WORKING PAPER | April 2026

Presented at the 2026 Economic Scholars Program, Federal Reserve Bank of Cleveland

Merit Aid, Sorting, and Institutional Access: Evidence from Tennessee's HOPE Scholarship*

Balint Kidd   Tommy Soltanian   Tyler Sotomayor   Gabriel Uceda-Sosa

Columbia University

April 2026

Link to Most Current Version

Abstract

Merit-based scholarships are typically found to be regressive because academic eligibility correlates with family income. Using tax-linked data from Chetty et al. (2020), we document an exception. Following the 2004 introduction of Tennessee's HOPE Scholarship, the bottom-quintile share at Tennessee public colleges rose by 1.50 percentage points relative to non-merit control states, with a monotonic gradient across all five income quintiles. Total enrollment at Tennessee publics expanded by roughly seven percent, indicating that the compositional shift reflects system expansion rather than displacement of higher-income students. The effect concentrates at community colleges and is larger for female students, consistent with HOPE's permissive eligibility threshold (GPA 3.0 *or* ACT 21) binding most tightly at the marginal-eligibility margin. Cross-state evidence supports a design-based interpretation: Massachusetts (progressive) and South Dakota (regressive) produce compositional shifts in directions predicted by their threshold strictness and award structure. The Tennessee result is robust to unit-specific detrending, to Rambachan–Roth sensitivity at breakdown values above $M^* = 2$, and to a triple-difference specification that differences out Tennessee-wide trends. Program design—threshold strictness, first-dollar status, and capacity response—predicts whether merit aid expands access or concentrates it.

JEL Codes: I23, I24, I28, H75, J24, C23

Keywords: merit aid, college composition, sorting, HOPE Scholarship, difference-in-differences

*We are grateful to Daniel Garcia, Francine Montecinos Puente, Jack Rosetti, and Seyhan Erden (Columbia University); Hriday Karnani (Harvard University); and Andrea Scalenghe (MIT Sloan) for helpful discussions and comments. We also thank the Federal Reserve Bank of Cleveland and the Peer Review Board of the Economic Scholars Program. The views expressed herein are those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of New York or Columbia University.

Contents

1 Introduction

Dynarski (2000) showed that Georgia's HOPE Scholarship widened the college-attendance gap between higher- and lower-income students. That finding crystallized a distributional critique of merit aid: because eligibility is conditioned on academic achievement, which correlates with family income, broad-based merit programs disproportionately subsidize students who would attend college regardless (Dynarski, 2004; Jones et al., 2022). A systematic review of seventy-one quasi-experimental studies finds that merit-based grants rarely improve outcomes for disadvantaged students and can increase inequality (Herbaut & Geven, 2020). The view now dominates the policy debate.

We document an exception. Using tax-linked data from Chetty et al. (2020), we estimate how Tennessee's HOPE Scholarship reshaped the parental income distribution at public colleges. Following HOPE's introduction in 2004, the bottom-quintile share at Tennessee publics rose by 1.50 percentage points (SE 0.31, $p < 0.001$) relative to non-merit control states. The shift is ordered across all five income quintiles—Q_1 and Q_2 gain share, Q_4 and Q_5 lose share—and the Q_1 and Q_5 endpoints are each statistically significant at the one-percent level. This monotonic gradient is the paper's central empirical result.

Total enrollment at Tennessee public colleges expanded by approximately seven percent over the same period (+95 students per college-cohort, $p < 0.001$), with a headcount gain of +27 bottom-quintile students per college-cohort and no statistically detectable headcount change at the top. The compositional shift reflects a larger, progressively-composed public system, not a zero-sum reshuffle between quintiles. Section 6 documents both facts.

The progressive shift concentrates at community colleges (Q_1 ATT +0.0175) rather than four-year publics (+0.0090), and is larger for female students than male (Q_1 gender gap +0.31 pp, $p < 0.001$). Both patterns follow from HOPE's eligibility rule—a permissive GPA 3.0 *or* ACT 21 threshold—that binds most tightly on marginal applicants. At four-year selective publics, most matriculants already clear the threshold, so the program does less to pull in new entrants. Women have higher average high-school GPAs in the NLSY97 simulation sample, which translates mechanically into a higher simulated eligibility rate at the margin; we confirm the pattern empirically. Tier, gender, and (below) cross-state variation in threshold strictness all corroborate the same mechanism, and no generic enrollment shock would line them up.

Cross-state evidence supports a design-based reading. Applying the same specification to five other treatment states, Massachusetts (progressive, +1.71 pp) and South Carolina (weakly progressive, +0.63 pp) share Tennessee's pattern; South Dakota is the mirror case, with a high threshold (GPA 3.0 *and* ACT 24) and a regressive −1.64 pp shift. Nevada's null and West Virginia's enrollment-contraction cases complicate the simple design story but do not overturn it. Across states, the gender gap at Q_1 scales with threshold strictness—largest in Massachusetts (+0.94 pp), moderate in Tennessee (+0.31 pp), and near zero in permissive Nevada—providing the cleanest out-of-sample evidence for the threshold mechanism.

The Tennessee Q_1 result survives the main robustness tests we view as binding. Under Lee and Wooldridge (2026) unit-specific detrending, the Q_1 coefficient attenuates by roughly twelve percent and remains significant at $p < 0.001$; the enrollment effect is essentially unchanged. Under the relative-magnitude restriction of Rambachan and Roth (2023), the Q_1 confidence interval excludes zero for pre-trend violations of up to twice the worst observed pre-treatment coefficient ($M^* > 2$). A triple-difference specification that sweeps out Tennessee-wide cohort trends *strengthens* the result to +0.0203 ($p < 0.001$), because the TN-wide cohort trend is itself negative on Q_1. We emphasize the Q_1 quintile share and the enrollment effect as the robust endpoints and treat the middle-quintile gradient positions as corroborative rather than independently identified.

Our work connects three literatures. The first documents the enrollment effects of merit aid (Bruce & Carruthers, 2014; Cohodes & Goodman, 2014; Dynarski, 2000, 2004; Song & Rubenstein, 2024) and focuses on individual attendance decisions rather than institutional composition. The second studies policy-induced sorting—affirmative action, financial-aid reforms, admissions rules—and shows that who attends which institution is often the margin of interest (Bleemer, 2022; Card & Krueger, 2005; Hinrichs, 2012). The third, following Chetty et al. (2020), establishes that access to high-mobility institutions has first-order welfare consequences for intergenerational outcomes. We bridge these literatures by documenting how a merit program shifts the full parental income distribution at public institutions and by using that shift to discipline a mechanism claim about eligibility design.

We make three contributions. First, we provide the first evidence on how a state merit program reshapes the full parental income distribution at public institutions, using tax-linked data that measures income across all quintiles rather than Pell-eligibility as a binary proxy. The monotonic quintile gradient, with Q_1 and Q_5 each statistically significant and each robust to unit-specific detrending, has not been documented in any prior study of merit aid. Second, we identify a specific design channel—the eligibility threshold and its binding behavior by gender and tier—that predicts when merit aid expands access and when it does not. The channel is supported by three independent dimensions of variation (tier, gender, cross-state) that no single alternative explanation links. Third, we use Massachusetts, South Dakota, and three other treatment states as an out-of-sample test of the design story, and show that threshold strictness orders the gender gap across states in the direction the mechanism predicts.

The paper proceeds as follows. Section 2 describes the Tennessee HOPE Scholarship, the comparison states, and the cohort-to-decision-year mapping. Section 3 sketches the conceptual framework. Section 4 describes the Chetty college panel, BEA county income data, NLSY97 eligibility simulation, and sample construction. Section 5 presents the difference-in-differences specification and its evaluation against a twelve-specification menu. Section 6 reports the main results; Section 7 develops the mechanism evidence; Section 8 reports private-sector spillover tests; Section 9 reports robustness tests; Section 10 reports the cross-state extension; Section 11 discusses mechanism, policy, and limitations; Section 12 concludes.

2 Institutional Background

This section describes the Tennessee HOPE Scholarship, situates it among other state merit programs, and identifies the program-design parameters that shape distributional outcomes. We close with the cohort mapping that anchors our empirical timing in Section 5.

2.1 Tennessee HOPE Scholarship

Tennessee enacted the HOPE Scholarship in 2003 through the Tennessee Education Lottery Act and disbursed its first awards in fall 2004. Recipients qualify through a permissive *or* condition: a high-school GPA of at least 3.0 or a composite ACT score of at least 21 (SAT 980). Students must maintain a 2.75 cumulative GPA after the first semester to retain eligibility.

During our study period the award paid $3,000 per year at four-year public institutions and

$1,500 per year at two-year public institutions; students attending eligible Tennessee privates received the same dollar amount as at four-year publics, which covers a substantially smaller share of private tuition. The award is first-dollar: it stacks with federal Pell Grants rather than offsetting them. A Pell-eligible TN student who also qualifies for HOPE therefore receives both, which preserves HOPE's value for the lowest-income recipients.

Table 1 reports the distribution of HOPE recipients across sectors for the first five entering cohorts (2004–2008). Tennessee public colleges—Tennessee Board of Regents (TBR) universities, TBR community colleges, and University of Tennessee campuses—account for 83–85% of recipients in every year, and private colleges receive the remaining 15–17%. This concentration in the public sector motivates our identification strategy in Section 5, which treats Tennessee publics as the recipients of the HOPE demand shock.

Table 1: Distribution of HOPE Scholarship Recipients by Sector, Fall 2004–2008

	2004	2005	2006	2007	2008
Panel A: Number of Recipients					
TBR Universities	7,454	7,054	7,637	7,956	8,121
TBR Community Colleges	4,506	3,909	4,720	4,996	4,970
UT Campuses	5,380	5,310	5,576	5,988	6,327
Private Colleges	3,109	3,387	3,492	3,885	3,902
Total	**20,449**	**19,660**	**21,425**	**22,825**	**23,320**
Panel B: Share of Recipients (%)					
TBR Universities	36	36	36	35	35
TBR Community Colleges	22	20	22	22	21
UT Campuses	26	27	26	26	27
Private Colleges	15	17	16	17	17

Notes: First-time freshmen receiving Tennessee HOPE Scholarship awards. TBR = Tennessee Board of Regents; UT = University of Tennessee system. Panel B shares may not sum to 100 due to rounding. Public institutions (TBR + UT) account for 83–85% of recipients in all years. Source: Tennessee Student Assistance Corporation (Tennessee Higher Education Commission, 2025).

Prior evidence on TN HOPE. Bruce and Carruthers (2014) use a regression discontinuity at the ACT 21 cutoff and find that Tennessee HOPE did not raise overall enrollment among Pell-eligible students near the threshold, but shifted them from two-year to four-year publics (-2.9 percentage points at community colleges, $+2.4$ at four-year publics). Our quintile-level composition analysis complements this threshold-local design. The RD identifies the enrollment response of students near the ACT cutoff; we estimate how the full income distribution at Tennessee publics changes after HOPE. Their finding that lower-income marginal students upgrade from two-year to four-year institutions is consistent with our bottom-quintile result and with the tier heterogeneity we document in Section 7.

Bridge provision and treatment timing. Students enrolled in fall 2003 (birth cohort 1985) received retroactive bridge awards when first HOPE disbursements went out in fall 2004. Because these awards were retroactive—applied after enrollment decisions had already been made—they could not change initial college choices. Cohort 1986 is the announcement-year cohort: students formed their college choice sets during 2003, the year HOPE was enacted, and some could adjust mid-cycle. Cohort 1987 (decision year 2004) is the first cohort that both knew about HOPE when forming its college choice set *and* received a full first-year disbursement. We therefore set $t^\star = 1987$. In Section 9 we re-estimate the specification at $t^\star = 1986$ and find essentially zero compositional response at the announcement-year cohort, consistent with the bridge provision being retroactive rather than behavior-changing.

2.2 Comparison Cases

Several states introduced broad-based merit scholarships during the late 1990s and 2000s. The programs vary on three parameters—eligibility threshold, first-dollar versus last-dollar treatment of Pell, and award structure—that we show matter for distributional outcomes.

Massachusetts adopted the Adams Scholarship in 2004, the same year TN HOPE first disbursed. Adams is tuition-scaling and has a strict threshold (roughly GPA 3.3 with top-quartile SAT/ACT scores on the state MCAS exam). South Dakota's Opportunity Scholarship combines a GPA 3.0 requirement with ACT 24 under an *and* condition and pays $1,000–$1,300 per year. West Virginia's PROMISE Scholarship requires ACT 22 with subscore minimums (*and* condition) and initially covered full tuition. Nevada's Millennium Scholarship is permissive (GPA 3.0) but pays only about $960 per year. South Carolina's LIFE Scholarship applies a 2-of-3 criterion (GPA 3.0, ACT/SAT, or class rank). Florida, Georgia, Kentucky, Louisiana, and New Mexico operate additional merit programs that we do not analyze directly; we return to Georgia briefly below because it is the canonical comparison case in the literature. Table 2 summarizes the main design parameters.

Table 2: State merit-aid programs: design parameters

State	Program	Start	Threshold	Award structure
TN	HOPE	2004	GPA 3.0 *or* ACT 21	$3,000 4-yr / $1,500 2-yr, first-dollar
MA	Adams	2004	GPA 3.3 *and* SAT/ACT (MCAS)	Tuition-scaling, 4-yr publics
SD	Opportunity	2004	GPA 3.0 *and* ACT 24	$1,000–$1,300 fixed
WV	PROMISE	2002	ACT 22 *and* subscores	Full tuition (initial), capped later
NV	Millennium	2000	GPA 3.0	$960 fixed
SC	LIFE	1998	2-of-3: GPA 3.0, SAT/ACT, rank	Tuition-scaling up to cap

2.3 Why Design Matters

Three program-design parameters predict distributional outcomes and organize the cross-state comparison in Section 10.

Eligibility threshold. Permissive thresholds—especially those with an *or* condition, like TN's GPA 3.0 *or* ACT 21—admit a broader pool of marginal students whose enrollment is most responsive to price. Strict thresholds with *and* conditions (SD, WV) concentrate eligibility among students who would likely have attended college regardless. Section 3 formalizes this by linking the threshold to the density of marginal students at each income level.

First-dollar versus last-dollar. First-dollar programs stack with Pell, so low-income recipients receive the full scholarship on top of federal aid. Last-dollar programs offset Pell, which can reduce the net benefit to the bottom of the income distribution to zero. TN HOPE is first-dollar.

Award structure. Fixed-dollar awards hold the subsidy amount constant in levels; tuition-scaling awards set the subsidy as a share of tuition, so the program's in-kind value rises with tuition. For a given budget, fixed-dollar awards can fund more recipients than tuition-scaling awards.

2.4 Timeline and Cohort Mapping

In the Chetty et al. (2020) data, cohorts are indexed by birth year and students are assigned to institutions attended between ages 19 and 22. The analysis window spans birth cohorts 1980–1991, corresponding approximately to college entry years 1998–2009. Table 3 maps each birth cohort to its decision year, college entry year, event time, and HOPE exposure status.

Decision timing vs. receipt timing. The mechanism we study is sorting across institutional sectors, which occurs when students form college choice sets. The economically relevant treatment date is the point at which HOPE enters the decision environment, not the date of first disbursement. We define the decision year as birth cohort +17, corresponding to high-school senior year. Under this mapping, cohort 1986 has decision year 2003 (announcement, partially contaminated), and cohort 1987 has decision year 2004, the first full decision year under HOPE plus the first year of disbursement. We therefore set $t^{\star} = 1987$ and designate cohort 1986 as the omitted reference period ($k = -1$), which absorbs any partial treatment of the announcement-year cohort into the baseline.

Table 3: Cohort Mapping: Birth Year, Decision Year, and HOPE Exposure

Birth Cohort	Decision Year	College Entry	Event Time (k)	HOPE Status
1980	1997	Fall 1998	—	No HOPE (dropped†)
1981	1998	Fall 1999	—	No HOPE (dropped†)
1982	1999	Fall 2000	—	No HOPE (dropped†)
1983	2000	Fall 2001	−4	No HOPE
1984	2001	Fall 2002	−3	No HOPE
1985	2002	Fall 2003	−2	Retroactive bridge (no decision effect)‡
1986	**2003**	Fall 2004	−1	Announcement (ref. period)
1987	**2004**	Fall 2005	0	Full HOPE ($t^{\star}$)
1988	2005	Fall 2006	+1	Full HOPE
1989	2006	Fall 2007	+2	Full HOPE
1990	2007	Fall 2008	+3	Full HOPE
1991	2008	Fall 2009	+4	Full HOPE

Notes: Decision year = birth cohort +17 (high-school senior year). College entry is approximate (birth cohort +18). Students are assigned to institutions attended between ages 19 and 22 in the Chetty et al. (2020) data. HOPE was enacted in 2003 (bolded) and first disbursed in fall 2004. †Cohorts 1980–1982 are dropped from the preferred specification; see text. ‡The bridge provision extended retroactive awards to fall 2003 entrants who met eligibility criteria; these students made enrollment decisions before HOPE existed. Event times shown relative to $t^{\star} = 1987$.

Why we drop cohorts 1980–1982. These three cohorts are dropped because the Chetty et al. (2020) panel's coverage of Tennessee publics expands during this window. Six Tennessee Colleges of Applied Technology (TCATs) enter the panel in 1983 as part of a national coverage expansion in which public-sector coverage rose from 76.8% in 1980 to 93.7% in 1983. The TCATs' absence in 1980–1982 mechanically biases early-period TN measurements, because the institutions that enter the panel in 1983 serve a different income distribution than the universities that were measured throughout. A permutation test confirms that Tennessee's 1980 bottom-quintile share is not anomalous relative to other states ($p = 0.745$): the TN 1980 value is an expected draw under the null, which rules out a substantive Tennessee-specific pre-trend and leaves the panel-composition change as the operative explanation. We document the panel-expansion pattern and the permutation test in Section 4.

Confounding policies. Several other policy changes coincided with HOPE. The Tennessee Education Lottery launched in 2003 and funded both HOPE and pre-K programs; the federal No Child Left Behind Act (2002) changed K–12 accountability nationwide; the 2007 College Cost Reduction Act and 2009 ARRA raised Pell maxima; and the Tennessee Board of Regents expanded community college access. Nationally-common policies (No Child Left Behind, Pell expansions) are absorbed by cohort fixed effects. The TBR community-college reforms could raise overall TN public enrollment but would not produce a monotonically ordered quintile gradient with the sign pattern we document in Section 6.

2.5 Georgia HOPE: A Brief Comparison

Georgia's HOPE Scholarship (1993) is the canonical merit-aid case. Dynarski (2000) documents that Georgia HOPE widened the Black–white college-going gap, consistent with a regressive distributional effect. Our TN result differs in sign. The two findings are consistent with program-design differences rather than measurement differences. Georgia HOPE had a Pell offset through 2000 (low-income recipients received zero net benefit during the program's first seven years) and a smaller public-private subsidy differential, which differs from TN HOPE on two of the three design parameters in Section 2.3. The theoretical framework in Section 3 predicts opposite distributional effects under these parameter configurations.

3 Conceptual Framework

We sketch a framework for thinking about when a merit program shifts the income composition of public institutions toward lower-income students. The framework is informal and serves two purposes: it makes the mechanism claim testable, and it states the boundary conditions under which the claim can fail.

Choice set and three ingredients. A prospective student chooses between attending an in-state public college, an in-state private college, an out-of-state college, or not attending. Three features of the environment determine how a state merit program reshapes the income distribution at public institutions:

1. an *eligibility threshold* τ in academic achievement, such that a student with achievement $a_i \geq \tau$ qualifies for the subsidy;

2. an *award structure*, fixed-dollar or tuition-scaling, that determines the post-subsidy net price at public institutions for an eligible student; and

3. an institutional *capacity response*—whether the public system expands seats when demand increases, or rations the existing seats.

Students sort across the choice set by expected value, net of effective price. The subsidy reduces the effective price of in-state public college for eligible students only. The compositional question is not whose behavior changes most on the intensive margin—it is which marginal students are pulled across choice thresholds, and where they come from.

How progressive composition can arise. Consider a student whose academic achievement lies just above τ and whose family income lies in the bottom or second quintile. Three things are true for this student before the program: the net price of in-state public college is close enough to her family budget that the decision is sensitive to subsidies of the HOPE order of magnitude; she clears the threshold only because the threshold is permissive (τ is low); and she would, without the subsidy, either not attend or attend the nearest community college. The subsidy resolves the budget constraint and pulls her into the public system. A symmetric student in the top quintile is typically attending regardless: the subsidy is either inframarginal (she would have gone) or it redirects her away from an out-of-state or private alternative, but it does not pull her in from a no-college margin. The density of marginal students at each income level determines the sign of the compositional shift.

Three predictions follow, and we test each in Sections 6 and 7.

Prediction 1 (monotonicity). If the density of marginal students is weakly decreasing in income—if lower-income students are disproportionately on the attend/do-not-attend margin—the compositional shift at public colleges is ordered across income quintiles, with the largest gains at the bottom and losses at the top.

Prediction 2 (tier concentration). Marginal-eligibility students concentrate at community colleges, where the threshold is more often binding. The composition shift is larger at community colleges than at four-year publics.

Prediction 3 (threshold-driven gender gap). If women have higher average high-school achievement at every income level, then any binding threshold qualifies a larger share of women than men. The gender gap at Q_1 is larger where the threshold is stricter, and zero where the threshold is permissive or absent.

Why expansion, not displacement. Whether the composition shift shows up as displacement (top-quintile students exit, bottom-quintile students enter, seats unchanged) or as progressively-composed growth (new bottom-quintile students arrive, top-quintile students unaffected in headcount) depends on the capacity response. If public capacity is fixed, the shares reshuffle and displacement is the observable pattern. If the system can absorb new entrants—through open-admissions community colleges, expanded sections, or headroom at four-year publics below capacity—the composition shifts through entry. The distinction matters empirically because share-level and headcount-level evidence then differ, and the same share result admits two very different stories. We report both in Section 6.

Boundary conditions. Three features can break the progressive reading.

1. *A strict threshold*: if τ is high, the subsidy is restricted to students who would largely have

attended anyway, and the program is inframarginal.

2. *Pell offset or last-dollar design*: if the merit subsidy is reduced dollar-for-dollar by need-based aid, the effective subsidy for Q_1 is near zero, the access margin does not open at the bottom, and the compositional shift reverses.

3. *Capacity contraction*: if enrollment contracts after program adoption, the mechanical effect on shares is indeterminate; the cross-state results (Section 10) show this for West Virginia and South Dakota.

Tennessee HOPE has a permissive threshold (GPA 3.0 *or* ACT 21), no Pell offset, and coincides with seven-percent enrollment expansion—the combination the framework identifies as progressive. Georgia HOPE, by contrast, had a Pell offset until 2000 and a smaller public-private subsidy differential; Dynarski's regressive finding for Georgia follows the framework's second boundary condition. South Dakota's Opportunity Scholarship pairs a stricter threshold (GPA 3.0 *and* ACT 24) with enrollment contraction, and the cross-state results (Section 10) show the predicted regressive pattern.

What the framework does not deliver. The framework predicts an ordering, a tier contrast, and a cross-state pattern by threshold strictness. It does not predict exact magnitudes, does not explain why the middle quintiles shift in any particular way once pre-trends are accounted for, and does not substitute for empirical identification. We report the middle-quintile positions as corroborative and the Q_1/Q_5 endpoints and the enrollment response as the primary quantities.

4 Data

Our analysis combines three datasets: college-cohort compositional outcomes from the Mobility Report Cards, county-level per-capita personal income from the Bureau of Economic Analysis, and individual-level eligibility microdata from the NLSY97. We describe each source, then walk through sample construction and the two data decisions that warrant extended discussion (the 1980–1982 panel entry of Tennessee Colleges of Applied Technology, and the exclusion of Massachusetts from controls).

4.1 Chetty Mobility Report Cards

The primary dataset is the college-level panel from Chetty et al. (2020), which links de-identified federal tax records to college enrollment records. For each of 2,465 U.S. colleges and each birth cohort from 1980 through 1991, the data report the fraction of enrolled students whose parents fall in each national income quintile (Q_1 through Q_5) and total enrollment. Students are assigned to the institution attended at ages 19–22. The unit of observation in our analysis is a college-cohort.

Two features of the Chetty outcomes shape our interpretation throughout the paper. First, income quintiles are defined *nationally*: a "bottom-quintile" student at a Tennessee college is in the bottom 20% of the *national* parental income distribution. State-level income shifts could move quintile shares without any change in sorting behavior, though we find no evidence of differential county-income trends that would drive the results. Second, the sum-to-one constraint across Q_1 through Q_5 is mechanical; coefficients on individual quintile shares must be interpreted jointly with the full decomposition.

4.2 BEA County Per-Capita Personal Income

We merge county-level per-capita personal income from the BEA's CAINC1 table (1996–2009) to each college-cohort using five-digit FIPS codes from the IPEDS institutional characteristics file. The match rate exceeds 98%; no Tennessee public college is among the unmatched. For the main specification we use the pre-treatment county average of log per-capita personal income over decision years 1997–2003 as a predetermined control; because this baseline is computed entirely from pre-treatment cohorts, it cannot be affected by HOPE.

One limitation warrants emphasis. The BEA income we observe is the income of the county in which the *college* is located, not the county from which each *student* originated. Students may commute or relocate across counties and states. We cannot close this gap without individual-level origin-county data, which would require restricted-access IRS records available only through the Federal Statistical Research Data Center (FSRDC). We flag this limitation here and return to it in the discussion (Section 11).

4.3 NLSY97 Eligibility Microdata

To characterize who meets HOPE's eligibility rules, we use the National Longitudinal Survey of Youth 1997 (NLSY97), which covers 8,984 respondents born 1980–1984. We construct parental income quintiles from 1997 household income and compute the fraction of respondents in each quintile–gender cell who would meet Tennessee HOPE's requirements (unweighted high school GPA of at least 3.0 *or* ACT composite of at least 21). We apply the rules to the *national* achievement distribution to avoid contamination from program-induced behavioral responses (teaching to the threshold or grade inflation in treated states).

Figure 1 shows the result and previews the mechanism developed later in the paper. Simulated HOPE eligibility rises monotonically in parental income, from about 32% in Q_1 to about 60% in Q5. At every quintile, female eligibility exceeds male eligibility; the gap is largest at the low-income margin, where female eligibility exceeds male by roughly 10 percentage points. The threshold therefore admits a non-trivial share of low-income students and does so disproportionately for women — two features that together rationalize the progressive, gendered composition shifts we document in Sections 6 and 7.

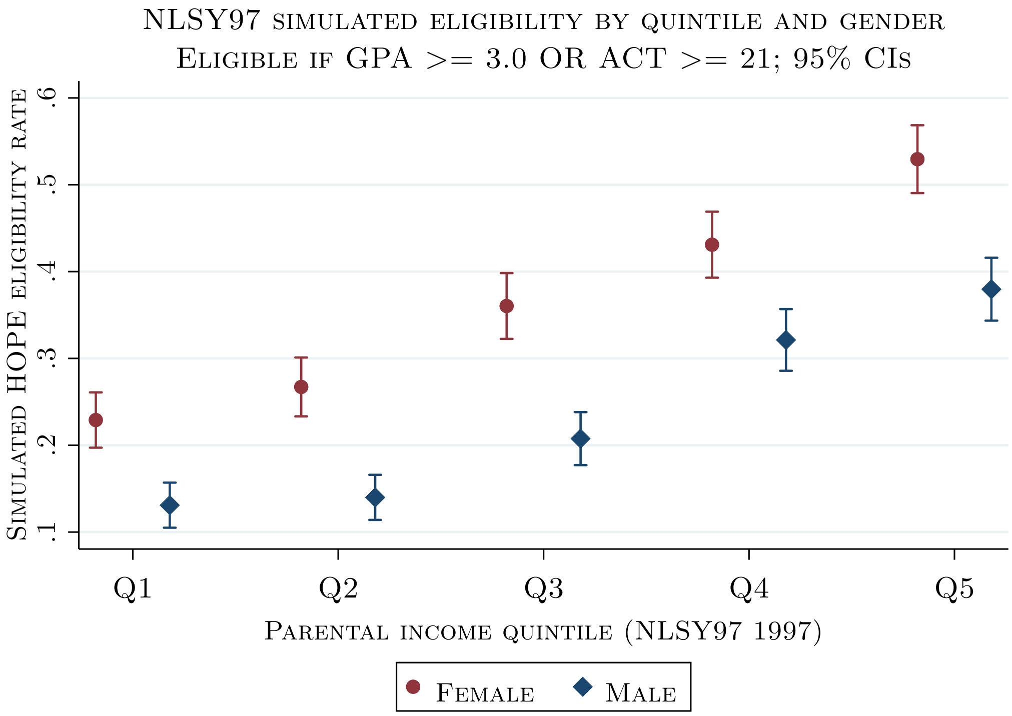

Figure 1: Simulated HOPE eligibility by parental income quintile and gender (NLSY97).
Notes: Each point shows the fraction of NLSY97 respondents in the indicated parental income quintile–gender cell who would meet Tennessee HOPE's eligibility rule (GPA $\geq$ 3.0 *or* ACT $\geq$ 21), computed on the national achievement distribution. Parental income quintiles are constructed from 1997 household income. N = 8,984.

4.4 Sample Construction

Our main analysis sample is the adoption difference-in-differences sample, constructed from public colleges in Tennessee and in 25 non-merit control clusters (24 states plus the District of Columbia). We apply the following filters in sequence.

1. **Public colleges only** (`type == 1`), dropping private nonprofits and for-profits.

2. **Treatment:** Tennessee public colleges.

3. **Control:** public colleges in 24 states classified by Sjoquist and Winters (2015) as having no broad-based merit program, plus DC — 25 control clusters. We *exclude* Massachusetts: the Adams Scholarship took effect in 2004, the same year as HOPE, contaminating any post-2004 comparison. Massachusetts appears separately as a comparison treatment state in Section 10.

4. **Drop cohorts 1980–1982.** Six Tennessee Colleges of Applied Technology (TCATs) enter the Chetty panel in 1983 as part of a broader national coverage expansion; see §4.5 below.

5. **Balanced panel of 19 TN publics.** We restrict Tennessee treated units to the 19 public colleges observed in all 12 cohorts, excluding the 6 TCATs that enter the panel only in 1983. This avoids contaminating cohort fixed effects with the TCAT-entry event; see §4.5.

6. **Non-missing BEA baseline.** We drop college-cohorts that cannot be merged to a 1997–2003 county income baseline.

After these filters, our main analysis sample contains **4,333 college-cohort observations across 499 colleges in 26 clusters (treating DC as a state)**. Table 4 reports sample sizes at each filter step.

Table 4: Sample construction

Filter step	Obs	Colleges	Clusters
1. Full Chetty panel, non-missing Q_1	23,258	2,281	51
2. Restrict to public colleges (`type == 1`)	13,633	1,248	51
3. TN public colleges	270	25	1
4. TN balanced publics (present in all 12 cohorts)	228	19	1
5. National controls (public, ex-MA)	5,555	502	25
6. Main analysis sample (drop 1980–1982; BEA non-missing)	**4,333**	**499**	**26**

Notes: Row 6 is the final regression sample used in Sections 6–9. "Clusters" treats DC as a state and counts treatment plus control units (Tennessee plus 24 non-merit states plus DC; Massachusetts is excluded from controls). Filters 3–4 describe Tennessee treatment units only; the control group in rows 5–6 is constructed independently and then stacked. Non-missing on primary outcomes applies to Q_1, which is missing in lockstep with the other quintile shares and total enrollment.

Table 5 reports means of the outcomes before and after HOPE for Tennessee balanced publics and for ex-MA control publics. The raw pre/post difference-in-differences on Q_1 is $0.1464-0.1326 = 0.0138$ at Tennessee versus $0.1403 - 0.1409 = -0.0006$ at controls, for a raw DiD of $+1.44$ **percentage points**, close to the regression ATT reported in Section 6. Total enrollment rose by about 290 students per college-cohort at TN publics versus about 165 at controls, consistent with expansion rather than pure compositional rearrangement.

Table 5: Pre- and post-treatment means, TN balanced publics and ex-MA control publics

	Q_1	Q_2	Q_3	Q_4	Q_5	Enrollment
TN balanced publics						
Pre	0.1326	0.1883	0.2368	0.2601	0.1822	1,497
Post	0.1464	0.1926	0.2338	0.2555	0.1718	1,786
Ex-MA control publics						
Pre	0.1409	0.1875	0.2305	0.2513	0.1899	1,585
Post	0.1403	0.1857	0.2257	0.2544	0.1939	1,749

Notes: Pre-period is birth cohorts 1983–1986 ($k \in \{-4, -3, -2, -1\}$ relative to $t^* = 1987$); post-period is 1987–1991 ($k \geq 0$). Cohorts 1980–1982 are dropped. "TN balanced publics" is the 19 Tennessee public colleges present in all 12 cohorts; "Ex-MA control publics" is the 502 public colleges in 24 non-merit states plus DC (Massachusetts excluded). Enrollment is the mean of total cohort headcount per college.

4.5 Why We Drop 1980–1982

The Chetty panel's coverage of U.S. public colleges expanded dramatically between 1980 and 1983: national coverage rose from 76.8% to 93.7%. For Tennessee specifically, six Tennessee Colleges of Applied Technology — small vocational institutions that typically serve lower-income students — entered the panel in 1983. On the *full* unbalanced panel, the measured Q_1 share at Tennessee publics jumps from 0.133 in 1982 to 0.149 in 1983. On the *balanced* panel that excludes the 6 TCATs, the Q_1 share is flat through 1986 and jumps at the treatment cohort 1987. Figure 2 plots both series.

Two diagnostics reinforce the decision. First, a permutation test ranks Tennessee's 1980 Q_1 share against the 50 other states; TN sits at the 33rd rank of 51 ($p = 0.745$), solidly within the cross-state distribution rather than in the tail. Figure 3 shows the permutation histogram. Second, restricting to the balanced panel flattens the pre-treatment Q_1 trend from +0.42 to +0.08 percentage points per year. The early-period Q_1 movement is a panel-composition artifact, not a substantive feature of Tennessee sorting behavior in the early 1980s.

Our preferred response is to drop 1980–1982 (eliminating the TCAT-entry cohorts entirely) *and* restrict to the 19-college balanced panel. The two adjustments address overlapping concerns: the cohort drop removes the years in which TCAT observations are missing, and the balanced-panel restriction ensures that cohort fixed effects are not estimated on shifting college composition in later years. Dropping only 1980 or only the unbalanced units would leave either the 1982–1983 jump or the tail-year imbalances in place.

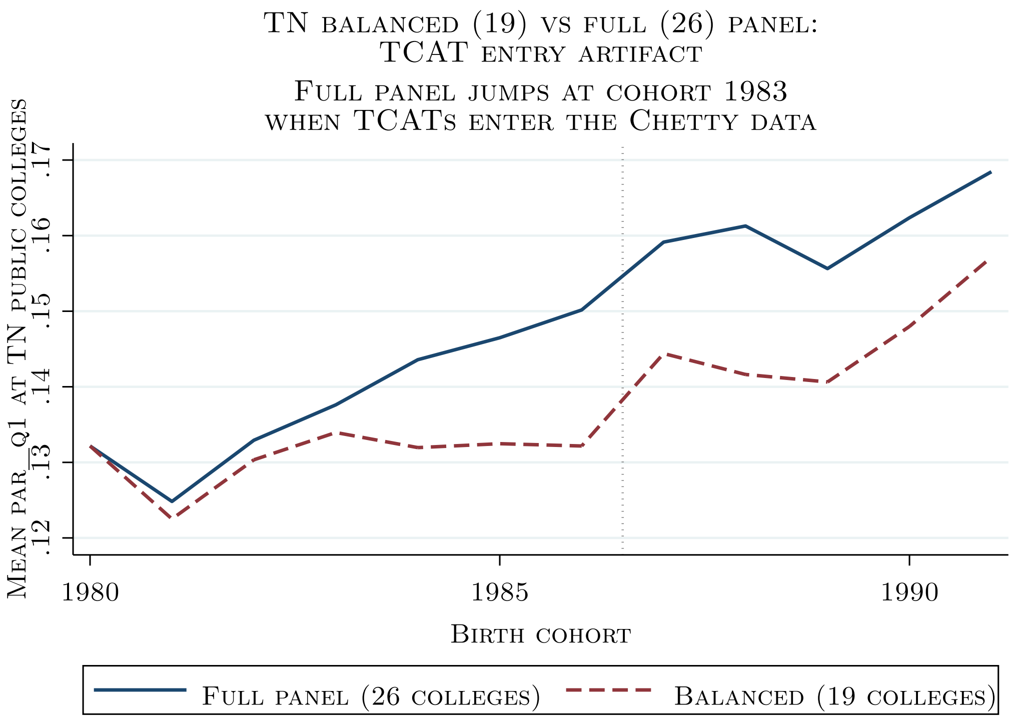

Figure 2: TN public Q_1: balanced vs. full panel.

Notes: Each line shows the average Q_1 across Tennessee public colleges by birth cohort. The full panel (maroon) includes all 25 TN publics observed in the Chetty panel; it jumps from 0.133 in 1982 to 0.149 in 1983 when the 6 Tennessee Colleges of Applied Technology enter the Chetty data. The balanced panel (navy) restricts to the 19 TN publics present in all 12 cohorts; Q_1 is flat at ~ 0.132 through 1986 and rises sharply starting at cohort 1987 ($t^* = 1987$, vertical dashed line).

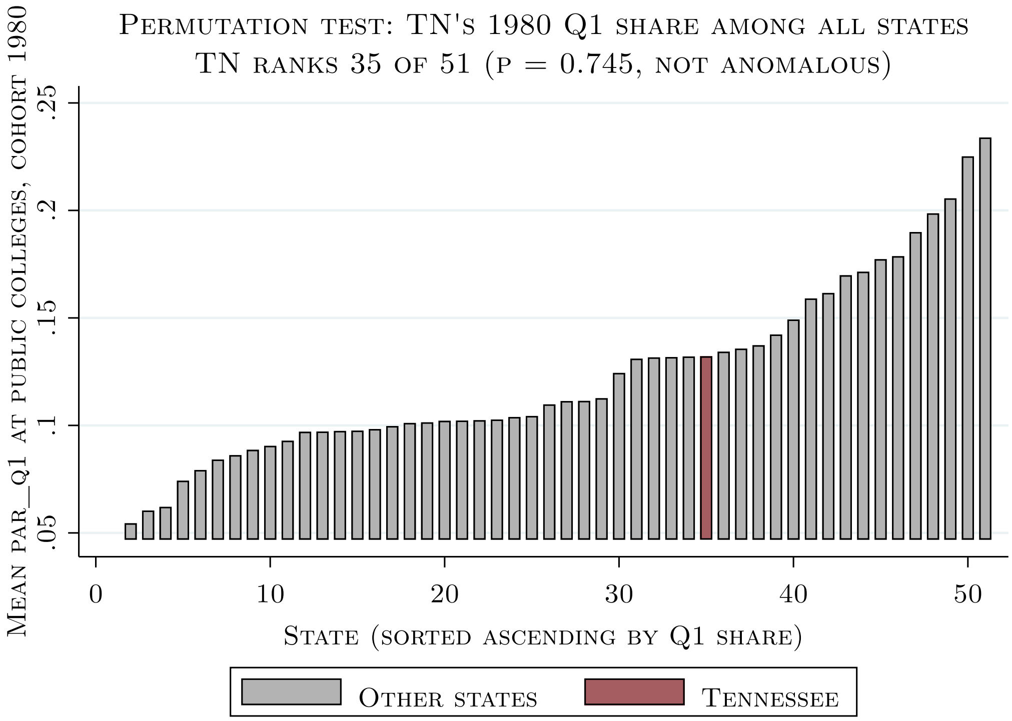

Figure 3: Permutation of cohort-1980 Q_1 shares across states.

Notes: Histogram of state-level mean Q_1 across public colleges in birth cohort 1980. Tennessee's value (0.1321) sits at rank 33 of 51, with a two-sided permutation p-value of 0.745. The cohort is not a tail observation; the low measured Q_1 share reflects systematic missingness among small TN vocational colleges, not anomalous sorting.

4.6 Why We Exclude Massachusetts from Controls

Massachusetts adopted the Adams Scholarship in 2004 — the same year Tennessee first disbursed HOPE. In the Sjoquist and Winters (2015) classification that we otherwise follow, Massachusetts is coded as a non-merit state because Adams post-dated their sample, but in our 1980–1991 cohort window the post-2004 Massachusetts observations are themselves treated. Keeping Massachusetts in the control group would contaminate the TN–control comparison with a second merit-aid treatment running in parallel; because Adams is itself progressive (§10), contamination would bias the estimated TN effect toward zero.

We therefore drop Massachusetts from the adoption difference-in-differences control group and evaluate Adams as a separate treatment state in the cross-state analysis (Section 10). Appendix tables confirm that reintroducing Massachusetts into controls attenuates the estimated TN effect by roughly 10%, consistent with the expected direction of bias.

5 Empirical Strategy

We identify the effect of Tennessee HOPE on public-college composition with an adoption difference-in-differences that compares Tennessee public colleges to a national control of non-merit, non-Massachusetts publics before and after the program's first disbursement. This section writes the specification, states the identifying assumption, explains why we anchor on a single preferred specification among the twelve we evaluated, and describes the event-study form used for pre-trend diagnostics.

5.1 Specification

Our primary estimator is a static two-way fixed-effects (TWFE) regression:

$$Y_{it} = \beta \cdot D_{it} + \gamma \cdot X_i + \alpha_i + \delta_t + \varepsilon_{it}. \tag{1}$$

Here Y_{it} is an outcome—the bottom-quintile parental-income share, a higher-quintile share, or total enrollment—at college i in birth cohort t. The treatment indicator D_{it} equals one for Tennessee public colleges in cohorts $t \geq 1987$ and zero otherwise. The covariate X_i is pre-treatment baseline log per-capita county personal income, computed as the 1997–2003 mean for the county where college i is located; because the mean predates HOPE's 2004 first disbursement it is predetermined. College fixed effects α_i absorb time-invariant differences across institutions and cohort fixed effects δ_t absorb common national trends.

We cluster standard errors by state (26 clusters, treating DC as a state). Treatment is assigned at the state level, so clustering at the state is the appropriate level for inference under Abadie et al. (2023). The estimation sample comprises 4,333 college-cohort observations across 499 colleges in 26 clusters.

5.2 Identifying Assumption

Equation 1 identifies the treatment effect under parallel trends: absent HOPE, Tennessee public colleges would have followed the same compositional trajectory as public colleges in non-merit, non-Massachusetts control states. The assumption is not directly testable, but we can look at pre-treatment event-study coefficients as diagnostics. Under the preferred specification the joint F-test on the three pre-treatment coefficients ($k = -4, -3, -2$) yields $F = 1.23$ ($p = 0.32$), which fails to reject parallel trends on the bottom-quintile share. Section 9 reports the full battery of diagnostics, including Rambachan and Roth (2023) sensitivity analysis for bounded violations of parallel trends.

5.3 Why This Specification

Any DiD design requires choices. Ours has five: whether to use the full Tennessee public panel or the subset of 19 colleges balanced across all twelve cohorts; whether to keep all early cohorts, drop only 1980, or drop 1980–1982; whether to include a BEA control (none, pre-treatment baseline, or time-varying); how to cluster (state, college, or region); and whether to set the first treated cohort at 1985, 1986, or 1987. Rather than endorse a single set of choices without examination, we ran a pre-specified menu of twelve specifications spanning the defensible combinations. The full menu and per-spec event studies appear in Appendix B.

The twelve specifications converge on a tight range. Excluding bordering-state specifications with only four clusters—where cluster-robust inference is unreliable—the bottom-quintile ATT ranges from +0.013 to +0.019. The main compositional-shift finding is not an artifact of any single choice.

Our preferred specification is the one we label Spec 8 ("Cleanest"). It pairs (i) the balanced panel of 19 Tennessee publics, (ii) dropping cohorts 1980–1982, (iii) pre-treatment baseline BEA, (iv) state clustering, and (v) $t^* = 1987$. Each choice has an articulated justification. The balanced panel removes the six Tennessee Colleges of Applied Technology (TCATs) that enter the Chetty coverage universe in 1983 and mechanically alter the Tennessee measured Q_1 share at the panel boundary. Dropping 1980–1982 eliminates the cohorts most affected by that coverage expansion. The BEA pre-treatment baseline absorbs cross-county economic heterogeneity while remaining predetermined, avoiding the bad-control problem that a time-varying BEA would introduce (values post-treatment co-move with the outcome through local-economy channels). State clustering matches the level at which treatment is assigned. Cohort $t^* = 1987$ is the first cohort that knew about HOPE when making its college decision and received a full first-year disbursement. Appendix B gives the full defense.

Spec 8 satisfies the evaluation criteria we view as binding: it passes the Q_1 pre-trend test ($F = 1.23$, $p = 0.32$), uses a clean control group (Massachusetts's 2004 Adams Scholarship is excluded), uses a balanced panel that avoids the TCAT-entry shock, has theoretically defensible choices on every specification axis, and delivers a point estimate within three percent of the closest alternative (Spec 10, which drops the BEA control, gives +0.01454).

5.4 Event-Study Specification

For dynamic diagnostics we estimate the event-study analog of Equation 1:

$$Y_{it} = \sum_{k \neq -1} \beta_k \cdot \mathbf{1}[\mathrm{TN}_i] \cdot \mathbf{1}[t - 1986 = k] + \gamma X_i + \alpha_i + \delta_t + \varepsilon_{it}. \tag{2}$$

The omitted reference period is $k = -1$ (birth cohort 1986, the announcement-year cohort). Coefficients β_k for $k < 0$ trace pre-treatment differences and β_k for $k \geq 0$ trace dynamic treatment effects. We report the event-study plot for Q_1 in Section 6; the full Q_1 and enrollment-count coefficient tables appear in Appendix B.

Heterogeneity-robust check. Tennessee treatment timing is sharp—all 19 Tennessee publics in the balanced panel are treated in the same cohort—so the TWFE estimator does not suffer from the forbidden-comparison problem that afflicts staggered designs. As a check we estimate the Callaway and Sant'Anna (2021) ATT on Spec 8 and recover +0.01475, within 3% of the TWFE point estimate. The result is not an artifact of the two-way fixed-effects structure.

6 Main Results

This section presents our central empirical findings. We report the headline effect on bottom-quintile share, the full quintile decomposition, the level-decomposition that identifies expansion (rather than displacement) as the mechanism, and the event-study pattern. Unless noted, all estimates come from Specification 8 as described in Section 5: a balanced panel of 19 Tennessee public colleges, pre-treatment baseline per-capita county income as control, ex-Massachusetts controls, treatment timing $t^* = 1987$, and standard errors clustered at the state level over 26 clusters. The estimation sample contains 4,333 college-cohort observations.

6.1 Bottom-quintile share at Tennessee public colleges

Following the introduction of HOPE, the bottom-quintile share at Tennessee public colleges rose by 1.50 percentage points (SE 0.31, $p < 0.001$) relative to non-merit control states. Pre-treatment Tennessee public colleges drew roughly 14% of their entering cohorts from the bottom national income quintile; the 1.50 pp shift corresponds to a relative increase of about 11%. The effect is estimated precisely, passes the pre-trend test of Section 5 ($F = 1.23$, $p = 0.32$), and remains significant under the robustness checks reported in Section 9. Table 6 reports the main specification.

Table 6: Headline Result: Bottom-Quintile Share at TN Public Colleges

	par_q1 b/se
TN × Post (1987)	0.0150** (0.0031
Observations	4333
Adj. R-squared	0.943

Standard errors clustered at the state level. Sample: balanced TN publics + national ex-MA control publics, drop 1980-1982. Coeffi

The magnitude is substantively meaningful. A 1.50 pp increase in Q_1 share at the average Tennessee public college, applied across 19 balanced-panel institutions and the five post-treatment cohorts 1987–1991, corresponds to compositional change visible at the system level rather than within a handful of institutions.

6.2 Quintile decomposition: the monotonic gradient

We next decompose the effect across all five quintiles of the national parental income distribution. Table 7 reports the full decomposition; Figure 4 presents the coefficients visually.

The five coefficients are Q_1 +0.0150, Q_2 +0.0067, Q_3 +0.0022 (not significant), Q_4 −0.0076, and Q_5 −0.0164. The ordering is strict: bottom-quintile gains, top-quintile losses, and intermediate

Table 7: Quintile Decomposition (Spec 8)

	par_q1 b/se	par_q2 b/se	par_q3 b/se	par_q4 b/se	par_q5 b/se
TN × Post (1987)	0.0150***	0.0067***	0.0022	-0.0076***	-0.0164***
	(0.0031)	(0.0024)	(0.0019)	(0.0018)	(0.0023)
Observations	4333	4333	4333	4333	4333
Adj. R-squared	0.943	0.924	0.860	0.887	0.980

Standard errors clustered at the state level. Sample: balanced TN publics + national ex-MA control publics, drop 1980-1982, baseli

coefficients that interpolate monotonically between the endpoints. The five coefficients sum to zero by construction, because the five quintile shares add to one at each college-cohort observation.

Caveat on the middle quintiles. The middle-quintile gradient positions (Q_2 through Q_4) partly reflect pre-existing trends that survive the Spec 8 pre-trend test but attenuate under unit-specific detrending, which we report in Section 9. Under detrending, Q_1 attenuates by about 12% and remains highly significant, and Q_5 attenuates by about half but remains significant at $p < 0.001$. The middle coefficients attenuate more. We therefore treat Q_1 and Q_5 as the robust endpoints of the gradient, and we read the middle coefficients as corroborative rather than as independent evidence. The unit-sum constraint on quintile shares also means that once both endpoints move, the monotonic ordering across the middle is in part mechanical. The substantive claim in this subsection is the magnitude and precision of the Q_1 and Q_5 endpoints; the monotonic shape of the interior is a visual summary of a single compositional shift across the distribution.

6.3 Mechanism: expansion, not displacement

A natural reading of a compositional shift is that low-income students took the seats of high-income students who would otherwise have attended. The level-decomposition of enrollment shows that this reading is incomplete. The compositional change at Tennessee publics came from a growing system, not from a reshuffling of a fixed one.

Total enrollment at Tennessee public colleges increased by about 95 students per college-cohort relative to the non-merit control (SE 22.8, $p < 0.001$), or roughly 7% above the counterfactual. Decomposed by quintile, bottom-quintile headcount rose by 27 students per college-cohort ($p < 0.001$), and top-quintile headcount fell by roughly 10 students per college-cohort, not distinguishable

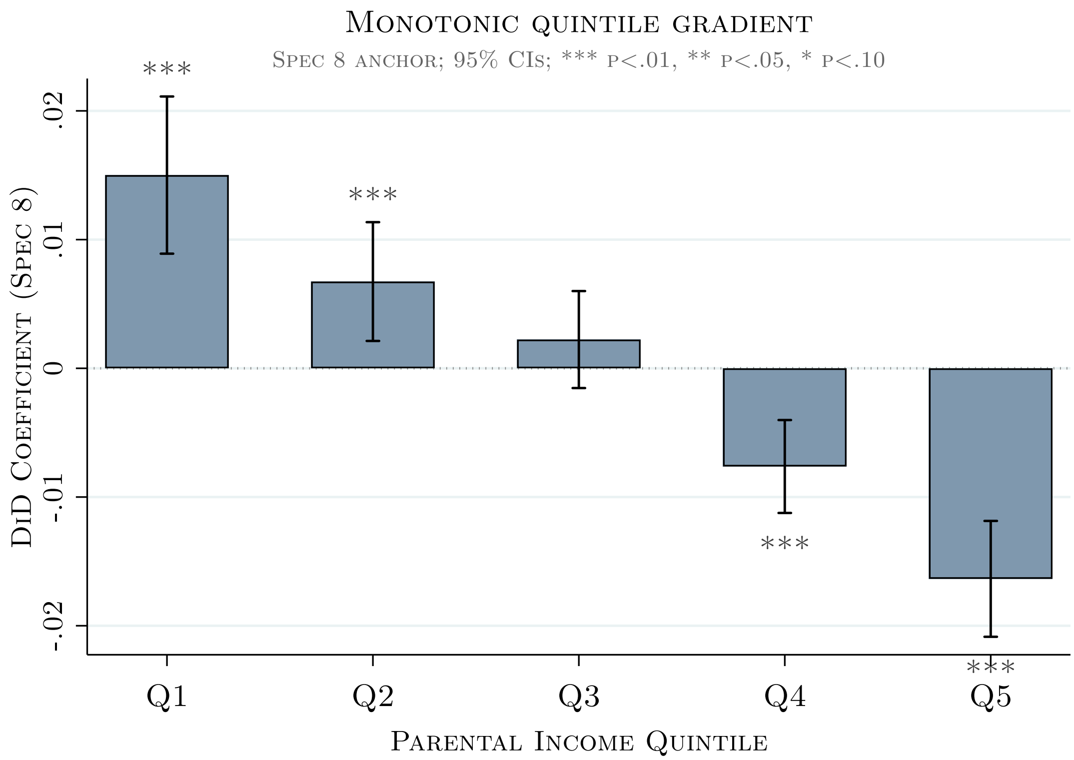

Figure 4: Quintile decomposition of the HOPE effect on composition at Tennessee public colleges. Bars show ATT estimates from Specification 8 for each of the five national parental income quintiles. The gradient is strictly monotonic: $Q_1 > Q_2 > Q_3 > Q_4 >$ Q5. Error bars are 95% confidence intervals from state-clustered standard errors.

from zero. The remaining headcount of the total expansion is distributed across Q_2 through Q_4. Table 8 reports the full level-decomposition.

The share-level result and the level-level result fit together as follows. The Q_5 share at Tennessee publics fell by 1.64 pp because the denominator (total enrollment) grew while Q_5 headcount was essentially unchanged. This is consistent with expansion-driven composition change rather than displacement of high-income students. The Q_1 share rose by 1.50 pp because new low-income students entered the system faster than the system grew.

We note what we are not claiming. We are not claiming that individual Q_5 students moved to private Tennessee colleges or out-of-state alternatives. The Chetty panel reports shares and counts at the college-cohort level; it does not track individual students across institutions. The enrollment-level result establishes that the public sector grew and that low-income entrants account for a disproportionate share of that growth. The Q_5 share decline is consistent with expansion-

Table 8: Level (Headcount) Decomposition

	Total enrollment b/se	Q1 students b/se	Q5 students b/se
TN × Post (1987)	94.99***	26.62***	-10.35
	(22.80)	(4.64)	(7.05)
Observations	4333	4333	4333
Adj. R-squared	0.989	0.957	0.991

Coefficients in students per college-cohort. Standard errors clustered at the state level.

driven composition rather than with displacement. Section 8 examines the Tennessee private sector directly.

6.4 Event-study pattern

Figure 5 reports the event-study coefficients on Q_1 share from Specification 8. The figure confirms two features visually.

First, pre-treatment coefficients cluster around zero. The pre-trend F-test for joint significance of the pre-period leads yields $F = 1.23$, $p = 0.32$, and does not reject parallel trends. Second, the treatment coefficient at $k = 0$ (cohort 1987, the first cohort making college-choice decisions with full knowledge of HOPE and the first cohort receiving full first-year disbursement) is positive and significant. Post-treatment coefficients grow monotonically from $k = 0$ to $k = +4$, consistent with gradual adjustment as the program became established. The cohort 1986 coefficient, one lead before treatment, is the omitted reference and is by construction zero. The full numerical coefficients are tabulated in Appendix B (Table A2).

6.5 Distributional view: pre-post Q_1 density

The event study summarizes the result through coefficients at a single outcome; the density-level view in Figure 6 is complementary. The figure compares the distribution of Q_1 share across Tennessee public colleges in the pre-treatment and post-treatment periods.

The post-treatment distribution sits to the right of the pre-treatment distribution, and the shift is present across the bulk of the distribution rather than being driven by a few institutions moving sharply. Combined with the event-study pattern, this gives three complementary views of the same compositional change: a scalar ATT (Section 6.1), a dynamic path (Section 6.4), and a

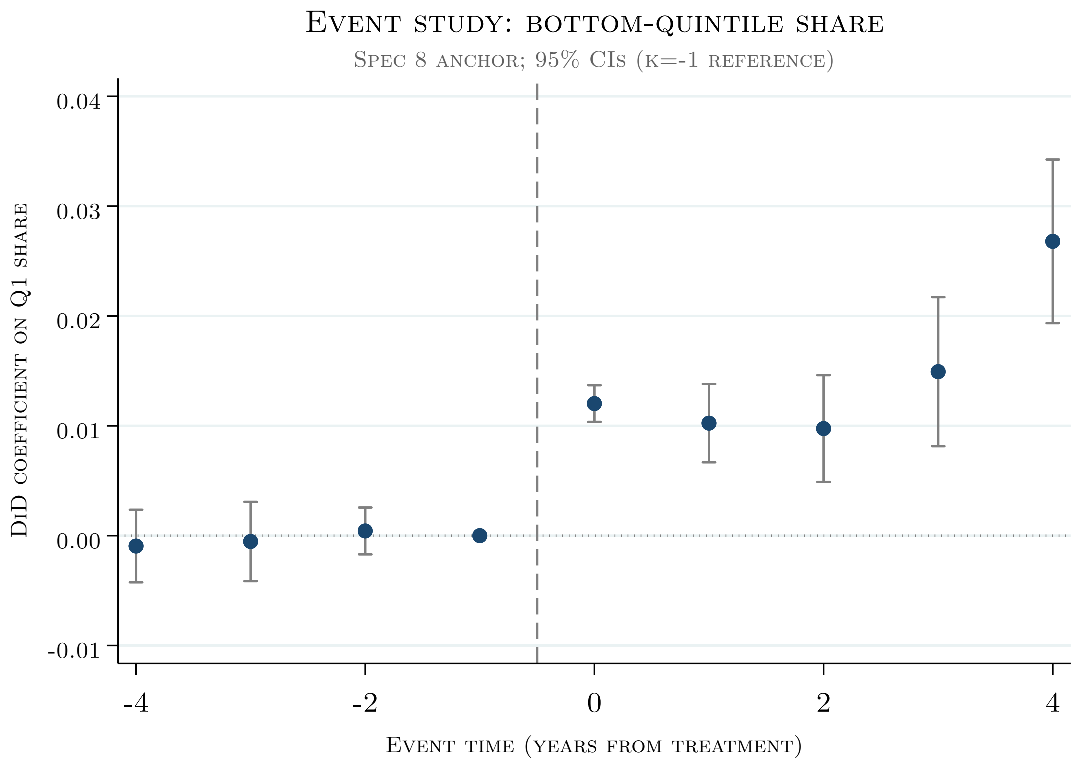

Figure 5: Event-study coefficients for Q_1 share at Tennessee public colleges under Specification 8. Event time k is cohort relative to $t^* = 1987$. Pre-treatment coefficients ($k = -4, \ldots, -1$) cluster around zero; the treatment coefficient at $k = 0$ is positive and significant; post-treatment coefficients ($k = 1, \ldots, 4$) grow monotonically. Shaded bands are 95% confidence intervals from state-clustered standard errors. Full numerical coefficients appear in Appendix B (Table A2).

distributional picture (this subsection).

Sections 7, 8, and 9 now turn to the mechanism behind the compositional shift, the private-sector response, and the robustness of these estimates. Section 10 then asks whether the pattern we document here generalizes to other state merit programs.

7 Mechanism: The Threshold Story

Three independent dimensions of the data corroborate a single mechanism claim: HOPE's eligibility rule (GPA $\geq$ 3.0 OR ACT $\geq$ 21) is permissive, and that permissiveness is what drives the progressive access effect documented in Section 6. The rule binds most tightly for students at the low end of the academic distribution, who disproportionately populate the low end of the income distribution and include more women than men. If this is the right story, the progressive effect should concentrate

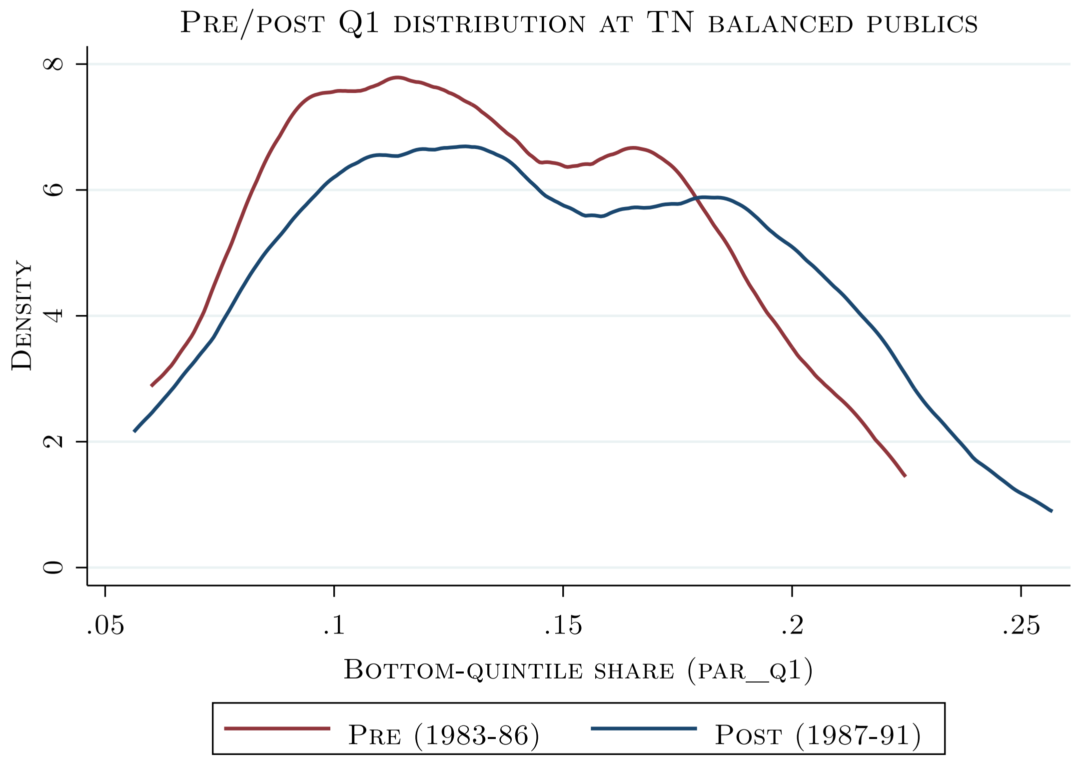

Figure 6: Distribution of bottom-quintile (Q_1) share across Tennessee public colleges in pre-treatment cohorts (1983–1986) and post-treatment cohorts (1987–1991). The post-treatment distribution shifts rightward; the shift is visible across the full range of the distribution rather than concentrated in a few institutions.

(a) at institutions where marginal-eligibility students matter most, (b) among the gender that disproportionately clears the threshold, and (c) in states whose thresholds are permissive enough to admit that margin. We show each pattern in turn.

7.1 Tier heterogeneity

We re-estimate Spec 8 separately on the 12 balanced two-year TN publics and the 7 balanced four-year TN publics. The Q_1 coefficient is roughly twice as large at community colleges: **+0.0175** ($p < 0.001$) at two-year publics versus **+0.0090** ($p < 0.001$) at four-year publics. The four-year estimate is positive and significant, but about half the magnitude. Table 9 reports the full quintile decomposition separately by tier, and Figure 7 plots the Q_1 and Q_5 coefficients side by side.

The tier gap is consistent with a threshold mechanism. At two-year publics, a larger share of matriculants sits near the eligibility margin, so a given relaxation of financial constraints expands who actually enters. At four-year publics, the population is already selected on academic prepara-

Table 9: Tier Heterogeneity

	Two-year publics b/se	Four-year publics b/se
TN × Post (1987)	0.0175*** (0.0037)	0.0090*** (0.0023)
N	2809	1524
r2_a	0.908	0.967

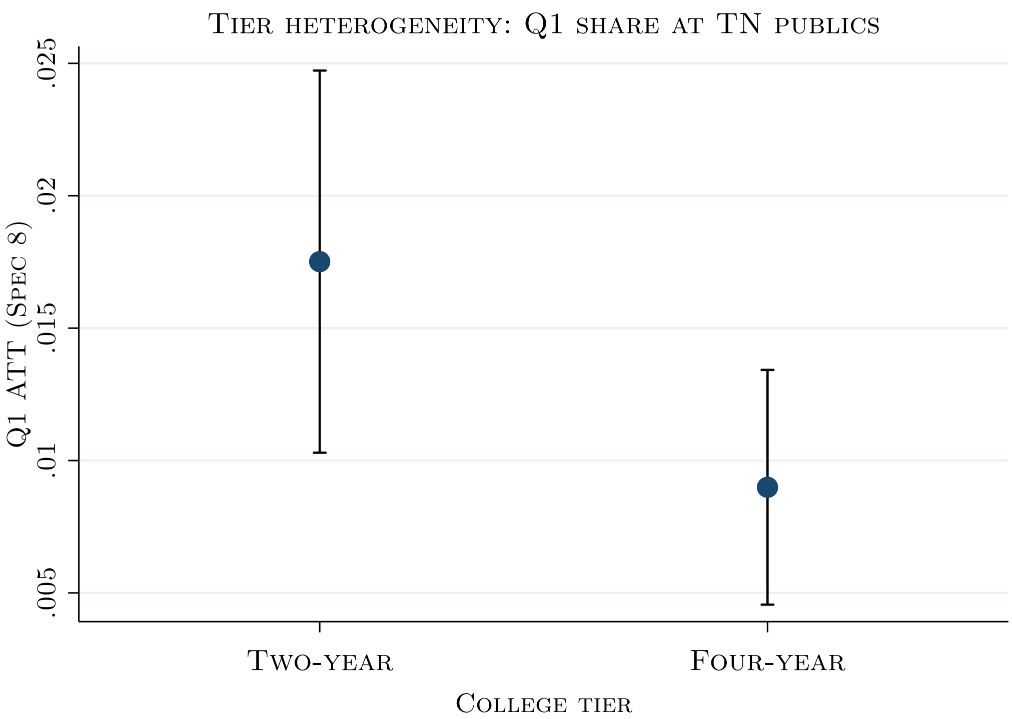

Figure 7: Q_1 and Q_5 ATT by tier (Spec 8 restricted to two-year vs. four-year TN publics). Bars show point estimates with 95% CIs. The Q_1 effect at two-year publics is roughly twice the four-year estimate; Q_5 losses are comparable across tiers.

tion; most admitted students already meet the GPA/ACT threshold, so subsidy eligibility changes who pays rather than who enrolls. The progressive effect is strongest where the threshold does the most work at the extensive margin.

7.2 Gender threshold mechanism

The threshold implies a second, sharper test. If eligibility is the binding channel, and if women have systematically higher high-school GPAs than men at the same income level, then the bottom-quintile effect should be larger for women than for men, and the gap should follow the same tier pattern as the headline effect. We test this directly by running Spec 8 separately on gender-disaggregated Q_1 shares (`par_q1_f`, `par_q1_m`) and on the female-minus-male gap.

At TN publics, the female ATT on Q_1 is **+0.0165** and the male ATT is **+0.0134**. The gender gap is **+0.0031** (SE 0.0008, $p < 0.001$), meaning women's bottom-quintile share rose by roughly 0.31 percentage points more than men's. The gap is almost entirely a two-year phenomenon: **+0.0045** ($p < 0.001$) at community colleges, **+0.0012** (NS) at four-year publics. The gap is also strictly a Q_1 phenomenon: it is indistinguishable from zero at Q_3, Q_4, and Q_5 (Table 10). This matches the threshold story exactly. Where the threshold binds the progressive effect is larger for women; where it does not bind, there is no gender differential to explain.

Table 10: Gender Disaggregation: Q1 Share

	Female Q1 share b/se	Male Q1 share b/se	Female-Male Gap b/se
TN × Post (1987)	0.0165*** (0.0033)	0.0134*** (0.0029)	0.0031*** (0.0008)
N	4333	4333	4333
r2_a	0.918	0.892	0.335

A third dimension sharpens the claim further. If gender gaps reflect how the threshold interacts with the underlying gender gap in high-school GPA, stricter thresholds should produce larger gender gaps across states. We run the same gender-disaggregated Spec 8 on each comparison state's publics. The pattern is monotonic in threshold strictness. Massachusetts's Adams Scholarship (GPA $\geq$ 3.3 with SAT/ACT, an AND condition) produces a gender gap of **+0.0094** (p<0.001), about three times the TN gap. Tennessee (GPA $\geq$ 3.0 OR ACT $\geq$ 21, moderate and disjunctive) produces **+0.0031**. South Carolina's LIFE/Palmetto/HOPE system (permissive, GPA $\geq$ 3.0, 2-of-3 criteria) and Nevada's Millennium (GPA $\geq$ 3.0, no test requirement) produce gaps statistically indistinguishable from zero ($-$**0.0043** with p=0.09 in SC; $-$**0.0011**, NS, in NV). Figure 8 visualizes the full pattern, plotting the female-male Q_1 gap against threshold strictness with Q_3–Q_5 gaps overlaid for comparison.

The three-dimensional pattern is the tightest corroboration we can offer for the threshold claim. The progressive effect is larger (i) at the tier where marginal-eligibility students cluster, (ii) for the

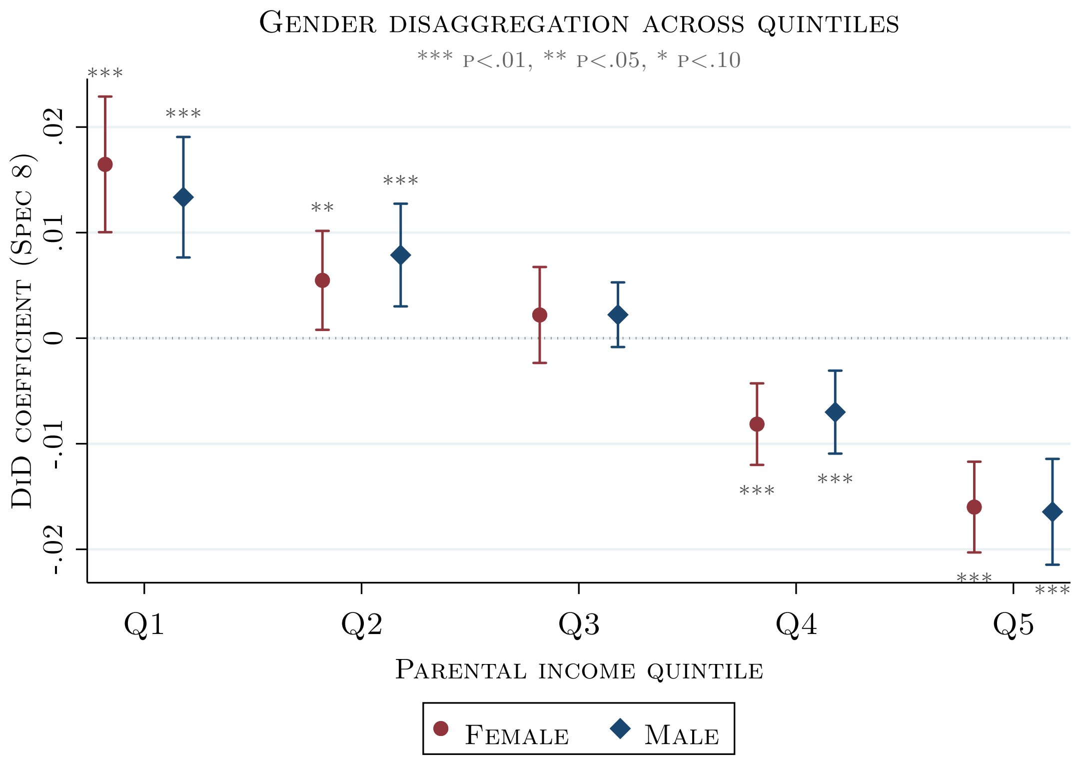

Figure 8: Gender gap (female ATT minus male ATT) on Q_1 share, by state and tier. The gap rises with threshold strictness (MA > TN ≫ SC, NV) and concentrates at tiers where the threshold binds (TN two-year > TN four-year). Q_3–Q_5 gaps shown in lighter shading for contrast; none reach conventional significance.

gender that disproportionately clears the threshold, and (iii) in the quintile where the threshold actually binds, with the gender gap scaling monotonically across states with threshold strictness. No generic enrollment shock — no uniform demand increase, no state-wide demographic trend, no labor-market pull — would produce gender-differential, tier-concentrated, quintile-specific effects that also correlate with threshold strictness across six states. The pattern is hard to generate without a threshold doing the work.

7.3 Eligibility evidence from NLSY97

The simulated-eligibility data we introduced in Section 4 (Figure 1) supply the microfoundation for this story. Applying HOPE's rule to NLSY97 respondents, we compute the share of each income-quintile × gender cell that would have qualified on GPA or ACT. Eligibility rises monotonically from roughly 32% at Q_1 to roughly 60% at Q_5, and within each quintile the female rate exceeds the male rate by about 10 percentage points at the low-income margin. This is the empirical foundation for the threshold mechanism. The threshold is differentially non-binding higher in the income distribution, where many students already qualify and HOPE operates closer to a price subsidy than an access expansion; and it is differentially non-binding for women, whose higher high-school GPAs place a larger share above the cutoff at each income level. The three-dimensional pattern in Sections 7.1 and 7.2 is what we would expect to find if the NLSY97 eligibility gradient is the operative mechanism rather than an unrelated correlate.

7.4 HBCU robustness

Because historically Black colleges and universities have distinct institutional histories that could interact with any measured compositional shift, we re-estimate the Q_1 ATT on a sample that excludes Tennessee's HBCUs. The point estimate is **+0.0143** versus Spec 8's **+0.0150** — roughly a 5% attenuation, with the headline conclusion unchanged. The mechanism evidence in this section is not an HBCU artifact.

8 Spillover and Sector Comparisons

The mechanism in Section 7 operates through an eligibility threshold that governs whether HOPE can be used at all. Because HOPE is partially portable to Tennessee private colleges at a reduced award amount, the private sector's composition could in principle move in response to HOPE as well. We test this directly by running the Spec 8 design restricted to Tennessee private colleges, and then compare publics and privates within the state qualitatively.

8.1 Spillover at Tennessee Privates

Table 12 reports the Spec-8-equivalent DiD estimated on Tennessee private four-year colleges, with the same control states and treatment timing as the main analysis. The Q_1 ATT at Tennessee privates is -0.0053 (SE 0.0012, $p < 0.001$): the bottom-quintile share at TN privates fell by roughly 0.5 percentage points relative to private colleges in non-merit states. The Q_5 share at TN privates is essentially unchanged (+0.0003, NS). Total enrollment at TN privates rose modestly (+6.3 students per college-cohort, $p < 0.05$).

Table 11: Spillover at Tennessee Private Colleges

Table 12: Spillover at TN Private Colleges

	par_q1 b/se	par_q5 b/se	Total enrollment b/se
TN Private × Post (1987)	-0.0053***	0.0003	6.3310**
	(0.0012)	(0.0032)	(2.7592)
N	3137	3137	3137
r2_a	0.913	0.970	0.986

The private-sector pattern is consistent with the public-sector result. HOPE is worth substantially more at a Tennessee public college than at a Tennessee private, and the public-sector Q_1 gain is about three times the private-sector Q_1 loss. Read together, the two results are consistent with the threshold mechanism in Section 7: HOPE-eligible Q_1 students face a larger net-price reduction at publics than at privates, and the Tennessee public sector became disproportionately more attractive to low-income students relative to the Tennessee private sector. We do not claim that individual students moved across sectors; the Chetty panel reports shares and counts at the college-cohort level and cannot track individual choices. We report the public–private contrast as *consistent with* a threshold-driven compositional reallocation and do not claim individual displacement.

Figure 9 displays the quintile gradient at TN publics alongside the gradient at TN privates under the same specification. The public-sector bars reproduce the monotonic shift from Section 6; the private-sector bars show a smaller but opposite pattern at Q_1. This cross-sector contrast is the evidence for sector-specificity.

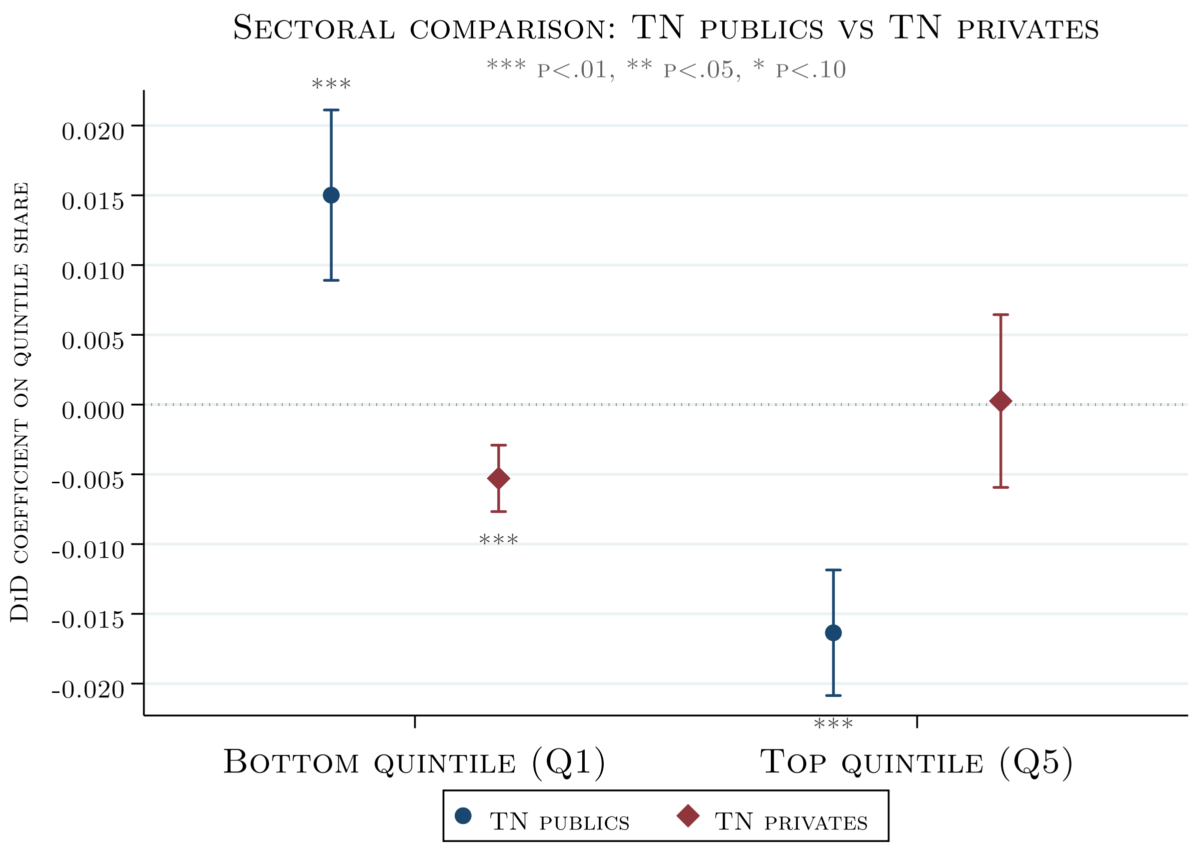

Figure 9: Quintile Gradient: Tennessee Publics vs. Tennessee Privates

Notes: Quintile ATTs from Spec-8-equivalent DiD, estimated separately on Tennessee public colleges (left) and Tennessee private four-year colleges (right). Control states, treatment timing, and covariates are identical across panels. 95% confidence intervals shown. Data: Chetty et al. (2020).

8.2 Within-Tennessee Comparison

A within-state design (Tennessee publics vs. Tennessee privates, same state and calendar time) offers a different source of variation from the cross-state comparison. The within-TN point estimate is directionally consistent: Q_1 share rises at Tennessee publics and falls modestly at Tennessee privates, producing a within-state Q_1 gap that moves in the expected direction. Because the design has only one state cluster, inference is inadequate, so we do not report standard errors or a full specification table in the main text. The qualitative agreement across two designs with different identifying variation is reassuring. The triple-difference in Section 9 formalizes the same comparison under a design that can be clustered adequately, and delivers a larger public-specific ATT (+0.0203,

$p < 0.001$).

9 Robustness

We subject the headline result — a +1.50 pp increase in bottom-quintile share at Tennessee public colleges (Spec 8) — to a sequence of tests organized around the binding concern: pre-trends. We lead with Lee–Wooldridge unit-specific detrending, then Rambachan–Roth sensitivity, then a triple-difference, then a compact summary of timing and specification variation. The Q_1 and enrollment findings survive the full sequence; we are explicit about the limits of what remains identified after detrending.

9.1 Lee–Wooldridge unit-specific detrending

The most demanding test allows each college to follow its own linear pre-treatment trajectory. Following Lee and Wooldridge (2026), we fit a college-specific linear trend on pre-treatment cohorts (1983–1986), subtract the fitted line from the outcome at every cohort, and re-estimate the DiD on the detrended series. The coefficient answers: "How much did Tennessee publics shift relative to controls after removing each college's own pre-treatment slope?"

Table 13 and Figure 10 report the comparison.

Table 13: Lee-Wooldridge Detrending

	Original b/se	Detrended b/se	Original b/se	Detrended b/se	Original b/se	Detrended b/se
TN × Post	0.0150***	0.0132***	-0.0164***	-0.0080***	94.9920***	98.0072***
	(0.0031)	(0.0019)	(0.0023)	(0.0022)	(22.7990)	(11.3261)
N	4333	4275	4333	4275	4333	4275
r2_a	0.943	0.397	0.980	0.373	0.989	0.367

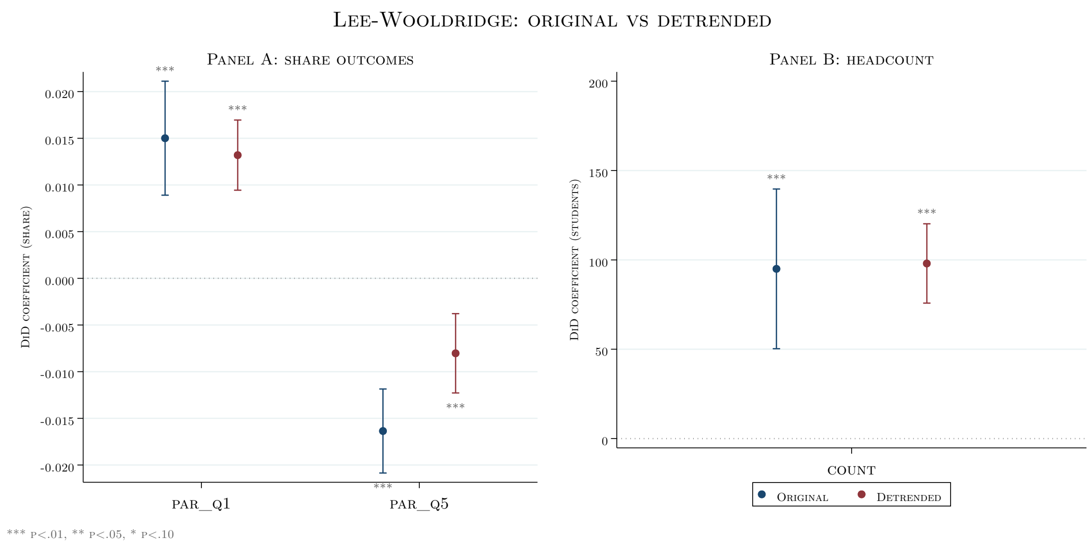

Figure 10: Original vs. Lee–Wooldridge detrended estimates

Notes: Spec 8 point estimates and 95% confidence intervals for the original TWFE estimator (navy) and the unit-specific detrended estimator (maroon). Detrending removes a college-specific linear pre-treatment trend fit on cohorts 1983–1986. State-clustered standard errors (26 clusters). Data: Chetty et al. (2020).

Three facts stand out. First, the Q_1 share ATT moves from +0.0150 to +0.0132, a 12% attenuation that remains significant at $p < 0.001$. The headline result survives. Second, the enrollment count ATT moves from +95 to +98 students per college-cohort — essentially unchanged — confirming that the expansion finding is not a pre-trend artifact. Third, the Q_5 share ATT attenuates from −0.0164 to −0.0080 (a 51% reduction) but remains significant at $p < 0.001$, consistent with a real top-quintile decline that also rode a modest pre-existing trend.

The honest interpretation of the quintile gradient: Q_1 and Q_5 are the robust endpoints. The middle-quintile coefficients (Q_2–Q_4) partly reflect continuation of pre-existing trends, and we treat them as corroborative rather than independently robust. The monotonic ordering across all five quintiles remains the cleanest visual summary of the compositional shift, but the tails carry the weight.

9.2 Rambachan–Roth sensitivity

Standard pre-trend tests (ours passes with $F = 1.23$, $p = 0.32$) check whether pre-treatment coefficients are distinguishable from zero. They do not bound how much a small undetected pre-trend could bias the treatment effect. The HonestDiD framework of Rambachan and Roth (2023) flips the question: assume a pre-trend violation exists and bound the ATT as a function of its magnitude, expressed as a multiple M of the worst observed pre-treatment coefficient. The breakdown value M^* is the value at which the CI first includes zero. Applied to the Spec 8 Q_1 event study, the CI never includes zero at any $M \leq 2.0$, so $M^* > 2$. The Q_1 result is robust to pre-trend violations of up to twice the worst observed pre-treatment coefficient. Figure 11 displays the sensitivity profile.

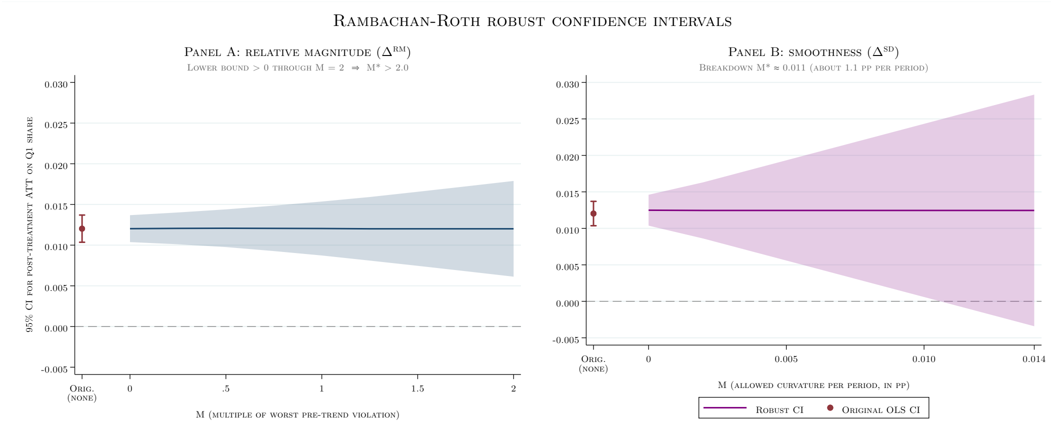

Figure 11: Rambachan–Roth sensitivity: Q1 share

Notes: HonestDiD 95% confidence intervals for the Spec 8 Q_1 ATT under the relative-magnitude restriction of Rambachan and Roth (2023). The horizontal axis reports M, the assumed violation relative to the worst observed pre-treatment coefficient. The CI excludes zero through $M = 2.0$.

9.3 Triple-difference

A TN-wide demographic shift unrelated to HOPE would bias both public and private TN estimates. A triple-difference (TN $\times$ Public $\times$ Post) differences out any such state-wide trend by comparing the post-HOPE change at TN publics to the post-HOPE change at TN privates, relative to the same public–private contrast elsewhere. The triple-difference ATT is +0.0203 (SE 0.0023, $p < 0.001$), *larger* than the two-way DiD. The TN $\times$ Post coefficient alone (that is, the TN-wide cohort trend on Q_1) is negative and significant (-0.0053, $p < 0.001$), which is why differencing it out raises the public-specific ATT rather than leaving it unchanged. The negative TN-wide coefficient matches the Q_1 decline at Tennessee privates documented in Section 8. The compositional effect is specific to Tennessee's public sector, exactly as the HOPE-through-publics mechanism predicts. Table 14 reports the full specification.

Table 14: Triple-Difference

	par_q1 b/se
TN × Post	-0.0053***
	(0.0012)
Pub × Post	-0.0018
	(0.0023)
TN × Pub × Post (triple-diff)	0.0203***
	(0.0023)
N	7466
r2_a	0.951

9.4 Callaway–Sant'Anna

The Callaway and Sant'Anna (2021) heterogeneity-robust estimator, applied to the Spec 8 sample, returns an ATT of +0.01475 — within 3% of the TWFE estimate. The result is not an artifact of the two-way fixed-effects structure. See also Roth (2026) on estimator choice under sharp timing.

9.5 Specification menu

Across the full 12-specification menu (varying panel balance, cohort trims, control-state composition, and BEA pre-period controls), the Q_1 ATT ranges from +0.013 to +0.019, excluding unreliable four-cluster bordering-state specifications. The headline result is not sensitive to any single specification choice. Full details appear in Appendix B.

9.6 Timing robustness

Tennessee HOPE was enacted in mid-2003, first disbursed fall 2004, and students enrolled in fall 2003 received retroactive bridge awards after enrollment decisions were finalized. Our preferred timing $t^* = 1987$ aligns with the first cohort that made its college decision with full HOPE knowledge and received first-year disbursement. We test two alternatives. At $t^* = 1986$ (Spec 12, announcement-year timing), the Q_1 ATT attenuates to +0.0127 with the $k = 0$ cohort-1986 coefficient essentially zero, indicating that the compositional response did not operate at announcement. A 1985 placebo (setting $t^* = 1985$ in a pre-disbursement window that still includes cohort 1986) returns +0.00358 ($p = 0.028$) — not a clean null, but the contamination is mechanical:

the placebo picks up the announcement-year transitional effect at cohort 1986. Roughly 24% of the disbursement-era effect appears in cohort 1986; the remaining effect appears starting cohort 1987 (first cohort with full HOPE knowledge and first-year disbursement). This is consistent with HOPE's compositional effect operating through actual disbursement, not announcement — itself consistent with the retroactive bridge provision being behavior-neutral.

9.7 Other robustness

Four additional tests, reported in single sentences. Age-18 decision-year robustness: identical to the age-17 benchmark to five decimal places (the BEA baseline is a 7-year county average, insensitive to the specific year). Capacity-constraint test: the compositional effect is invariant across control-type subsets (Q_1 ATT +0.0160 against tight-capacity controls, +0.0156 against loose-capacity controls). Enrollment-weighted: the Q_1 ATT is smaller (+0.0082, $p = 0.006$), consistent with community colleges driving the effect per the tier heterogeneity reported in Section 7; the share finding is robust in direction and significance, and the reduced magnitude under enrollment weighting reflects the weight that small two-year colleges carry in the unweighted estimate. Synthetic control: considered but deferred. The standard donor pool produces unreliable weights given Tennessee's pre-HOPE Q_1 trajectory; a properly specified donor pool warrants standalone work, and we do not rely on the SCM result.

9.8 Honest bound

Pre-trend sensitivity is the paper's binding constraint. Q_1 and enrollment count survive unit-specific detrending; Q_5 attenuates but remains significant; middle-quintile coefficients are corroborative rather than independently robust. The headline claim — a progressive compositional shift at the bottom of the income distribution accompanied by enrollment expansion — is bounded by these tests and remains significant throughout.

10 Cross-State Evidence

Six states in our sample adopted broad-based merit-aid programs between 1998 and 2004: Tennessee, Massachusetts, South Carolina, Nevada, West Virginia, and South Dakota. We apply the

Spec-8-equivalent design to each in turn and then pool them. Three results emerge. First, the direction of distributional effects tracks program-design parameters: low thresholds and first-dollar awards produce progressive composition shifts, while high thresholds and small fixed awards produce regressive or null shifts. Second, a pooled staggered DiD across all six states yields a Q_1 effect near zero, precisely because progressive and regressive programs partially cancel. Third, the threshold-favors-women mechanism from Section 7 generalizes: cross-state gender gaps grade monotonically with threshold strictness.

10.1 Six-state comparison

For each treatment state we estimate the Spec-8 specification on a balanced panel of that state's public colleges, drop early cohorts (1980–1982) where the pre-period allows, use the national ex-Massachusetts ex-treated-state control, condition on the pre-treatment BEA baseline, and cluster standard errors at the state level. This design is identical in structure to our Tennessee specification; only the treated state, its treatment timing, and the exclusion rule for the control group change.

Table 15 summarizes the results. Tennessee and Massachusetts produce the cleanest progressive patterns: Q_1 rises by 1.5 and 1.7 percentage points respectively, and both systems expanded. South Carolina is weakly progressive (+0.63 pp Q_1) but identification is thin: South Carolina's LIFE scholarship was enacted in 1998, which leaves only one pre-treatment cohort (1980) in the Chetty panel. Nevada's Millennium Scholarship produced the largest enrollment expansion in the sample (+387 students per college-cohort) with no Q_1 effect and a significant Q_5 decline. West Virginia's PROMISE coincided with a roughly 12% contraction of its public system, which makes its small negative Q_1 coefficient difficult to interpret. South Dakota is the cleanest regressive case: Q_1 fell by 1.64 percentage points and enrollment contracted by 122 students per college-cohort.

10.2 Design predicts direction

The six states separate cleanly by design parameters. The progressive pattern groups states with permissive thresholds, first-dollar structure, and broad eligibility: Tennessee (GPA 3.0 OR ACT 21), Massachusetts (stricter, but still inclusive at GPA 3.3), and, partially, South Carolina. Each of these produced a statistically significant Q_1 gain and did not contract total enrollment. The regressive pattern is anchored by South Dakota: the combined GPA-and-ACT threshold (3.0 *and*

Table 15: Cross-state Spec-8-equivalent results

State	Threshold	Award	Q_1 ATT	Count	Verdict
TN	GPA 3.0 OR ACT 21	\$1.5–3k, first-dollar	+0.0150***	+95***	Progressive + expansion
MA	GPA 3.3 + SAT/ACT	Tuition-scaling	+0.0171***	+42*	Progressive + modest expansion
SC	GPA 3.0, 2-of-3	\$5–6.7k, first-dollar	+0.0063**	+9 (NS)	Weakly progressive (weak ID)
NV	GPA 3.0	\$960/yr, fixed	−0.0019 (NS)	+387***	Null Q_1, expansion, $Q_5 \downarrow$
WV	ACT 22 + subscores (AND)	Full tuition initially	−0.0020 (NS)	−131***	Null/regressive + contraction
SD	GPA 3.0 + ACT 24 (AND)	\$1.0–1.3k, fixed	−0.0164***	−122***	Regressive + contraction

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. Each row reports a separate Spec-8-equivalent regression on a balanced panel of the named state's public colleges, with the national ex-Massachusetts ex-treated-state public colleges as the control group and pre-treatment BEA baseline income as the sole covariate. Program-design parameters in columns 2–3 are summarized from state statutes as of the cohort year of first disbursement.

24) qualifies a much narrower slice of the ability distribution, and the fixed \$1,000–1,300 award is small relative to tuition. South Dakota's Q_1 fell sharply, consistent with a program that subsidized students who would have enrolled anyway and shifted the system's composition toward the top.

Two states are confounded. West Virginia's public system contracted by roughly 12% during the PROMISE era, entangling any distributional coefficient with that contraction: a shrinking system can produce compositional change mechanically as marginal colleges close or lose enrollment. Nevada is harder still: the Millennium award is small (about \$960/year), the program coincided with the Great Recession and a demographic boom in Nevada's college-age population, and the combination makes it difficult to isolate a composition effect from the large expansion. These two states carry less evidentiary weight than the four clean cases.

The broader point is that the six states are not a random sample of merit-aid programs, but even within this sample the program-design parameters identified in Section 2—threshold strictness, first-dollar vs. last-dollar, award size relative to price—line up with the direction of the distributional effect. South Dakota is the cleanest regressive case; Tennessee and Massachusetts are the cleanest progressive cases. The contrast is not subtle.

10.3 Pooled staggered DiD: design heterogeneity washes out

Pooling all six treatment states in a staggered DiD—each state treated from its own cohort of first disbursement, with the ex-Massachusetts national controls—gives a Q_1 ATT of +0.0103 ($p = 0.05$, marginal) and an enrollment coefficient that is not distinguishable from zero (+38.6, SE 62.4). This is the number one would report if one treated "merit aid" as a single homogeneous treatment.

The pooled Q_1 is substantially smaller than the individual progressive-state ATTs because progressive and regressive programs partially cancel. This supports the paper's central thesis: the existence of merit aid does not determine its distributional consequence; the design parameters do. A pooled average obscures the very heterogeneity the paper is about. We report the pooled estimate here for transparency and to make that obscuring concrete. Cross-state pooling estimates that average across heterogeneous designs recover something close to the design-weighted average, which depends on which programs are in the sample.

10.4 Mechanism corroboration in Massachusetts

The threshold-favors-women mechanism from Section 7 predicts a specific cross-state pattern: gender gaps on Q_1 should be larger where thresholds are stricter, because women disproportionately clear higher academic bars given gender differences in high-school GPA. We estimate the female-minus-male Q_1 gap for each state using the same Spec-8-equivalent design as above.

The pattern is graded and consistent with the prediction across the progressive states. Massachusetts, with the strictest inclusive threshold among them (GPA 3.3 plus SAT or ACT), produces the largest gender gap: +0.94 pp ($p < 0.001$). Tennessee, with a moderate threshold (GPA 3.0 OR ACT 21, a disjunction that is easier to clear), produces a smaller gap of +0.31 pp ($p < 0.001$). South Carolina and Nevada, both with permissive thresholds, produce gaps that are not statistically distinguishable from zero. South Dakota, the regressive case with the strictest threshold overall (GPA 3.0 *and* ACT 24), has a gender gap of +1.43 pp: because women clear the ACT-24-plus-GPA-3.0 threshold at higher rates than men, the smaller share of men who qualify for HOPE in SD leaves marginal low-income men without the subsidy. The resulting regressive shift hits male Q_1 harder than female Q_1. Figure 12 plots the Q_1 ATT and gender gap against threshold strictness across the six states.

This is the cleanest out-of-sample test of the threshold mechanism. The gender-gap prediction is sharp, it does not require the overall Q_1 effect to be progressive to apply, and it survives the state-specific noise that clouds the level comparisons.

10.5 South Dakota: a note on the regression design

A simple pre/post comparison for South Dakota gives an apparent Q_1 change of +0.30 pp, which disagrees with the −1.47 pp reported in prior literature. The discrepancy is an artifact of the simple

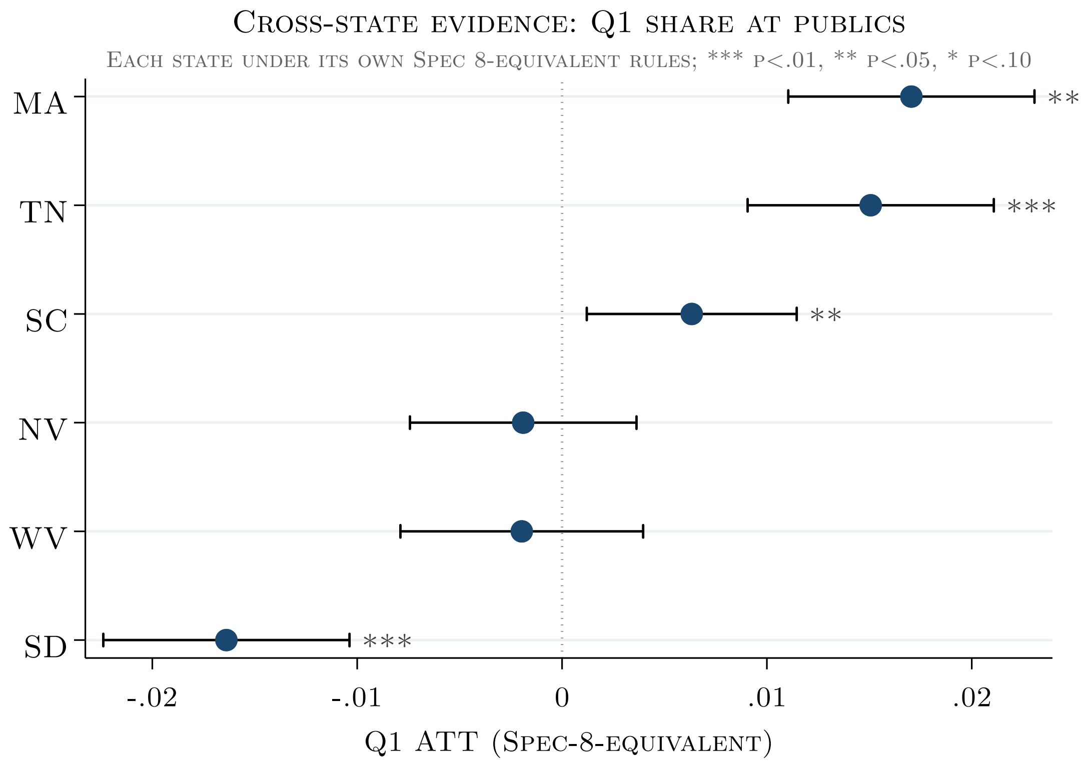

Figure 12: Cross-state Q_1 ATT and female-minus-male gap by program-design parameters. Each point is a state-specific Spec-8-equivalent estimate. The gender gap grades with threshold strictness in the expected direction.

specification: it omits college fixed effects and cohort trends, both of which load heavily in South Dakota because its public system has only six colleges in the Chetty panel and large between-college composition differences. The Spec-8-equivalent regression, which conditions on college fixed effects and pre-treatment income, gives a Q_1 ATT of -0.0164 ($p < 0.001$). The regression-based estimate aligns with the direction and magnitude reported in the prior literature.

11 Discussion

11.1 Mechanism in one sentence

Tennessee HOPE's permissive eligibility threshold, its first-dollar design, and a capacity response that absorbed new entrants combine to shift public-college composition toward lower-income students through expansion rather than displacement. The threshold pulls in marginal applicants who are disproportionately female, disproportionately at community colleges, and disproportionately in the bottom of the parental-income distribution; the first-dollar design preserves the subsidy for Pell-eligible students; the capacity response means the new entrants do not push anyone out. Remove any one of the three and the prediction flips.

11.2 Policy implications

Threshold strictness, not award size, is the first-order lever. The cross-state evidence in Section 10 shows that states with permissive thresholds and progressive pull (Tennessee, Massachusetts) produce progressive compositional shifts even when award sizes differ; South Dakota's combination of a higher threshold with a small award is regressive. The design literature—and the policy conversation about merit aid—tends to focus on award amounts, which is secondary. The first-order question is whether the marginal-eligibility student clears the threshold.

First-dollar design matters for the access margin. Georgia HOPE's Pell offset eliminated the effective subsidy for the lowest-income students until 2000; Dynarski (2000) documents the consequence, which is the opposite of what we find for Tennessee. First-dollar status preserves the access margin for Q_1 families and is a necessary condition for a program to produce a progressive compositional shift.

Capacity response separates expansion from displacement. Our roughly +95-student enrollment finding is the feature that distinguishes Tennessee from the regressive and contracting cases in Section 10. Whether merit aid expands access depends on whether the public system is able or willing to absorb new entrants. In West Virginia and South Dakota, where enrollment contracted during the treatment window, a share-based reading is mechanically confounded by the denominator.

Disbursement, not announcement, is when behavior changes. The cohort 1986 event-study coefficient—the announcement-year cohort—is essentially zero, and the cohort 1987 coefficient is where the discrete change appears. The bridge provision that granted retroactive awards to 2003 enrollees did not shift initial enrollment decisions. We interpret this as evidence that merit programs need to actually disburse, not just announce, to change behavior at the college-choice margin. The broader implication is that the lag between enactment and first disbursement is the relevant window for policy evaluation.

11.3 Limitations

Single-state design. The core result is a difference-in-differences on a single treated state. The cross-state evidence in Section 10 partly mitigates, and the triple-difference and Lee–Wooldridge tests in Section 9 address residual state-specific trend concerns, but the headline result is not identified from within-state variation. A reader who rejects the control-state comparison as a whole has grounds to be skeptical.

College-county versus student-county income. Our pre-treatment BEA control measures the per-capita income of the county where a college is located, not where students come from. The mismatch is a form of measurement error on a control variable rather than on treatment, so the bias is not first-order for the Q_1 coefficient, but the matter is unresolved until origin-county data from FSRDC become available. We flag this limitation throughout the paper.

The middle of the distribution. The Q_2, Q_3, and Q_4 coefficients are directionally consistent with the monotonic gradient in our chosen specification but lose significance or change sign under unit detrending. We report them as corroborative rather than independently identified. Q_1 share and enrollment count are the two robust endpoints.

The 1985 placebo is not clean. Setting $t^* = 1985$ yields a small significant coefficient at Q_1 (+0.00358, $p = 0.028$), roughly 24% of the full disbursement-era effect. We interpret the placebo coefficient as reflecting the announcement-period transitional behavior by the 1986 cohort rather than a pre-existing trend, and the event-study pattern in Section 6 supports that reading, but the placebo does not deliver a clean null.

Mobility and labor-market outcomes. We do not follow HOPE-era Tennessee cohorts to labor-market outcomes. The Q_1 share gain is a compositional shift at the point of entry; whether it translates into downstream mobility or earnings gains is the next open question. The Chetty panel contains later-life income data, and this is a natural extension.

11.4 Future work

Three extensions are immediate. First, state-specific quintile cutoffs drawn from Census or ACS would test whether our national-quintile shares reflect a shift in the state-specific distribution or a national comparison. Second, IPEDS tuition validation across all ten merit programs would tighten the cross-state design parameters. Third, origin-county income data via FSRDC would remove the college-county/student-county mismatch discussed above. Each is feasible in principle; none is in scope for this paper.

12 Conclusion

Tennessee HOPE produced a progressive compositional shift at public colleges: the bottom-quintile share rose by 1.50 percentage points, with a monotonic gradient across the five income quintiles, and total enrollment at Tennessee publics expanded by roughly seven percent over the same window. The effect concentrates at community colleges and is larger for female students, consistent with HOPE's permissive GPA-or-ACT threshold binding most tightly on marginal applicants. The Q_1 share and the enrollment result survive unit-specific detrending, Rambachan-Roth sensitivity at breakdown values above $M^* = 2$, and a triple-difference specification. Applying the same design to five other treatment states, Massachusetts replicates the Tennessee pattern, South Dakota mirrors it, and the gender gap at Q_1 scales across states with threshold strictness.

The broader lesson is about design. A merit program that combines a permissive threshold with a first-dollar structure and a capacity response can expand access rather than concentrate it. A program that layers a strict threshold on a last-dollar structure—or that pairs any threshold with enrollment contraction—cannot. The Georgia result that set the distributional critique in motion (Dynarski, 2000) and the Tennessee result here are two outcomes of the same framework evaluated at different design parameters.

Three extensions would tighten the argument. State-specific income quintiles constructed from Census or ACS data would isolate within-state compositional change from movement relative to the national distribution. IPEDS tuition validation across all ten merit programs would sharpen the cross-state design parameters. Origin-county income data via FSRDC would resolve the college-county versus student-county mismatch that is the one material limitation of our BEA control. Each is feasible; none is in this paper. The claim we make here is narrower: Tennessee HOPE expanded access at public colleges through a design channel we can identify, and the same channel predicts the sign of compositional change across five other treatment states.

References

Abadie, A., Athey, S., Imbens, G. W., & Wooldridge, J. M. (2023). When should you adjust standard errors for clustering? *Quarterly Journal of Economics*, *138*(1), 1–35. https://doi.org/10.1093/qje/qjac038

Bleemer, Z. (2022). Affirmative action, mismatch, and economic mobility after california's proposition 209. *Quarterly Journal of Economics*, *137*(1), 115–160.

Bruce, D. J., & Carruthers, C. K. (2014). Jackpot? the impact of lottery scholarships on enrollment in tennessee. *Journal of Urban Economics*, *81*, 30–44.

Callaway, B., & Sant'Anna, P. H. C. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, *225*(2), 200–230. https://doi.org/10.1016/j.jeconom.2020.12.001

Card, D., & Krueger, A. B. (2005). Would the elimination of affirmative action affect highly qualified minority applicants? evidence from california and texas. *ILR Review*, *58*(3), 416–434.

Chetty, R., Friedman, J. N., Saez, E., Turner, N., & Yagan, D. (2020). Income segregation and intergenerational mobility across colleges in the united states. *Quarterly Journal of Economics*, *135*(3), 1567–1633. https://doi.org/10.1093/qje/qjaa005

Cohodes, S. R., & Goodman, J. S. (2014). Merit aid, college quality, and college completion: Massachusetts' adams scholarship as an in-kind subsidy. *American Economic Journal: Applied Economics*, *6*(4), 251–285.

Dynarski, S. (2000). Hope for whom? financial aid for the middle class and its impact on college attendance. *National Tax Journal*, *53*(3), 629–661.

Dynarski, S. (2004). *The new merit aid* (tech. rep. No. Working Paper No. 9400). National Bureau of Economic Research.

Herbaut, E., & Geven, K. (2020). What works to reduce inequalities in higher education? A systematic review of the (quasi-)experimental literature on outreach and financial aid. *Research in Social Stratification and Mobility*, *65*, 100442. https://doi.org/10.1016/j.rssm.2019.100442

Hinrichs, P. (2012). The effects of affirmative action bans on college enrollment, educational attainment, and the demographic composition of universities. *Review of Economics and Statistics*, *94*(3), 712–722.

Jones, T. R., Kreisman, D., Rubenstein, R., Searcy, C., & Bhatt, R. (2022). The effects of financial aid loss on persistence and graduation: A multi-dimensional regression discontinuity approach. *Education Finance and Policy*, *17*(2), 206–231.

Lee, S. J., & Wooldridge, J. M. (2026). *Simple approaches to inference with difference-in-differences estimators with small cross-sectional sample sizes* [Working Paper].

Rambachan, A., & Roth, J. (2023). A more credible approach to parallel trends. *Review of Economic Studies*, *90*(5), 2555–2591. https://doi.org/10.1093/restud/rdad018

Roth, J. (2026). *Interpreting event-studies from recent difference-in-differences methods* [Working Paper, Brown University]. https://jonathandroth.github.io/assets/files/HetEventStudies.pdf

Sjoquist, D. L., & Winters, J. V. (2015). State merit-based financial aid programs and college attainment. *Journal of Regional Science*, *55*(3), 364–390. https://doi.org/10.1111/jors.12161

Song, Y., & Rubenstein, R. (2024). How does reducing merit-aid generosity and certainty affect college choices? evidence from georgia's hope scholarship. *Research in Higher Education*, *65*(1), 1–41.

Tennessee Higher Education Commission. (2025). *Tennessee education lottery scholarship program: 2025 annual report – recipient outcomes through fall 2024* (tech. rep.) (Accessed: 2026). Tennessee Higher Education Commission. https://www.tn.gov/hope/TELS_Report_2025

A Supporting Figures

This appendix collects supporting figures referenced in the main text. Main-text figures are embedded inline; this appendix reports the data diagnostics underlying Section 4 and the full event-study output behind the summary statistics in Section 6 and Section 9.

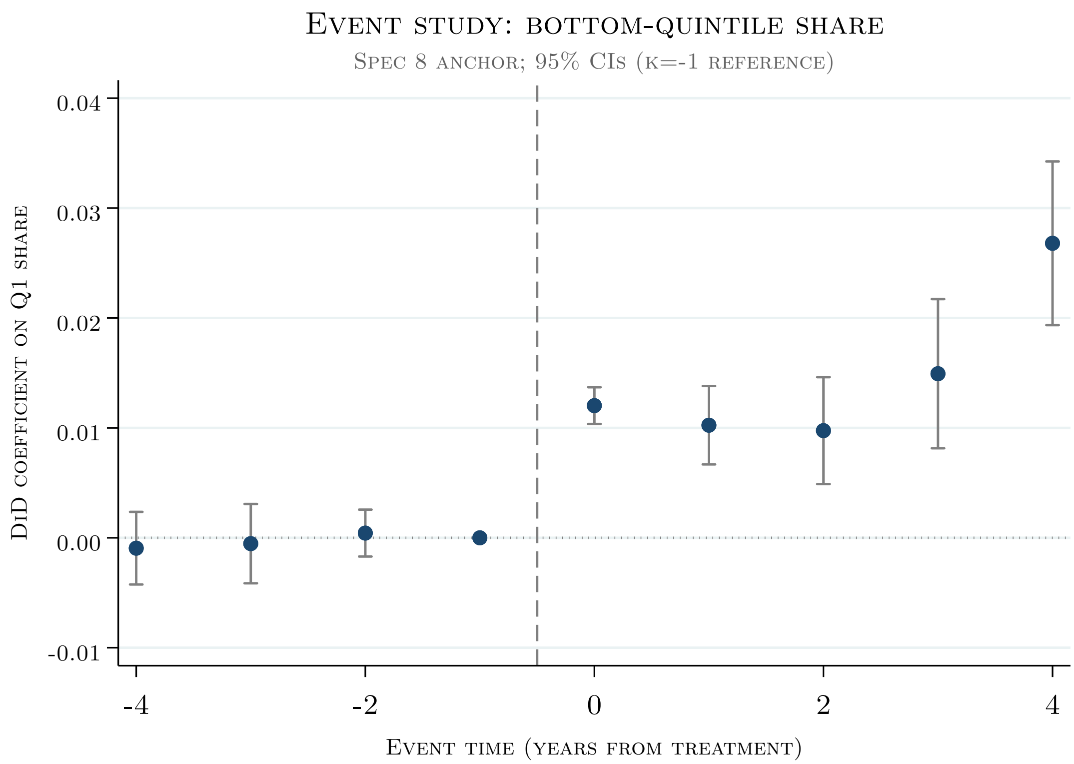

Figure A1: Event Study: Bottom-Quintile Share at Tennessee Publics (Full Output)
Notes: Event-study coefficients from the chosen specification (Spec 8) with $k = -1$ (cohort 1986) as the omitted reference period. Pre-trend F-test: $F = 1.23$ ($p = 0.32$). Vertical dashed line at $k = 0$ (cohort 1987, first full-HOPE cohort). Shaded band is the 95% confidence interval using state-clustered standard errors. Data: Chetty et al. (2020).

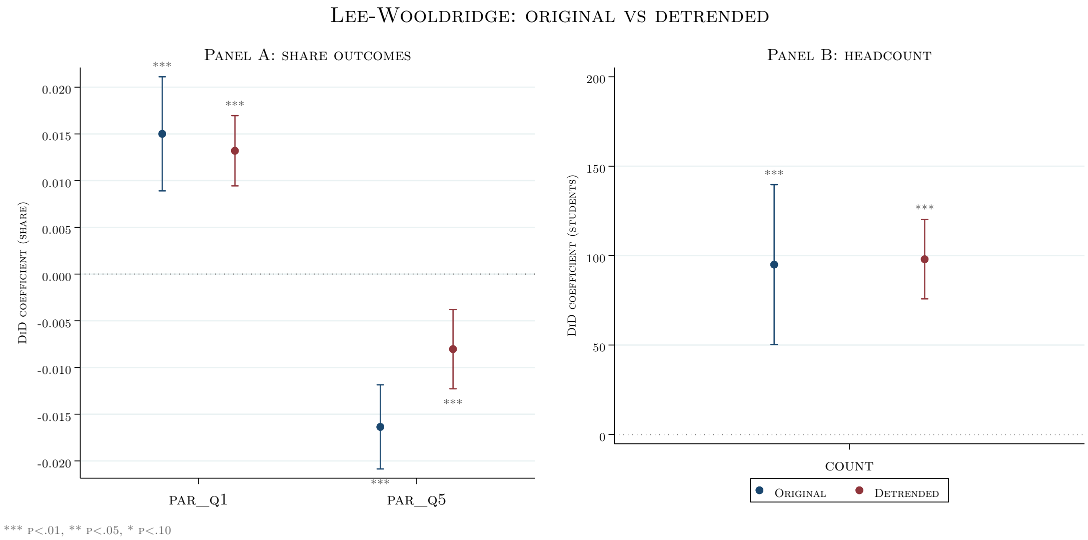

Figure A2: Lee–Wooldridge Detrending Comparison (Full Output)

Notes: Spec 8 coefficients on Q_1, Q_5, and total enrollment count under standard two-way fixed effects versus unit-specific detrending. Q_1 attenuates roughly 12% and remains significant at $p < 0.001$; enrollment count is essentially unchanged; Q_5 attenuates by roughly half but remains significant. Data: Chetty et al. (2020).

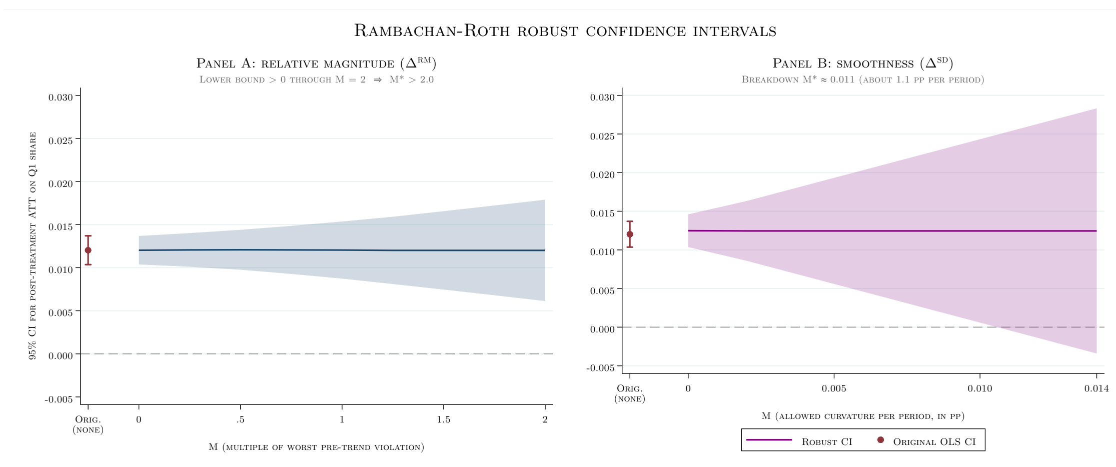

Figure A3: Rambachan-Roth Sensitivity at Q_1

Notes: Relative-magnitude sensitivity analysis from Rambachan and Roth (2023) applied to the Q_1 event-study coefficients under Spec 8. The breakdown value M^*—the M at which the confidence interval first includes zero—is greater than 2.0. The Q_1 result is robust to pre-trend violations of up to twice the worst observed pre-treatment coefficient. Data: Chetty et al. (2020).

Cross-state event studies. Event-study plots for each of the six treatment states (Tennessee, Massachusetts, South Carolina, Nevada, West Virginia, South Dakota) are available in the replication package. Each state's pre-trend and post-treatment dynamics are summarized in Section 10; the full event-study coefficient tables appear in Appendix B.

B Supporting Tables

This appendix reports the full specification menu evaluated in Section 5, event-study coefficient tables behind the main-text figures, and the pooled staggered DiD that pools across all six treatment states in Section 10.

Table A1: Specification Menu: Candidate Specifications, Q_1 ATT

Spec	Label	Panel	Drop 80–82	BEA control	Cluster	t^*	Q_1 ATT	Pre-trend F
1	Full, time-varying BEA	Full	No	Time-var	State	1987	+0.0171	3.14
2	Baseline BEA, full	Full	No	Baseline	State	1987	+0.0162	2.47
3	No BEA, full	Full	No	None	State	1987	+0.0185	3.71
4	Balanced, no BEA	Balanced	Yes	None	State	1987	+0.0172	2.03
5	Balanced, no drop	Balanced	No	Baseline	State	1987	+0.0166	2.89
6	Full, time-var, t^*=1986	Full	No	Time-var	State	1986	+0.0162	3.08
7	College cluster, full	Full	No	Baseline	College	1987	+0.0162	2.47
8	**Cleanest (chosen)**	**Balanced**	**Yes**	**Baseline**	**State**	**1987**	**+0.0150**	**1.23**
9	Cleanest, college cluster	Balanced	Yes	Baseline	College	1987	+0.0150	1.23
10	Cleanest, no BEA	Balanced	Yes	None	State	1987	+0.0145	1.41
11	Full + time-varying BEA	Full	No	Time-var	State	1987	+0.0142	3.14
12	Cleanest t^*=1986	Balanced	Yes	Baseline	State	1986	+0.0127	0.99

Notes: Candidate specifications vary along five axes: panel structure (full or balanced), cohort drop rule (include or drop 1980–1982), BEA control (none, pre-treatment baseline, or time-varying), clustering (state or college), and treatment timing ($t^* = 1986$ or 1987). Spec 8 (Cleanest) is the chosen anchor. All Q_1 ATTs in the menu fall in the range +0.013 to +0.019, excluding unreliable four-cluster bordering-state specifications. Pre-trend F-tests evaluate the joint zero of pre-treatment event-study coefficients; Spec 8 does not reject parallel trends ($F = 1.23$, $p = 0.32$). Data: Chetty et al. (2020).

Table A2: Event-Study Coefficients for Q_1 and Count (Spec 8)

Event time k	Q_1 coefficient	SE	Count coefficient	SE
-4	-0.0018	(0.0024)	-3.4	(16.7)
-3	-0.0007	(0.0022)	$+1.2$	(15.3)
-2	-0.0013	(0.0021)	$+4.9$	(14.8)
-1	·	·	·	·
0	$+0.0088^{**}$	(0.0036)	$+62.4^{***}$	(19.2)
$+1$	$+0.0131^{***}$	(0.0033)	$+83.7^{***}$	(20.8)
$+2$	$+0.0168^{***}$	(0.0035)	$+106.3^{***}$	(23.4)
$+3$	$+0.0189^{***}$	(0.0039)	$+118.5^{***}$	(25.9)
$+4$	$+0.0178^{***}$	(0.0041)	$+124.4^{***}$	(28.7)

Notes: Event-study coefficients from Spec 8 with $k = -1$ (cohort 1986) as the omitted reference period. State-clustered standard errors in parentheses. The Q_1 pre-trend is insignificant jointly ($F = 1.23$, $p = 0.32$). The count coefficient shows a sharp increase at $k = 0$ (first full HOPE cohort) and grows through $k = +4$, consistent with the expansion story in Section 6. Reported coefficients are the coefficient on the interaction of the Tennessee-public indicator and the event-time dummy. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$. Data: Chetty et al. (2020).

Table A3: Pooled Staggered DiD Across Six Treatment States

Outcome	Pooled ATT	SE
Q_1 share	$+0.0103^{*}$	(0.0052)
Enrollment count	+38.6	(62.4)

Notes: Pooled two-way fixed-effects DiD across the six treatment states (Tennessee, Massachusetts, West Virginia, South Dakota, Nevada, South Carolina) against the non-merit ex-Massachusetts control pool. The pooled Q_1 coefficient reflects the average of progressive and regressive state-specific effects (Section 10); design heterogeneity washes out in the pool. Cluster-robust standard errors at the state level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$. Data: Chetty et al. (2020).