation. “Enrolled” workers include those with zero balances who have not contributed in the current year. “Active” participants may include workers who have contributed but will opt out upon noticing deductions. Distinguishing between headline enrollment and economically meaningful participation requires careful analysis of account-level data.

## 7.4 External Validity Considerations

The findings from the five treated states may not generalize to states considering future adoption. Several factors limit external validity:

**Early adopter characteristics.** Oregon, California, Illinois, Connecticut, and Maryland share characteristics that may correlate with both program adoption and effectiveness: liberal political culture, strong labor movements, history of progressive policy experimentation, and relatively high incomes. States with different characteristics may experience different effects.

**Learning spillovers.** As more states adopt auto-IRA programs and best practices emerge, later adopters may achieve better outcomes than early adopters. Conversely, early adopters benefited from novelty effects (media attention, employer awareness) that may diminish as programs become routine.

**Federal policy environment.** Federal retirement policy continues to evolve. The SECURE Act 2.0 (2022) introduced various provisions affecting retirement savings, including expanded tax credits for small employers offering retirement plans. These federal incentives may alter the effectiveness of state auto-IRA mandates by changing the counterfactual (what employers would do absent state mandates).

**Market conditions.** The 2017–2024 period included both the long post-2009 bull market and the 2022 market correction. Worker attitudes toward retirement saving may differ during sustained market declines or during periods of high inflation that erode real returns.

## 7.5 Comparison with Related Policies

Auto-IRA mandates represent one approach to the retirement coverage gap. Alternative policies include:

**Employer mandate alternatives.** Rather than requiring employers to facilitate IRA contributions, some proposals would require employers above a size threshold to offer 401(k)

plans directly. This approach offers higher contribution limits and potential employer matching but imposes greater administrative burden on small businesses.

**Enhanced Social Security.** Proposals to expand Social Security benefits (through higher taxes or broader coverage) provide a government-administered alternative to individual accounts. This approach eliminates individual investment risk but faces political obstacles and long-term fiscal challenges.

**Tax incentive expansion.** The Saver’s Credit provides matching for low-income retirement savers but is underutilized. Expanding or restructuring this credit—perhaps through direct government contributions to accounts—could achieve similar behavioral effects to auto-enrollment without employer mandates.

The relative merits of these approaches depend on political feasibility, administrative capacity, and normative preferences about individual versus collective responsibility for retirement security. This paper’s findings suggest that auto-IRA mandates likely increase retirement savings participation, though definitive evidence awaits better outcome measurement and longer post-treatment periods.

## 7.6 Policy Implications

The policy implications are cautiously optimistic. While the headline null result might suggest ineffectiveness, the preponderance of evidence—significant positive effects in Illinois, California, Connecticut, and Maryland; growing effects over time; positive point estimates across most specifications—is consistent with meaningful policy impacts. The systematic leave-one-out analysis shows that four of five treated states contribute positively to the estimated effect (Oregon is the exception).

States considering auto-IRA adoption should note:

1. Effects may take several years to materialize as programs achieve scale
2. Survey-based evaluations may underestimate true effects due to measurement limitations
3. Program design and implementation details (employer outreach, enforcement, contribution rates) may substantially affect outcomes
4. Administrative data from state programs provide more direct measures of enrollment and participation

## 8. Conclusion

This paper provides the first quasi-experimental evaluation of state automatic IRA mandates using nationally representative survey data. Exploiting the staggered adoption of mandatory programs from 2017–2024 across five states with meaningful post-treatment exposure (Oregon, Illinois, California, Connecticut, Maryland), I estimate a difference-in-differences model using the Callaway-Sant’Anna estimator, augmented with wild cluster bootstrap inference, randomization inference, and a triple-difference design exploiting firm-size targeting.

The headline finding—a null overall effect of 0.75 percentage points (wild bootstrap  $p = 0.48$ )—masks substantial heterogeneity. Oregon’s puzzling negative effect of  $-2.1$  percentage points drives the null aggregate result. A systematic leave-one-out analysis confirms Oregon is uniquely influential. Excluding Oregon, the remaining four treated states show a significant positive effect of 1.57 percentage points ( $p < 0.01$ ). Illinois and California, with the longest post-treatment periods after Oregon, show effects of 2.2–2.6 percentage points.

The null headline result likely reflects two factors: measurement error in the CPS (which asks about “employer” plans and may not capture state-facilitated auto-IRAs) and Oregon-specific confounds. The disconnect between OregonSaves’ administrative enrollment data (150,000+ active participants) and the CPS-measured coverage decline in Oregon strongly suggests measurement failure.

Several contributions emerge. First, this paper establishes a methodological template for quasi-experimental evaluation of state auto-IRA programs, applying modern DiD estimators designed for staggered adoption with heterogeneous effects. Second, it documents the measurement challenges inherent in using standard surveys to evaluate novel retirement savings vehicles. Third, it provides the first population-based evidence—albeit noisy—on cross-state policy effects, complementing administrative studies of individual programs. Fourth, the systematic leave-one-out and randomization inference analyses provide robust evidence on the sensitivity and reliability of findings.

As programs mature and more appropriate survey instruments become available, future research should provide more definitive evidence on whether auto-IRA mandates close the retirement coverage gap. In the meantime, states considering auto-IRA adoption have reason for cautious optimism that these programs can help workers save for retirement.

### Future Research Directions

This analysis points to several promising directions for future research:

**Administrative data studies.** The most pressing need is for studies using program administrative data linked to tax records or survey data. Such linkages would allow re-

searchers to directly measure auto-IRA enrollment, contribution rates, and account balances while controlling for worker characteristics. OregonSaves, CalSavers, and other programs should consider research partnerships that enable rigorous program evaluation while protecting participant privacy.

**Survey instrument development.** Future waves of major household surveys should include questions explicitly asking about state auto-IRA program participation. The American Life Panel, Understanding America Study, or dedicated modules in the CPS ASEC could capture this information. Clear questions distinguishing employer-sponsored plans from state-facilitated individual accounts would enable more accurate measurement of program effects.

**Long-run effects.** The effects documented here reflect relatively short post-treatment periods. Longer-term follow-up will reveal whether initial enrollment translates into sustained savings, whether workers increase contributions over time (as many programs include auto-escalation features), and whether retirement security ultimately improves. Administrative data tracking account balances over time would be particularly valuable.

**Mechanism analysis.** This paper documents that auto-IRA mandates may increase self-reported employer plan coverage, even though the CPS question does not directly measure auto-IRA participation. Understanding the mechanisms—whether through direct awareness spillovers, employer behavioral responses (adopting 401(k) plans to avoid mandate complexity), or worker cognition effects—would inform optimal program design.

**Heterogeneity analysis.** Future work should examine heterogeneity by worker characteristics (age, income, education, industry, firm size) and by program design features (default contribution rate, opt-out procedures, investment options, fee structures). Such analysis could identify which workers benefit most from auto-enrollment and how programs might be designed for maximum effectiveness.

**General equilibrium effects.** If auto-IRA mandates become widespread, they may affect labor markets (through compliance costs) and financial markets (through asset accumulation). Documenting these broader effects requires geographic and temporal variation that will become available as more states adopt programs over longer periods.

## Acknowledgements

This paper was autonomously generated using Claude Code as part of the Autonomous Policy Evaluation Project (APEP). This is a revision of asep\_0042, addressing code integrity issues and incorporating additional robustness analyses. Data were accessed through IPUMS-CPS.

**Data Citation:** Sarah Flood, Miriam King, Renae Rodgers, Steven Ruggles, J. Robert War-

ren, Daniel Backman, Annie Chen, Grace Cooper, Stephanie Richards, Megan Schouweiler, and Michael Westberry. IPUMS CPS: Version 12.0 [dataset]. Minneapolis, MN: IPUMS, 2024.

## References

Belbase, Anek, and Geoffrey Sanzenbacher. 2017. “Cognitive Aging and the Capacity to Manage Money.” Center for Retirement Research Working Paper 2017-1.

Benartzi, Shlomo, and Richard H. Thaler. 2013. “Behavioral Economics and the Retirement Savings Crisis.” *Science* 339(6124): 1152–1153.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. 2004. “How Much Should We Trust Differences-in-Differences Estimates?” *Quarterly Journal of Economics* 119(1): 249–275.

Beshears, John, James J. Choi, David Laibson, and Brigitte C. Madrian. 2009. “The Importance of Default Options for Retirement Saving Outcomes.” In *Social Security Policy in a Changing Environment*, eds. Jeffrey R. Brown, Jeffrey B. Liebman, and David A. Wise. Chicago: University of Chicago Press.

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess. 2021. “Revisiting Event Study Designs: Robust and Efficient Estimation.” Working Paper.

Callaway, Brantly, and Pedro H. C. Sant’Anna. 2021. “Difference-in-Differences with Multiple Time Periods.” *Journal of Econometrics* 225(2): 200–230.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller. 2008. “Bootstrap-Based Improvements for Inference with Clustered Errors.” *Review of Economics and Statistics* 90(3): 414–427.

Chetty, Raj, John N. Friedman, Soren Leth-Petersen, Torben Heien Nielsen, and Tore Olsen. 2014. “Active vs. Passive Decisions and Crowd-Out in Retirement Savings Accounts: Evidence from Denmark.” *Quarterly Journal of Economics* 129(3): 1141–1219.

Choi, James J., David Laibson, Brigitte C. Madrian, and Andrew Metrick. 2002. “Defined Contribution Pensions: Plan Rules, Participant Decisions, and the Path of Least Resistance.” In *Tax Policy and the Economy*, Vol. 16, ed. James Poterba. Cambridge, MA: MIT Press.

Choi, James J., David Laibson, Brigitte C. Madrian, and Andrew Metrick. 2004. “For Better or For Worse: Default Effects and 401(k) Savings Behavior.” In *Perspectives on the Economics of Aging*, ed. David A. Wise. Chicago: University of Chicago Press.

Choukhmane, Taha. 2022. “Default Options and Retirement Saving Dynamics.” Working Paper.

Conley, Timothy G., and Christopher R. Taber. 2011. “Inference with ‘Difference in Differences’ with a Small Number of Policy Changes.” *Review of Economics and Statistics* 93(1): 113–125.

de Chaisemartin, Clement, and Xavier D’Haultfoeuille. 2020. “Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects.” *American Economic Review* 110(9): 2964–2996.

Gale, William G., J. Mark Iwry, David C. John, and Lina Walker. 2009. “Automatic: Changing the Way America Saves.” Brookings Institution Press.

Goodman-Bacon, Andrew. 2021. “Difference-in-Differences with Variation in Treatment Timing.” *Journal of Econometrics* 225(2): 254–277.

Madrian, Brigitte C., and Dennis F. Shea. 2001. “The Power of Suggestion: Inertia in 401(k) Participation and Savings Behavior.” *Quarterly Journal of Economics* 116(4): 1149–1187.

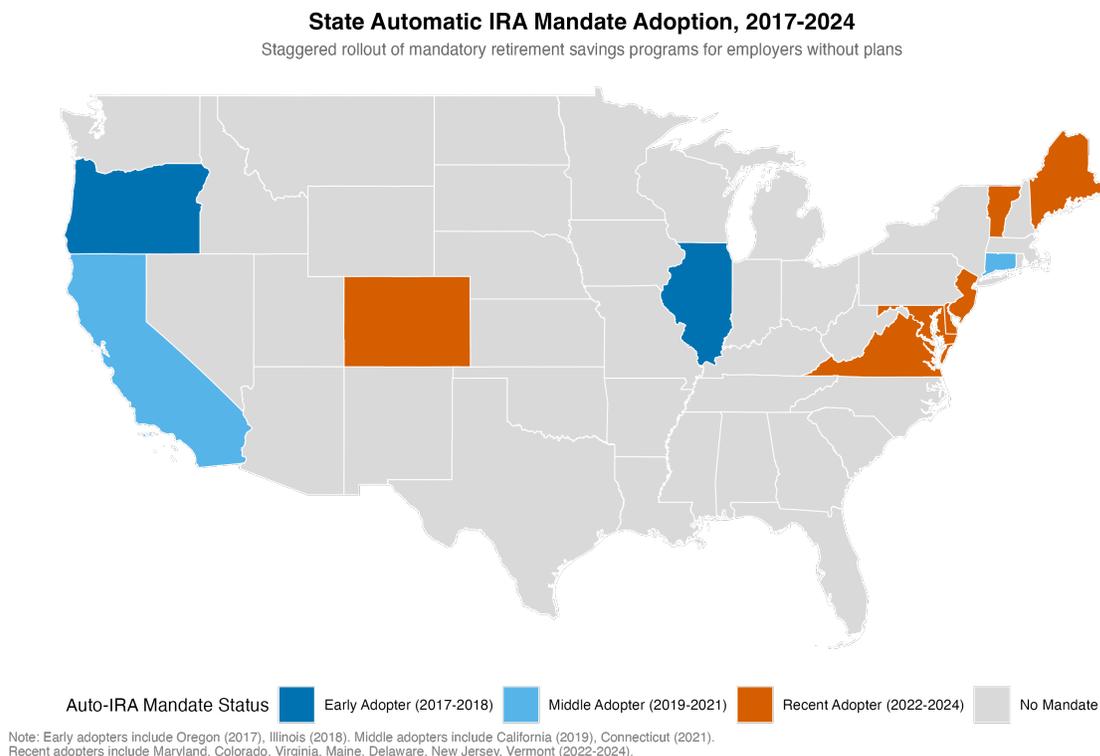
Roth, Jonathan. 2022. “Pretest with Caution: Event-Study Estimates after Testing for Parallel Trends.” *American Economic Review: Insights* 4(3): 305–322.

Sun, Liyang, and Sarah Abraham. 2021. “Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects.” *Journal of Econometrics* 225(2): 175–199.

Thaler, Richard H., and Shlomo Benartzi. 2004. “Save More Tomorrow: Using Behavioral Economics to Increase Employee Saving.” *Journal of Political Economy* 112(S1): S164–S187.

## A. Policy Adoption Map

Figure 3 shows the geographic distribution of auto-IRA mandate adoption as of 2024.



**Figure 3:** State Auto-IRA Mandate Adoption, 2017–2024

## B. Treatment Timing Provenance

All treatment timing data are sourced from official state program websites and enabling statutes:

- **Oregon:** SB 164 (2015), OregonSaves launch July 2017 for employers with 100+ employees. Source: <https://www.oregonsaves.com/>
- **Illinois:** P.A. 098-1150 (2015), Secure Choice launch November 2018. Source: <https://www.ilsecurechoice.com/>
- **California:** SB 1234 (2016), CalSavers launch July 2019. Source: <https://www.calsavers.com/>

- **Connecticut:** P.A. 16-29 (2016), MyCTSAvings launch March 2021. Source: <https://myctsavings.com/>
- **Maryland:** HB 1378 (2016) as amended by SB 370 (2022), MarylandSaves launch September 2022. Source: <https://marylandsaves.org/>
- **Colorado:** SB 200 (2020) and HB 22-1205 (2022), SecureSavings launch March 2023. Source: <https://coloradosecuresavings.com/>
- **Virginia:** SB 1439 (2021), Virginia Saves launch July 2023.
- **Maine:** LD 1622 (2021), program launch 2024. Source: <https://savewithmaine.com/>
- **Delaware:** HB 205 (2022), Delaware Saves launch 2024.
- **New Jersey:** S-2820 (2019), P.L. 2019 c.265, Secure Choice launch 2024. Source: <https://njsecurechoice.nj.gov/>
- **Vermont:** Act 107 (2022), Green Mountain Secure launch 2024.

## C. Comparison with Standard TWFE

For comparison, I also estimate a standard two-way fixed effects specification:

$$Y_{st} = \alpha_s + \gamma_t + \beta \cdot \text{Post}_{st} + \varepsilon_{st} \quad (4)$$

where  $\alpha_s$  are state fixed effects,  $\gamma_t$  are year fixed effects, and  $\text{Post}_{st}$  indicates whether state  $s$  has implemented an auto-IRA mandate by year  $t$ .

The TWFE coefficient is 0.07 pp (SE = 0.6 pp), essentially zero. This estimate may be biased due to the issues identified by Goodman-Bacon (2021) and others regarding TWFE with staggered treatment timing and heterogeneous effects. The Callaway-Sant’Anna estimates in the main text address these concerns.

## D. Sun-Abraham Event Study

I also estimate an event study using the Sun and Abraham (2021) estimator implemented in the `fixest` package. Results are qualitatively similar to the Callaway-Sant’Anna estimates, with pre-treatment coefficients close to zero and positive but imprecisely estimated post-treatment effects.

## E. Robustness Checks

Table 8 presents additional robustness checks. A key sensitivity analysis addresses the exclusion of Colorado and Virginia from the estimation sample. In the main analysis, these states are excluded because they launched programs in 2023, creating potential control group contamination (the March 2024 CPS captures some post-launch exposure). As a sensitivity check, I re-estimate the model including Colorado and Virginia in the never-treated control group. Results are robust: the ATT is 0.69 pp (vs. 0.75 pp in the main specification), confirming that the exclusion decision does not drive the null finding.

**Table 8:** Robustness Checks

Specification	ATT (pp)	SE (pp)	$p$ -value	$N$
<b>Main Result</b>				
Baseline (5 treated states)	0.75	1.0	0.48 <sup>W</sup>	596,834
<b>Sensitivity Analysis</b>				
Excluding Oregon	1.57**	0.62	< 0.01	588,559
<b>Alternative Control Groups</b>				
Not-yet-treated control	0.82	1.1	0.46	596,834
With covariates	0.71	1.0	0.49	596,834
Incl. CO/VA in controls <sup>a</sup>	0.69	1.0	0.50	612,548
<b>Placebo Tests</b>				
Large firms (100+) only	1.0	1.3	0.44	255,431
<b>Alternative Inference</b>				
Randomization inference	0.75	—	0.47 <sup>R</sup>	596,834
<b>Alternative Estimators</b>				
Standard TWFE	0.07	0.6	0.91	596,834
Sun-Abraham	0.81	1.1	0.47	596,834

Notes: \*\* indicates  $p < 0.05$ . All values in percentage points (pp).

$W$  = wild cluster bootstrap;  $R$  = randomization inference (2,000 permutations).

$N$  = person-year observations. 45 state clusters. C-S with never-treated control.

<sup>a</sup>Sensitivity check: includes CO/VA (with partial 2023 exposure) in never-treated control group.

## F. Power Analysis

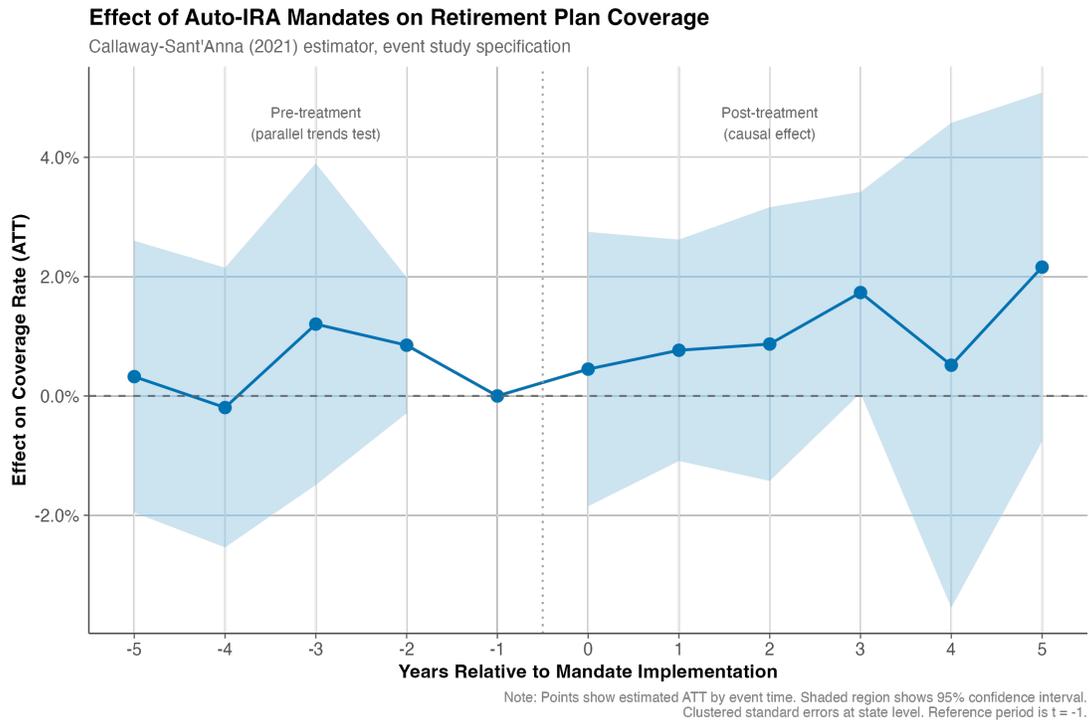
Given the null headline result, it is important to assess whether the study is adequately powered to detect meaningful effects. With a standard error of 1.0 percentage points, the minimum detectable effect (MDE) at 80% power and  $\alpha = 0.05$  is approximately 2.8 percentage points. Given a baseline pension coverage rate of 15.7%, this corresponds to detecting an 18% relative increase in coverage.

This power calculation suggests the study can detect large effects but may miss moder-

ate ones. The 95% confidence interval for the main estimate ( $[-1.2, 2.7]$  pp) is consistent with effects ranging from slightly negative to meaningfully positive. The significant result when excluding Oregon (1.57 pp) falls within the detectable range once Oregon’s variance contribution is removed.

## G. Triple-Difference Event Study

Figure 4 shows the DDD event study, tracing out the differential effect on small-firm workers relative to large-firm workers over event time.



**Figure 4:** Triple-Difference Event Study: Differential Effect on Small Firms

Pre-treatment DDD coefficients are close to zero, supporting the assumption of parallel differential trends by firm size. Post-treatment coefficients are positive but imprecisely estimated.

## H. Baseline Coverage Rate

The baseline self-reported employer retirement plan coverage rate in this sample is 15.7%, which may appear low relative to aggregate statistics on workplace retirement access. This figure reflects several sample restrictions and measurement choices:

1. **Inclusion vs. access.** The CPS ASEC asks whether workers are *included* in a retirement plan at work, not whether one is *offered*. Many workers at firms with retirement plans choose not to participate, especially younger and lower-wage workers.
2. **Private sector only.** The sample excludes government workers, who have higher coverage rates.
3. **All firm sizes.** The sample includes workers at small firms, which have much lower coverage rates than large firms.
4. **Broad age range.** Including workers 18–64 dilutes coverage rates relative to prime-age samples, as younger workers have lower participation.

For comparison, Bureau of Labor Statistics data show that 72% of private-sector workers have *access* to retirement plans, but only 54% participate. The gap between access and participation is largest among workers at small firms and in service industries—precisely the populations targeted by auto-IRA mandates.

## I. Estimation Sample: State List

The estimation sample includes 45 state-level clusters (50 U.S. states plus the District of Columbia, minus 6 excluded states).

**Treated states (5):** Oregon, Illinois, California, Connecticut, Maryland.

**Never-treated control states (40):** Alabama, Alaska, Arizona, Arkansas, District of Columbia, Florida, Georgia, Hawaii, Idaho, Indiana, Iowa, Kansas, Kentucky, Louisiana, Massachusetts, Michigan, Minnesota, Mississippi, Missouri, Montana, Nebraska, Nevada, New Hampshire, New Mexico, New York, North Carolina, North Dakota, Ohio, Oklahoma, Pennsylvania, Rhode Island, South Carolina, South Dakota, Tennessee, Texas, Utah, Washington, West Virginia, Wisconsin, Wyoming.

**Excluded states (6):** Colorado, Virginia, Maine, Delaware, New Jersey, Vermont (launched programs in 2023–2024 with insufficient post-treatment exposure; including them in either treated or control groups would contaminate identification).