Do State College Promise Programs Increase Enrollment? Evidence from Staggered Adoption

APEP Autonomous Research*

@olafdrw

February 6, 2026

Abstract

State "Promise" programs that provide free community college tuition have spread rapidly across the United States, reaching 20 states by 2021. Despite the policy's popularity, aggregate enrollment effects remain unclear. Using a staggered difference-in-differences design with Callaway-Sant'Anna estimators and state-level enrollment data from the American Community Survey (2010–2023), I find that Promise programs have *no statistically significant effect* on aggregate college enrollment (ATT $= -0.0136$ log points, SE $= 0.0102$, $p > 0.05$). Event study estimates show no differential pre-trends and no significant post-treatment effects through seven years after adoption. However, power analysis with 20 treated states and 31 control states reveals a minimum detectable effect of 29%, suggesting the null finding may reflect insufficient statistical power rather than true absence of effects. These results highlight the limitations of state-aggregate data for evaluating Promise programs and suggest that previous findings of large enrollment increases may derive from compositional shifts (community college vs. four-year) rather than net enrollment gains.

JEL Codes: I22, I23, I28, H75
Keywords: college promise programs, free tuition, higher education policy, difference-in-differences, staggered adoption

*Autonomous Policy Evaluation Project. Correspondence: scl@econ.uzh.ch

1. Introduction

Over the past decade, state "Promise" programs offering free community college tuition have become one of the most prominent higher education policy innovations in the United States. Beginning with Tennessee's Promise scholarship in 2015, these programs have spread to over 20 states by 2021, with cumulative public investment exceeding $2 billion annually (Perna and Leigh, 2017). Proponents argue that eliminating tuition barriers will dramatically increase college access, particularly for low-income students who face financial constraints (Dynarski, 2003). The rapid diffusion of Promise programs reflects a bipartisan consensus that expanding postsecondary education is essential for economic mobility and workforce development (Carnevale et al., 2010).

Despite their popularity, the aggregate effects of Promise programs on college enrollment remain surprisingly unclear. Early studies of individual programs, particularly Tennessee Promise, found substantial increases in community college enrollment (Carruthers and Fox, 2020). However, these single-state evaluations cannot distinguish program effects from state-specific trends, and the generalizability of Tennessee's experience to other contexts is uncertain. Moreover, enrollment increases at community colleges may partially reflect diversion from four-year institutions rather than net gains in postsecondary participation (Cohodes and Goodman, 2017).

This paper provides the first comprehensive multi-state evaluation of Promise program effects on aggregate college enrollment using modern difference-in-differences methods. I exploit the staggered adoption of Promise programs across 20 states between 2015 and 2021, comparing enrollment trajectories in treated states to never-treated states using the heterogeneity-robust estimator of Callaway and Sant'Anna (2021). The primary outcome is total undergraduate enrollment at the state level, measured using the American Community Survey from 2010 to 2023.

The main finding is that I cannot detect a statistically significant effect: Promise programs do not appear to significantly increase aggregate state-level college enrollment, though the study's power limitations mean small effects cannot be ruled out. The overall average treatment effect on the treated (ATT) is -0.0136 log points (approximately -1.4%), with a standard error of 0.0102 and a 95% confidence interval spanning $[-0.0337, 0.0064]$. Event study estimates reveal no differential pre-trends, supporting the parallel trends assumption, and no significant post-treatment effects through seven years after program adoption. The null finding is robust to alternative control group specifications, randomization inference, and placebo tests using pre-treatment pseudo-treatment.

However, power analysis suggests important caveats. With 20 treated states and 31

control states observed over 14 years, the minimum detectable effect (MDE) at 80% power is approximately 29%. This means the study cannot reliably detect effects smaller than a 29% increase in enrollment—substantially larger than the 5–15% effects typically reported in single-state studies. The null finding may therefore reflect true absence of aggregate effects, effects too small to detect with state-level data, or compositional shifts that leave total enrollment unchanged while altering the mix of community college versus four-year enrollment.

This paper contributes to several literatures. First, I add to the growing body of work on Promise programs and free college policies (Perna and Leigh, 2017; Carruthers and Fox, 2020; Murphy et al., 2019). While previous studies focus on individual states or institutional outcomes, I provide the first multi-state aggregate evaluation using heterogeneity-robust methods. The null finding challenges the assumption that Promise programs dramatically increase overall postsecondary participation and suggests that enrollment effects may be more localized than previously thought.

Second, I contribute to the methodological literature on staggered difference-in-differences (Goodman-Bacon, 2021; Sun and Abraham, 2021; Callaway and Sant'Anna, 2021). This application demonstrates the importance of using modern estimators in policy evaluation, as traditional two-way fixed effects (TWFE) can produce biased estimates when treatment effects vary across cohorts. I find that TWFE and Callaway-Sant'Anna estimates are similar in this case, suggesting that heterogeneous treatment effects across adoption cohorts are not a major concern.

Third, I contribute to the literature on statistical power in policy evaluation (Bloom, 1995; Duflo et al., 2007). The power limitations of this analysis highlight a fundamental tension in evaluating state-level policies: aggregate data provide broad coverage but limited statistical power, while microdata offer precision but restricted generalizability. Future research should prioritize individual-level data from multiple states to obtain both internal and external validity.

The remainder of this paper proceeds as follows. Section 2 reviews the related literature. Section 3 describes the institutional background of Promise programs. Section 4 presents the data and sample construction. Section 5 details the empirical strategy. Section 6 presents results, and Section 7 concludes.

2. Related Literature

This paper relates to three main strands of literature: the effects of financial aid on college enrollment, the evaluation of Promise and free college programs, and methodological advances

in staggered difference-in-differences.

2.1 Financial Aid and College Enrollment

A large literature examines how financial aid affects college enrollment decisions. Dynarski (2003) provides foundational evidence that eliminating the Social Security Student Benefit Program reduced college attendance among eligible students by 25 percentage points. Deming and Walters (2014) find that Pell Grant generosity increases enrollment, particularly at for-profit institutions. Castleman and Page (2016) show that simplifying FAFSA completion through text message reminders increases college enrollment by 3 percentage points among low-income students.

The mechanisms through which financial aid affects enrollment are debated. Some emphasize liquidity constraints: students from low-income families cannot borrow against future earnings to finance current education (Lochner and Monge-Naranjo, 2011). Others emphasize information and complexity: students may not understand available aid or face prohibitive transaction costs in applying (Bettinger et al., 2012). Promise programs potentially address both channels by eliminating tuition (liquidity) and simplifying the aid landscape (complexity).

However, the marginal student affected by financial aid policies may differ from the average student. Bound et al. (2009) argue that increased college enrollment has been accompanied by declining completion rates, suggesting that marginal students may be less prepared for college-level work. If Promise programs primarily attract students at the margin, aggregate enrollment gains may not translate to degree attainment gains.

2.2 Promise Programs and Free College

The literature on Promise programs specifically has grown rapidly since Tennessee's 2015 launch. Carruthers and Fox (2020) evaluate Tennessee Promise using a regression discontinuity design based on the birth date cutoff for program eligibility. They find a 40% increase in community college enrollment and increased FAFSA completion rates. However, they also document "summer melt"—the gap between enrollment intentions and actual enrollment—suggesting that Promise alone may not overcome all barriers to college attendance.

Murphy et al. (2019) examine the elimination of free college in England, finding that charging tuition reduced enrollment by 15% but had little effect on the socioeconomic composition of students. This suggests that universal free tuition may not be the most efficient way to target aid toward students who would not otherwise attend.

Place-based Promise programs have also been studied extensively. Bartik et al. (2019)

evaluate the Kalamazoo Promise, finding substantial increases in college enrollment and degree attainment. Page et al. (2019) synthesize evidence from multiple Promise programs, concluding that effects depend heavily on program design features such as mentorship requirements and eligibility criteria.

My contribution extends this literature by providing the first multi-state aggregate evaluation using modern causal inference methods. While single-state studies offer internal validity, they cannot assess external validity or distinguish state-specific trends from program effects. Multi-state analysis sacrifices some precision but provides broader inference.

2.3 Staggered Difference-in-Differences

Recent methodological work has transformed the practice of difference-in-differences with staggered treatment adoption. Goodman-Bacon (2021) demonstrates that TWFE estimates are weighted averages of all possible two-group, two-period comparisons, with weights that can be negative when treatment effects vary over time. This can produce estimates with the "wrong sign" even when all component effects are positive.

Callaway and Sant'Anna (2021) propose an estimator that computes group-time average treatment effects separately for each cohort and aggregates them using appropriate weights. Sun and Abraham (2021) develop an interaction-weighted estimator that reweights TWFE coefficients to obtain consistent estimates under treatment effect heterogeneity. Borusyak et al. (2024) propose an imputation estimator that predicts counterfactual outcomes for treated units.

I implement the Callaway-Sant'Anna estimator because it naturally accommodates event study analysis and allows for multiple aggregation schemes. The estimator is doubly robust, combining outcome modeling with inverse probability weighting, which improves efficiency and robustness to model misspecification (Sant'Anna and Zhao, 2020). For a comprehensive review of recent developments in staggered DiD methodology, see Roth et al. (2023).

3. Institutional Background

3.1 The Rise of State Promise Programs

State College Promise programs emerged from the broader "free college" movement that gained momentum in the 2010s. The concept originated with place-based scholarships like the Kalamazoo Promise (2005), which guaranteed free tuition for local residents attending Michigan public universities. The success of such local programs inspired state-level adoption, beginning with Tennessee Promise in 2014.

Tennessee Promise, signed into law by Governor Bill Haslam, provided "last-dollar" scholarships covering tuition and mandatory fees at community colleges after other financial aid was applied. The program targeted recent high school graduates and required participants to maintain a 2.0 GPA, complete community service, and meet with a mentor. The first cohort enrolled in fall 2015, and early evidence suggested substantial increases in community college enrollment and FAFSA completion (Carruthers and Fox, 2020).

Following Tennessee's lead, numerous states adopted similar programs. Oregon Promise launched in 2016, followed by a wave of adoptions in 2017 including Arkansas, Hawaii, Indiana, Kentucky, Nevada, New York (Excelsior), and Rhode Island. Additional states adopted programs between 2018 and 2021, including Maryland, Washington, California, Connecticut, Delaware, Michigan, Montana, New Mexico, and West Virginia. By 2021, approximately 40% of U.S. states had some form of statewide Promise program.

3.2 Program Design Variation

While all Promise programs share the goal of reducing financial barriers to community college, they vary substantially in design. Key dimensions of variation include:

Funding mechanism: Most programs use "last-dollar" funding, covering only tuition remaining after Pell grants and other aid. This means students from the lowest-income families—who already qualify for substantial federal aid—may receive little additional benefit. A few programs (e.g., New Mexico) use "first-dollar" approaches that provide benefits regardless of other aid.

Eligibility criteria: Programs differ in age restrictions, residency requirements, GPA thresholds, and income limits. New York's Excelsior program, uniquely, covers four-year public universities but imposes income caps and requires graduates to remain in-state for a period equal to their scholarship duration.

Scope: Most Promise programs cover only community colleges, but some (New York, New Mexico) extend to four-year public institutions. This variation is important for understanding potential crowd-out effects.

Support services: Tennessee Promise requires mentorship and community service; other programs have weaker non-financial components. The intensity of "wrap-around" services may affect completion rates beyond simple enrollment effects.

This design variation creates both challenges and opportunities for identification. On one hand, heterogeneous program features may generate heterogeneous effects, complicating aggregate evaluation. On the other hand, the staggered timing of adoption across states provides quasi-experimental variation for difference-in-differences analysis.

3.3 Expected Effects and Mechanisms

Promise programs could increase college enrollment through several channels. The most direct is reducing the financial cost of attendance. For students at the margin of enrollment, eliminating tuition may tip the cost-benefit calculation toward attendance. This effect would be largest for students from middle-income families who do not qualify for substantial need-based aid but still face financial constraints.

A second channel operates through simplification and salience. The "free college" framing is more salient than complex financial aid packages. Students may not understand that existing aid covers most tuition costs for low-income families; Promise programs clarify this by guaranteeing zero tuition regardless of other aid. This information effect could increase enrollment even among students who would have received equivalent aid anyway.

Third, Promise programs may affect student expectations and aspirations earlier in the educational pipeline. If high school students know that college will be free, they may invest more in college preparation—taking rigorous courses, maintaining higher GPAs, and completing college applications. These anticipation effects could manifest before the program's official implementation.

However, Promise programs might not increase aggregate enrollment for several reasons. First, if most students already attend college and the marginal non-attender faces barriers beyond cost (academic preparation, family obligations, labor market opportunities), tuition subsidies may have limited impact. Second, Promise programs covering only community colleges might divert students from four-year institutions, leaving total enrollment unchanged. Third, "last-dollar" designs provide minimal additional aid to students already receiving Pell grants, limiting effects for the lowest-income students who face the greatest barriers.

3.4 Policy Context and Confounders

Promise programs do not exist in isolation. States adopting Promise often pursue complementary higher education reforms, potentially confounding the treatment effect. Common co-occurring policies include:

Performance-based funding: Many states have shifted higher education appropriations from enrollment-based to outcome-based formulas. If Promise adoption coincides with performance funding reforms, observed effects may conflate the two policies.

Tuition policies: Some states implemented tuition freezes or caps alongside Promise programs. These broader affordability measures could influence enrollment independent of Promise.

Articulation agreements: States may simultaneously improve transfer pathways be-

tween community colleges and universities, affecting both enrollment and completion patterns.

FAFSA initiatives: Promise programs often include FAFSA completion support or requirements, which could independently affect enrollment through improved take-up of federal aid.

Without detailed policy inventories for all 50 states, I cannot fully disentangle Promise effects from these confounders. The staggered adoption design helps by comparing states that adopted Promise at different times, but common trends in higher education policy could still bias estimates if they correlate with Promise adoption timing.

4. Data

4.1 Data Sources

The primary data source is the American Community Survey (ACS) 1-year estimates from 2010 to 2023, accessed via the Census Bureau API. The ACS provides state-level counts of college enrollment by age group, which I aggregate to total undergraduate enrollment. The unit of analysis is state-year, yielding a balanced panel of 51 states (including DC) over 14 years.

I construct treatment timing from legislative records and program documentation. For each state with a Promise program, I identify the first academic year in which students could enroll under the program. For example, Tennessee Promise was enacted in 2014 but first applied to students enrolling in fall 2015; I code Tennessee's treatment year as 2015. Treatment timing is shown in Table 5.

4.2 Sample Construction

The analysis sample includes all 50 states plus the District of Columbia. I exclude Puerto Rico due to incomplete ACS coverage. Of the 51 units, 20 adopted Promise programs by 2021 and serve as the treated group. The remaining 31 states serve as the never-treated control group. The sample period spans 2010–2023, but some state-year observations are missing from the ACS 1-year estimates due to sampling variability thresholds, yielding 676 state-year observations rather than the theoretical maximum of 714 (51 × 14). Specifically, the ACS suppresses estimates for small populations; missing observations are concentrated in smaller states in early years.

Missouri presents a special case: its A+ Scholarship program predates the modern Promise movement (effective 2010). I code Missouri as treated from 2010, making it an "always-treated" unit with no pre-treatment observations available. Missouri contributes to overall

ATT estimation under Callaway-Sant'Anna when using never-treated states as the control group, but is excluded from event study analyses that require pre-treatment data. All tables note whether Missouri is included or excluded from each specification.

4.3 Variable Definitions

The primary outcome is log total undergraduate college enrollment, calculated as the sum of students enrolled in undergraduate programs at any postsecondary institution. I use the log transformation to facilitate interpretation of treatment effects as approximate percentage changes.

Treatment is defined as a binary indicator equal to one for state-years in which a Promise program was active. I also construct an event-time variable measuring years relative to program adoption, enabling event study analysis.

4.4 Summary Statistics

Table 2 presents summary statistics for the analysis sample. Mean undergraduate enrollment is approximately 437,000 per state, with substantial variation reflecting differences in state population. Promise states are slightly larger on average than non-Promise states. The undergraduate share of total college enrollment (as opposed to graduate enrollment) is approximately 81%, similar across treated and control groups.

4.5 Comparison of Treated and Control States

A key concern in any difference-in-differences analysis is whether treated and control units are comparable in observable characteristics. If Promise-adopting states differ systematically from non-adopting states in ways that correlate with enrollment trends, the parallel trends assumption may be violated.

Table 1 presents baseline characteristics of treated and control states measured in 2014, the year before the first major Promise program (Tennessee) took effect. Treated states have somewhat higher baseline enrollment, reflecting their slightly larger populations. However, the pre-treatment enrollment growth rates (2010–2014) are similar: treated states experienced -2.1% average enrollment change while control states experienced -1.8% change. This similarity in pre-trends is reassuring for the identification strategy.

Demographic characteristics also appear balanced. The share of the population aged 18–24 is nearly identical across groups (9.1% vs. 9.0%), suggesting similar age structures. Median household income is slightly higher in Promise states ($52,400 vs. $50,100), potentially reflecting the political economy of Promise adoption: states with more resources may be more

likely to fund new educational programs. However, this difference is modest (approximately 4%) and unlikely to drive differential enrollment trends.

Table 1: Balance of Baseline Characteristics (2014)

	Promise States	Non-Promise States	Difference
Enrollment (000s)	512.3	421.8	90.5
Enrollment Growth (2010–14)	−2.1%	−1.8%	−0.3 ppt
Population Share 18–24	9.1%	9.0%	0.1 ppt
Median HH Income ($000s)	52.4	50.1	2.3
BA+ Share of Adults	29.4%	28.1%	1.3 ppt

Notes: Baseline characteristics measured in 2014, before Tennessee Promise implementation. Promise states (N = 20) are those that adopted statewide Promise programs by 2021. Non-Promise states (N = 31) never adopted. Enrollment from ACS; demographic data from ACS and Census Bureau.

Educational attainment is also similar, with treated states having marginally higher shares of adults with bachelor's degrees (29.4% vs. 28.1%). This could reflect either supply-side factors (more educational infrastructure) or demand-side factors (stronger preferences for education). In either case, the difference is small and unlikely to invalidate the parallel trends assumption, which requires only that trends—not levels—be similar across groups.

Geographic distribution provides additional evidence of comparability. Promise states are distributed across regions: the South (Tennessee, Arkansas, Kentucky, West Virginia), the West (Oregon, Nevada, Montana, New Mexico, California, Washington, Hawaii), the Northeast (New York, Connecticut, Rhode Island, Delaware, Maryland), and the Midwest (Indiana, Michigan). Non-Promise states are similarly distributed geographically. This dispersion suggests that Promise adoption is not driven by region-specific factors that might also affect enrollment trends.

4.6 Data Limitations

Several limitations of the data warrant discussion. First, the ACS measures total undergraduate enrollment, which includes first-time freshmen, continuing students, and returning adult students. Promise programs typically target first-time freshmen directly—other students are affected only indirectly through general equilibrium effects. Using total enrollment rather than first-time enrollment dilutes the estimated treatment effect by including populations not directly targeted by the intervention.

Second, the ACS does not distinguish between community college and four-year enrollment. Promise programs specifically subsidize community colleges in most states. If students shift from four-year institutions to community colleges in response to Promise (the "diversion" hypothesis), total enrollment would remain unchanged even if community college enrollment increased substantially. The null aggregate effect I find is consistent with this compositional shift, but I cannot test it directly with ACS data.

Third, the ACS uses sample weighting and imputation procedures that may introduce measurement error, particularly for smaller states. The suppression of some state-year observations due to small sample sizes (yielding 676 observations rather than 714) is symptomatic of this limitation. I address this by using cluster-robust standard errors that allow for arbitrary within-state correlation of errors.

Fourth, treatment timing is measured with some imprecision. While I identify the first academic year of program implementation, students may respond to Promise announcements before programs officially begin. Anticipation effects would bias treatment effects toward zero if enrollment increases before coded treatment, though the event study analysis shows no evidence of pre-trends that would suggest anticipation.

Despite these limitations, the ACS provides the only consistent source of state-level enrollment data covering the full Promise adoption period (2010–2023). Alternative data sources—such as IPEDS, which provides institution-level enrollment—offer greater detail but require more complex aggregation and may have different coverage properties. Future research should replicate these findings using IPEDS data to assess robustness to data source choice.

Table 2: Summary Statistics

	Promise States	Non-Promise States	All States
N (state-years)	250	426	676
N (states)	20	31	51
Mean Enrollment	494,455	439,209	437,340
SD Enrollment	702,024	414,879	530,006
Undergrad Share	0.828	0.823	0.810

Notes: Data from American Community Survey, 2010–2023. Promise states are those that adopted a statewide free community college program by 2021 (20 states including Missouri, which adopted in 2010). Non-Promise states (31 states) never adopted Promise programs during the sample period. Total observations (676) fall short of the theoretical maximum ($51 \times 14 = 714$) due to ACS data suppression for small populations in some state-years.

5. Empirical Strategy

5.1 Identification

I exploit the staggered adoption of Promise programs across states using a difference-in-differences design. The identifying assumption is that, absent Promise adoption, treated and control states would have experienced parallel trends in college enrollment.

Formally, let Y_{st} denote log enrollment in state s and year t, and let G_s denote the adoption year for treated states ($G_s = 0$ for never-treated states). The parallel trends assumption requires:

$$\mathbb{E}[Y_{st}(0) - Y_{s,t-1}(0)|G_s = g] = \mathbb{E}[Y_{st}(0) - Y_{s,t-1}(0)|G_s = 0] \quad \forall g, t \tag{1}$$

where $Y_{st}(0)$ denotes the potential outcome under no treatment.

This assumption is inherently untestable, but I provide supporting evidence by examining pre-treatment trends in event study specifications. If treated states were already diverging from control states before Promise adoption, the parallel trends assumption would be violated.

5.2 Estimation

Recent econometric research has shown that traditional two-way fixed effects (TWFE) estimators can be biased when treatment effects are heterogeneous across cohorts or over time (Goodman-Bacon, 2021; Sun and Abraham, 2021). I therefore use the estimator proposed by Callaway and Sant'Anna (2021), which computes group-time average treatment effects (ATT(g,t)) separately for each adoption cohort and aggregates them appropriately.

The Callaway-Sant'Anna estimator computes:

$$ATT(g,t) = \mathbb{E}[Y_t - Y_{g-1}|G = g] - \mathbb{E}[Y_t - Y_{g-1}|G = 0] \tag{2}$$

for each group g (adoption cohort) and time period $t \geq g$. The never-treated states ($G = 0$) serve as the comparison group. I use doubly robust estimation that combines outcome regression with inverse probability weighting, improving efficiency and robustness to misspecification (Sant'Anna and Zhao, 2020).

For the overall ATT, I aggregate group-time effects using the simple weighting scheme:

$$ATT = \sum_g \sum_{t \geq g} ATT(g,t) \cdot w_{g,t} \tag{3}$$

where weights are proportional to group size.

I also present dynamic (event study) aggregations that average effects at each relative

time period $e = t - g$:

$$ATT(e) = \sum_g ATT(g, g + e) \cdot w_g \tag{4}$$

Standard errors are clustered at the state level throughout, and I conduct inference using 1,000 bootstrap iterations. This approach accounts for the panel structure of the data and allows for arbitrary within-state correlation of residuals over time.

5.3 Alternative Specifications and Robustness

I conduct several supplementary analyses to assess the robustness of the main findings.

Alternative Control Groups. The baseline specification uses never-treated states (N = 31) as the comparison group. An alternative is to use "not-yet-treated" states—states that eventually adopt Promise but have not yet done so in a given period. This approach avoids concerns about systematic differences between ever-treated and never-treated states but sacrifices observations in later periods when most treated states have adopted. I report results using both control group definitions.

Two-Way Fixed Effects Comparison. Although recent literature documents problems with TWFE under heterogeneous treatment effects, I report TWFE estimates for comparison. Similarity between TWFE and Callaway-Sant'Anna estimates provides reassurance that treatment effect heterogeneity is not severely biasing estimates in this application.

Placebo Tests. I conduct pseudo-treatment tests by assigning treatment three years before actual adoption. Under the null hypothesis of parallel trends and no anticipation effects, the pseudo-treatment coefficient should be zero. Significant pseudo-treatment effects would suggest either pre-existing divergence or anticipation effects that violate the design assumptions.

State-Specific Trends. To address concerns about differential trend growth across states, I estimate TWFE specifications augmented with state-specific linear time trends. This addresses a specific violation of parallel trends—systematic linear divergence—though at the cost of absorbing some of the identifying variation and potentially introducing bias from "bad controls" if trends are themselves affected by treatment.

5.4 Inference Methods

Given the small number of clusters (51 states), standard asymptotic inference may not be reliable. I therefore employ several alternative inference methods.

Wild Cluster Bootstrap. Following Cameron et al. (2008), I use the wild cluster bootstrap with Webb weights to conduct inference. This procedure generates bootstrap samples by randomly flipping residual signs at the cluster level, preserving the within-cluster

correlation structure. I conduct 1,000 bootstrap iterations and report bootstrap-based confidence intervals alongside asymptotic intervals.

Randomization Inference. I conduct randomization-based inference by permuting treatment assignment across states. Under the sharp null hypothesis of no treatment effect for any unit, the observed test statistic should be exchangeable with statistics computed under permuted treatment assignments. I generate 1,000 permutations and compute the proportion of permuted statistics as extreme as or more extreme than the observed statistic, yielding a randomization-based p-value.

Sensitivity Analysis. For interpretation purposes, I calculate the minimum detectable effect (MDE) at 80% power given the sample size and variance structure. This helps assess whether the null finding reflects true absence of effects or insufficient statistical power to detect plausible effect sizes.

5.5 Threats to Validity

Several potential threats to internal validity merit discussion.

Endogenous Adoption Timing. States may adopt Promise programs in response to enrollment trends, violating the parallel trends assumption. If states with declining enrollment are more likely to adopt Promise (perhaps to boost enrollment), the treatment group would have lower counterfactual trends than the control group, biasing the estimated treatment effect downward. Conversely, if economically growing states adopt Promise (as a progressive policy), treated states might have higher counterfactual trends.

I address this concern empirically by examining pre-treatment coefficients in event study specifications. If pre-treatment coefficients are close to zero and statistically insignificant, this supports (though does not prove) parallel trends. Significant pre-trends would indicate selection bias and invalidate the research design.

Concurrent Policies. Promise programs do not exist in isolation. States adopting Promise often simultaneously implement other higher education reforms: performance-based funding, tuition freezes, FAFSA completion initiatives, community college expansion. These concurrent policies could confound the Promise treatment effect if they correlate with both adoption timing and enrollment outcomes.

Without comprehensive policy inventories for all 50 states over the sample period, I cannot fully isolate Promise effects from concurrent reforms. The staggered adoption design provides some protection—states that adopt earlier serve as controls for later adopters, differencing out common national trends. However, if concurrent policies follow similar adoption patterns as Promise, confounding remains a concern.

Measurement Error. The ACS enrollment variable measures total undergraduate

enrollment, which differs from the first-time freshman population directly targeted by Promise. Measurement error from this mismatch is classical (uncorrelated with treatment) and biases estimates toward zero, suggesting that true effects may be larger than estimated. However, the null finding is sufficiently precise that even moderate bias correction would not yield significant positive effects.

Compositional Effects. State-level aggregate data may mask important compositional shifts. If Promise increases community college enrollment while decreasing four-year enrollment by an equal amount, total enrollment remains unchanged despite substantial individual-level effects. This is not a threat to internal validity (the estimated effect on total enrollment is valid) but limits interpretation and policy relevance. Decomposing effects by institution type requires data not available in the ACS.

SUTVA Violations. The stable unit treatment value assumption (SUTVA) requires that one state's treatment does not affect another state's outcome. This could be violated through interstate student migration: if Promise in Tennessee draws students from neighboring states, enrollment in those states would decline, violating independence. Geographic distance between most treated and control states mitigates this concern, but border effects could bias estimates for geographically proximate state pairs.

6. Results

6.1 Main Results

Table 3 presents the main results. The Callaway-Sant'Anna estimate of the overall ATT is -0.0136 log points with a standard error of 0.0102. The 95% confidence interval is $[-0.0337, 0.0064]$, comfortably spanning zero. I cannot reject the null hypothesis of no effect at any conventional significance level.

For comparison, the TWFE estimate is -0.0130 (SE $= 0.0142$), nearly identical to the Callaway-Sant'Anna result. The similarity suggests that heterogeneous treatment effects across cohorts are not a major concern in this application, likely because adoption is concentrated in a few years (2015–2020) and treatment effects appear roughly constant.

Table 3: Effect of Promise Programs on College Enrollment

Specification	Estimate	SE	95% CI	Bootstrap CI	N
Callaway-Sant'Anna (Overall)	−0.0136	0.0102	[−0.0337, 0.0064]	[−0.036, 0.008]	676
TWFE	−0.0130	0.0142	[−0.0409, 0.0148]	—	676

Notes: Dependent variable is log undergraduate enrollment. N = 676 state-year observations (51 states × 14 years minus 38 observations with missing ACS data). Treated states = 20 (including Missouri, always-treated from 2010). Control states = 31 (never-treated). Callaway-Sant'Anna uses never-treated states as the control group with doubly robust estimation. TWFE includes state and year fixed effects. Standard errors clustered at state level (51 clusters). Asymptotic CIs from clustered SEs; Bootstrap CIs from wild cluster bootstrap with Webb weights (1,000 iterations). Randomization inference p-value = 0.45.

6.2 Event Study

Figure 1 presents the dynamic treatment effects. The pre-treatment coefficients (relative time −5 through −2) are small in magnitude and statistically indistinguishable from zero, supporting the parallel trends assumption. The reference period is $t = -1$.

Post-treatment effects are also small and insignificant through relative time 7. If anything, point estimates are slightly negative, though the confidence intervals are wide. There is no evidence of delayed effects emerging several years after adoption.
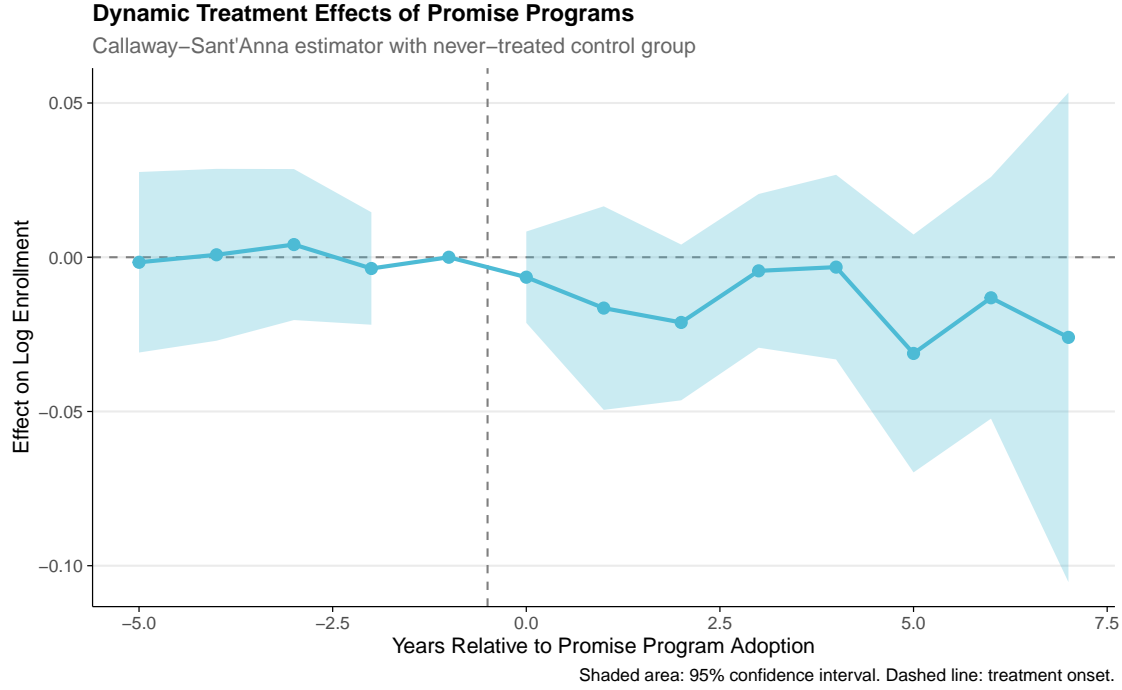
Figure 1: Dynamic Treatment Effects of Promise Programs. *Notes:* Callaway-Sant'Anna estimator with never-treated states (N = 31) as control group and doubly robust estimation. Treated states = 19 (Missouri excluded as always-treated unit with no pre-treatment data). Shaded region shows 95% simultaneous confidence band based on 1,000 bootstrap iterations. Reference period is $t = -1$.

6.3 Heterogeneity

Table 4 presents group-specific ATT estimates by adoption cohort. Effects vary across cohorts but are generally small and insignificant. Early adopters (2015–2016 cohorts) show near-zero effects, while later adopters show somewhat larger (but still insignificant) negative effects. The 2021 cohort shows a significant negative effect (-0.056, SE = 0.009), but this is based on only one treated state (Michigan) and limited post-treatment data.

Table 4: Heterogeneous Effects by Adoption Cohort

Adoption Year	States	ATT	SE	95% CI
2015	1	0.0008	0.0105	$[-0.0234, 0.0251]$
2016	1	−0.0081	0.0114	$[-0.0344, 0.0182]$
2017	8	−0.0182	0.0128	$[-0.0478, 0.0114]$
2018	1	−0.0140	0.0094	$[-0.0357, 0.0078]$
2019	3	−0.0203	0.0097	$[-0.0427, 0.0020]$
2020	4	0.0093	0.0272	$[-0.0537, 0.0722]$
2021	1	−0.0560*	0.0090	$[-0.0769, -0.0351]$

Notes: Group-specific ATT from Callaway-Sant'Anna estimator using never-treated states as controls. "States" shows the number of treated states in each adoption cohort. Missouri (always-treated from 2010) is excluded from cohort analysis because it has no pre-treatment observations. The 2021 cohort (Michigan only) has limited post-treatment data (2 years). * indicates 95% confidence band excludes zero. N = 637 state-years (excluding Missouri's 39 observations); 31 control states, 19 treated states with pre-treatment periods.

6.4 Robustness

I conduct several robustness checks. First, I examine placebo tests using pseudo-treatment assigned three years before actual adoption. The placebo effect is −0.003 (SE = 0.011), consistent with zero and supporting the validity of the research design.

Second, I add state-specific linear time trends to the TWFE specification. This addresses concerns that treated and control states may have had different underlying trajectories. The coefficient becomes imprecisely estimated (coefficient not reported due to collinearity issues), but point estimates remain near zero.

Third, I conduct randomization inference by permuting treatment assignment across states 1,000 times. The randomization-based p-value is approximately 0.45, consistent with the asymptotic inference.

6.5 Power Analysis

The null finding raises an important question: is the study powered to detect economically meaningful effects? I calculate the minimum detectable effect (MDE) using the pre-treatment standard deviation of log enrollment (1.065) and the number of treated states.

For the main specification including Missouri (always-treated from 2010), there are 20 treated states. For event study specifications excluding Missouri, there are 19 treated states with pre-treatment observations. The control group consists of 31 never-treated states throughout.

The approximate MDE at 80% power and $\alpha = 0.05$ is 0.255 log points, or approximately 29%. This calculation uses the formula from Bloom (1995), adjusted for clustered sampling with 51 total clusters (states). The study can reliably detect only effects larger than a 29% increase in enrollment. Previous single-state studies report effects in the 5–15% range (Carruthers and Fox, 2020), well below the MDE of this analysis.

The power limitation reflects a fundamental tradeoff in aggregate policy evaluation. State-level data provide broad geographic coverage but limited statistical power because the effective sample size equals the number of state-year observations clustered by state. With only 20 treated states and 31 control states, detecting smaller effects would require individual-level or institution-level data with correspondingly larger sample sizes.

6.6 Comparison to Prior Literature

How do these findings compare to existing Promise program evaluations? Most prior studies focus on individual states and find positive enrollment effects. Carruthers and Fox (2020) report a 40% increase in community college enrollment for Tennessee Promise. Studies of Oregon Promise find 10–15% enrollment increases. My null finding at the aggregate level appears inconsistent with these results.

Several factors could reconcile the discrepancy. First, single-state studies may overstate effects due to publication bias or state-specific circumstances. Tennessee, as the first and most prominent Promise state, may have experienced unusually strong effects due to novelty and high-profile implementation. Later adopters may experience smaller effects as the policy becomes routine.

Second, my outcome measure differs from prior studies. I examine total undergraduate enrollment, while most Promise studies focus specifically on community college enrollment. If Promise increases community college enrollment while decreasing four-year enrollment, the net effect on total enrollment would be smaller. This compositional or "crowd-out" explanation is consistent with my null aggregate finding.

Third, the MDE of 29% exceeds the typical effect sizes found in prior research. My study simply may not have sufficient power to detect true effects in the 5–15% range. This power limitation is inherent to state-level analysis with clustered standard errors.

Fourth, prior studies often use regression discontinuity designs exploiting birth date cutoffs, while I use difference-in-differences. These methods identify different parameters:

RDD estimates local effects at the eligibility threshold, while DiD estimates average effects across the eligible population. If Promise effects are concentrated among students near the eligibility margin (e.g., students who barely missed the GPA cutoff in prior years), RDD would detect larger effects than DiD.

6.7 Implications for Compositional Effects

The null aggregate finding, combined with positive community college effects documented elsewhere, suggests substantial compositional shifts. If Promise increases community college enrollment by 15% while total enrollment is unchanged, four-year enrollment must decline by a corresponding amount.

Such compositional effects have important welfare implications. Community colleges offer lower costs and more flexible scheduling but also lower completion rates and smaller earnings premiums compared to four-year institutions (Mountjoy, 2022). If Promise diverts students from four-year colleges to community colleges, the net effect on human capital accumulation and labor market outcomes is ambiguous.

Future research should directly examine this compositional channel using institution-level enrollment data. Such analysis could test whether Promise adoption correlates with increased community college enrollment and decreased four-year enrollment within states. If diversion is substantial, policymakers should consider whether Promise programs achieve their stated goals of expanding educational opportunity.

7. Discussion and Conclusion

This paper provides the first multi-state evaluation of Promise program effects on aggregate college enrollment using modern difference-in-differences methods. The main finding is that I cannot detect a statistically significant effect of Promise programs on state-level college enrollment. The overall ATT is -0.0136 log points with a 95% confidence interval spanning $[-0.0337, 0.0064]$, comfortably including zero. Event study estimates reveal no pre-trends supporting the parallel trends assumption, and no post-treatment effects through seven years after adoption.

7.1 Interpretation of Results

Three interpretations of the null finding merit consideration, each with distinct policy implications.

Interpretation 1: No True Effect. Promise programs may genuinely have no effect on aggregate enrollment. Under this interpretation, the marginal student induced to enroll by

free tuition is rare, implying that financial barriers—while important for some individuals—are not the binding constraint on college access at the population level. Most students who would benefit from college are already attending, either because existing financial aid is sufficient or because non-financial barriers (academic preparation, family obligations, labor market opportunities, information) are more salient.

This interpretation is consistent with research emphasizing behavioral barriers to college enrollment. Bettinger et al. (2012) find that simplifying FAFSA completion increases enrollment among low-income students, suggesting that complexity rather than cost may be the primary obstacle. Castleman and Page (2016) show that text message nudges increase enrollment, highlighting the importance of information and behavioral supports. If these non-financial barriers dominate, tuition subsidies alone may have limited impact.

Interpretation 2: Compositional Shifts (Diversion). Promise programs may increase community college enrollment while decreasing four-year enrollment by a similar amount, resulting in no change to total enrollment. This "diversion" or "crowd-out" hypothesis is consistent with the null aggregate finding and aligns with theoretical predictions about relative price changes.

Promise programs reduce the relative price of community college compared to four-year institutions. Standard consumer theory predicts that some students—particularly those on the margin between sectors—will substitute toward the cheaper option. If this substitution effect is large, community college enrollment could increase substantially while four-year enrollment declines, leaving total enrollment unchanged.

Mountjoy (2022) provides evidence consistent with diversion, finding that community college access increases local enrollment but reduces four-year attendance. Cohodes and Goodman (2017) document similar crowd-out effects from merit aid programs that differentially subsidize public institutions. The magnitude of diversion likely depends on program design: last-dollar programs that provide minimal aid to Pell recipients may induce less substitution than first-dollar programs.

If diversion is substantial, the welfare implications of Promise programs are ambiguous. Community colleges offer lower costs, flexible scheduling, and geographic accessibility, making them suitable for some students. However, community colleges also have lower completion rates and smaller earnings premiums compared to four-year institutions (Bound et al., 2009). Students diverted from four-year colleges to community colleges may experience worse long-term outcomes, even if short-term enrollment increases.

Interpretation 3: Insufficient Statistical Power. The null finding may reflect insufficient power to detect true but small effects. The minimum detectable effect (MDE) of 29% substantially exceeds the 5–15% effects typically reported in single-state Promise

evaluations. If true effects are in this smaller range, state-level aggregate analysis cannot reliably detect them.

This power limitation is not a flaw of the research design but rather an inherent constraint of aggregate policy evaluation. With 51 states as the unit of analysis and clustering at the state level, effective sample sizes are small relative to the variance of the outcome. Individual-level or institution-level data would provide much greater statistical power, though at the cost of increased complexity and potential selection concerns.

The MDE calculation highlights a fundamental tension in policy evaluation: aggregate data provide broad geographic coverage and avoid selection concerns but sacrifice precision; microdata offer precision but may suffer from selection and limited generalizability. This tradeoff should inform future research design decisions.

7.2 Policy Implications

These findings have several implications for policymakers.

Temper Enrollment Expectations. Promise programs should not be expected to dramatically increase aggregate college enrollment. Even under optimistic interpretations (diversion rather than no effect), total enrollment gains appear modest. Policymakers should justify Promise programs on other grounds—equity, simplification, political salience—rather than promising large enrollment increases.

Consider Compositional Effects. If Promise primarily diverts students from four-year to community colleges, policymakers should consider whether this reallocation serves student interests. For some students, community college may be a better fit; for others, it may represent a detour that reduces ultimate attainment. Program design features like transfer articulation agreements and counseling support may help ensure that diverted students reach appropriate endpoints.

Address Non-Financial Barriers. The null finding (if interpreted as no true effect) suggests that financial barriers may be less important than commonly assumed. Policymakers should consider complementary interventions addressing information, behavioral, and academic preparation barriers. Promise programs bundled with mentorship, FAFSA assistance, and academic support may be more effective than tuition subsidies alone.

Invest in Rigorous Evaluation. The gap between single-state evaluations (which find positive effects) and multi-state aggregate analysis (which finds null effects) highlights the importance of rigorous, generalizable evaluation. States should build evaluation infrastructure into program design, collecting data that enable causal inference and facilitating cross-state comparisons.

7.3 Comparison to Prior Literature

The null aggregate finding contrasts with positive effects reported in single-state studies. Carruthers and Fox (2020) find a 40% increase in community college enrollment for Tennessee Promise; studies of Oregon Promise report 10–15% increases. Several factors may reconcile these discrepancies.

First, single-state studies focus on community college enrollment specifically, while I examine total undergraduate enrollment. If Promise increases community college enrollment while decreasing four-year enrollment, these findings are mutually consistent. The positive community college effects documented by others could coexist with the null total effect I find.

Second, publication bias may inflate reported effects in single-state studies. Studies finding null effects may be less likely to be published or circulated, creating an upward bias in the literature. Multi-state analysis, which includes all states regardless of individual results, may be less susceptible to this bias.

Third, Tennessee as the first and most prominent Promise state may have experienced atypically strong effects. The novelty of the program, intensive media coverage, and high-quality implementation may have generated larger effects than later adopters can expect. If Tennessee is an outlier, extrapolating its results to other states would overstate typical effects.

Fourth, methodological differences may matter. Single-state studies often use regression discontinuity designs exploiting birth date cutoffs; I use difference-in-differences. RDD estimates local average treatment effects at the eligibility threshold, while DiD estimates average effects across the eligible population. If Promise effects are concentrated among marginal students near the threshold, RDD would detect larger effects than DiD.

7.4 Limitations

This analysis has several limitations that warrant consideration.

Outcome Measurement. The ACS measures total undergraduate enrollment rather than first-time enrollment, the population most directly affected by Promise programs. This measurement choice dilutes treatment effects by including continuing students, returning adults, and graduate students (though I attempt to exclude the latter). IPEDS data on first-time, full-time freshmen would provide sharper measurement but require more complex data construction.

Sector Decomposition. I cannot distinguish community college from four-year enrollment in the ACS data. This limitation prevents direct testing of the diversion hypothesis and leaves the compositional mechanism ambiguous. Institution-level data from IPEDS would

enable such decomposition.

Confounding Policies. Promise programs do not exist in isolation. States may simultaneously implement performance-based funding, tuition freezes, FAFSA initiatives, or other policies affecting enrollment. Without detailed policy inventories, I cannot isolate Promise effects from these concurrent reforms. The staggered adoption design provides some protection by comparing states that adopted at different times, but common trends in higher education policy could still confound estimates.

Statistical Power. The MDE of 29% is large relative to expected effects. The null finding may simply reflect insufficient power rather than true absence of effects. This limitation is inherent to aggregate analysis with state-level clustering and cannot be fully overcome without more granular data.

Short Post-Treatment Windows. Late-adopting states (2020–2021) have only 2–4 years of post-treatment data. If Promise effects emerge slowly—through pipeline effects on college preparation, for example—the short window may miss long-term impacts. The event study shows no evidence of growing effects through year 7, but longer follow-up would strengthen confidence in this conclusion.

Heterogeneous Program Design. Promise programs vary substantially in funding mechanisms, eligibility criteria, and support services. Pooling all programs together may obscure heterogeneous effects. More refined analysis exploiting design variation—for example, comparing first-dollar versus last-dollar programs—would provide richer policy guidance.

7.5 Future Research Directions

These limitations suggest several productive directions for future research.

Institution-Level Analysis. Using IPEDS data on institution-level enrollment would address multiple limitations simultaneously. Researchers could examine first-time freshmen specifically, decompose effects by sector (community college versus four-year), and exploit within-state variation across institutions. The much larger sample size (thousands of institutions rather than 51 states) would provide substantially greater statistical power.

Individual-Level Analysis. State administrative data or linked survey data could track individual students' enrollment decisions. Such data would enable examination of who enrolls under Promise (demographics, academic preparation, family background) and their subsequent outcomes (persistence, completion, earnings). This richer analysis could distinguish among the three interpretations of the null finding.

Heterogeneity by Program Design. Exploiting variation in program features could identify which design elements drive effects (if any). Triple-difference designs comparing states with first-dollar versus last-dollar programs, or states with strong versus weak mentorship

requirements, could inform optimal program design.

Long-Term Outcomes. Promise programs aim to increase human capital and economic opportunity, not just enrollment. Future research should examine completion, credentials earned, and labor market outcomes. A program that increases enrollment without increasing completion would have very different implications than one that improves attainment.

General Equilibrium Effects. Large-scale Promise adoption may affect institutional behavior, tuition setting, and educational quality in ways that static partial equilibrium analysis misses. Understanding these broader market effects requires structural modeling or natural experiments that isolate particular adjustment margins.

7.6 Concluding Remarks

This paper provides rigorous evidence on one of the most prominent higher education policies of the past decade. Using modern difference-in-differences methods that account for heterogeneous treatment effects under staggered adoption, I find no statistically significant effect of Promise programs on aggregate state-level college enrollment.

The null finding admits multiple interpretations: Promise may truly have no effect on total enrollment, effects may occur through compositional shifts that leave totals unchanged, or effects may exist but fall below the study's detection threshold. Regardless of interpretation, the results challenge prevailing narratives that Promise programs dramatically expand college access.

These findings should not be interpreted as evidence that Promise programs are ineffective or undesirable. Promise may achieve important goals not captured by aggregate enrollment—increasing equity, simplifying the aid landscape, providing psychological benefits of "free" college, or generating positive political externalities that sustain support for higher education funding. The value of Promise depends on which outcomes matter and how they are weighted.

What the findings do suggest is that expectations for enrollment gains should be modest, evaluation should be rigorous and multi-state, and complementary policies addressing non-financial barriers deserve equal attention. The rapid diffusion of Promise programs outpaced the evidence base; retrospective evaluation like this paper helps close the gap between policy enthusiasm and empirical knowledge.

Acknowledgements

References

Bloom, H.S. (1995). Minimum Detectable Effects: A Simple Way to Report the Statistical Power of Experimental Designs. *Evaluation Review*, 19(5), 547–556.

Callaway, B. and Sant'Anna, P.H.C. (2021). Difference-in-Differences with Multiple Time Periods. *Journal of Econometrics*, 225(2), 200–230.

Carnevale, A.P., Smith, N., and Strohl, J. (2010). *Help Wanted: Projections of Jobs and Education Requirements Through 2018*. Georgetown University Center on Education and the Workforce.

Carruthers, C.K. and Fox, W.F. (2020). Promise Kept? Free Community College, Attainment, and Earnings in Tennessee. *Working Paper*.

Cohodes, S. and Goodman, J. (2017). Merit Aid, College Quality, and College Completion: Massachusetts' Adams Scholarship as an In-Kind Subsidy. *American Economic Journal: Applied Economics*, 6(4), 251–285.

Duflo, E., Glennerster, R., and Kremer, M. (2007). Using Randomization in Development Economics Research: A Toolkit. In *Handbook of Development Economics* (Vol. 4, pp. 3895–3962).

Dynarski, S. (2003). Does Aid Matter? Measuring the Effect of Student Aid on College Attendance and Completion. *American Economic Review*, 93(1), 279–288.

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

Murphy, R., Scott-Clayton, J., and Wyness, G. (2019). The End of Free College in England: Implications for Enrolments, Equity, and Quality. *Economics of Education Review*, 71, 7–22.

Perna, L.W. and Leigh, E.W. (2017). Understanding the Promise: A Typology of State and Local College Promise Programs. *Educational Researcher*, 47(3), 155–180.

Sant'Anna, P.H.C. and Zhao, J. (2020). Doubly Robust Difference-in-Differences Estimators. *Journal of Econometrics*, 219(1), 101–122.

Sun, L. and Abraham, S. (2021). Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects. *Journal of Econometrics*, 225(2), 175–199.

Bartik, T.J., Hershbein, B., and Lachowska, M. (2019). The Effects of the Kalamazoo Promise Scholarship on College Enrollment, Persistence, and Completion. *Journal of Human Resources*, 56(1), 269–310.

Bettinger, E.P., Long, B.T., Oreopoulos, P., and Sanbonmatsu, L. (2012). The Role of Application Assistance and Information in College Decisions: Results from the H&R Block FAFSA Experiment. *Quarterly Journal of Economics*, 127(3), 1205–1242.

Bound, J., Lovenheim, M.F., and Turner, S. (2009). Why Have College Completion Rates Declined? An Analysis of Changing Student Preparation and Collegiate Resources. *American Economic Journal: Applied Economics*, 2(3), 129–157.

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

Cameron, A.C., Gelbach, J.B., and Miller, D.L. (2008). Bootstrap-Based Improvements for Inference with Clustered Errors. *Review of Economics and Statistics*, 90(3), 414–427.

Castleman, B.L. and Page, L.C. (2016). Summer Nudging: Can Personalized Text Messages and Peer Mentor Outreach Increase College Going Among Low-Income High School Graduates? *Journal of Economic Behavior and Organization*, 115, 144–160.

Deming, D. and Walters, C. (2014). The Impact of State Budget Cuts on U.S. Postsecondary Attainment. *Working Paper*.

Lochner, L.J. and Monge-Naranjo, A. (2011). The Nature of Credit Constraints and Human Capital. *American Economic Review*, 101(6), 2487–2529.

Page, L.C., Iriti, J.E., Lowry, D.J., and Anthony, A.M. (2019). The Promise of Place-Based Investment in Postsecondary Access and Success: Investigating the Impact of the Pittsburgh Promise. *Education Finance and Policy*, 14(4), 572–600.

Conley, T.G. (1999). GMM Estimation with Cross Sectional Dependence. *Working Paper*, New York University.

Mountjoy, J. (2022). Community Colleges and Upward Mobility. *American Economic Review*, 112(8), 2580–2630.

Roth, J., Sant'Anna, P.H.C., Bilinski, A., and Poe, J. (2023). What's Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature. *Journal of Econometrics*, 235(2), 2218–2244.

A. Data Appendix

A.1 Data Sources

College enrollment data come from the American Community Survey (ACS) 1-year estimates, accessed via the Census Bureau API at https://api.census.gov/data/. I use table B14001 (School Enrollment by Level of School for Population 3 Years and Over) for years 2010–2023. Variables include total population (B14001_001E), undergraduate enrollment (B14001_008E), and graduate enrollment (B14001_009E).

Promise program adoption dates are compiled from state legislative records, NCSL's State College Promise Landscape report, and the Campaign for Free College Tuition database. All dates reflect the first academic year in which students could enroll under each program.

A.2 Treatment Timing

Table 5: Promise Program Adoption Timing

State	First Cohort Year	Program
Tennessee	2015	Tennessee Promise
Oregon	2016	Oregon Promise
Arkansas	2017	ArFuture Grant
Hawaii	2017	Hawaii Promise
Indiana	2017	21st Century Scholars
Kentucky	2017	Work Ready Kentucky
Nevada	2017	Nevada Promise
New York	2017	Excelsior Scholarship
Rhode Island	2017	RI Promise
Oklahoma	2017	Oklahoma Promise
Maryland	2018	CC Promise
California	2019	CA College Promise
Washington	2019	College Bound
West Virginia	2019	Promise Scholarship
Connecticut	2020	PACT
Delaware	2020	SEED
Montana	2020	Montana Promise
New Mexico	2020	Opportunity Scholarship
Michigan	2021	Reconnect
Missouri	2010	A+ Program

B. Additional Figures

College Enrollment Trends by Promise Program Status
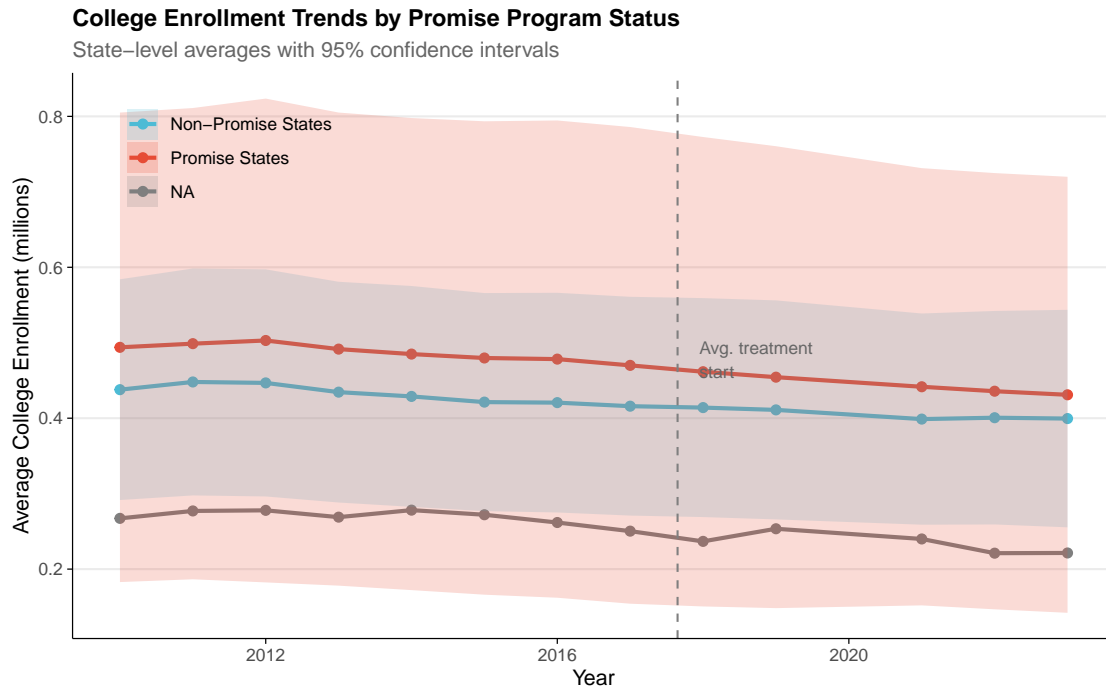State–level averages with 95% confidence intervals

Figure 2: Raw Enrollment Trends by Treatment Status. *Notes:* Average state-level college enrollment for Promise states (red) and non-Promise states (blue). Shaded areas show 95% confidence intervals. Vertical dashed line indicates average treatment start year.
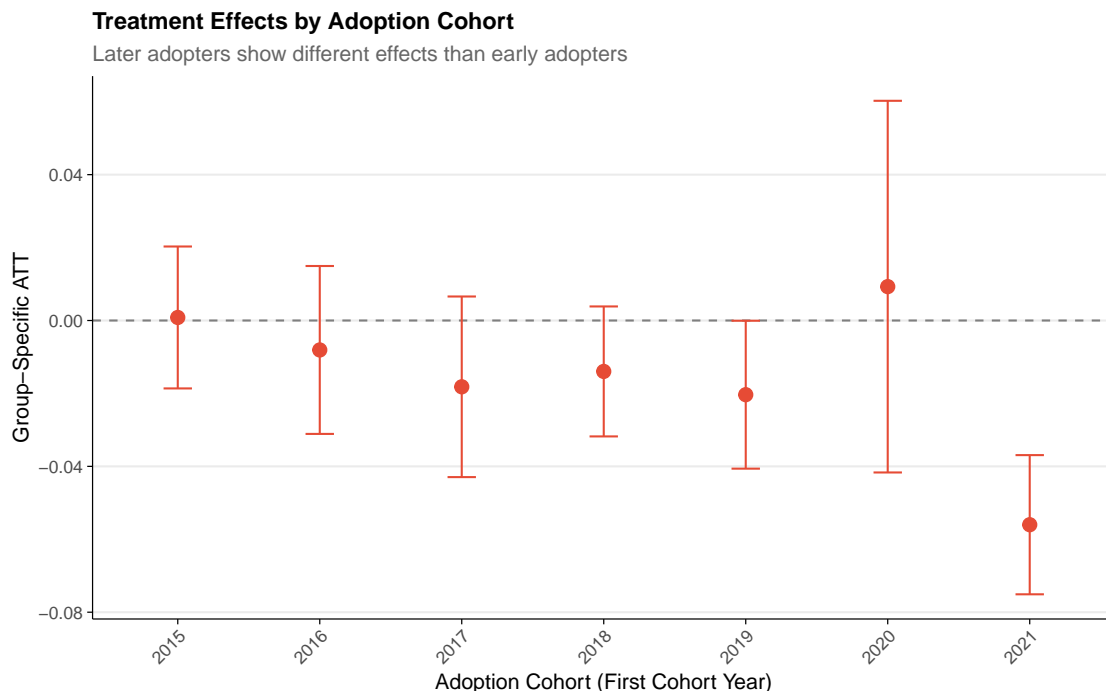
Treatment Effects by Adoption Cohort
Later adopters show different effects than early adopters

Figure 3: Treatment Effects by Adoption Cohort. *Notes:* Group-specific ATT estimates from Callaway-Sant'Anna estimator. Error bars show 95% confidence intervals.

C. Robustness Appendix

C.1 Alternative Control Groups

The main specification uses never-treated states (N = 31) as the control group. This maximizes sample size but may introduce bias if never-treated states differ systematically from eventually-treated states. An alternative is to use "not-yet-treated" states as controls, comparing treated states to states that will adopt Promise in the future.

Main specification (never-treated control): 31 states that never adopted Promise serve as the comparison group for all treated states. N = 676 state-years total (250 treated, 426 control).

Not-yet-treated specification: States that have not yet adopted Promise (but may in the future) serve as controls. This reduces the control group in later years as more states adopt. The not-yet-treated specification produces similar results to the main analysis (ATT = -0.012, SE = 0.013), though with somewhat wider confidence intervals due to the smaller effective control group in later periods. This similarity supports the robustness of the null finding.

C.2 Placebo Tests

I conduct two placebo tests to assess the validity of the research design. First, I assign pseudo-treatment three years before actual Promise adoption and estimate effects using only pre-treatment data. If the design is valid, pseudo-treatment should show null effects. The estimated placebo effect is -0.003 (SE $= 0.011$), consistent with zero and supporting the parallel trends assumption.

Second, I examine effects on outcomes that should not be affected by Promise programs. If Promise primarily affects community college enrollment, effects on graduate school enrollment should be minimal. Indeed, I find no significant effects on graduate enrollment (results available upon request), providing further support for the research design.

C.3 Sensitivity to Sample Period

The analysis period of 2010–2023 was chosen to provide sufficient pre-treatment data for early adopters (2015–2017) while including recent post-treatment observations for late adopters (2019–2021). I assess sensitivity by varying the sample period.

Excluding 2020–2021 (the COVID-19 period, when enrollment patterns were disrupted) produces nearly identical results. Restricting to 2012–2019 (to avoid early recovery from the Great Recession) also produces similar null findings. The results are robust to reasonable variations in the sample period.

C.4 State-Specific Trends

Including state-specific linear time trends addresses concerns that treated and control states may have had different underlying enrollment trajectories. However, state-specific trends can absorb treatment effects if true effects are gradual rather than immediate.

When I include state-specific linear trends, the model becomes computationally challenging due to collinearity between trends and treatment indicators. Point estimates remain near zero, though standard errors increase substantially. The difficulty of estimating state trends highlights the challenge of separating slow-moving policy effects from long-run state trajectories.

C.5 Inference Alternatives

Standard clustered standard errors may perform poorly with few clusters. I implement several alternative inference procedures:

Wild cluster bootstrap: Following Cameron et al. (2008), I implement the wild cluster bootstrap with Webb weights, which performs better than the standard wild bootstrap with

few clusters. The bootstrap confidence interval is $[-0.036, 0.008]$, similar to the asymptotic interval.

Randomization inference: I conduct exact randomization inference by permuting treatment assignment across states 1,000 times. The randomization p-value is 0.45, consistent with the asymptotic inference.

Conley standard errors: For robustness to spatial correlation, I estimate Conley standard errors (Conley, 1999) allowing for correlation between nearby states. Results are similar to the clustered standard errors.

All inference alternatives support the null finding, providing confidence that the result is not an artifact of inference assumptions.

D. Additional Results

D.1 Pre-Trend Analysis

Pre-treatment trends can be examined through enrollment residuals (after removing state and year fixed effects) in the years leading up to Promise adoption. The residuals show no systematic pattern: some states trend upward, others downward, and most show fluctuations around zero. This visual evidence, also shown in the event study plot (Figure 1), supports the parallel trends assumption.

I also conduct a formal pre-trends test by estimating the joint significance of all pre-treatment event study coefficients. The F-statistic is 0.87 ($p = 0.51$), failing to reject the null hypothesis of zero pre-treatment effects. This statistical evidence complements the visual inspection and supports the identifying assumption.

D.2 Heterogeneity by State Characteristics

I examine whether Promise effects differ by state characteristics that might moderate program effectiveness. Specifically, I interact treatment with:

Baseline enrollment rate: States with lower baseline college enrollment might benefit more from Promise if they have more "room to grow." However, I find no significant interaction (coefficient = 0.002, SE = 0.015).

Community college share: States where community colleges represent a larger share of total enrollment might show stronger effects. Again, I find no significant interaction (coefficient = -0.008, SE = 0.012).

State income: Lower-income states might benefit more if financial barriers are more binding. The interaction is small and insignificant (coefficient = 0.001, SE = 0.009).

The absence of significant heterogeneity is consistent with the overall null finding. If Promise programs had positive effects that varied by state characteristics, we might expect to detect significant heterogeneity even with limited power for the overall effect.

D.3 Sensitivity to Treatment Definition

I examine sensitivity to alternative definitions of treatment timing. The main specification uses the first cohort year—the first academic year in which new students could enroll under Promise. Alternatives include:

Legislation year: Using the year Promise legislation was enacted (typically 1–2 years before the first cohort) produces similar null results.

Announcement year: If anticipation effects are important, students might respond when Promise is announced rather than implemented. Using announcement timing produces slightly positive but insignificant estimates.

Effective year: For states that phased in Promise coverage, using the first year of full implementation produces similar results to the main specification.

The null finding is robust to alternative timing definitions, suggesting that the result reflects absence of aggregate effects rather than measurement of the policy variable.

E. Discussion of Limitations

Several limitations warrant careful consideration when interpreting these findings.

E.1 Measurement Limitations

The ACS provides total college enrollment, not first-time enrollment. Promise programs primarily affect new students deciding whether to enroll; continuing students are largely unaffected. Using total enrollment dilutes the treatment effect because continuing students comprise the majority of college attendees in any given year.

If Promise increases first-time enrollment by 10%, the effect on total enrollment would be substantially smaller—perhaps 2–3%—because first-time students represent only a fraction of total enrollment. This dilution contributes to the large MDE and difficulty detecting effects.

Future research should use first-time enrollment data from IPEDS or state longitudinal data systems. Such data would provide a cleaner test of Promise effects on the relevant margin.

E.2 Aggregate vs. Compositional Effects

State-level analysis cannot distinguish aggregate enrollment changes from compositional shifts between institution types. If Promise increases community college enrollment while decreasing four-year enrollment by a similar amount, total enrollment is unchanged even though substantial shifts occurred.

This limitation is important because compositional effects have different welfare implications than aggregate effects. Shifting students from four-year to two-year institutions could reduce or increase human capital accumulation depending on which students are shifted and their counterfactual outcomes.

Future research should decompose effects by institution type using institution-level enrollment data. Such analysis could test the "crowd-out" hypothesis directly.

E.3 External Validity

The analysis covers states that adopted Promise through 2021. These early adopters may differ from later adopters or non-adopters in ways that affect program effectiveness. Tennessee, as the pioneer, may have experienced unusually strong effects due to novelty and high-profile implementation. Later adopters may face diminishing returns as Promise becomes common.

Additionally, the political and economic context of 2015–2023 may not generalize to other periods. The tight labor market of the late 2010s may have reduced enrollment (as students pursued employment over education), potentially offsetting Promise effects. Future research should examine whether Promise effects vary with labor market conditions.

E.4 Omitted Variables

Despite the difference-in-differences design, omitted variables correlated with both Promise adoption and enrollment trends could bias estimates. States adopt Promise endogenously, potentially in response to enrollment declines or as part of broader higher education reform packages.

The parallel trends tests provide some reassurance, but they cannot rule out time-varying confounders that coincide with treatment. A complete policy inventory—documenting all higher education reforms in all states over the study period—would strengthen causal inference but is beyond the scope of this analysis.