The Returns to College(s): Evaluating College Value-Added*

Jack Mountjoy† Brent R. Hickman‡

March 26, 2026

Preliminary. Latest version available here.

Abstract

We empirically evaluate the prospects of credible value-added measures (VAMs) for individual U.S. colleges. Using linked administrative data spanning the state of Texas, we first estimate a variety of college-specific VAMs using prominent observational methods. We then use quasi-experimental enrollment variation from admission cutoffs to evaluate these VAMs, testing their ability to forecast causal cutoff-crossing effects on applicants' eventual educational attainment and earnings. The most prominent measures, which simply average the outcomes of students who attend each college, substantially overpredict causal impacts. For educational attainment outcomes, typical regression-adjusted VAMs with controls for test scores and demographics also overpredict, while more stringent VAMs incorporating application and admission controls predict well. These more stringent VAMs underpredict earnings gains, however, especially when comparing students choosing among the same admission offers. We find that students who decline a more selective admission offer tend to have relatively high earnings at their less-selective choice, consistent with self-selection on unobservable earnings potential. These results provide new evidence on the prospects of credible college value-added measures and on the nature of self-selection into college choices.

*This paper completely overhauls our earlier working paper, "The Returns to College(s): Relative Value-Added and Match Effects in Higher Education." We are grateful for feedback from many interlocutors and seminar participants, especially Josh Angrist, Zach Bleemer, Chris Campos, Raj Chetty, David Deming, John Friedman, Caroline Hoxby, Chinhui Juhn, Larry Katz, Thomas Lemieux, Chris Walters, and Seth Zimmerman. We also thank Holly Kosiewicz, Greg Branch, Trey Miller, Mark Lu, Greg Phelan, John Thompson, Janie Jury, and the memory of Rodney Andrews at the UT-Dallas Education Research Center. We are grateful for financial support from the Robert H. Topel Faculty Research Fund at Chicago Booth and the Industrial Relations Section at Princeton University. The conclusions of this research do not necessarily reflect the opinions or official positions of the Texas Education Research Center, the Texas Education Agency, the Texas Higher Education Coordinating Board, the Texas Workforce Commission, or the State of Texas. JEL codes: C52, I23, I24, I26, J24, J31. Keywords: returns to college, value-added measures, model evaluation.

†Corresponding author. University of Chicago and NBER. jack.mountjoy@chicagobooth.edu.

‡Washington University in St. Louis. hickmanbr@gmail.com.

1 Introduction

Enrolling in college tends to pay off for students on the margin of non-attendance (Zimmerman, 2014; Smith et al., 2025; Mountjoy, 2022, 2026). Student outcomes vary dramatically across different colleges, however, suggesting that the choice of *where* to enroll may be as consequential as *whether* to enroll at all. Relative to students who attend Alabama State, for example, those who attend MIT are nearly four times as likely to graduate (93% vs. 25%) and earn nearly four times as much after college ($104,700 vs. $27,700). Attributing these divergent outcomes to differing "value-added" from each college, however, is confounded by systematic student selection; the average MIT matriculant scored 1530 out of 1600 on the SAT, compared to 860 at Alabama State. Different colleges likely enroll students with systematically different unobservable traits as well, like ambition and work ethic, that influence a student's outcomes regardless of where she attends (Bodoh-Creed and Hickman, 2024).

Credible measures of individual colleges' value-added on student outcomes, purged of such selection, could help students make informed enrollment decisions, incentivize colleges to focus on impact over selection, and improve the targeting of state and federal policies like performance-based funding and gainful employment rules (Cunha and Miller, 2014). In pursuit of these goals, a wide variety of observational methods have recently been developed to generate college-level value-added measures (VAMs). The simplest and most prominent method, exemplified by the U.S. Department of Education's College Scorecard, simply reports the raw mean outcomes of students who attend each school. More sophisticated methods adjust for institutional differences in students' observable characteristics like test scores and family income (*The Economist*, 2015; The Brookings Institution, 2015; Chetty et al., 2020). The most stringent methods incorporate individual-level application or admissions data to narrow the comparisons to matched applicants with similar admission prospects (Hoxby, 2019) or realized admission sets (Mountjoy and Hickman, 2021). While most of these methods explicitly aim to generate credible VAMs, they produce discordant results that are difficult to evaluate without some kind of causal benchmark.

At the same time, a separate literature increasingly employs quasi-experimental research designs to identify causal impacts of shifting students across U.S. colleges, like the admission RDs in Hoekstra (2009), Zimmerman (2014), Smith et al. (2025), and Mountjoy (2026), and the affirmative action regime changes in Hinrichs (2014),

Bleemer (2022), and Black et al. (2023). These quasi-experimental approaches, however, do not directly identify a college-by-college schedule of VAMs relative to a common baseline. Instead, they typically yield a single treatment effect of enrolling in a target school (or set of target schools) relative to a composite counterfactual of the affected students' many next-best alternatives. Even when exogenous instruments like admission cutoffs are available at multiple target schools, as in Mountjoy (2026), each target-specific instrument still draws students from a composite counterfactual that varies across targets. The number of available instruments, moreover, is typically far fewer than the number of college options (as many are non-selective), preventing a multivariate IV approach to identifying college-specific VAMs.

In this paper, we empirically evaluate the prospects of credible college value-added measures by generating observational VAMs from several prominent methods, then subjecting those VAMs to quasi-experimental validation, all in a unified setting. We adapt the K-12 school VAM testing framework of Angrist et al. (2017) to higher education, beginning with the idea that a desirable property of a set of college VAMs would be their ability to accurately predict the effects of exogenously shifting students across schools on those students' outcomes. College admissions practices that involve cutoffs in SAT/ACT scores can provide exactly such exogenous shifts when comparing applicants just above versus just below a given cutoff. We therefore test whether the actual effects of crossing an admission cutoff on students' long-term outcomes are accurately predicted by the corresponding change in the observational VAM of those students' initial college (including no college). Observational methods that produce highly predictive VAMs gain some credibility for practical use, while we show how prediction failures can reveal interesting patterns of economic self-selection of students into colleges.

The paper proceeds as follows. In Section 2, we introduce our setting and data. We use linked administrative records spanning the state of Texas, tracking all public high school students from their pre-college characteristics into their college applications, admissions, enrollments, degrees, and earnings. The breadth of this data allows us to study the measurement of value-added across a vast diversity of postsecondary institutions, from local community colleges to regional public universities to highly selective private institutions, with the statistical precision afforded by observing the entire population of students who attend each school. The richness of this data, moreover, allows us to implement all of the most prominent VAM estimation methods

in one unified setting, including those requiring individual-level data on applications and admissions, along with the admission RD design used to evaluate those VAMs.

In Section 3, we describe and implement the various VAM estimation methods. We focus on four prominent and distinct approaches. The first simply compares the raw mean outcomes of students who attend each college, as is common in college guides, government statistics, popular media, and increasingly in state postsecondary funding formulas. The second approach regression-adjusts these mean outcomes with "typical controls" for observable confounders, including high school academic preparation and demographics, reminiscent of the workhorse K-12 teacher value-added model and exemplified in higher education by the baseline specification of Chetty et al. (2020). The third approach follows Hoxby (2015, 2019) by narrowing the comparisons to matched applicants who apply to the same schools, have identical SAT/ACT scores, and are either inferred to be "on-the-bubble" for admission to a given school, delivering "vertical" comparisons, or are facing a choice between two similarly selective schools, delivering "horizontal" comparisons. The fourth approach implements the widespread but controversial "matched admit" strategy inspired by Dale and Krueger (2002), which compares the outcomes of admitted students choosing among the same set of admission offers. Our earlier working paper (Mountjoy and Hickman, 2021) focused on generating VAMs for Texas public universities using this matched admit method, with a large number of papers using related types of admission controls.[1]

These four methods generate substantially different schedules of college VAMs. Raw mean outcomes differ dramatically across colleges, but much of that variation is explained by selection on observable student characteristics. We show this both by the large shift in the VAM schedule when moving from the raw means to the "typical controls" VAMs, and by a balance test that reveals substantial differences across colleges in their students' predicted outcomes based on their pre-college covariates. The typical controls VAMs, of course, mechanically balance student covariates across colleges; more interestingly, both the the Hoxby (2015)-style matched applicant method and the Dale and Krueger (2002)-style matched admit method achieve rough balance in covariate-predicted earnings across most of the treatment college distribution.[2] These

[1]These include college-specific VAM estimation in Cunha and Miller (2014), Arcidiacono et al. (2016), Ciocca Eller (2023), Bleemer (2021, 2022), Bleemer and Quincy (2025), Edwards et al. (2025), and Chetty et al. (2026), along with more aggregated treatment comparisons in Arcidiacono (2005), Long (2008), Fryer and Greenstone (2010), Broecke (2012), Smith (2013), and Ge et al. (2022).

[2]We add covariate controls to these specifications when evaluating their VAMs, but their rough

methods still generate divergent VAM schedules for actual outcomes, however. The matched admit VAMs are tightly distributed with relatively small differences across colleges, as shown in Mountjoy and Hickman (2021) for Texas public universities and now expanded to include private institutions and community and technical colleges.[3] The Hoxby (2015) matched applicant VAMs, meanwhile, tend to lie between those of the "typical controls" method and the matched admit method.

In Section 4, we empirically evaluate these VAMs by adapting the testing framework of Angrist et al. (2017), originally developed for K-12 schools using centralized assignment algorithms and randomized admission lotteries, to postsecondary institutions with admission cutoffs. Specifically, we draw on the admission discontinuities in SAT/ACT scores identified in Mountjoy (2026), which generate exogenous enrollment variation among marginal applicants to twenty-eight Texas public universities. The VAM testing procedure has a natural visual representation: for each target university with an admission cutoff, we plot the reduced-form effect of crossing that cutoff on applicants' actual outcomes against the first-stage-esque effect of crossing the same cutoff on the VAM of the college option they are induced into (including no college). If the VAMs are accurate, all of these points (one for each target college with an admission cutoff) should lie along the 45-degree line up to sampling error, which we test with an omnibus test statistic. That omnibus test further decomposes into a forecast test of whether the average relationship has a slope of one, plus an overidentification test of whether all of the points lie on the forecast line up to sampling error.

We find four main patterns of test results. First, raw mean outcome VAMs overpredict causal impacts across all outcomes by roughly a factor of two, consistent with substantial selection bias in popular raw outcome comparisons. Second, for educational attainment outcomes, the "typical controls" VAMs also overpredict causal impacts, improving only modestly over the raw means. Third, the more stringent matched applicant and matched admit VAMs are highly accurate for predicting impacts on educational attainment, with forecast coefficients statistically indistinguishable from one. Fourth, these more stringent VAM methods underpredict earnings gains, especially the Dale and Krueger (2002)-inspired matched admit method comparing students with identical admission offers.

balance makes the evaluation results largely insensitive to the inclusion or exclusion of such controls.

[3]Chetty et al. (2026) replicate the public university results of Mountjoy and Hickman (2021) using THECB admissions data linked to nationwide IRS earnings records.

To understand this divergence across methods and outcomes, in Section 5 we theoretically and empirically characterize the types of students involved in the observational and quasi-experimental comparisons. We find that students in the matched admit comparison who decline a more selective admission offer tend to have relatively high earnings at their less-selective choice, both relative to rejected admission cutoff compliers and relative to what their own observable characteristics would predict, consistent with self-selection on unobservable earnings potential. Together, these results provide new evidence on the prospects of credible college value-added measures and on the nature of self-selection into college choices.

2 Setting and Data

The data for this analysis come from linked administrative registries that span the entire state of Texas, maintained by the University of Texas at Dallas Education Research Center (UTD-ERC, 2025). Texas is the second largest U.S. state by land area and population, with 32 million residents and an economy that, taken alone, would rank as the eighth largest in the world ($2.8 trillion GDP). The state's postsecondary ecosystem enrolls 1.7 million college students, about 9 percent of the U.S.total, across a diverse array of flagship and regional public universities, private institutions, and local community and technical colleges.

Our analysis sample begins with the universe of students who graduate from a Texas public high school between 2004 and 2014.[4] The 2004 cohort is the first to have SAT and ACT scores recorded in the college admission records, and we stop at the 2014 cohort to observe a balanced panel of ten years of outcomes after high school for all sample members. We link several registries provided by the Texas Education Agency (TEA) to measure these students' pre-college characteristics, including demographics, standardized test scores, high school coursework, attendance, disciplinary infractions, and high school campus.

We link these high school graduates to their application and admission records from all 35 Texas public universities, provided by the Texas Higher Education Coordinating Board (THECB). These application-level records include the admission decision and the SAT/ACT score used in that decision, which together enable the re-

[4]Private high school students are not observed in this data; they account for only 5 percent of all Texas high school graduates (National Center for Education Statistics, 2018).

gression discontinuity research design described in Section 4. We also follow each high school graduate longitudinally through additional THECB administrative registries of college enrollment spells, credit accumulation, and degree completion spanning all public and private non-profit postsecondary institutions in the state. Importantly, these data allow us to observe educational outcomes regardless of dropout or transfer across institutions. For the 2008-2014 high school graduating cohorts, National Student Clearinghouse records are also available, allowing us to observe initial college enrollments nationwide and thus distinguish enrollment outside of Texas from true non-enrollment. Finally, to study earnings outcomes, we merge in quarterly earnings records from the Texas Workforce Commission (TWC) that cover all Texas employees subject to the state unemployment insurance system.[5]

3 Value-Added Measures

All of the prominent observational methods for generating college-specific value-added measures (VAMs) can be described as variations on the regression equation

$$Y_i = \alpha + \sum_{j=1}^{J} \beta_j D_{ij} + X_i'\theta + \varepsilon_i, \tag{1}$$

where Y_i is the outcome of individual student i (e.g. degree completion, earnings), D_{ij} is a binary indicator for whether student i initially enrolls in college j after high school, and X_i is a vector of covariate controls. The β_j coefficients are the VAMs of interest, measuring the difference in mean outcomes between students who enroll in college j versus an omitted reference college $j = 0$, adjusted for observable differences in X_i. In all of the results that follow, we set UT-Austin as the omitted reference college, and we include not attending any college as one of the college options. For students in the 2008-2014 high school graduating cohorts, we also use their National Student Clearinghouse data to construct a single out-of-state college option for the small share of Texas high school graduates who enroll outside the state.

[5]Stevens (2007) estimates that 90% of the civilian labor force is captured in state UI records; excluded are the self-employed, independent contractors, some federal employees including military personnel, and workers in the informal sector.

3.1 Raw Means

The most prominent outcome-based comparisons of colleges—pervasive in popular media, college guides, government statistics, and in many states' postsecondary funding formulas—simply compare the raw mean outcomes of students who attend each school. We implement this method by estimating Equation (1) without any control variables in X_i except cohort fixed effects.

3.2 Typical Controls

The second most prominent method adds covariates like test scores and demographics to Equation (1) via the control set X_i. This "typical controls" specification is ubiquitous in the literature on the return to college quality, and reminiscent of the workhorse value-added model measuring teacher effectiveness in K-12 classrooms. Chetty et al. (2020)'s nationwide estimation of college-level VAMs uses a similar baseline specification, and aggregate versions of this specification using publicly available institution-level statistics also appear in several college ranking methodologies like those of *The Economist* (2015) and The Brookings Institution (2015).

We implement a relatively rich version of the "typical controls" method by estimating Equation (1) with the following student-level pre-college covariates in the control set X_i: categorical indicators for gender, race, cohort, free/reduced price lunch eligibility, gifted status, and an indicator for being at-risk of dropping out of high school, plus cubic polynomials in 10th grade math and English standardized test scores, number of advanced high school courses taken, percent of school days in attendance, and number of suspension days for disciplinary infractions.

3.3 Matched Applicants

Hoxby (2015, 2019) develops a VAM estimation method that incorporates data on students' college application behavior to isolate "vertical" and "horizontal" comparisons of matched applicants. Specifically, vertical comparisons involve applicants to the same college with the same SAT score who are inferred to be "on-the-bubble" of admission into that college, defined as having an estimated admission probability between 40 and 60 percent as a univariate function of applicant SAT scores. The motivating idea behind this approach is that "within the bubble range, students are

fairly likely to be either admitted or rejected and the decision may turn on a minor factor. Essentially, we make use of unintentional randomization conducted by admissions staff" (Hoxby 2015, p. 16). Meanwhile, "horizontal" comparisons involve applicants to the same pair of colleges with the same SAT scores, with the pair of colleges required to be of similar selectivity (defined as their respective 25th and 75th percentiles of enrolled students' scores being within 3 percentiles of each other), and with the applicants required to be likely admitted, defined as scoring between the 65th and 80th percentiles of the school's enrolled students. In both cases, the method compares the outcomes of these matched applicants across the different colleges they end up enrolling in, summarizing the outcome differences with fixed effect parameters for each treatment college.

Our implementation of this matched applicant approach stays close to the method described in Hoxby (2015), with three differences driven by the data and setting. First, Hoxby (2015) uses nationwide data on SAT test takers and the colleges to which they send their scores, merged to National Student Clearinghouse records of college enrollment. This allows for broad measurement of students' application behavior, but with score sends acting as a proxy for actual applications. We observe Texas high school graduates' actual applications, but only to Texas public universities. Second, Hoxby (2015) does not observe admissions decisions, instead proxying for the "on-the-bubble" admission range at each target school by estimating enrollment at that school as a function of test scores and finding the smoothed values with enrollment probabilities between 40 and 60 percent of the maximum smoothed enrollment probability. We use our data on actual admissions to each Texas public university to directly estimate these bubble ranges. Finally, Hoxby (2015) excludes individuals who do not attend any college, focusing on identifying relative value-added across colleges. We must include the non-enrollment option in our VAM evaluation, given that a small but nonzero share of our admission cutoff compliers in Section 4 end up in that non-enrollment option when falling just below the cutoff. Their small share, however, ensures that a range of imputations of the VAM for non-enrollment would have little impact on our VAM evaluation results.

Hoxby (2015) estimates each college's VAM in a two-step paired comparisons procedure, first taking all pairwise differences in mean outcomes between matched applicants who enroll in different colleges, then regressing those pairwise differences on college indicators to yield college-level VAMs, weighting by the number of students

involved in each pairwise difference. We streamline this into a one-step student-level regression in the form of Equation (1) by limiting the estimation sample to vertical and horizontal matched applicants and including fixed effects for each matched applicant cell in the control set X_i.[6]

3.4 Matched Admits

The fourth method, pioneered by Dale and Krueger (2002) to study linear returns to college selectivity, incorporates student-level data on realized admissions to compare the outcomes of applicants choosing among the same set of admission offers. Our earlier working paper (Mountjoy and Hickman, 2021) focused on this "matched admit" method as a way of generating VAMs for each Texas public university, with related VAM implementations appearing in Cunha and Miller (2014), Arcidiacono et al. (2016), Ciocca Eller (2023), Bleemer (2021, 2022), Bleemer and Quincy (2025), Edwards et al. (2025), and Chetty et al. (2026), along with many papers using a related control strategy to estimate impacts of more aggregated treatments than individual colleges, including Arcidiacono (2005), Long (2008), Fryer and Greenstone (2010), Broecke (2012), Smith (2013), and Ge et al. (2022).

The motivating idea behind this matched admit approach is that a student's decisions about which colleges to apply to, and those colleges' decisions about whether to admit or reject, reveal important information about that student's abilities, ambitions, and other unobserved advantages that may not be sufficiently captured by typical control variables like test scores and demographics. Despite its proliferation, this approach remains controversial, with Hoxby (2009) arguing that it relies on "the small share of students who make what is a very odd choice. These are students who know that they could choose a much more-selective college and who have already expressed interest in a much more-selective college (they applied), yet, they choose differently than 9 out of 10 students. Almost certainly, these odd students are characterized by omitted variables that affect both their college decision and their later life outcomes" (p. 115).

To produce and evaluate college-level VAMs from the matched admit approach, we build on the main specification from Mountjoy and Hickman (2021), now expanding the relevant treatment options beyond public universities to include all public and

[6]We have verified that the one-step student-level regression yields results similar to the two-step procedure.

private non-profit postsecondary institutions in Texas, along with the no-college option and the composite out-of-state option. This expansion is necessary for our VAM evaluation using the admission discontinuities described in Section 4, as many barely rejected applicants fall out of the public four-year sector into community and technical colleges, with smaller shares going to private institutions, out-of-state institutions, and non-enrollment. We estimate matched admit VAMs for each of these treatment options using a student-level regression in the form of Equation (1) where the estimation sample is limited to applicants who are admitted to at least two Texas public universities, and the control set X_i includes fixed effects for each unique portfolio of public university admissions observed in that sample.

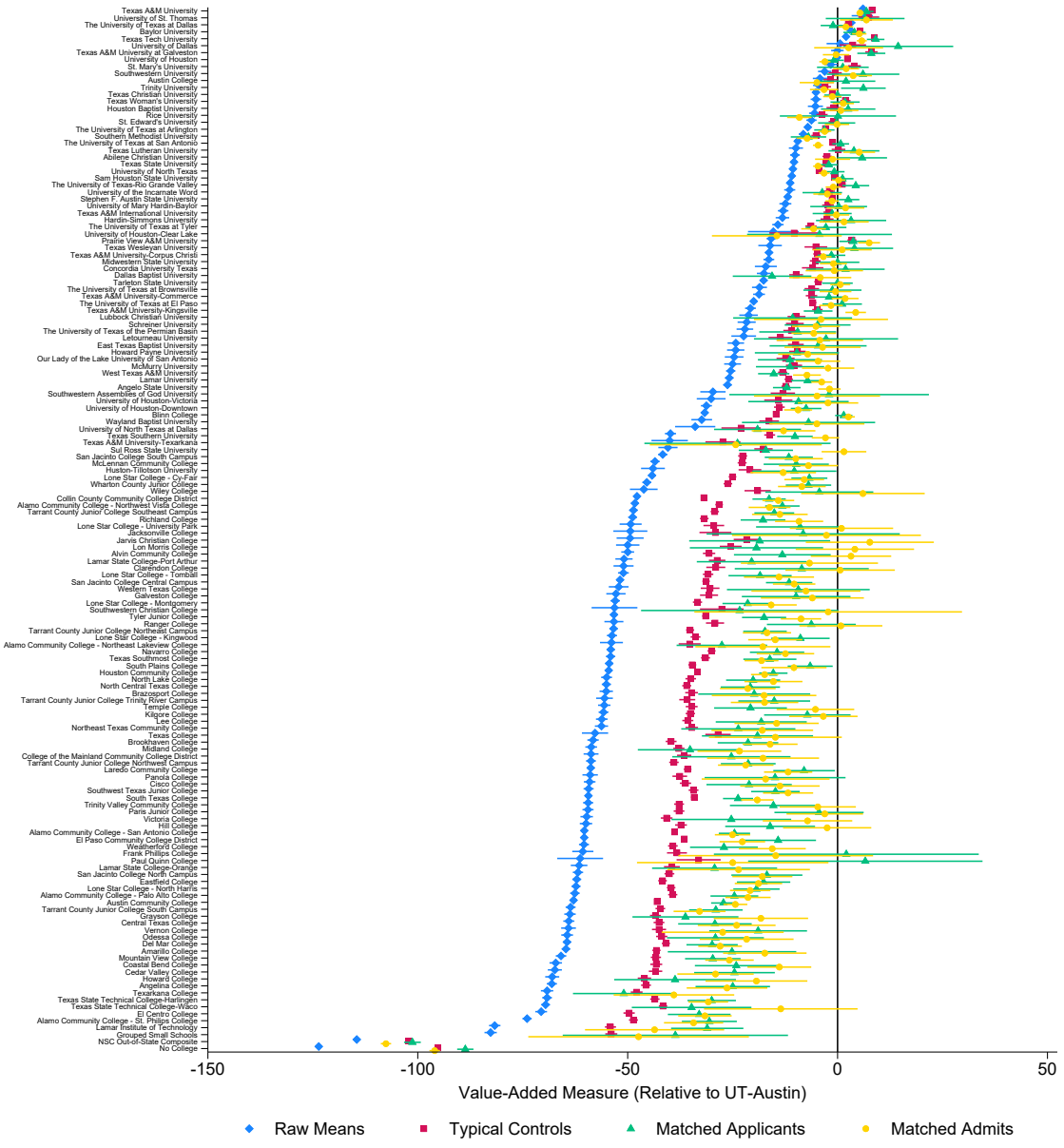
3.5 VAM Estimates

Figures 1-4 present the estimated VAM schedules from each method for four different outcomes: cumulative college credits completed (Figure 1), completion of any postsecondary credential (Figure 2), completion of a bachelor's degree (Figure 3), and earnings (Figure 4). The education outcomes are measured at year ten since high school graduation, and annualized positive quarterly earnings are averaged over years 8-12. Colleges (including "no college") are ranked along the vertical axis by their raw outcome means, and each school's VAM estimate (relative to UT-Austin) is plotted along the horizontal axis, with different markers for each method.

The four methods generate substantially different schedules of college VAMs. The raw mean outcomes, visualized on the far left of each plot, differ dramatically across colleges. Much of that variation is explained by selection on observable student characteristics, however, as shown when moving from the raw means to the "typical controls" VAMs. Another way to gauge the amount of selection on observables in the raw means is the balance test in Figure 5. The solid diamonds in that figure plot the actual raw mean outcomes at each school, while the hollow diamonds replace each student's actual outcome with their *predicted* outcome based on their pre-college covariates.[7] The outcomes that each college's incoming students are predicted to have, based entirely on their pre-college characteristics, follow a remarkably similar pattern

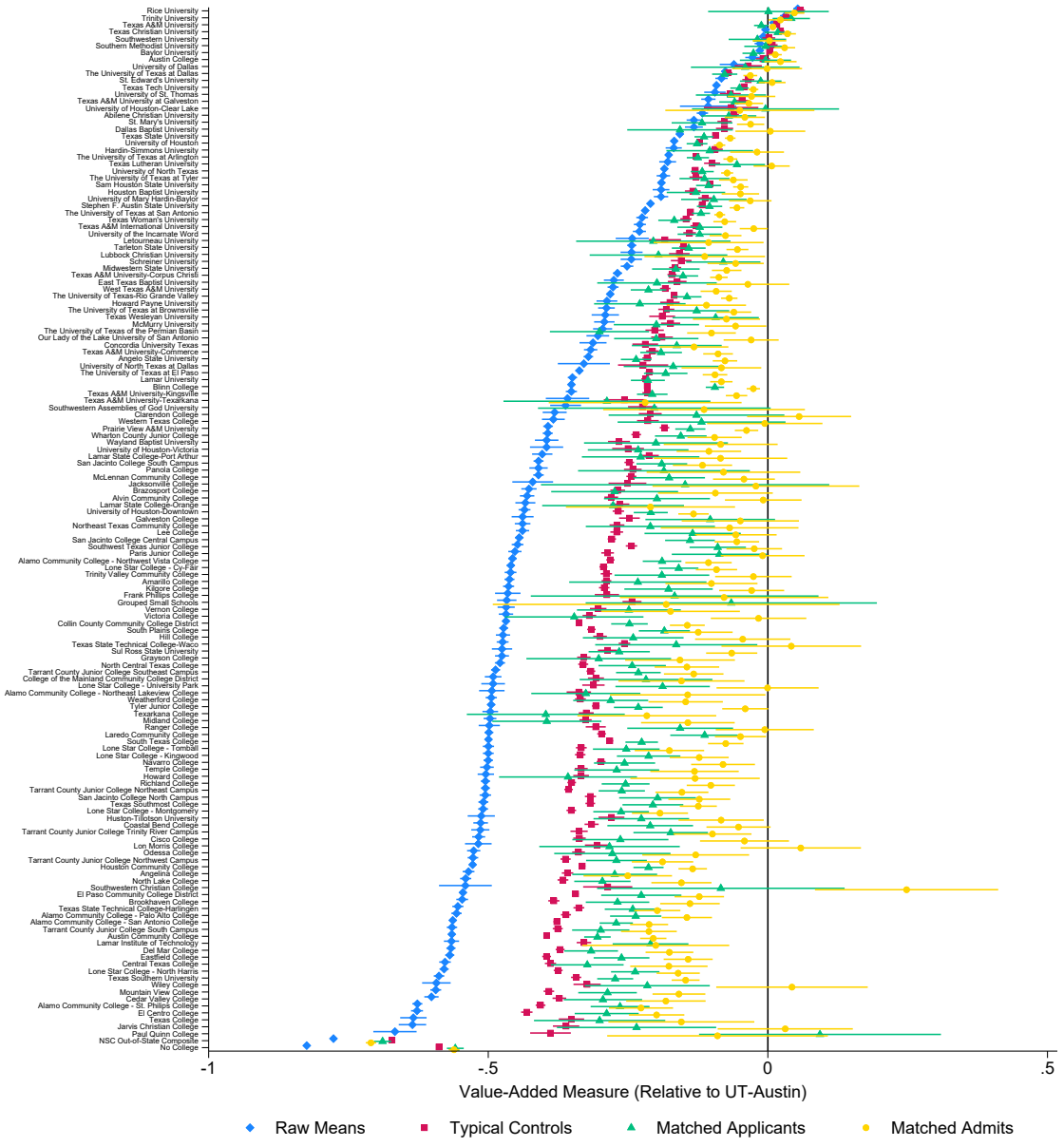
[7]Specifically, we use the predicted values from an OLS regression of earnings on cubic polynomials in 10th grade math and reading scores, attendance, days of disciplinary suspension, and number of advanced high school courses, along with categorical indicators for free/reduced price lunch eligibility, gender, at-risk status, race, and high school cohort.

Figure 1: Value-Added Measures: Cumulative Credits



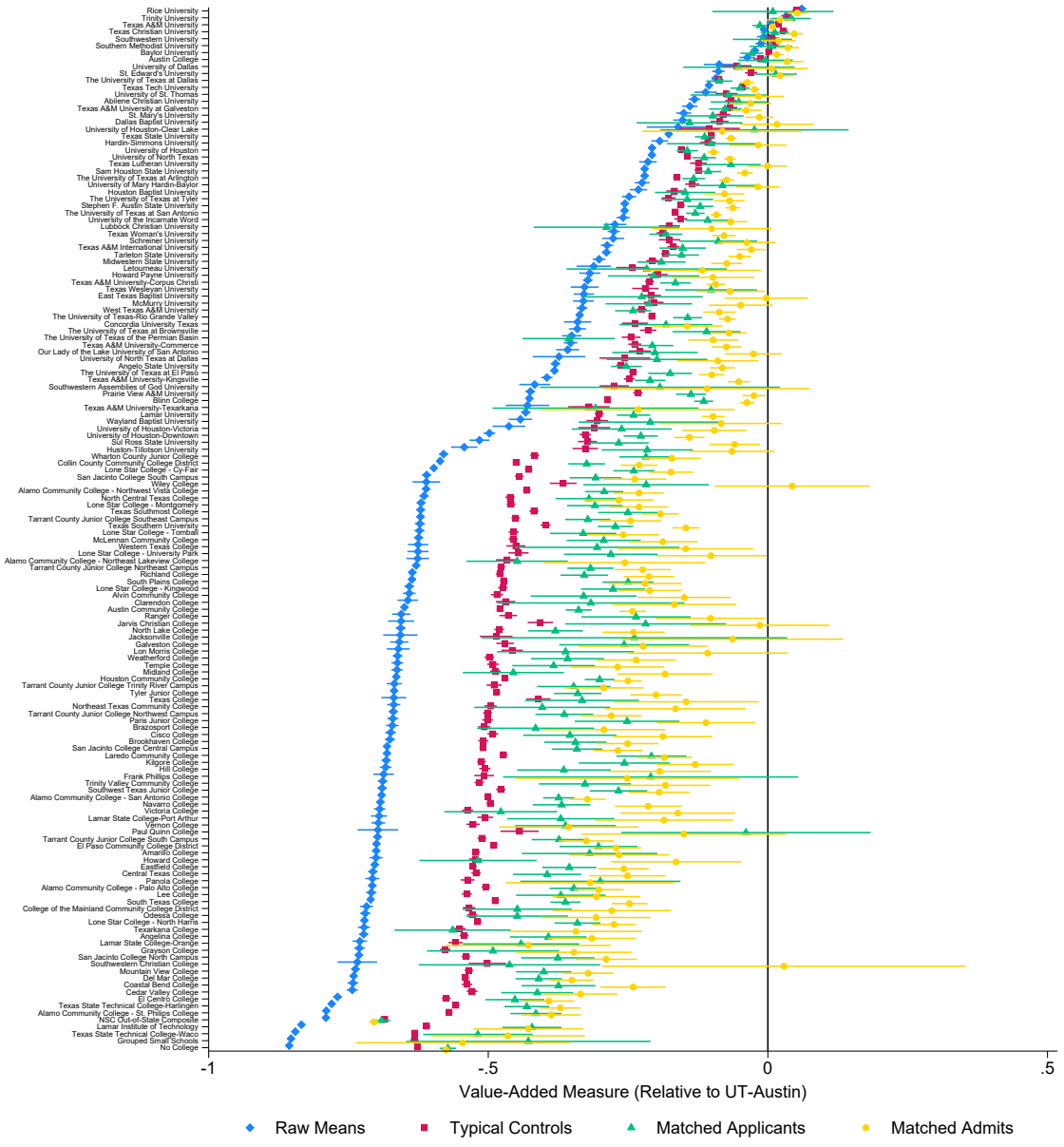
Notes: Each point and 95% confidence interval represents one college's estimated value-added relative to the omitted reference school (UT-Austin, represented by the vertical line at zero). Different markers denote the different VAM estimation methods described in the text.

Figure 2: Value-Added Measures: Any Degree Completion



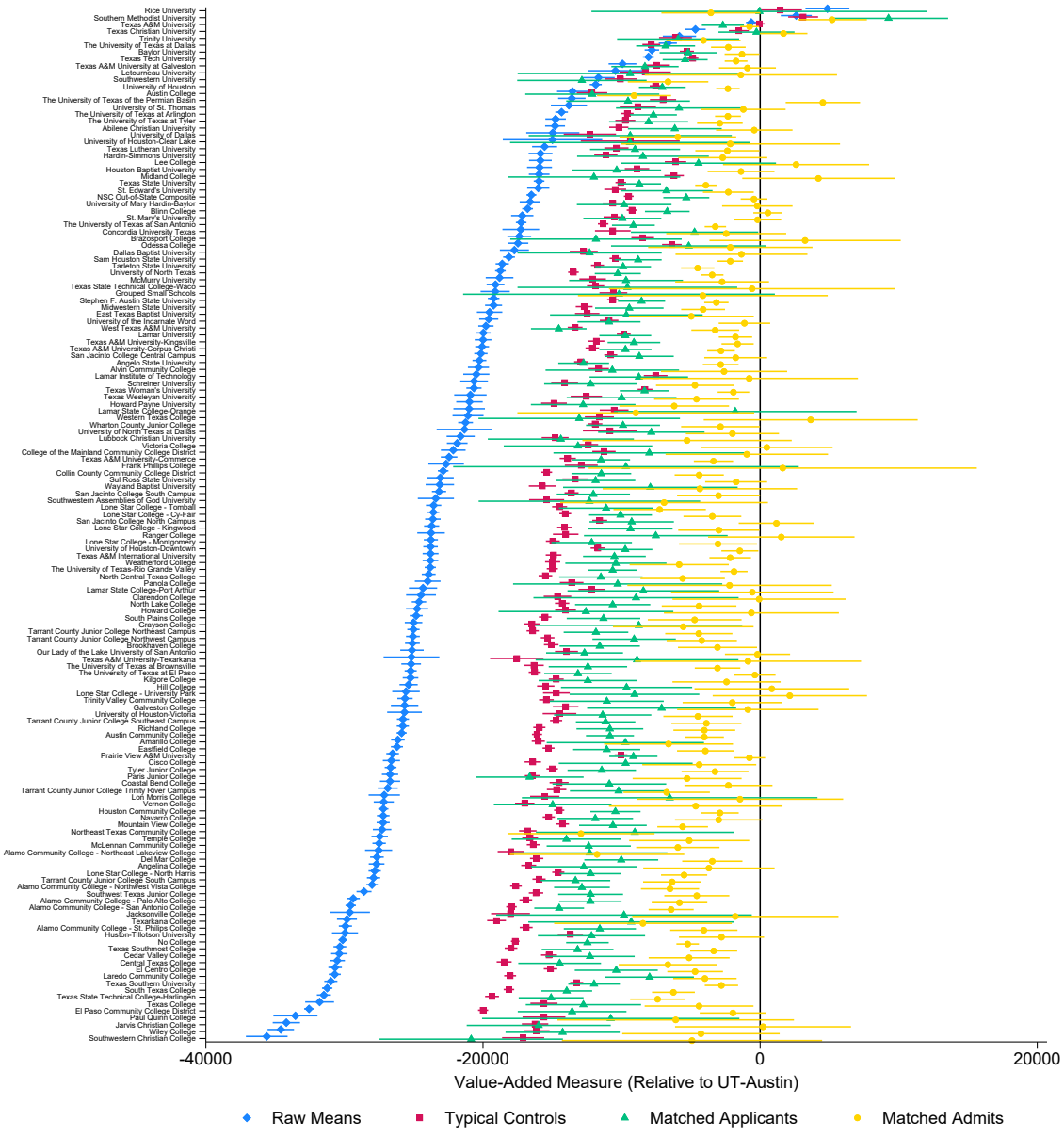
Notes: Each point and 95% confidence interval represents one college's estimated value-added relative to the omitted reference school (UT-Austin, represented by the vertical line at zero). Different markers denote the different VAM estimation methods described in the text.

Figure 3: Value-Added Measures: Bachelor's Degree Completion



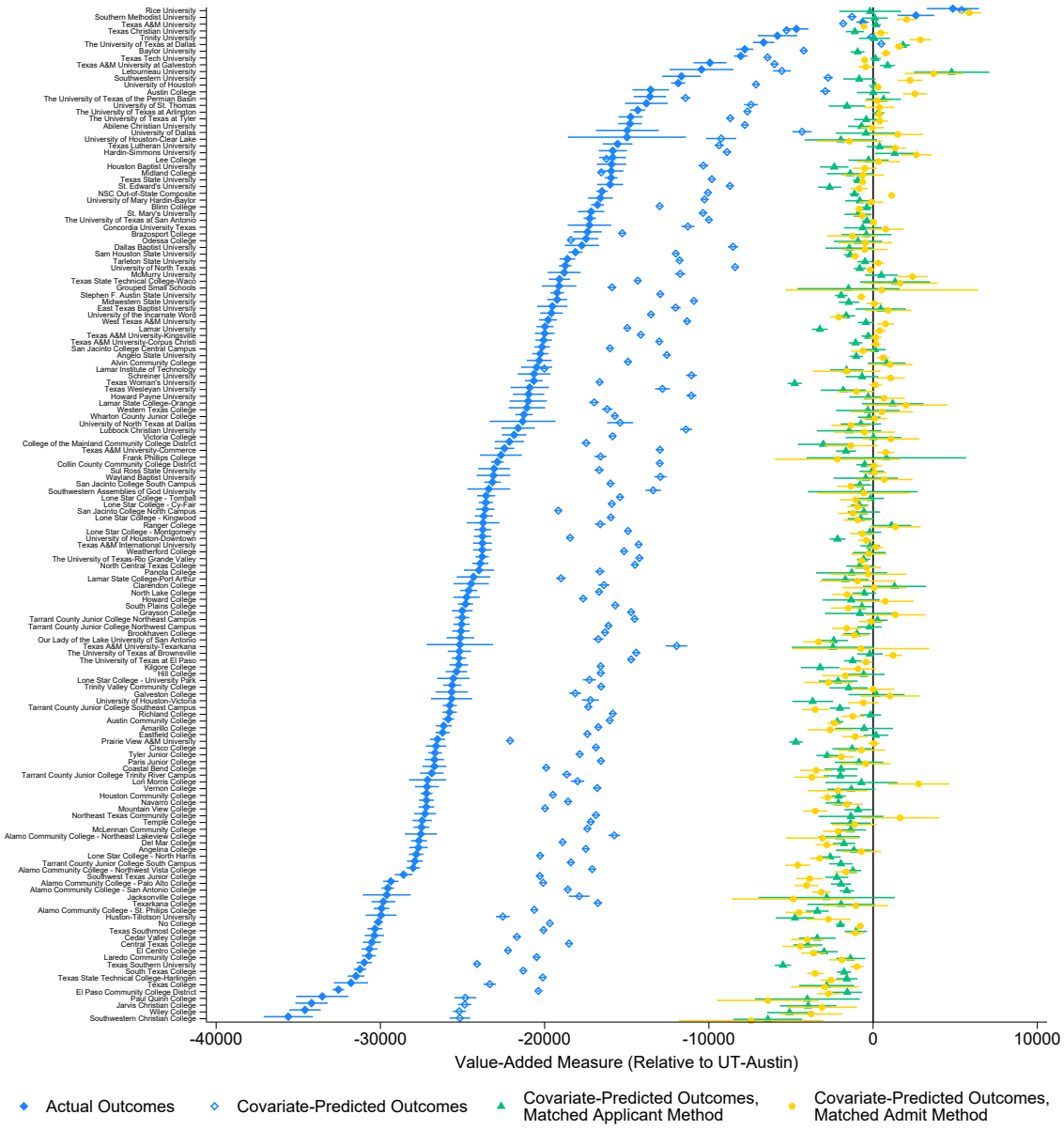
Notes: Each point and 95% confidence interval represents one college's estimated value-added relative to the omitted reference school (UT-Austin, represented by the vertical line at zero). Different markers denote the different VAM estimation methods described in the text.

Figure 4: Value-Added Measures: Earnings



Notes: Each point and 95% confidence interval represents one college's estimated value-added relative to the omitted reference school (UT-Austin, represented by the vertical line at zero). Different markers denote the different VAM estimation methods described in the text.

Figure 5: Value-Added Measures: Covariate-Predicted Earnings (Balance Test)



Notes: Each point and 95% confidence interval represents one college's estimated value-added relative to the omitted reference school (UT-Austin, represented by the vertical line at zero). Different markers denote the different VAM estimation methods described in the text.

as the actual outcomes. This clearly illustrates how different colleges enroll systematically different types of students, with such selection playing a dominant role in the raw outcome differences across colleges.

The typical controls VAMs, of course, mechanically balance students' covariates across colleges; more interestingly, Figure 5 shows that both the the Hoxby (2015)-style matched applicant method and the Dale and Krueger (2002)-style matched admit method achieve rough balance in covariate-predicted earnings across most of the treatment college distribution. We add covariate controls to these two specifications when estimating and evaluating their VAMs, but the rough covariate balance across most treatment colleges depicted in Figure 5 makes those controls largely irrelevant for our subsequent results.

These methods still generate divergent VAM schedules for actual outcomes, however. The Dale and Krueger (2002)-style matched admit VAMs tend to be more tightly distributed, especially for earnings. The Hoxby (2015)-style matched applicant VAMs, meanwhile, tend to lie between those of the "typical controls" method and the matched admit method, hewing closer to the typical controls VAMs for four-year institutions (which roughly comprise the the top half of the figures) but with more distinction for sub-baccaluareate institutions (which roughly comprise the bottom half of the figures).

4 Evaluating Value-Added

The previous section demonstrated that prominent observational methods for estimating college value-added measures produce substantially different VAM schedules. This section brings in quasi-experimental variation to evaluate these VAMs, assessing their ability to accurately predict the effects of exogenously shifting applicants across colleges via admission discontinuities.

4.1 Admission Discontinuities

We draw on the collection of admission discontinuities identified and deployed in Mountjoy (2026), which describes them in detail. In short, twenty-eight Texas public universities exhibit discontinuities in admission and enrollment probabilities as a function of applicants' SAT/ACT scores. Figure 6 illustrates these discontinuities,

pooled over applicants to the twenty-eight "target" universities with cutoffs. The probability of admission to a given target university jumps by an average of 27 percentage points at the cutoff, and the probability of enrollment jumps by an average of 15 percentage points.

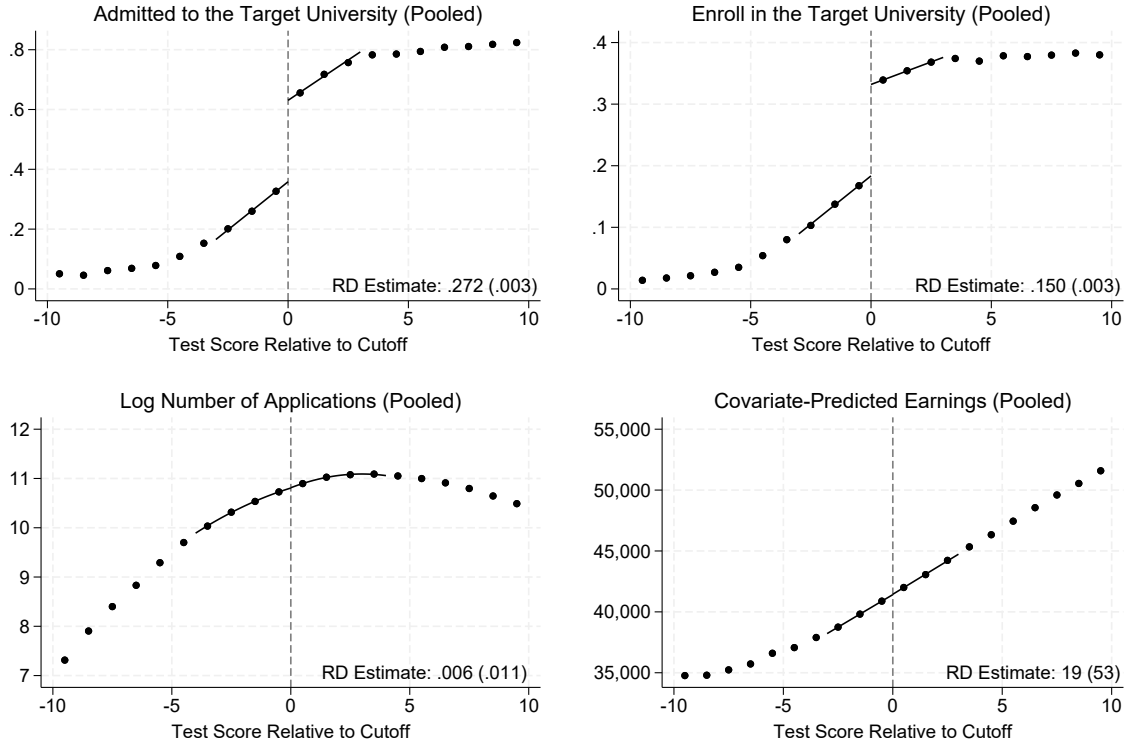
The validity of the regression discontinuity (RD) design as a source of exogenous variation in college enrollments rests on the assumption that students cannot precisely manipulate their test score values or their decision to apply to the target school immediately around these cutoffs. We reproduce two pieces of supporting evidence from Mountjoy (2026) in the bottom panels of Figure 6. First, the log number of applications is smooth through the cutoffs, showing no evidence of bunching. Second, applicants' pre-college characteristics are smooth through the cutoffs, confirming that students just above and just below are observationally indistinguishable.

Figure 7 describes the 28 target schools with cutoffs. The vertical axis measures the average location of each school's admission cutoff in concorded SAT points, and the horizontal axis measures each school's selectivity as the average SAT score among enrolled students. The comparison to the dashed 45 degree line shows that the marginal applicants around each school's cutoff tend have scores around 150 SAT points lower than their average potential peer at the target school. The size of the circles reflects the relative mass of cutoff compliers that each target school contributes; this distribution is diverse and diffuse, with the largest target schools (University of North Texas and University of Houston) each contributing only 11 percent of the total mass of cutoff compliers.

Figure 8 visualizes where these target schools lie within the entire spectrum of Texas postsecondary institutions, ranked according to their raw mean earnings VAMs. The figure also visualizes the weighted distribution of marginal applicants' fallback options, i.e. the schools that applicants attend instead of the target school (including no college) when they fall just below the target school's admission cutoff. Together, the distributions of target schools and fallback options illustrate the breadth of the college treatments spanned by the quasi-experimental variation, with that breadth ensuring that our evaluation exercise is not focused on only limited portions of the college VAM schedules.

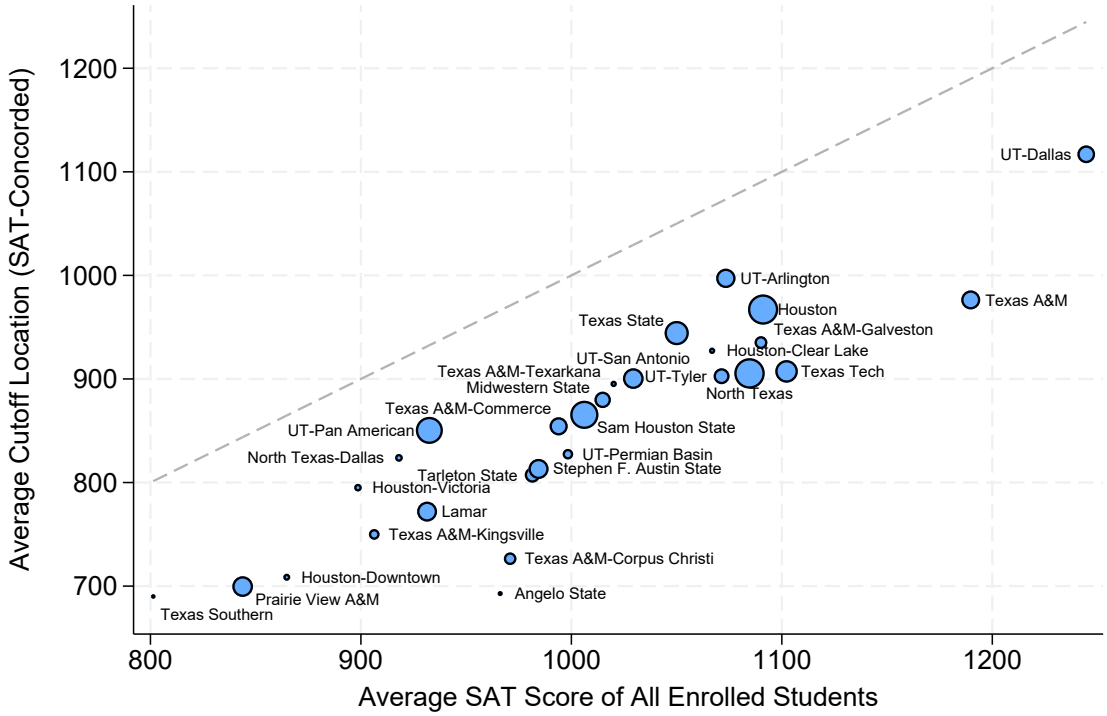
Going forward, we group the nine smallest target schools into a single target for the target-specific RD estimates, yielding 19 binary cutoff-crossing instruments. Our implementation of the RD design follows the main specification in Mountjoy (2026),

Figure 6: Admission RDs from Mountjoy (2026)



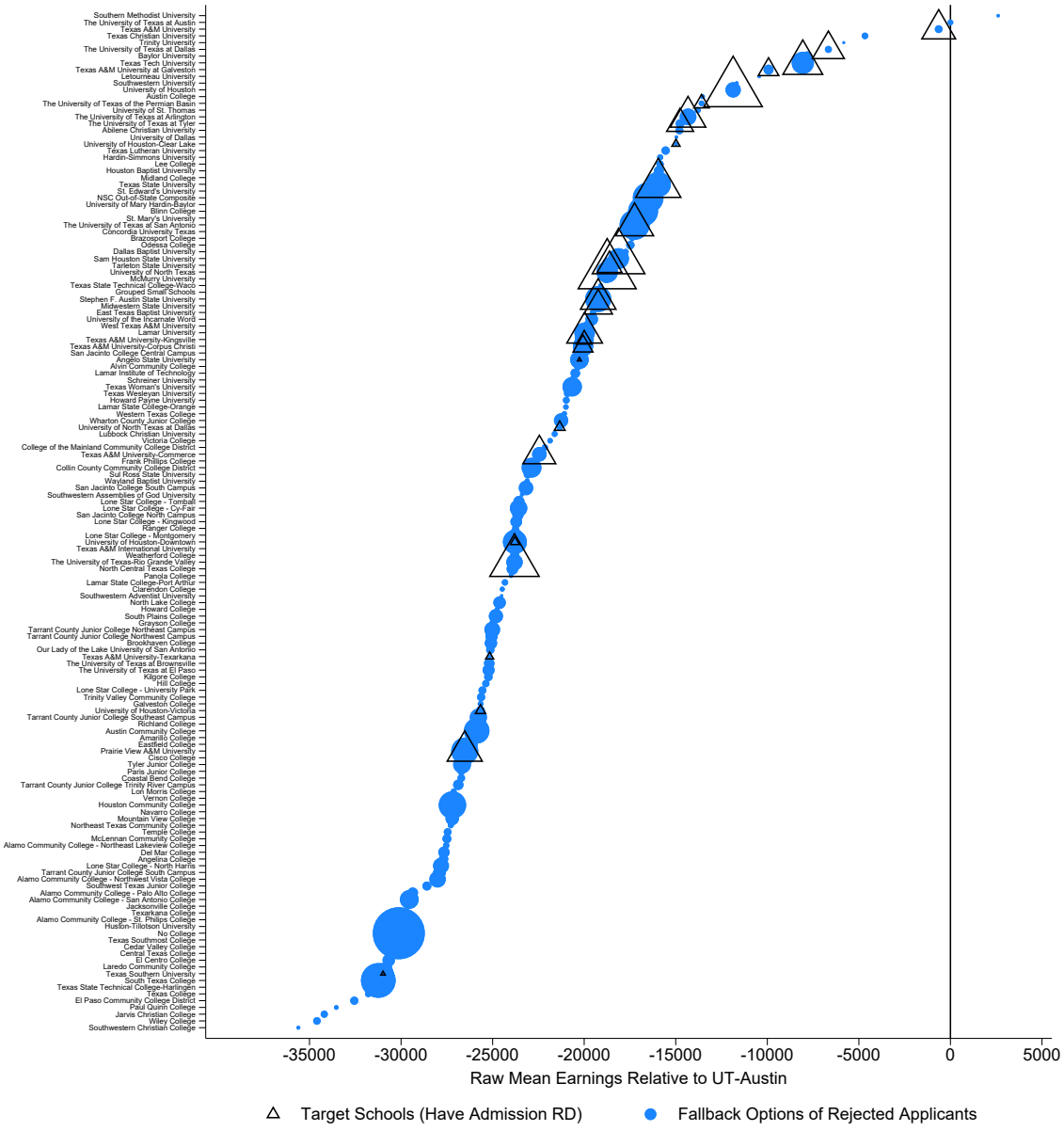
Notes: This figure plots the pooled admission discontinuities from Mountjoy (2026); see that paper for additional details. The top left panel plots the share of applicants admitted to the target university by test score relative to the cutoff, pooling across all 28 target universities. The top right panel plots the share initially enrolling in the target university. The bottom left panel conducts a McCrary (2008) test of manipulation by taking the log number of applications at each value of the concorded running variable and estimating the discontinuity at the cutoff with separate local quadratic functions on each side of the cutoff and a bandwidth of four concorded ACT points. The bottom right panel plots covariate-predicted earnings averaged over 8-12 years after application, predicted via linear regression using the following covariates: categorical indicators for gender, race, free or reduced price lunch eligibility, gifted program participation, and ever being classified as at risk of dropping out of high school, as well as cubic polynomials in 10th grade math and English standardized test scores, high school graduation year, the number of advanced courses taken in high school, the percentage of days absent, and the number of days suspended for disciplinary infractions. The running variable in each plot pools SAT and ACT submitters by dividing each SAT submitter's running variable by 40 and grouping with the nearest ACT running variable value. The RD estimates come from the main specification described in Mountjoy (2026).

Figure 7: Describing the RD Target Schools



Notes: Each circle represents one of the 28 target schools with admission discontinuities. The horizontal axis plots the average SAT score of all enrolled students. The vertical axis plots the average SAT-concorded cutoff location. Circle sizes are proportional to each school's complier-weighted share of the RD testing sample.

Figure 8: Distributions of RD Target Schools and Fallback Options



Notes: Colleges are ranked by raw mean earnings relative to UT-Austin. Triangles denote target schools with admission cutoffs, weighted by their complier share of the RD estimation sample. Circles denote fallback options that rejected applicants attend instead of their target school when they fall just below the target school's admission cutoff.

which uses local linear slopes in the running variable (the applicant's SAT/ACT score relative to the admission cutoff) that are allowed to differ arbitrarily on either side of the cutoff and differ across application cells, defined as the combination of target school, application year, type of test submitted (SAT or ACT), and whether the applicant was in the top GPA quartile of their high school class (as many target schools use different cutoffs for such applicants). We use a bandwidth of 3 concorded ACT points on either side of the cutoff and weight with a triangular kernel. Standard errors are clustered at the applicant level, following Kolesár and Rothe (2018).

4.2 The VAM Prediction Test

We adapt the VAM testing framework of Angrist et al. (2017) to evaluate each set of college VAMs against the RD variation, asking whether the VAMs accurately predict the causal effects of admission cutoff-crossing on student outcomes. To fix ideas, consider a single target school's admission cutoff. Crossing the cutoff exogenously shifts students' enrollment patterns: some students who would have attended fallback schools are instead induced to enroll at the target school (fuzzy RD compliers). These exogenous enrollment shifts affects those students' eventual outcomes like degree attainment and earnings. For each VAM estimation method, we can assess the VAMs' predictive ability by replacing each student's actual outcome Y_i with the VAM estimate of the college they initially enroll in (including no college),

$$\widehat{\text{VAM}}_i = \sum_{j=0}^{J} \hat{\beta}_j D_{ij}, \tag{2}$$

and re-running the same natural experiment. The effect of cutoff-crossing on the VAM-predicted outcome $\widehat{\text{VAM}}_i$ captures how much the VAMs predict that student outcomes should change. We then compare that prediction with the effect of cutoff-crossing on the students' actual outcomes Y_i. If the VAMs are accurate, these two effects should be equal at every target school, up to estimation error.

We assess VAM accuracy both visually and formally. Visually, in Figures 9–12, for each target university with an admission cutoff, we plot the reduced-form effect of crossing that cutoff on applicants' actual outcomes Y_i on the vertical axis, and the first-stage-esque effect of crossing the same cutoff on $\widehat{\text{VAM}}_i$ on the horizontal axis. If the VAMs are accurate, all of these points (one for each target college with an

admission cutoff) should lie along the 45-degree line, up to estimation error. Formally, we test this null hypothesis with the Angrist et al. (2017) omnibus test, which can be implemented in a one-step student-level regression of the prediction residual ($Y_i - \widehat{\text{VAM}}_i$) on the 19 binary cutoff-crossing instruments and the RD controls. The null of the omnibus test is that all 19 coefficients on the cutoff-crossing instruments are jointly equal to zero, which would be the case if the change in $\widehat{\text{VAM}}_i$ equals the change in Y_i at every cutoff. Angrist et al. (2017) show that the omnibus test can be usefully decomposed into a forecast test of whether the average relationship has a slope of one, plus an overidentification test of whether the target-school-specific deviations from the estimated forecast line are solely due to estimation error.

4.3 Test Results

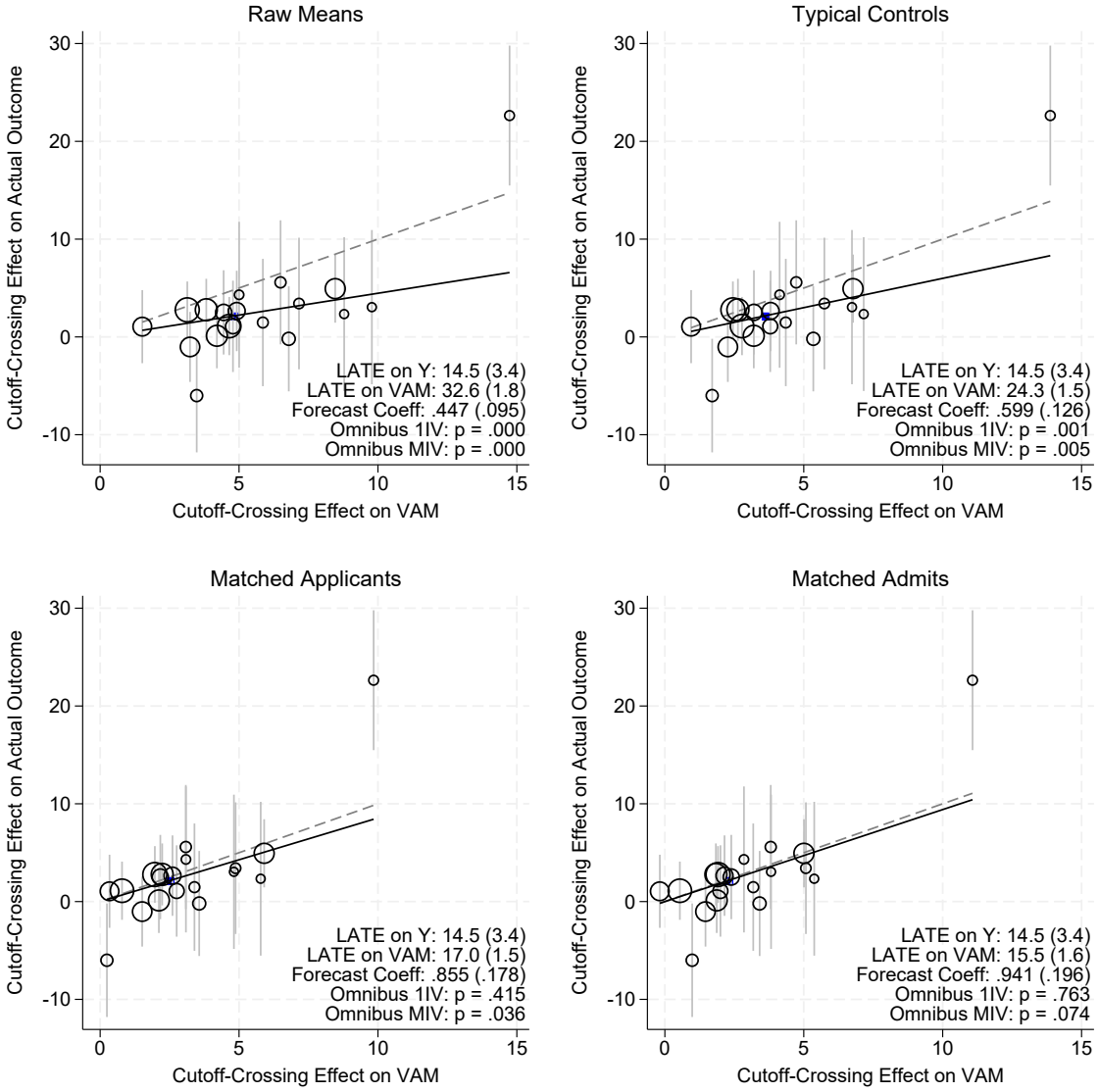
Figures 9–12 present the testing results for each outcome, with the four observational VAM methods displayed as sub-panels within each figure. Each circle represents one target school, sized in proportion to the precision of the reduced-form estimates on the vertical axes. The slope of the solid line equals the estimated forecast coefficient, while the dashed line marks the 45-degree benchmark.

We find four main patterns of test results. First, the raw means in the top left panel of each figure consistently deliver forecast coefficients around 0.40-0.45 across all outcomes. Each one-unit increase in $\widehat{\text{VAM}}_i$ is thus associated with only a 0.40-0.45 unit increase in Y_i, implying that the raw mean VAMs overpredict differences in causal impacts across colleges by over a factor of two. This pattern is consistent with substantial selection bias in popular raw mean outcome comparisons of colleges, both for educational attainment outcomes and for earnings.

Second, for the educational attainment outcomes, the "typical controls" VAMs also overpredict causal impacts, with forecast coefficients around 0.60. This is a modest improvement over the raw means, but consistent with typically observable covariates being insufficient controls for addressing differential selection on student achievement potential across colleges.

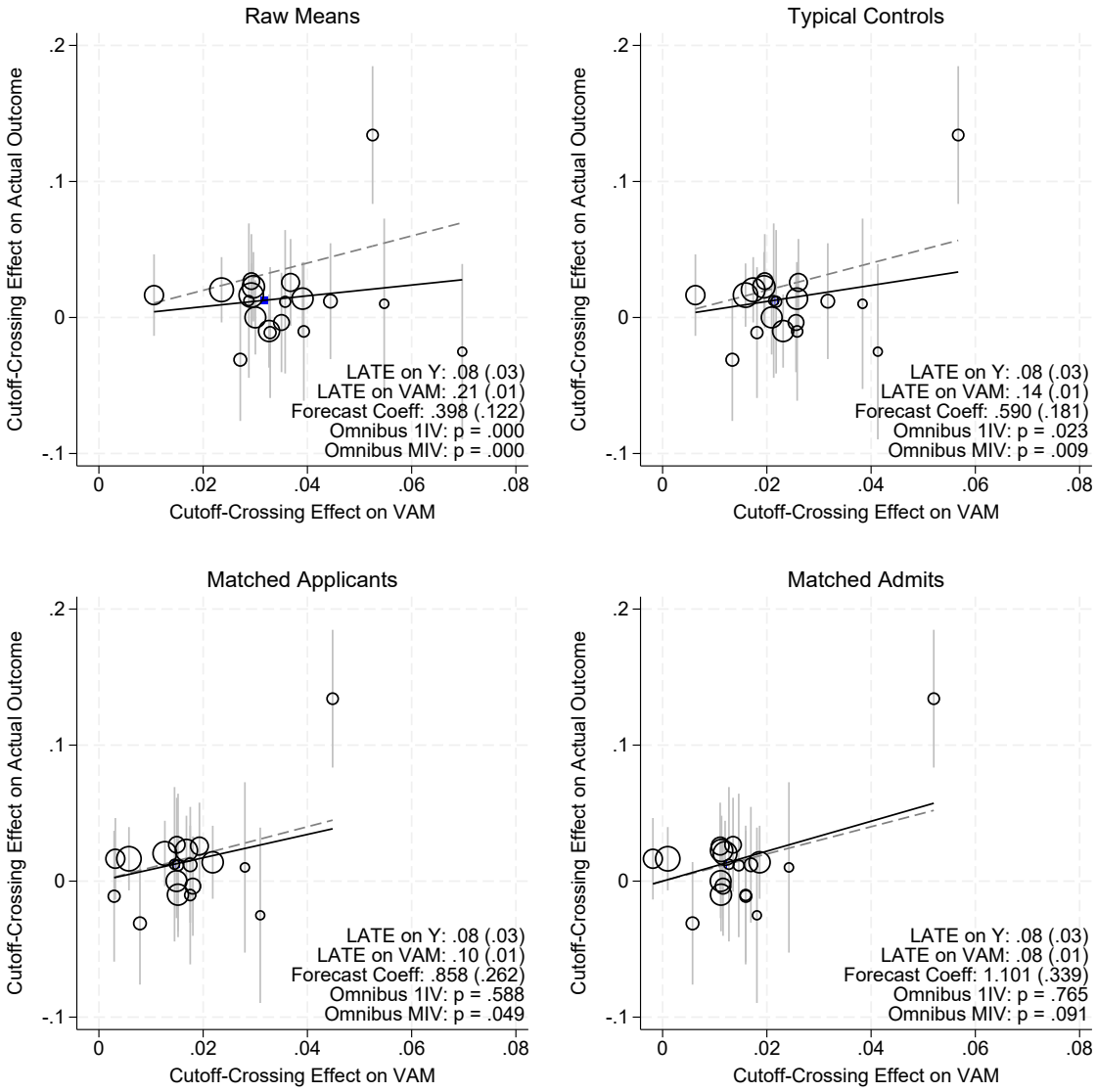
Third, the more stringent matched applicant and matched admit VAMs are highly accurate for predicting impacts on educational attainment, with forecast coefficients statistically indistinguishable from one. There is one outlier target college, Prairie View A&M, that is predicted to have large value-added on attainment outcomes

Figure 9: VAM Prediction Tests: Cumulative Credits



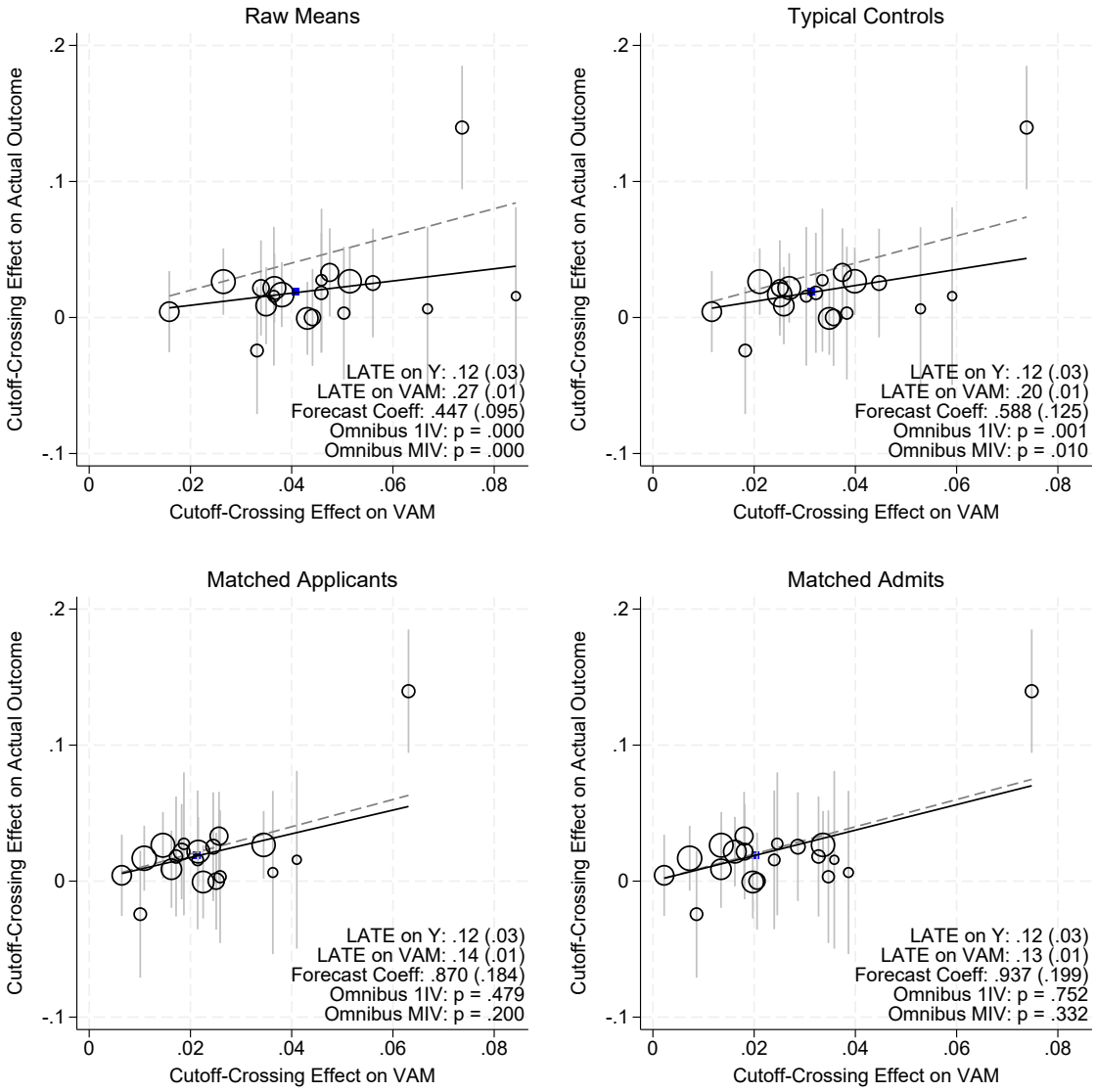
Notes: Each panel displays the VAM prediction test for a different VAM estimation method. Each point represents a target university with an admission cutoff. The vertical axis value is the reduced-form effect of crossing that target university's admission cutoff on the actual outcome. The horizontal axis value is the first-stage-esque effect of cutoff-crossing on the VAM of the college enrolled in (including no college). The circles are sized by the precision of the reduced form. The dashed line is the 45-degree line, corresponding to perfect forecast. The solid line is the estimated forecast line.

Figure 10: VAM Prediction Tests: Any Degree



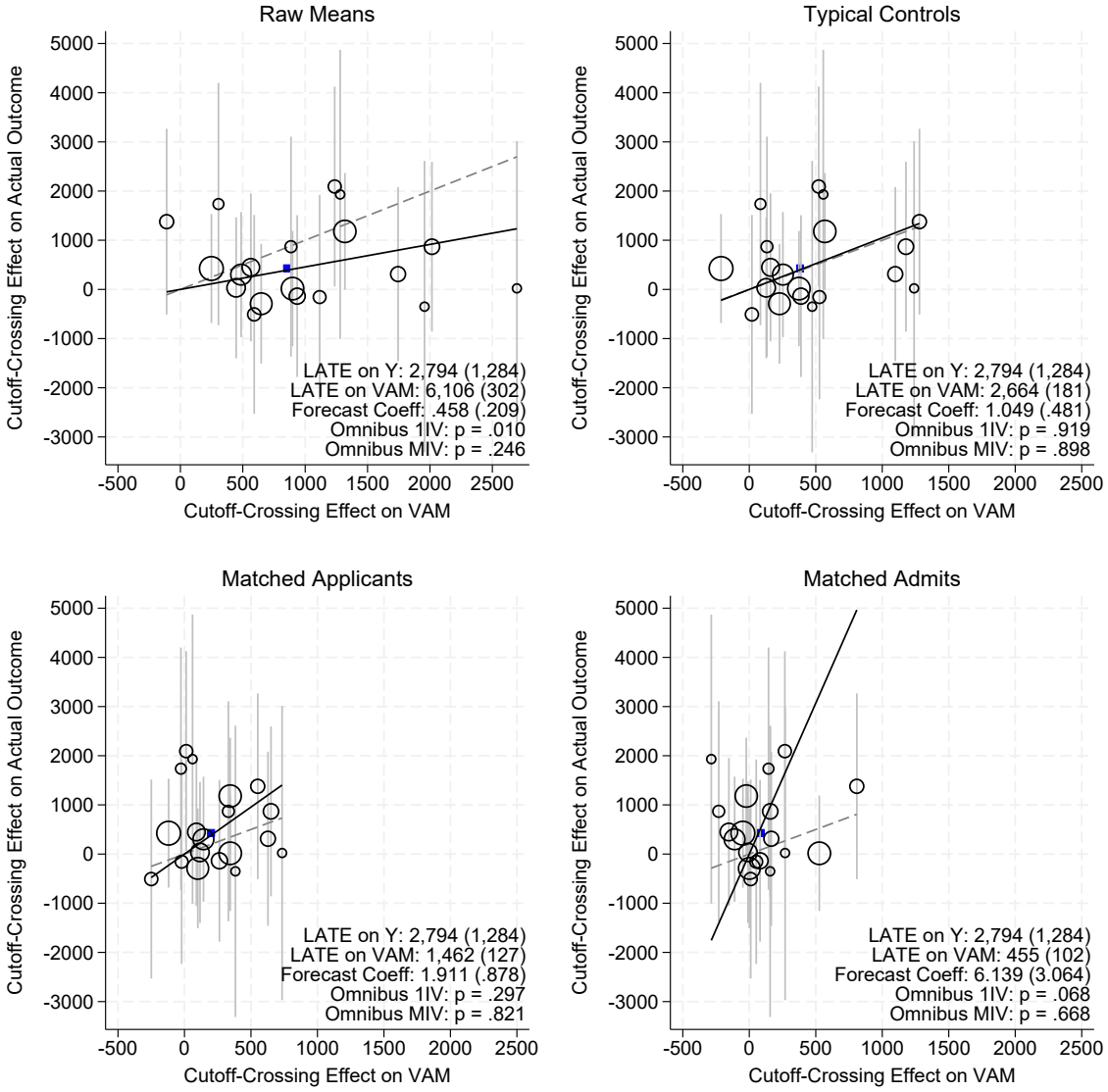
Notes: Each panel displays the VAM prediction test for a different VAM estimation method. Each point represents a target university with an admission cutoff. The vertical axis value is the reduced-form effect of crossing that target university's admission cutoff on the actual outcome. The horizontal axis value is the first-stage-esque effect of cutoff-crossing on the VAM of the college enrolled in (including no college). The circles are sized by the precision of the reduced form. The dashed line is the 45-degree line, corresponding to perfect forecast. The solid line is the estimated forecast line.

Figure 11: VAM Prediction Tests: Bachelor's Degree



Notes: Each panel displays the VAM prediction test for a different VAM estimation method. Each point represents a target university with an admission cutoff. The vertical axis value is the reduced-form effect of crossing that target university's admission cutoff on the actual outcome. The horizontal axis value is the first-stage-esque effect of cutoff-crossing on the VAM of the college enrolled in (including no college). The circles are sized by the precision of the reduced form. The dashed line is the 45-degree line, corresponding to perfect forecast. The solid line is the estimated forecast line.

Figure 12: VAM Prediction Tests: Earnings



Notes: Each panel displays the VAM prediction test for a different VAM estimation method. Each point represents a target university with an admission cutoff. The vertical axis value is the reduced-form effect of crossing that target university's admission cutoff on the actual outcome. The horizontal axis value is the first-stage-esque effect of cutoff-crossing on the VAM of the college enrolled in (including no college). The circles are sized by the precision of the reduced form. The dashed line is the 45-degree line, corresponding to perfect forecast. The solid line is the estimated forecast line.

but has even higher actual impacts, driving the omnibus test to marginally reject for cumulative credits and any degree completion even when using the matched applicant and matched admit VAMs. Prairie View A&M is the third largest historically black university (HBCU) in the U.S., the second oldest public university in Texas, and one the nation's largest producers of black STEM graduates. These results indicate that Prairie View A&M's actual causal impacts on educational attainment are even larger than what the available observational VAM methods would predict.

Finally, for earnings, the test results deliver a rather different pattern across VAM methods compared to educational attainment. The raw mean VAMs still overpredict, but the typical controls VAMs have a forecast coefficient close to one. That forecast coefficient is imprecisely estimated, however, thanks to large residual variance in earnings, preventing any precise conclusions about the predictive accuracy of the typical controls VAMs for earnings effects. Strikingly, the more stringent VAM methods underpredict earnings gains, with the matched applicant method delivering a noisy forecast coefficient of 1.9 (SE .88), and the matched admit method delivering a dramatically large forecast coefficient estimate of 6.1 (SE 3.1). That is, the actual earnings gains for cutoff-crossers tend to be twice as large as what the matched applicant VAMs would predict, and six times as large as what the matched admit VAMs would predict. We investigate the drivers of this phenomenon in Section 5.

5 Learning about Selection

The forecast test results of Section 4 offer a clear lesson about selection in raw outcome means as college value-added measures: they consistently overpredict quasi-experimental effects, confirming that "better" schools enroll better students. The VAMs produced by the stringent matched applicant and match admit methods, on the other hand, perform well for educational attainment outcomes but underpredict earnings effects, especially the Dale and Krueger (2002)-inspired VAMs that compare the outcomes of admitted students choosing among the same set of admission offers. In this section, we investigate potential drivers of this result, finding evidence consistent with admitted students self-selecting into college choices based on their own idiosyncratic potential outcomes.

5.1 What Types of Students Are Compared in Each Method?

One common conjecture about the divergence between large earnings effects of selectivity for admission cutoff-crossers versus small effects for matched admits is that the former are marginally admitted students, while the latter are inframarginal, and the treatment effect of attending a more selective college may differ across these different groups. To put the different observational VAM comparisons and the quasi-experimental RD comparisons on more equal footing, we begin by narrowing the focus to the subpopulation of students whose SAT/ACT scores and application decisions put them right around an admission cutoff at one of our target universities. That is, we ignore students who are clearly inframarginal and focus on the identfying variation of each method *within* the subpopulation of marginal students.

To characterize the types of marginal students who are compared in each method, consider the different "response types" around the cutoff. *Always-Takers* enroll at the target school regardless of which side of the cutoff they are on; this is possible below the cutoff given that the SAT/ACT cutoffs are "fuzzy" rather than sharp determinants of admission. *Compliers* enroll at the target school if and only if they fall above the cutoff; if below, they attend their fallback option instead. *Never-Takers* do not enroll at the target school regardless of cutoff status. We can further divide this group into *Admitted Never-Takers*, who do not enroll in the target school even if both above the cutoff and actually admitted, versus *Rejected Never-Takers*, who were not admitted despite being above the cutoff (which again is possible due to the fuzziness of the admission RD). We rule out the existence of *Defiers*, who would enroll if below the cutoff but not enroll if above the cutoff.

Which types are involved in each comparison? The local average treatment effect (LATE) identified by the RD design only compares potential outcomes within the compliers:

$$\mathbb{E}[Y_1 - Y_0|\text{Compliers}],$$

where Y_1 is the potential outcome at the target school, which we'll refer to as "treated" ($T = 1$) and Y_0 is the potential outcome at the (generally less selective) fallback option, which we'll refer to as "untreated" ($T = 0$).

In contrast, the Hoxby (2015) matched applicant approach (focusing on the "vertical" comparisons that make up the vast majority of the identifying variation in our setting) compares *all* treated students vs. *all* untreated students at the cutoff

("on-the-bubble"), which involves multiple response types:

$$E[Y_1|T=1] - E[Y_0|T=0] = E[Y_1|\text{Treated Compliers \& Always-Takers}] - E[Y_0|\text{Untreated Compliers \& Never-Takers}].$$

Finally, the Dale and Krueger (2002) matched admit approach compares *admitted* treated vs. untreated students at the cutoff:

$$\begin{aligned} &E[Y_1|T=1, \text{Admitted}=1] - E[Y_0|T=0, \text{Admitted}=1] \\ =&E[Y_1|\text{Treated Compliers \& Always-Takers}] - E[Y_0|\text{Admitted Never-Takers}]. \end{aligned}$$

Note that the matched applicant and matched admit approaches share the same treatment group, comprised of treated compliers and always-takers. They differ in their untreated groups: untreated matched applicants include untreated compliers and all never-takers, whereas untreated matched admits, by nature of being admitted to the treatment option, exclude untreated compliers and rejected never-takers, leaving only admitted never-takers.

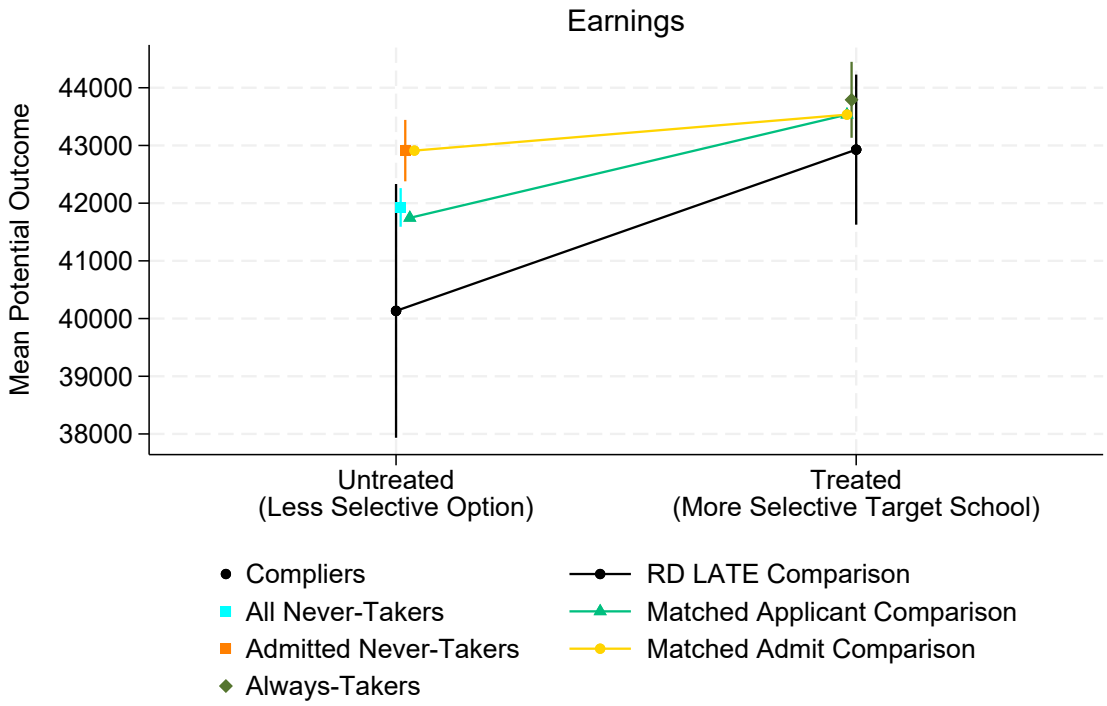
The key takeaway from this derivation is that the RD LATE's untreated group isolates compliers who would enroll at the target school if admitted, but they get exogenously denied the opportunity. In contrast, the untreated Dale and Krueger (2002) matched admits are admitted never-takers who have the opportunity to attend the target school, but voluntarily chose not to enroll. The untreated Hoxby (2015) matched applicants are a mix of these different response types. To the extent that these distinct types of students have substantively different potential outcomes, these methods will involve substantively different comparisons.

5.2 Describing the Students Compared in Each Method

Having established the distinct types of marginal students compared in each method, we now use the RD specification (pooled across target schools) and the identification results of Abadie (2002) to empirically describe the characteristics and potential outcomes of these types, unpacking the sources of disagreement between the methods and revealing interesting selection patterns.

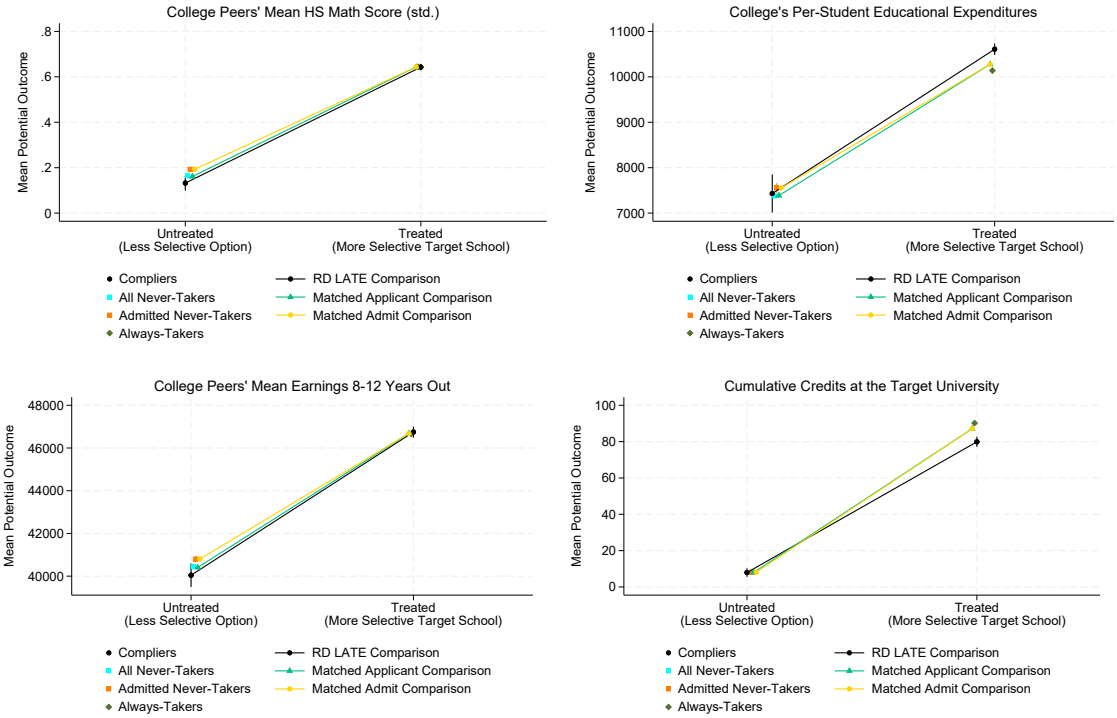
Figure 13 presents the key results. We plot the mean potential outcomes of each response type of marginal student involved in the RD, matched applicant, and matched

Figure 13: Unpacking the Methods and Revealing Self-Selection



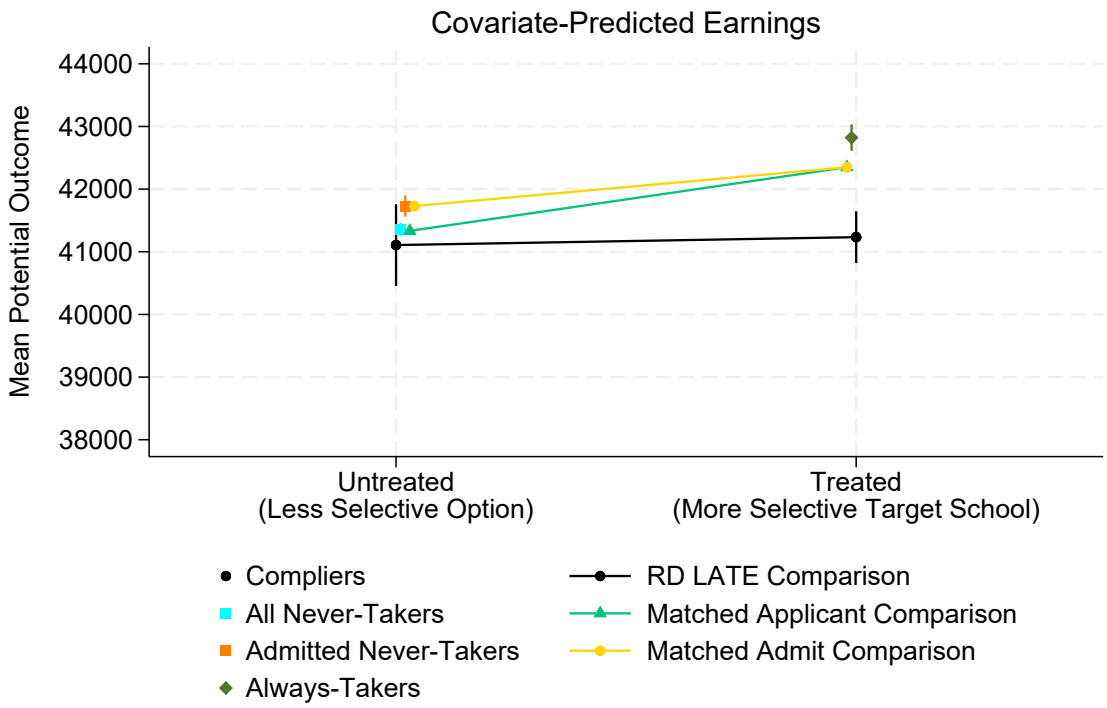
Notes: This figure plots the mean potential outcomes of each response type of marginal student involved in the RD, matched applicant, and matched admit comparisons. The horizontal axis distinguishes between the treated state (enrolling at the target school) versus the untreated state (enrolling at the less selective alternative). Horizontal lines represent the treated vs. untreated comparisons involved in each method. Vertical lines are 95% confidence intervals around each point estimate.

Figure 14: Similar Treatment Dosage across Methods



Notes: This figure plots the mean potential outcomes of each response type of marginal student involved in the RD, matched applicant, and matched admit comparisons. The horizontal axis distinguishes between the treated state (enrolling at the target school) versus the untreated state (enrolling at the less selective alternative). Horizontal lines represent the treated vs. untreated comparisons involved in each method. Vertical lines are 95% confidence intervals around each point estimate.

Figure 15: Observables vs. Unobservables: Covariate-Predicted Earnings



Notes: This figure plots the mean potential outcomes of each response type of marginal student involved in the RD, matched applicant, and matched admit comparisons. The horizontal axis distinguishes between the treated state (enrolling at the target school) versus the untreated state (enrolling at the less selective alternative). Horizontal lines represent the treated vs. untreated comparisons involved in each method. Vertical lines are 95% confidence intervals around each point estimate.

admit comparisons. Each method compares a group of treated students who enroll in the more selective target school, depicted on the right side of the plot, to a group of untreated students on the left side of the plot who enroll in their less selective alternative to the target school. Three important patterns emerge. First, all of the treated types experience fairly similar earnings outcomes when attending the target school. Second, and crucially, untreated matched admit "decliners" who choose their less selective option have substantially higher mean earnings in that untreated state relative to untreated compliers. That is, admitted students who had the opportunity to attend the more selective school but voluntarily chose not to earn substantially more at their less selective alternative compared to the rejected compliers who would have attended the target school if admitted. Third, the matched applicant mean untreated potential outcome lies between these two extremes, as it includes both of these types (along with rejected never-takers).

In Figure 14, we rule out the explanation that the matched admit comparison simply involves weaker treatment "dosage" relative to the other methods. The treated versus untreated contrasts are large and very similar across methods when considering selectivity (as measured by peer average high school test scores), per-student educational spending, peer mean earnings, and cumulative credits completed at the target school. The small estimated earnings effect for matched admits is therefore not attributable to weaker treatment dosage along these dimensions.

Finally, to assess how much of the higher earnings potential of matched admit "decliners" is captured by their observable characteristics, Figure 15 replaces each response type's actual earnings with their covariate-predicted earnings. The vertical gap between untreated compliers and untreated matched admits is much smaller. This suggests that matched admit "decliners" have idiosyncratically high earnings potential largely along unobservable dimensions not captured by our battery of pre-college covariates.

6 Conclusion

In this paper, we empirically evaluated the prospects of credible value-added measures (VAMs) for individual U.S. colleges. We used linked administrative data spanning the state of Texas, a setting well-suited to this exercise given its large size, diverse post-secondary landscape, and rich data. We estimated a variety of college-specific VAMs

using the most prominent observational methods from the literature, including raw mean outcomes, regression adjustment for typically available controls, the matched applicant method of Hoxby (2015, 2019), and the matched admit method inspired by Dale and Krueger (2002). To evaluate these VAMs, we used quasi-experimental enrollment variation from admission cutoffs, testing the ability of each set of VAMs to forecast causal cutoff-crossing effects on applicants' eventual educational attainment and earnings.

Our results indicate that the most prominent measures, which simply average the outcomes of students who attend each college, substantially overpredict causal impacts. For educational attainment outcomes, typical regression-adjusted VAMs with controls for test scores and demographics also overpredict, while more stringent VAMs incorporating application and admission controls predict well. These more stringent VAMs underpredict earnings gains, however, especially when comparing students choosing among the same admission offers.

To understand this divergence, we theoretically and empirically characterized the types of students involved in the observational and quasi-experimental comparisons. We found that students in the matched admit method who decline a more selective admission offer tend to have relatively high earnings at their less-selective choice, both relative to rejected admission cutoff compliers and relative to what their own observable characteristics would predict, consistent with self-selection on unobservable earnings potential. These results provide new evidence on the prospects of credible college value-added measures and on the nature of self-selection into college choices.

References

Abadie, Alberto (2002). "Bootstrap Tests for Distributional Treatment Effects in Instrumental Variable Models". *Journal of the American Statistical Association* 97.457, pp. 284–292.

Angrist, Joshua D, Peter D Hull, Parag A Pathak, and Christopher R Walters (2017). "Leveraging Lotteries for School Value-Added: Testing and Estimation". *The Quarterly Journal of Economics* 132.2, pp. 871–919.

Arcidiacono, Peter (2005). "Affirmative Action in Higher Education: How Do Admission and Financial Aid Rules Affect Future Earnings?" *Econometrica* 73.5, pp. 1477–1524.

Arcidiacono, Peter, Esteban M. Aucejo, and V. Joseph Hotz (2016). "University Differences in the Graduation of Minorities in STEM Fields: Evidence from California". *American Economic Review* 106.3, pp. 525–562.

Black, Sandra E, Jeffrey T Denning, and Jesse Rothstein (2023). "Winners and Losers? The Effect of Gaining and Losing Access to Selective Colleges on Education and Labor Market Outcomes". *American Economic Journal: Applied Economics* 15.1, pp. 26–67.

Bleemer, Zachary (2021). "Top Percent Policies and the Return to Postsecondary Selectivity", Working paper.

— (2022). "Affirmative Action, Mismatch, and Economic Mobility after California's Proposition 209". *The Quarterly Journal of Economics* 137.1, pp. 115–160.

Bleemer, Zachary and Sarah Quincy (2025). "Changes in the College Mobility Pipeline since 1900", NBER Working Paper No. 33797.

Bodoh-Creed, Aaron L. and Brent R. Hickman (2024). "Pre-College Human Capital Investment and Affirmative Action: A Structural Policy Analysis of US College Admissions", Working paper, Washington University in St. Louis.

Broecke, Stijn (2012). "University Selectivity and Earnings: Evidence from UK Data on Applications and Admissions to University". *Economics of Education Review* 31.3, pp. 96–107.

Chetty, Raj, David J Deming, and John N Friedman (2026). "Diversifying Society's Leaders? The Determinants and Causal Effects of Admission to Highly Selective Private Colleges". *The Quarterly Journal of Economics* 141.1, pp. 51–145.

Chetty, Raj, John N. Friedman, Emmanuel Saez, Nicholas Turner, and Danny Yagan (2020). "Income Segregation and Intergenerational Mobility across Colleges in the United States". *The Quarterly Journal of Economics* 135.3, pp. 1567–1633.

Ciocca Eller, Christina (2023). "What Makes a Quality College? Reexamining the Equalizing Potential of Higher Education in the United States". *American Journal of Sociology* 129.3, pp. 637–714.

Cunha, Jesse M. and Trey Miller (2014). "Measuring Value-Added in Higher Education: Possibilities and Limitations in the Use of Administrative Data". *Economics of Education Review* 42, pp. 64–77.

Dale, Stacy Berg and Alan B. Krueger (2002). "Estimating the Payoff to Attending a More Selective College: An Application of Selection on Observables and Unobservables". *The Quarterly Journal of Economics* 117.4, pp. 1491–1527.

Edwards, Ashley, Justin Ortagus, Jonathan Smith, and Andria Smythe (2025). "HBCU Enrollment and Longer-Term Outcomes". *American Economic Journal: Economic Policy* 17.3, pp. 197–235.

Fryer, Roland G. and Michael Greenstone (2010). "The Changing Consequences of Attending Historically Black Colleges and Universities". *American Economic Journal: Applied Economics* 2.1, pp. 116–148.

Ge, Suqin, Elliott Isaac, and Amalia Miller (2022). "Elite Schools and Opting in: Effects of College Selectivity on Career and Family Outcomes". *Journal of Labor Economics* 40.S1, S383–S427.

Hinrichs, Peter (2014). "Affirmative Action Bans and College Graduation Rates". *Economics of Education Review* 42, pp. 43–52.

Hoekstra, Mark (2009). "The Effect of Attending the Flagship State University on Earnings: A Discontinuity-Based Approach". *The Review of Economics and Statistics* 91.4, pp. 717–724.

Hoxby, Caroline M (2009). "The Changing Selectivity of American Colleges". *The Journal of Economic Perspectives* 23.4, pp. 95–118.

— (2015). "Computing the Value-Added of American Postsecondary Institutions", Internal Revenue Service Statistics of Income Division working paper.

— (2019). "The Productivity of US Postsecondary Institutions". *Productivity in Higher Education*. Ed. by Caroline M Hoxby and Kevin Stange. University of Chicago Press, pp. 31–66.

Kolesár, Michal and Christoph Rothe (2018). "Inference in Regression Discontinuity Designs with a Discrete Running Variable". *American Economic Review* 108.8, pp. 2277–2304.

Long, Mark C. (2008). "College Quality and Early Adult Outcomes". *Economics of Education Review* 27.5, pp. 588–602.

McCrary, Justin (2008). "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test". *Journal of Econometrics* 142.2, pp. 698–714.

Mountjoy, Jack (2022). "Community Colleges and Upward Mobility". *American Economic Review* 112.8, pp. 2580–2630.

— (2026). "Marginal Returns to Public Universities". *The Quarterly Journal of Economics* 141.1, pp. 429–497.

Mountjoy, Jack and Brent R. Hickman (2021). "The Returns to College(s): Relative Value-Added and Match Effects in Higher Education", NBER Working Paper No. 29276.

National Center for Education Statistics (2018). *Digest of Education Statistics*. Tech. rep.

Smith, Jonathan (2013). "Ova and out: Using Twins to Estimate the Educational Returns to Attending a Selective College". *Economics of Education Review* 36, pp. 166–180.

Smith, Jonathan, Joshua Goodman, and Michael Hurwitz (2025). "The Economic Impact of Access to Public Four-Year Colleges". *Journal of Human Resources* Forthcoming.

Stevens, David W. (2007). "Employment That Is Not Covered by State Unemployment Insurance Laws", U.S. Census Bureau Technical Paper No. TP–2007–04.

UTD-ERC (2025). *UT Dallas Education Research Center Data Holdings*.

Zimmerman, Seth D (2014). "The Returns to College Admission for Academically Marginal Students". *Journal of Labor Economics* 32.4, pp. 711–754.