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

Jack Mountjoy† Brent R. Hickman‡

September 18, 2026

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 the most prominent observational methods. We then use quasi-experimental enrollment variation from admission discontinuities to evaluate these VAMs. The most popular VAMs, which simply average the outcomes of students from each college, overpredict causal impacts on all outcomes by a factor of two. Regression-adjusted VAMs with typical controls like test scores and demographics overpredict causal impacts on educational attainment, but accurately predict earnings gains. More stringent VAMs that incorporate application and admission controls accurately predict educational gains, but underpredict earnings gains, especially when comparing admitted students choosing among the same admission offers. We find that such students who willingly decline a more selective offer tend to have unusually high earnings at their less selective choice, consistent with self-selection. These results provide new evidence on college VAM credibility and student self-selection into colleges.

*This paper 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, Attila Lindner, Chris Walters, and Seth Zimmerman. We also thank Holly Kosiewicz, Greg Branch, Trey Miller, Mark Lu, Lynn Bergeland Morgan, Thomas Brunell, 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, college value-added, value-added measures, admission cutoffs, self-selection.

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

‡University of Utah. hickmanbr@gmail.com.

1 Introduction

College tends to pay off for marginal enrollees (Zimmerman, 2014; Heckman et al., 2018; 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) (U.S. Department of Education, 2024). 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 (National Center for Education Statistics, 2024). 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-specific 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; 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.

Meanwhile, a separate literature increasingly employs quasi-experimental research designs to identify causal impacts of shifting students across U.S. colleges, like the admission discontinuities in Hoekstra (2009), Zimmerman (2014), Kozakowski (2023), Bleemer (2024), Smith et al. (2025), and Mountjoy (2026), the affirmative action regime changes in Hinrichs (2014), Bleemer (2022), Black et al. (2023), and Antman et al. (2024), and the off-the-waitlist design in Chetty et al. (2026). 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 discontinuities 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 institutions (including all community colleges) are non-selective.

In this paper, we bridge these literatures and empirically evaluate the prospects of credible value-added measures for individual colleges. We use rich administrative data to first generate observational VAMs spanning the spectrum of prominent methods, then subject those VAMs to quasi-experimental validation using admission discontinuities, all in a unified setting. We adapt the K-12 school VAM testing framework of Angrist et al. (2017) to higher education, rooted in the LaLonde (1986)-esque idea that credible observational VAMs should accurately forecast the causal effects of exogenously shifting students across schools. We find that the forecast accuracy of observational VAMs varies substantially across prominent methods, with different patterns for educational attainment outcomes versus earnings. We also go beyond the Angrist et al. (2017) VAM testing framework to show how forecast 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, credit accumulation, 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 and technical colleges to regional public universities to selective flagships and 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 a unified setting, including those requiring individual-level data on college applications and admissions, along with the admission discontinuity design used to evaluate those VAMs.

In Section 3, we describe and implement the various observational VAM methods. We focus on four distinct approaches that span the range of prominent methods. The first, which we call Raw Means, 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, Typical Controls, regression-adjusts for typically available confounders like test scores and demographics, exemplified by the baseline college VAM specifications of Chetty et al. (2020, 2026) and rem-

iniscent of the workhorse K-12 value-added model for teachers and schools. Unlike in the K-12 setting, however, the Typical Controls approach to college VAMs does not have the luxury of including lagged outcomes as proxies for student ability, like controlling for prior test scores when estimating VAMs on current test scores.[1] Such lagged outcome controls have been shown to improve the performance of K-12 test score VAMs (Chetty et al., 2014; Angrist et al., 2017; Staiger et al., 2025), but we cannot logically construct lagged controls for postsecondary outcomes like bachelor's degree completion and adult earnings among our traditional-aged college-goers.[2]

To address potentially remnant selection bias in the Typical Controls approach, our third VAM method, Matched Applicants, 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 and final approach, Matched Admits, implements the widespread but controversial strategy inspired by Dale and Krueger (2002) that compares the outcomes of admitted students choosing among the same set of realized 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.[3]

These four methods generate substantially different schedules of college VAMs. Raw Mean VAMs differ dramatically across colleges, but much of that variation is explained by students' pre-college characteristics (Mabel et al., 2020). 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 solely on their pre-college covariates. The Typical Controls VAMs, of course, mechanically balance these observable student covariates across colleges; more interestingly, both the Hoxby (2015)-style Matched Applicant method and the Dale and Krueger (2002)-style Matched Admit method achieve rough balance in covariate-predicted

[1] Riehl et al. (2019) offer a rare counterexample by estimating college VAMs on exam scores in Colombia, where students take standardized exams both at college entry and at college exit.

[2] Dynarski et al. (2018) discuss lagged outcome approaches to identifying earnings returns to community college for older adults, who do have meaningful pre-treatment earnings histories.

[3] These include college-specific VAM estimation in Cunha and Miller (2014), Arcidiacono et al. (2016), Ciocca Eller (2023), Bleemer (2022, 2024), 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). Related methods also appear in the literature on K-12 school VAM estimation, including Abdulkadiroglu et al. (2020) and Angrist et al. (2024).

earnings across most of the treatment college distribution.[4] These methods still generate divergent VAM schedules for actual outcomes, however. The Dale and Krueger (2002)-style Matched Admit VAMs for earnings 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.[5] The Hoxby (2015)-style Matched Applicant VAMs, meanwhile, tend to lie between those of the Typical Controls method and the Matched Admit method.

In Section 4, we evaluate these discordant VAM schedules by adapting the VAM 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 discontinuities. Specifically, we draw on the admission discontinuities in SAT/ACT scores identified in Mountjoy (2026), which generate exogenous enrollment variation among applicants to twenty-eight Texas public universities spanning a wide range of selectivity. The VAM testing procedure has an intuitive visual representation: for each target university with an admission cutoff, we plot the causal effect of crossing that cutoff on applicants' actual outcomes against the effect of crossing the same cutoff on the observational VAM of the college they end up attending (including no college). If the VAMs from a given observational method are accurate, all of these points (one for each target college with an admission cutoff) should lie along the 45-degree line with a slope of one, up to sampling error.

We find three main patterns of test results. First, the Raw Mean VAMs consistently overpredict causal impacts on all outcomes by roughly a factor of two, implying substantial selection bias in popular comparisons of raw outcomes across colleges. Second, the Typical Controls VAMs also overpredict causal impacts on educational attainment outcomes like credit accumulation and degree completion, but accurately predict earnings gains. Third, the more stringent Matched Applicant and Matched Admit VAMs are highly accurate for predicting impacts on educational attainment, but they underpredict earnings gains, especially the Dale and Krueger (2002)-inspired Matched Admit method comparing students with identical admission offers. Statistically, we can conclude that the Raw Mean VAMs substantially overpredict earnings effects and that the Matched Admit VAMs substantially underpredict them, while we cannot reject the earnings accuracy of the Typical Controls and Matched Applicant VAMs.

To understand this divergence across methods and outcomes and learn about student se-

[4]We add covariate controls to these specifications when evaluating their VAMs, but their rough balance makes the evaluation results largely insensitive to the inclusion or exclusion of such controls.

[5]Chetty et al. (2026) replicate the Texas public university results of Mountjoy and Hickman (2021) using nationwide IRS earnings records.

lection behavior, in Section 5 we go beyond the Angrist et al. (2017) VAM testing framework to theoretically and empirically characterize the types of students involved in the observational and quasi-experimental comparisons. To ensure a level playing field, we ignore clearly inframarginal students and focus on the identfying variation of each method *within* the policy-relevant subpopulation of academically marginal students, i.e. those whose SAT/ACT scores and application decisions render them on the cusp of an admission cutoff at one of our 28 target universities. Interestingly, we find that admitted students in the Dale and Krueger (2002)-style Matched Admit comparison who voluntarily decline their more selective offer tend to have unusually high earnings after attending their less-selective choice, both relative to involuntarily rejected admission cutoff compliers and relative to the Matched Admit decliners' observationally identical peers at their less-selective choice. This pattern is consistent with admitted students in Dale and Krueger (2002)-style comparisons self-selecting into colleges based on their own idiosyncratic earnings potential à la Roy (1951), and doing so primarily on unobserved dimensions not captured by pre-college observables like test scores and demographics. As Dale and Krueger (2002) briefly speculated, but did not investigate empirically given their lack of quasi-experimental benchmark: "the students who chose to attend a less selective school even though they were admitted to a more selective one might be the students with attributes that lead them to benefit more from attending a less selective school" (p. 1524). We identify majors and industries as potential mechanisms driving this phenomenon, as the Dale and Krueger (2002) selectivity "decliners" end up completing more lucrative majors and working in higher-paying industries than their peers.

Together, these results provide new evidence on the prospects of credible college value-added measures and on the nature of self-selection into colleges. A handful of prior quasi-experimental studies in higher education have explored whether their main causal estimates align with a limited set of observational VAMs as an informal secondary exercise, yielding little consensus about which college VAM methods may be more or less credible for a given type of outcome.[6] We provide the first systematic evaluation of college VAMs by testing the whole spectrum of prominent observational VAM methods against rich quasi-experimental variation that exogenously shifts students across a wide diversity of institutions. Our results show that the accuracy of observational VAMs in predicting causal effects varies substantially across different VAM methods, and differentially so across educational versus earnings outcomes.

Furthermore, our explicit theoretical and empirical characterization of the different types

[6]See Cohodes and Goodman (2014), Chetty et al. (2020), Bleemer (2022), Bleemer and Mehta (2022), Bleemer (2024), Londoño-Vélez et al. (2023), Campos et al. (2026), Chetty et al. (2026), and Mountjoy (2026).

of students involved in these comparisons suggests some caution in the pursuit of a one-size-fits-all college value-added schedule applicable to all prospective students, as different students experience different outcomes at the same college and may self-select on this heterogeneity (Heckman et al., 2006). We find novel evidence consistent with such behavior among admitted college applicants deciding where to attend, contributing to a growing literature investigating self-selection into other educational choices like whether to attend college at all (Willis and Rosen, 1979; Carneiro et al., 2011), college sector (Mountjoy, 2022), field of study (Kirkeboen et al., 2016; Kirkeboen et al., 2026; Campos et al., 2026), high school campus (Abdulkadiroglu et al., 2026), high school track (Silliman and Virtanen, 2022), private schools (Bau, 2022), charter schools (Walters, 2018), school districts (Bruhn, 2023), remote schooling (Bruhn et al., 2025), academic effort (Bodoh-Creed and Hickman, 2024; Cotton et al., 2026), and early childcare (Kline and Walters, 2016; Cornelissen et al., 2018).

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, 2026). Texas is the second largest U.S. state, 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.[7] The 2004 cohort is the first to have SAT/ACT scores available, and we stop at the 2014 cohort to observe a balanced panel of twelve 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, and disciplinary infractions.

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 regression discontinuity design described in Section 4. We also follow each high school graduate longitudinally through ad-

[7]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).

ditional 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.[8]

3 Value-Added Measures

All of the prominent observational methods for generating college-specific value-added measures (VAMs) can be described as variations of 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 in the first academic year after high school graduation, 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 composite out-of-state college option for the small share of Texas high school graduates who enroll outside the state.

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 report the raw mean outcomes of students who attend each school. We implement

[8]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.

this method by estimating Equation (1) without any control variables in X_i except high school graduation year 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 for K-12 teachers and schools. Chetty et al. (2020)'s nationwide estimation of college-specific 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 (Kulkarni and Rothwell, 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 program participation, and an indicator for being labeled 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. We also consider a version that interacts all of these covariates with the college treatment indicators, allowing for student-college match effects along observable dimensions.

3.3 Matched Applicants

Hoxby (2015, 2019) develops a college VAM estimation method that incorporates data on students' application behavior to isolate potentially more credible "vertical" and "horizontal" comparisons of matched applicants. Specifically, vertical comparisons involve applicants to the same college with the same SAT/ACT 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 test 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/ACT score, 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 with 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 and SAT/ACT scores, 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 non-enrollment as a treatment option in our VAM evaluation, given that a small but nonzero share of our admission cutoff compliers in Section 4 end up not enrolling in any college 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 individual college indicators to yield college-specific 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.

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 admitted 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 (2022, 2024), 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 college selectivity or other institutional characteristics, including Arcidiacono (2005), Long (2008), Fryer and Greenstone (2010), Broecke (2012), Smith (2013), and Ge et al. (2022). Related methods also appear in the literature on K-12 school VAM estimation, including Abdulkadiroglu et al. (2020) and Angrist et al. (2024).

The motivating idea behind the 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...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-specific 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 private non-profit institutions and community and technical colleges 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 public university 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 treatment option 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 across our four main outcomes: cumulative college credits completed (Figure 1), completion of any postsecondary credential (Figure 2), completion of a bachelor's degree (Figure 3), and annual earnings (Figure 4). The three educational attainment outcomes are measured ten years after high school graduation, giving students a long horizon to complete their postsecondary education. The earnings outcome averages each individual's real positive quarterly earnings over years 10-12 since high school graduation (roughly ages 28-30), multiplied by four to annualize.[9] The roughly 150 college treatment options, including no college and the out-of-state composite, are ranked along the vertical axis by their Raw Means, and each school's VAM estimate (relative to UT-Austin) is measured on the horizontal axis, with different markers for each method.

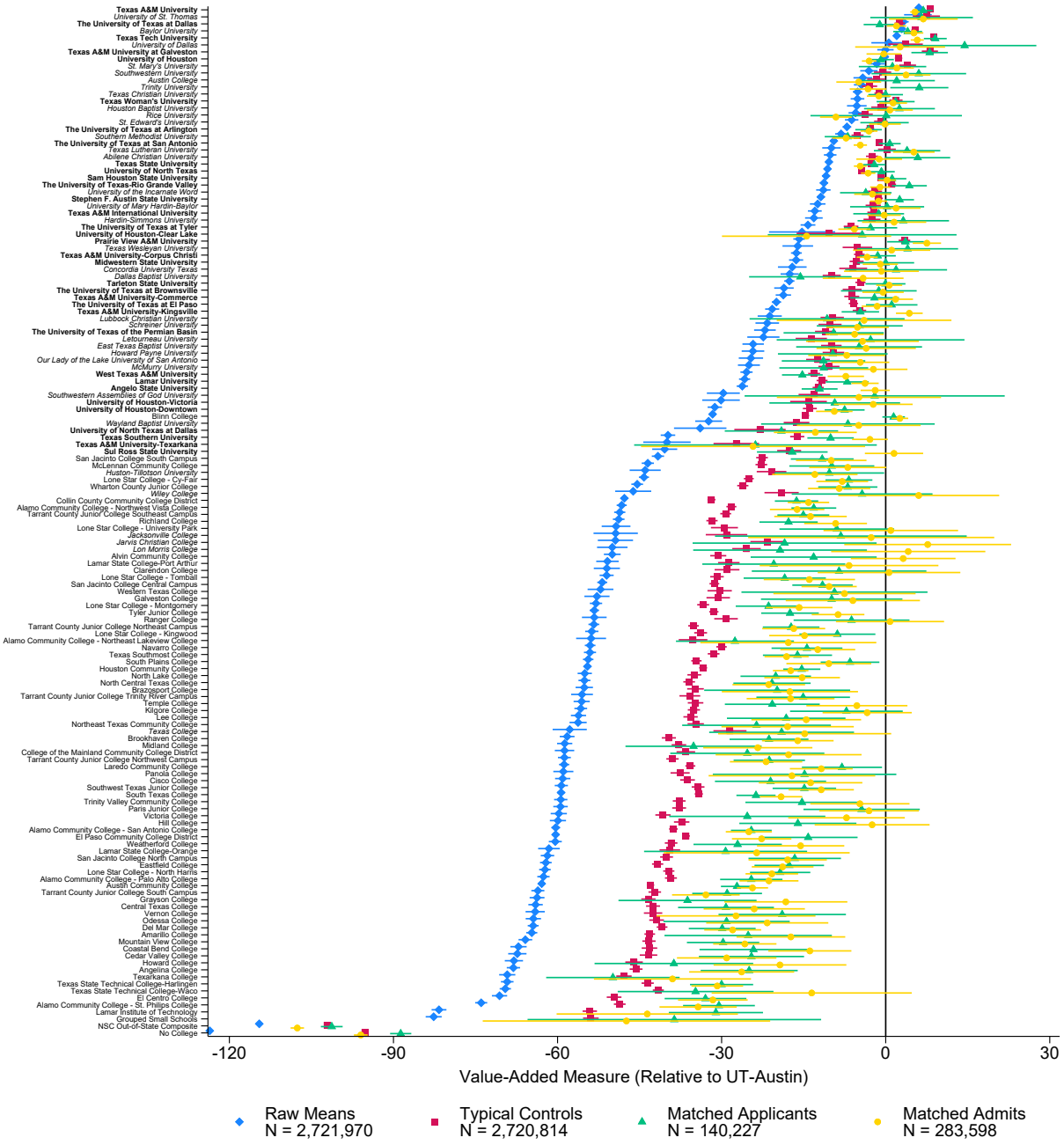
The four methods generate substantially different schedules of college VAMs. The Raw Mean outcomes, visualized on the far left of each plot, differ dramatically across colleges: those at the bottom have 60 percentage point lower degree completion rates and $30,000 lower mean earnings compared to those at the top. 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.[10] The outcomes that each college's incoming students are predicted to have, based entirely on their pre-college characteristics, follow a similar pattern as the actual outcomes. This clearly illustrates how different colleges enroll systematically different types of students, with such selection accounting for much of the raw outcome differences across colleges.

The Typical Controls VAMs, of course, mechanically balance students' observable 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

[9] We use the "sandwich" earnings measure common in analyses of quarterly state earnings data (e.g., Sorkin, 2018; Card et al., 2025), which keeps the quarters with positive earnings that are "sandwiched" between other quarters with positive earnings, thus ignoring the high-variance transition quarters between spells with earnings and spells without earnings.

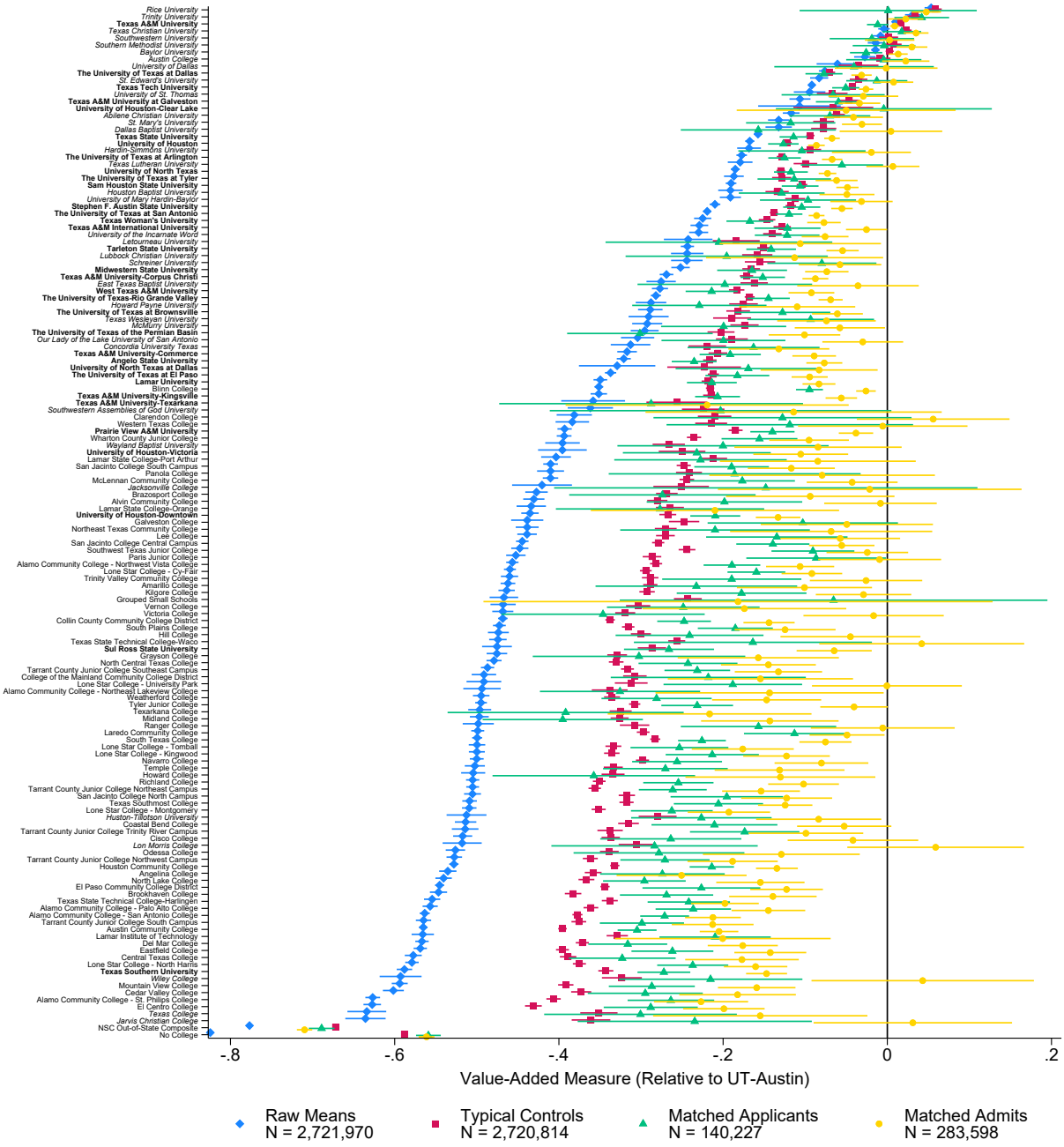
[10] Specifically, we use the predicted values from an OLS regression of earnings on cubic polynomials in 10th grade math and reading scores, attendance, days of disciplinary suspension, and number of advanced high school courses, along with categorical indicators for free/reduced price lunch eligibility,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

Figure 1: Value-Added Measures: Cumulative Credits



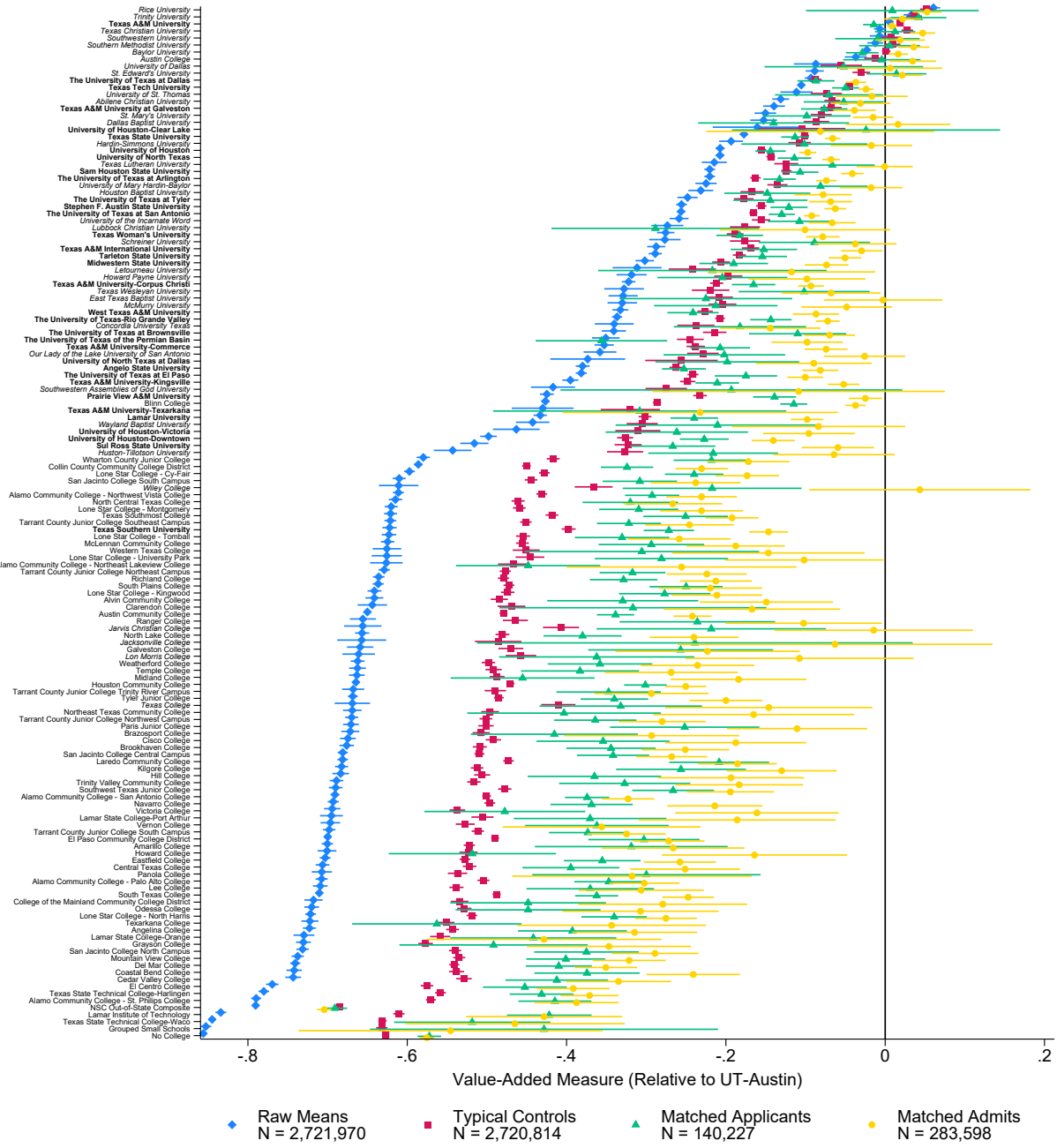
Notes: Each point and 95% confidence interval represents one college's estimated value-added relative to the omitted reference school (UT-Austin, represented by the vertical line at zero). Public four-year universities' names are in bold, private institutions are in italics, and community/technical colleges are in normal font. Different markers denote the different VAM estimation methods described in the text. Cumulative credits are measured ten years after high school graduation.

Figure 2: Value-Added Measures: Any Degree Completion



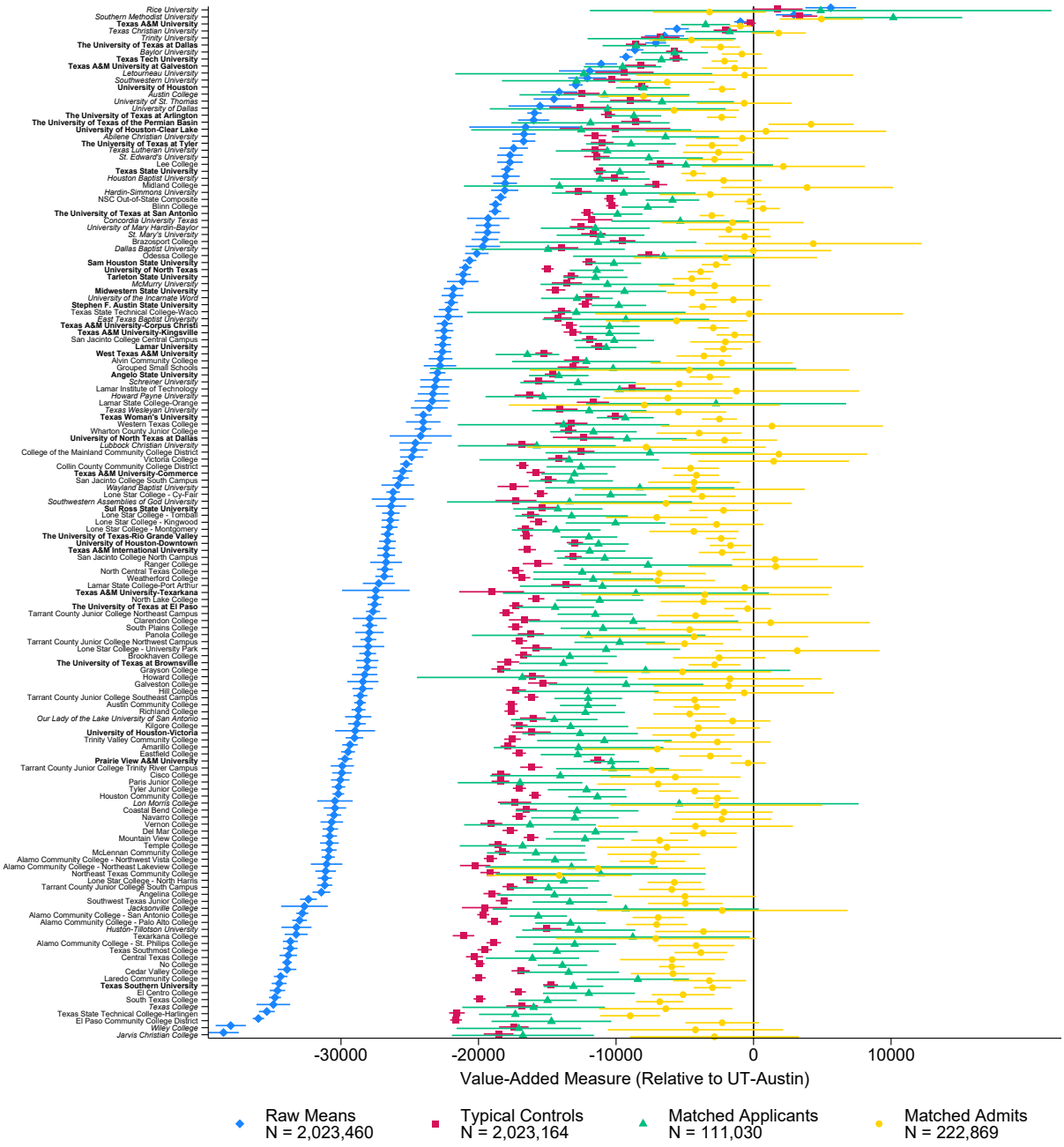
Notes: Each point and 95% confidence interval represents one college's estimated value-added relative to the omitted reference school (UT-Austin, represented by the vertical line at zero). Public four-year universities' names are in bold, private institutions are in italics, and community/technical colleges are in normal font. Different markers denote the different VAM estimation methods described in the text. Degree completion is measured ten years after high school graduation.

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



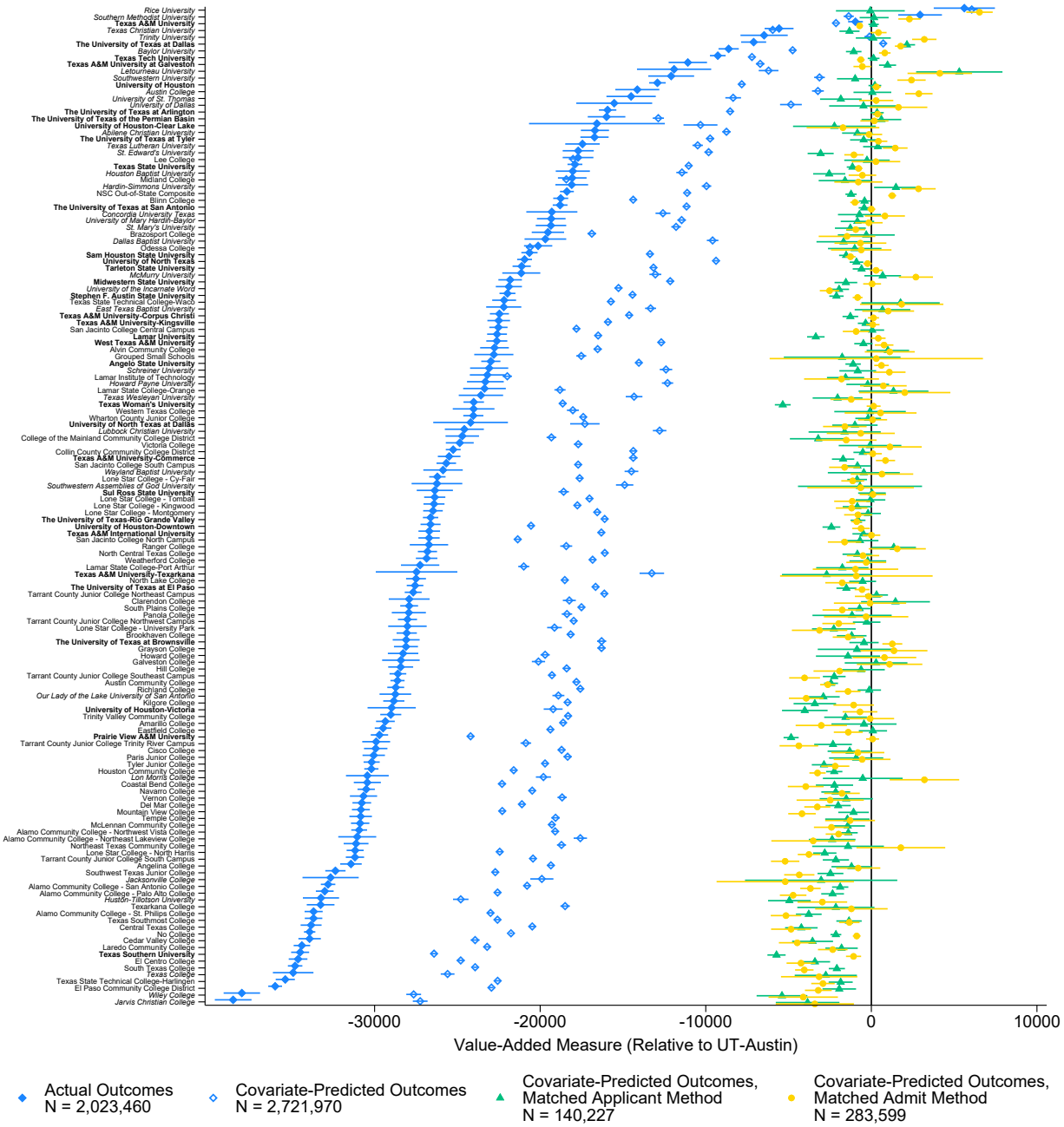
Notes: Each point and 95% confidence interval represents one college's estimated value-added relative to the omitted reference school (UT-Austin, represented by the vertical line at zero). Public four-year universities' names are in bold, private institutions are in italics, and community/technical colleges are in normal font. Different markers denote the different VAM estimation methods described in the text. Degree completion is measured ten years after high school graduation.

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). Public four-year universities' names are in bold, private institutions are in italics, and community/technical colleges are in normal font. Different markers denote the different VAM estimation methods described in the text. Earnings are measured in 2015 USD and averaged within individual over years 10-12 since high school graduation.

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). Public four-year universities' names are in bold, private institutions are in italics, and community/technical colleges are in normal font. The solid diamonds plot the actual Raw Mean outcomes at each school, while the hollow diamonds replace each student's actual outcome with their covariate-predicted outcome based on 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 graduation year. The remaining two markers plot the placebo value-added of each college on these predicted outcomes using the Matched Applicant and Matched Admit methods as a balance test.

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

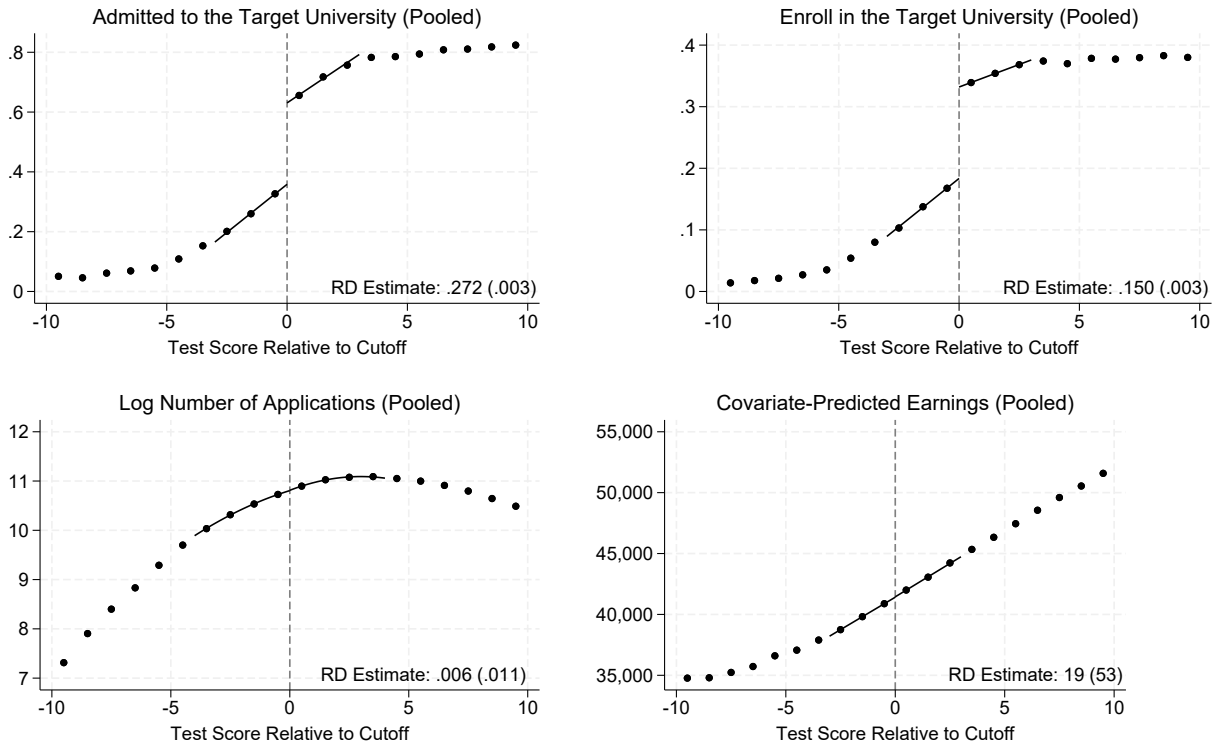
4.1 Admission Discontinuities

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 during our sample.[11] Figure 6 illustrates these discontinuities, pooled over all 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 when crossing that university's 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 identical at baseline.

[11] Figure A.1 shows the diverse and diffuse distribution of selectivity across these 28 schools. The two largest schools (University of North Texas and University of Houston) each contribute only 11 percent of the total mass of cutoff compliers, ensuring that our RD results are not driven by any single school.

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 with cutoffs. 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.

Figure 7 visualizes where the 28 target schools with admissions cutoffs lie within the entire spectrum of Texas postsecondary institutions, ranked according to their Raw Mean earnings. The figure also visualizes the weighted distribution of marginal applicants' fallback options, i.e. the schools they attend (including no college) instead of the target school 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 quasi-experimental enrollment variation generated by the admission cutoffs, with that breadth ensuring that our evaluation exercise is not focused on only limited portions of the college VAM schedules.

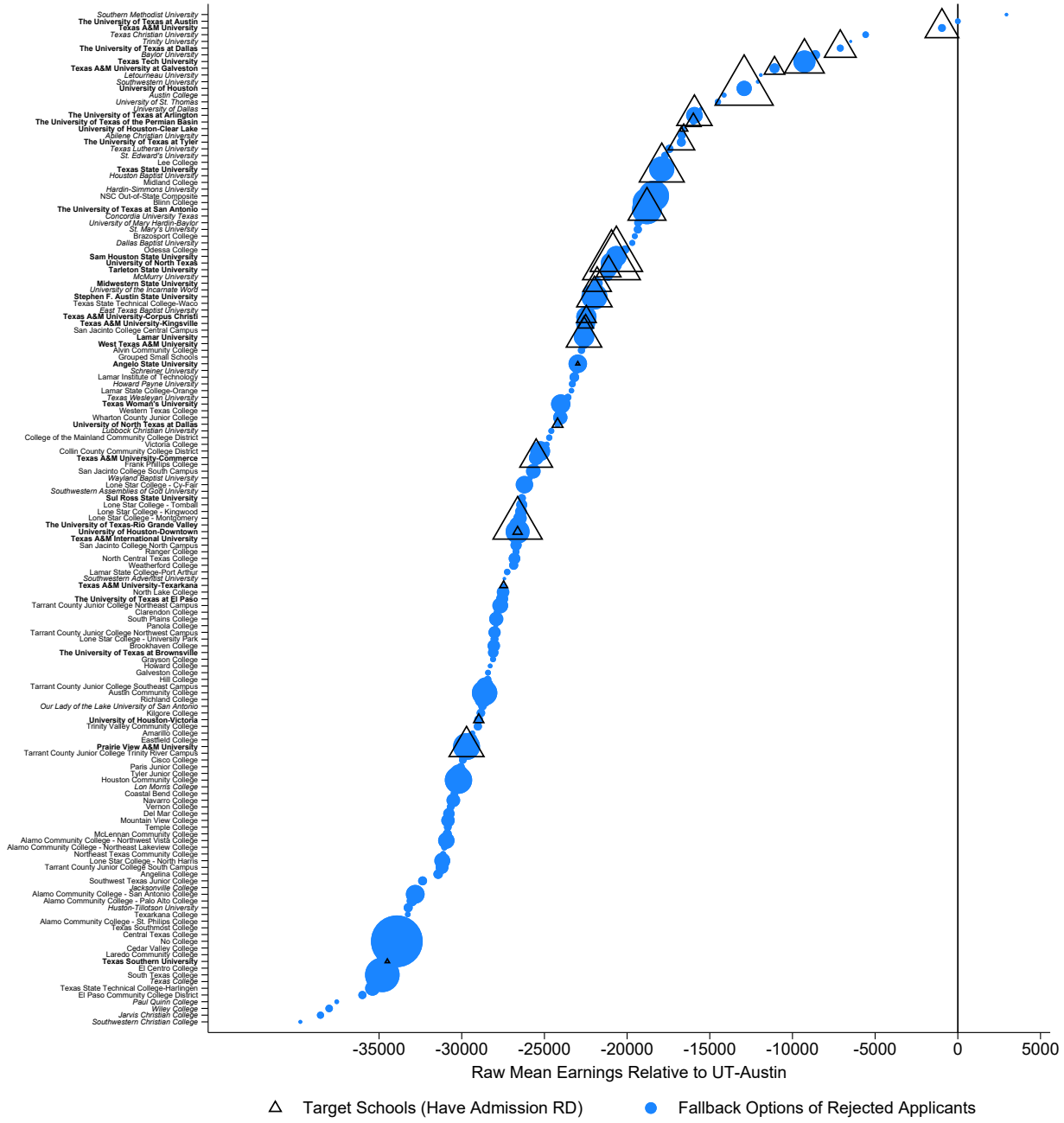
Going forward, we group the nine smallest target schools into a single target for the target-specific RD estimates, yielding 19 binary cutoff-crossing instruments. Our implementation of the RD design uses the same estimation sample and main specification as Mountjoy (2026); we simply update the earnings outcome to include newly available quarters. The regression includes 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.[12] Standard errors are clustered at the applicant level, following Kolesár and Rothe (2018). See Mountjoy (2026) for more details and robustness across a battery of alternative specifications.

4.2 The VAM Prediction Test

We adapt the K-12 school VAM testing framework of Angrist et al. (2017) to evaluate each observational VAM method against the quasi-experimental RD variation, asking whether the observational 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 some applicants into enrolling at the target school and out of their next best alternatives (fuzzy RD compliers). These exogenous enrollment shifts affect those students' eventual outcomes like degree attainment and earnings. For each observational VAM method, we can assess the VAMs' ability to predict these effects by

[12] As noted in Mountjoy (2026), the running variable (especially for ACT score submitters) is discretely distributed, so data-driven methods of optimal bandwidth determination and inference that require a continuous running variable like Calonico et al. (2014) may be inappropriate. In practice, Mountjoy (2026) shows that the bandwidths selected for each outcome by the Calonico et al. (2014) method, acting as if the running variable were continuous, are always very close to the main specification bandwidth of three concorded ACT points, and the RD estimates are similar across a wide range of alternative bandwidths.

Figure 7: 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. Public four-year universities' names are in bold, private institutions are in italics, and community/technical colleges are in normal font.

replacing each student's actual outcome Y_i with the VAM 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 $\widehat{\text{VAM}}_i$ is the VAMs' collective prediction about how student outcomes should change. We then compare that prediction with the effect of cutoff-crossing on students' actual outcomes Y_i. If the VAMs are accurate, these two effects should be equal at every target school, up to sampling error.

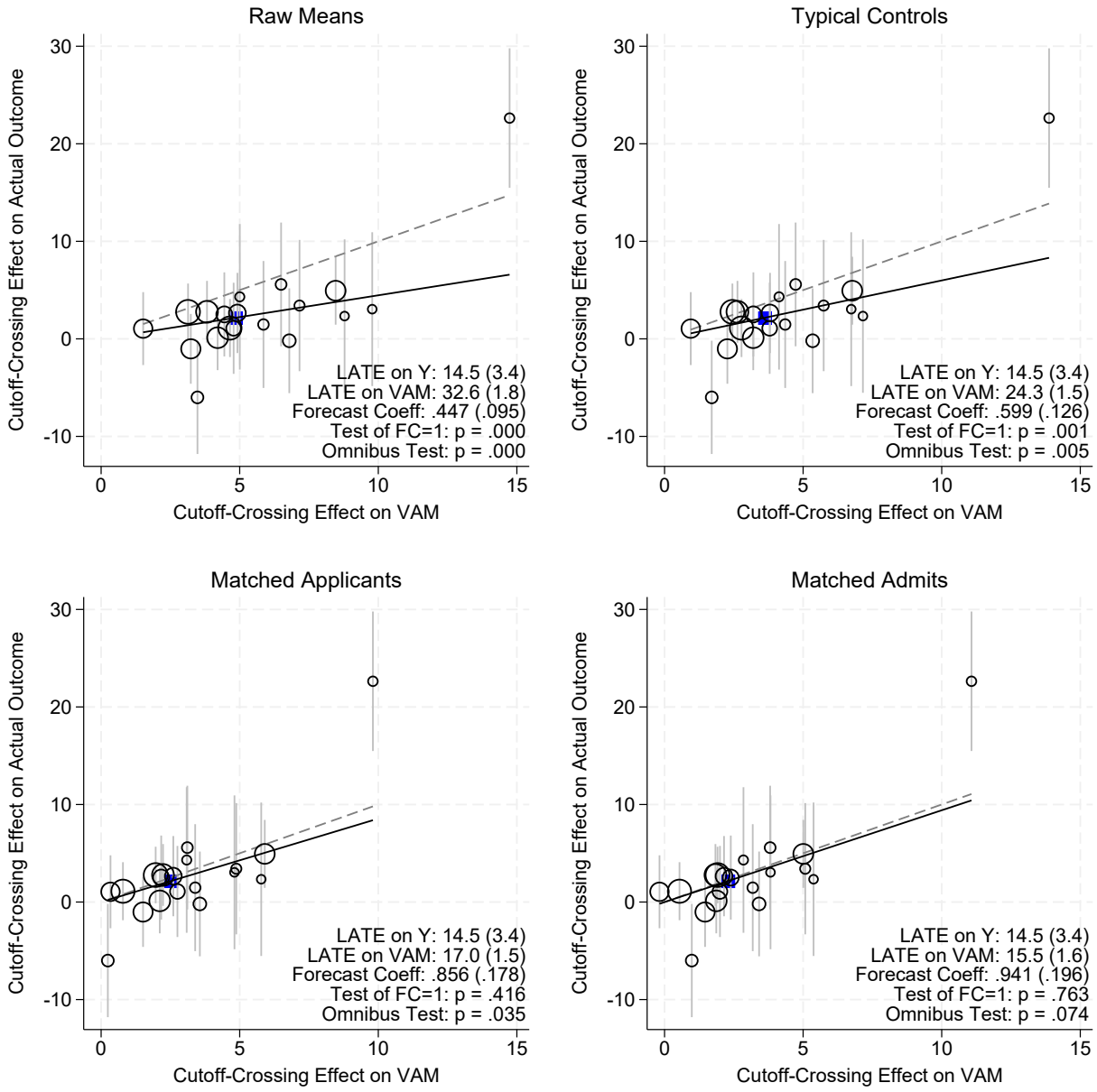
We assess VAM accuracy both visually and formally. Visually, in Figures 8–11, for each target university with an admission cutoff, we plot the reduced-form effect of crossing that cutoff on applicants' actual outcomes Y_i on the vertical axis. On the horizontal axis is the reduced-form effect of crossing the same cutoff on $\widehat{\text{VAM}}_i$. VAM accuracy is the extent to which these points lie along the 45-degree line, up to sampling error. Formally, we test this null hypothesis with the Angrist et al. (2017) omnibus test, which involves 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. We also estimate the forecast coefficient from a 2SLS regression of Y_i on $\widehat{\text{VAM}}_i$ and the RD controls, instrumenting for $\widehat{\text{VAM}}_i$ with a single binary instrument that indicates cutoff-crossing at the school targeted by the application.[13] This forecast coefficient is nearly equivalent to the ratio of the pooled LATE on Y_i—the actual effect of enrolling in the target school among cutoff-compliers—and the pooled LATE on $\widehat{\text{VAM}}_i$—the VAMs' prediction of that effect—which we also report in the figures.

4.3 Test Results

Figures 8–11 present the test results for each outcome, with the four observational VAM methods displayed as sub-panels within each figure. Each circle represents one target school with an admission cutoff (with the smallest schools grouped together), 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.

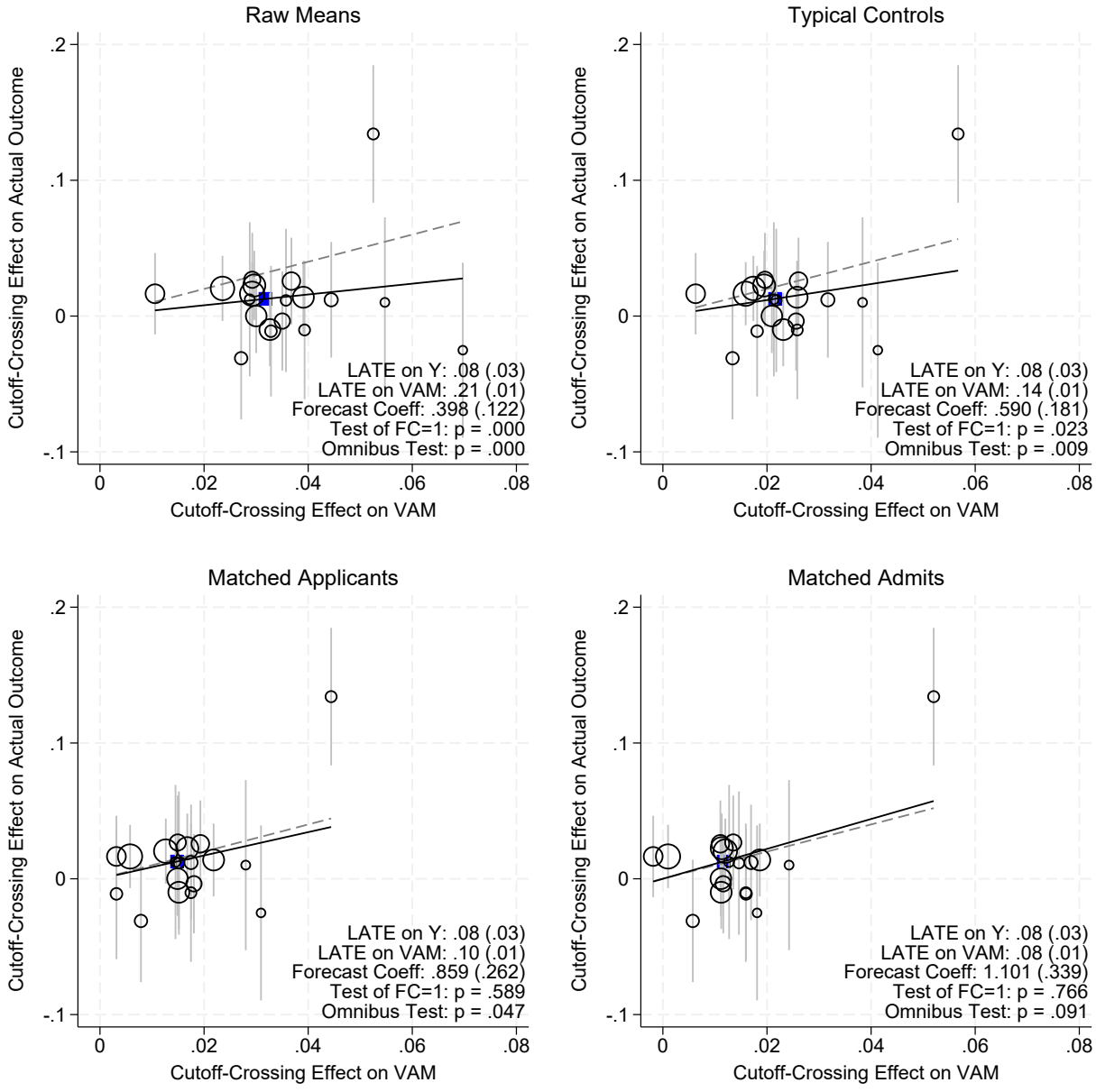
[13] We estimate similar forecast coefficients with similar precision when using all 19 cutoff-crossing instruments, instead of one pooled instrument, in the UJIVE procedure of Kolesár (2013).

Figure 8: VAM Prediction Tests: Cumulative Credits



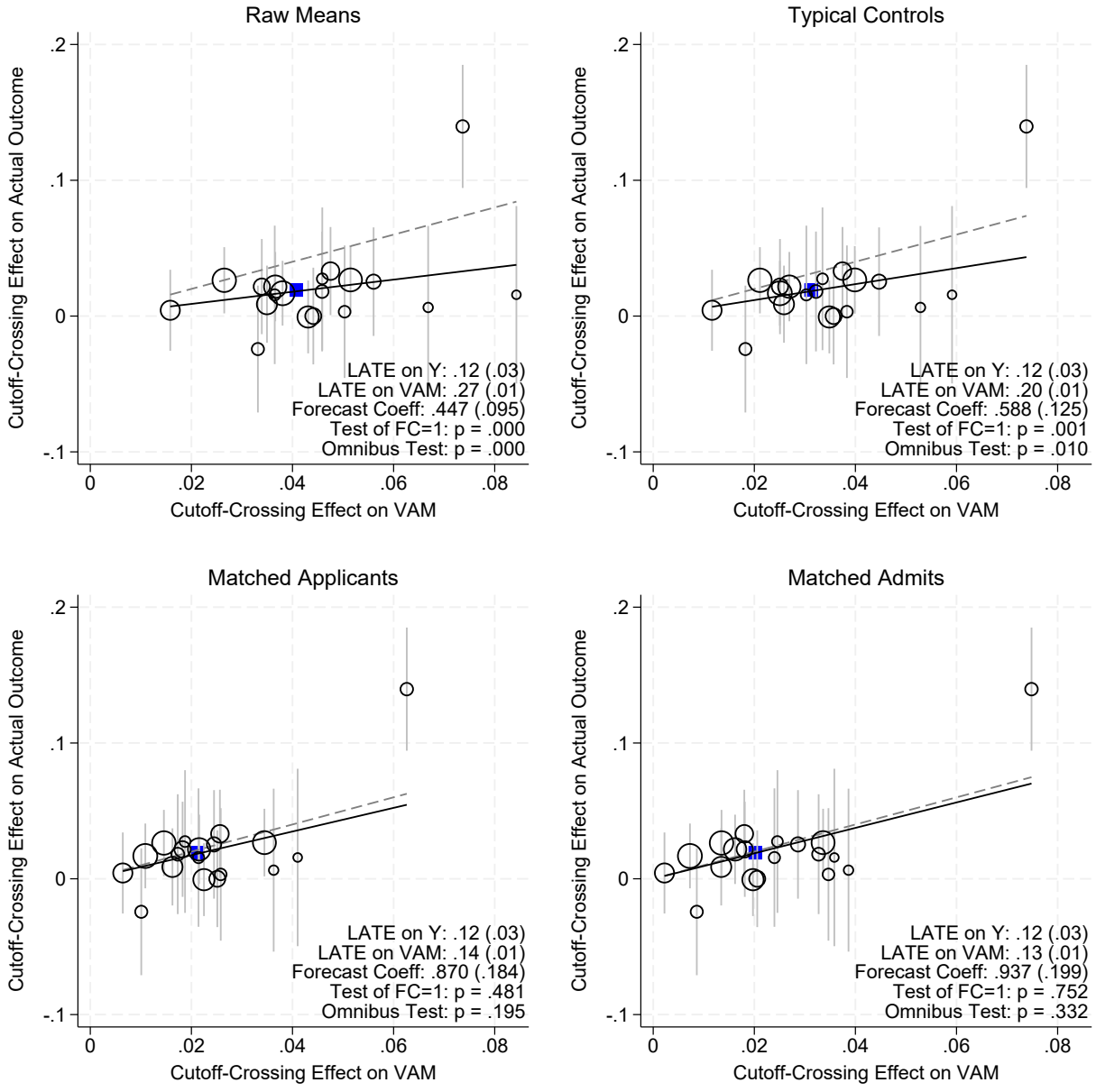
Notes: Each panel displays the VAM prediction test for a different VAM estimation method. Each point represents a target university with an admission cutoff. The vertical axis value is the reduced-form effect of crossing that target university's admission cutoff on the actual outcome. The horizontal axis value is the reduced form effect of cutoff-crossing on the VAM of the college enrolled in (including no college). The circles are sized by the precision of the vertical axis reduced form. The dashed line is the 45-degree line, corresponding to perfect forecast. The solid line is the estimated forecast line. The LATEs reported in the captions come from a pooled 2SLS regression across all the target universities with admission cutoffs, instrumenting for enrolling at the applicant's target university with a single cutoff-crossing indicator. The omnibus test involves 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. N applicants = 234,271.

Figure 9: VAM Prediction Tests: Any Degree



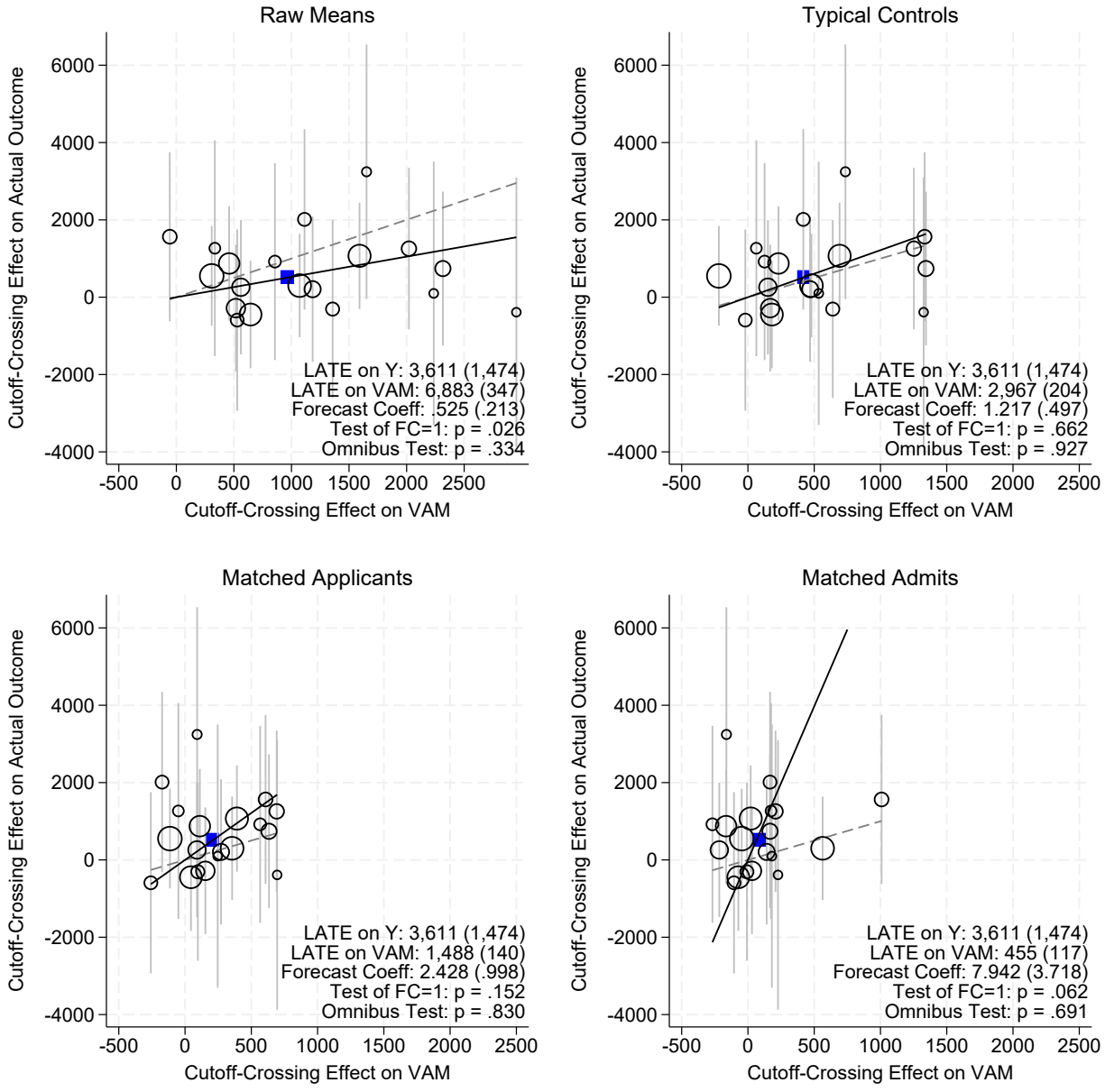
Notes: See the notes to Figure 8.

Figure 10: VAM Prediction Tests: Bachelor's Degree



Notes: See the notes to Figure 8.

Figure 11: VAM Prediction Tests: Earnings



Notes: See the notes to Figure 8. N applicants with earnings = 188,275.

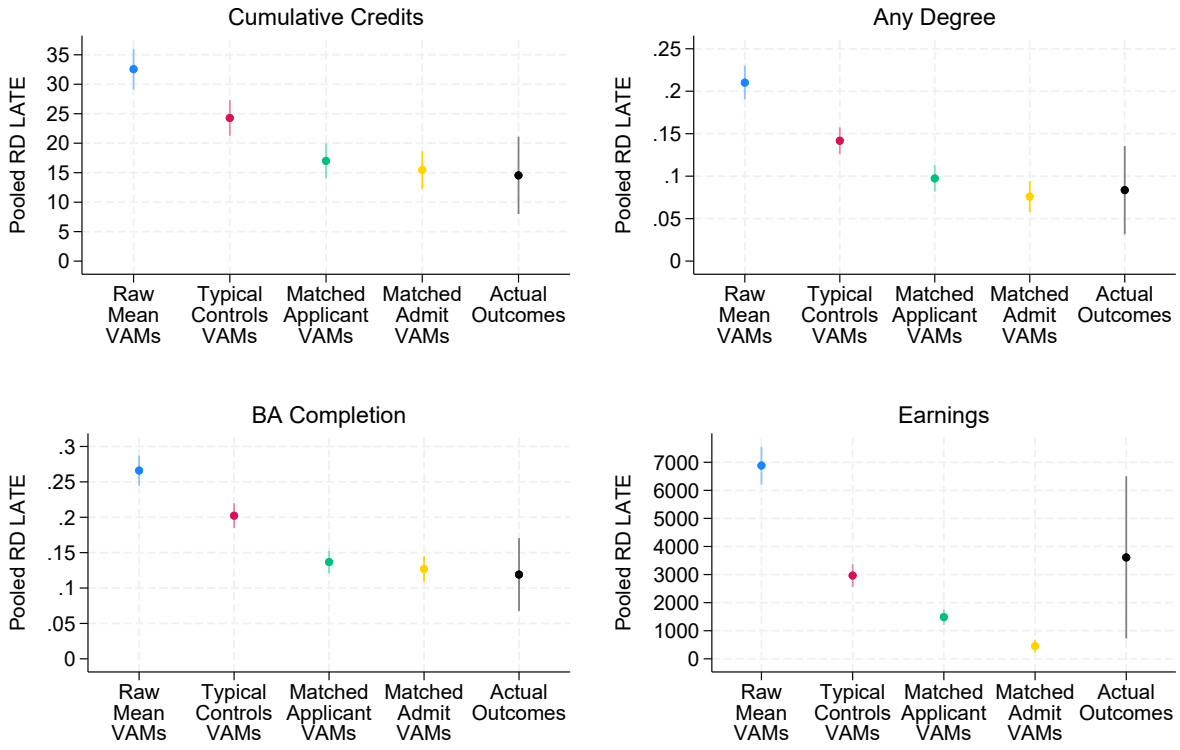
We find four main patterns of test results. First, the Raw Means in the top left panel of each figure consistently deliver forecast coefficients around 0.4-0.5 across all outcomes. Each one-unit increase in $\widehat{\text{VAM}}_i$ is thus associated with only a 0.4-0.5 unit increase in Y_i, implying that the Raw Mean VAMs overpredict differences in causal impacts across colleges by roughly a factor of two. This pattern is consistent with substantial selection bias in popular raw 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 0.6. This is a modest improvement over the Raw Means, but consistent with typically observable covariates being insufficient controls for addressing differential selection on educational attainment potential across colleges. Figure A.2 shows only slightly improved test results when moving from the additive Typical Controls model to the version that interacts all of the covariates with all of the college treatment options. Allowing for match effects between colleges and observable student characteristics thus does not meaningfully improve the ability of the Typical Controls VAMs to predict educational attainment gains for admission cutoff compliers.

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 but has even higher actual impacts, driving the omnibus test to 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. Unlike all of the other target colleges, Prairie View A&M's actual causal impacts on educational attainment are consistently *larger* than what the available observational VAM methods would predict. Figures A.3-A.6 show that excluding Prairie View A&M from the set of RD target schools (by dropping applications to it from the RD sample) leads the omnibus test to fail to reject the accuracy of the Matched Applicant and Matched Admit VAMs for predicting effects on educational attainment, while still rejecting the accuracy of the Raw Means and Typical Controls VAMs.

Finally, for earnings, the test results deliver a rather different pattern across VAM methods compared to educational attainment. The Raw Means still substantially overpredict causal impacts by roughly a factor of two, but the Typical Controls VAMs now have a forecast coefficient indistinguishable from one, both in Figure 11 and for the version with college-covariate interactions in Figure A.2. The more stringent VAMs, meanwhile, *under-*

Figure 12: Evaluating College VAMs: Summarizing with Pooled LATEs



Notes: Each panel summarizes the VAM prediction test results for a different outcome. Each point is an RD LATE of enrolling in a target university among cutoff compliers, pooled across all target universities with admission cutoffs. The far right point in each panel uses the applicant's actual outcome, while the other points replace the applicant's actual outcome with the observational VAM of their initial college (including no college).

predict earnings gains. The Matched Applicant VAMs predict a roughly $1,500 gain for cutoff compliers, but the LATE on actual earnings is nearly $3,700, delivering a noisy forecast coefficient of 2.4. More strikingly, the Dale and Krueger (2002)-inspired Matched Admit VAMs predict only a $450 gain relative to the actual gain of $3,700, delivering a large forecast coefficient of roughly 8 that is statistically distinguishable from 1 ($p = .062$). Figure A.7 confirms this statistical distinction with more power by doubling the RD bandwidth from 3 concorded ACT points to 6—which Mountjoy (2026) shows delivers very similar RD LATE point estimates but greater precision—yielding a forecast coefficient of roughly 5.4 that is distinguished from 1 at $p = .014$. Figures A.8 and A.9 show that these test results remain very similar when measuring individual earnings in logs and ranks rather than dollars.

Figure 12 summarizes these results by visualizing all the pooled RD estimates together in the same plot. The far right point in each panel uses the applicant's actual outcome, while the other points replace the applicant's outcome with the observational VAM of their

initial college (including no college). For all outcomes, the Raw Mean VAMs consistently overpredict the actual outcome effects by roughly a factor of two. The Typical Controls, Matched Applicant, and Matched Admit VAMs in turn deliver progressively smaller predicted effects, with that trend successfully arriving at the actual outcome effect for the three educational attainment outcomes, but not for earnings. For earnings, we cannot reject the accuracy of the Typical Controls and Matched Applicant VAMs, while the Dale and Krueger (2002)-inspired Matched Admit VAMs significantly underpredict the actual earnings effects by nearly an order of magnitude.[14] We investigate the drivers of this phenomenon in Section 5.

5 Learning about Selection

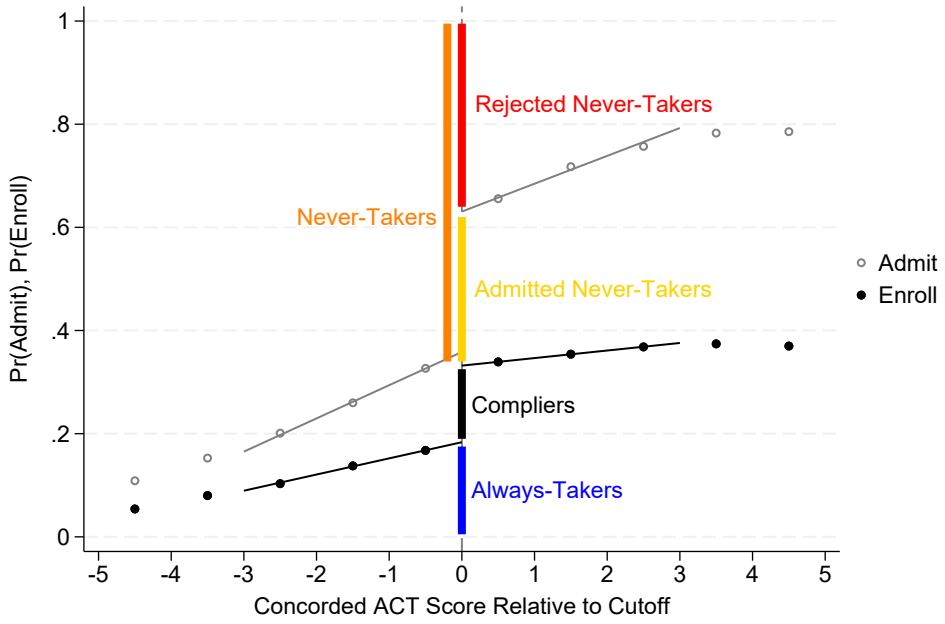
The results of Section 4 show that popular comparisons of raw outcome means across colleges consistently overpredict quasi-experimental effects, empirically confirming that "better" schools enroll better students. The VAMs produced by the stringent Matched Applicant and Match Admit methods, on the other hand, perform well for educational attainment outcomes but underpredict earnings effects, especially the Dale and Krueger (2002)-inspired Matched Admit 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. We find evidence consistent with admitted students self-selecting into college choices based on their own idiosyncratic gains, with these gains mediated by major choice and industry of employment.

5.1 What Types of Students Are Compared in Each Method?

One common conjecture about the divergence between large earnings effects of college selectivity for admission RD cutoff-crossers versus small effects for Dale and Krueger (2002)-style 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 groups. To ensure a level playing field when comparing these methods, we focus on the subpopulation of students whose SAT/ACT scores and application decisions render them on the cusp of 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.

[14] See Figure A.10 for the similar but more precise LATEs from doubling the RD bandwidth. Figure A.11 plots the dynamic earnings LATEs each year since application, showing that the main patterns across methods become apparent by year 7 and remain stable through year 12, the last year available for all cohorts.

Figure 13: Response Types at the RD Cutoff



Notes: This figure visualizes the types of marginal students present at the RD admission cutoff.

To characterize the types of marginal students who are compared in each method, consider the different "response types" around the cutoff, visualized in Figure 13. *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 less selective alternative. *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 at the target school if below its cutoff but not if above.

Which of these types are involved in each method? 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 applicant's less selective alternative, which we'll refer to as "untreated" ($T = 0$).

In contrast, the Hoxby (2015)-style 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:

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

Finally, the Dale and Krueger (2002)-style 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)-style Matched Admits are Admitted Never-Takers who have the opportunity to attend the target school, but voluntarily decline. The untreated Hoxby (2015)-style Matched Applicants, meanwhile, are a mixture of these types along with Rejected Never-Takers. To the extent that these theoretically distinct types of marginal students have empirically different potential outcomes, these methods will compare different quantities and potentially induce different conclusions about the value-added of more selective colleges.

5.2 Describing the Students Compared in Each Method

Having theoretically established the distinct types of marginal students compared in each method, we now use the main RD specification (pooled across target schools with one cutoff-crossing instrument and a binary treatment indicator for attending the application's target school) and the identification results of Abadie (2002) to empirically describe the potential outcomes of these types, unpacking the sources of disagreement between the methods and

revealing interesting selection patterns.

In Figure 14, we plot the mean potential outcomes of each type of marginal applicant involved in the RD LATE and Matched Admit comparisons. Figure A.12 adds the Matched Applicant comparison. Each method compares a group of treated students who enroll in the more selective target school, depicted on the inner right side of the plot, to a group of untreated students who enroll in their less selective alternative to the target school, depicted on the inner left side of the plot. In the outer points, we compare marginal students' own outcomes to the mean outcomes of their observationally identical peers at the same institution (including no college), with peers in the target school (treated state) on the far right and peers in the less selective alternative (untreated state) on the far left.[15]

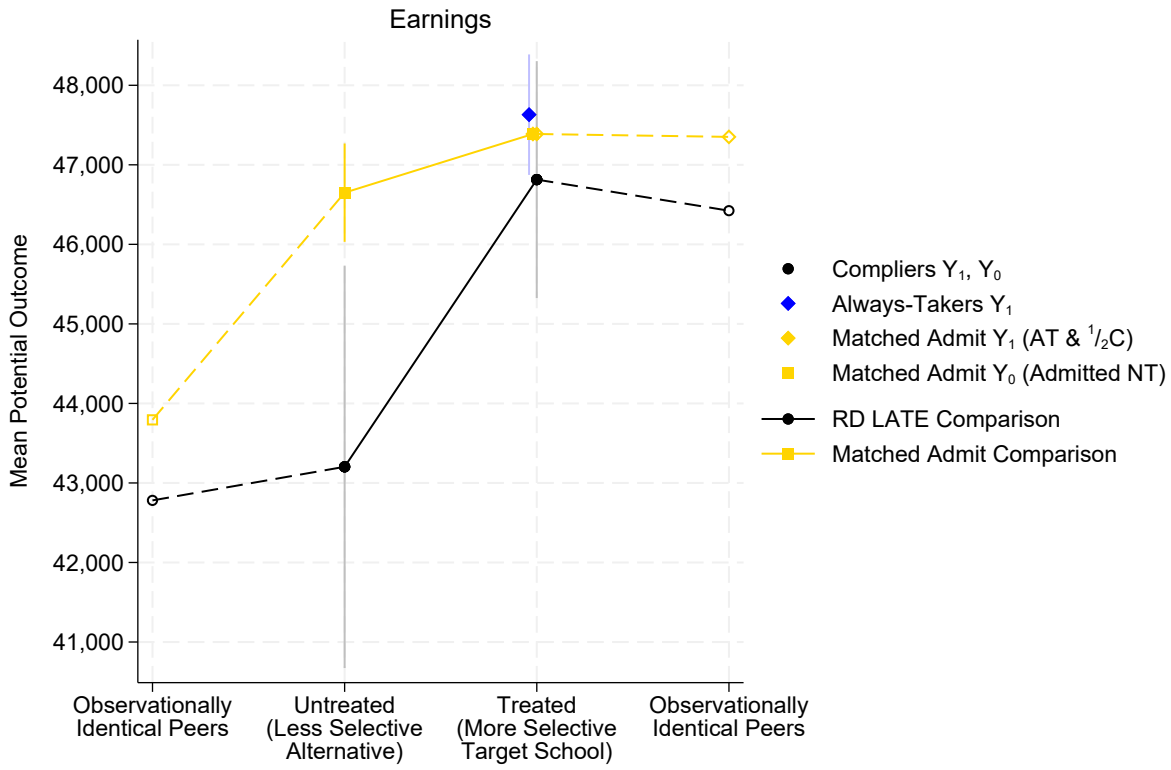
Four notable patterns emerge. First, all of the treated types experience fairly similar earnings outcomes when attending the target school, as seen in the inner right "treated" column. Second, the inner left "untreated" column shows that the untreated students in the Matched Admit comparison, i.e. the Admitted Never-Takers, have substantially higher mean earnings than untreated Compliers. That is, admitted marginal students who have the opportunity to attend a more selective target school, but voluntarily decline, earn substantially more at their less selective choice compared to the marginal applicants who involuntarily end up in that same untreated state by getting rejected. Third, the outer left comparison shows that the Matched Admit selectivity "decliners" (Admitted Never-Takers) also have substantially higher mean earnings relative to their *observationally identical peers* at their less selective choice. Matched Admit decliners thus have idiosyncratically high earnings potential at their less selective choice chiefly along *unobservable* dimensions not captured by typical pre-college covariates like test scores and demographics. Fourth, Figure A.12 adds the Matched Applicants, whose mean untreated potential outcome lies between the untreated outcomes of Compliers and Matched Admit decliners (Admitted Never-Takers), as they are a mixture of these two types and Rejected Never-Takers.

5.3 Mechanisms and Discussion

The results above provide novel evidence consistent with Dale and Krueger (2002)'s unexplored speculation that "the students who chose to attend a less selective school even though

[15]Observationally identical peers' outcomes are the predicted values from a regression of the earnings of all high school graduates on all college treatment indicators interacted with all observable 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.

Figure 14: Unpacking the Methods and Revealing Self-Selection



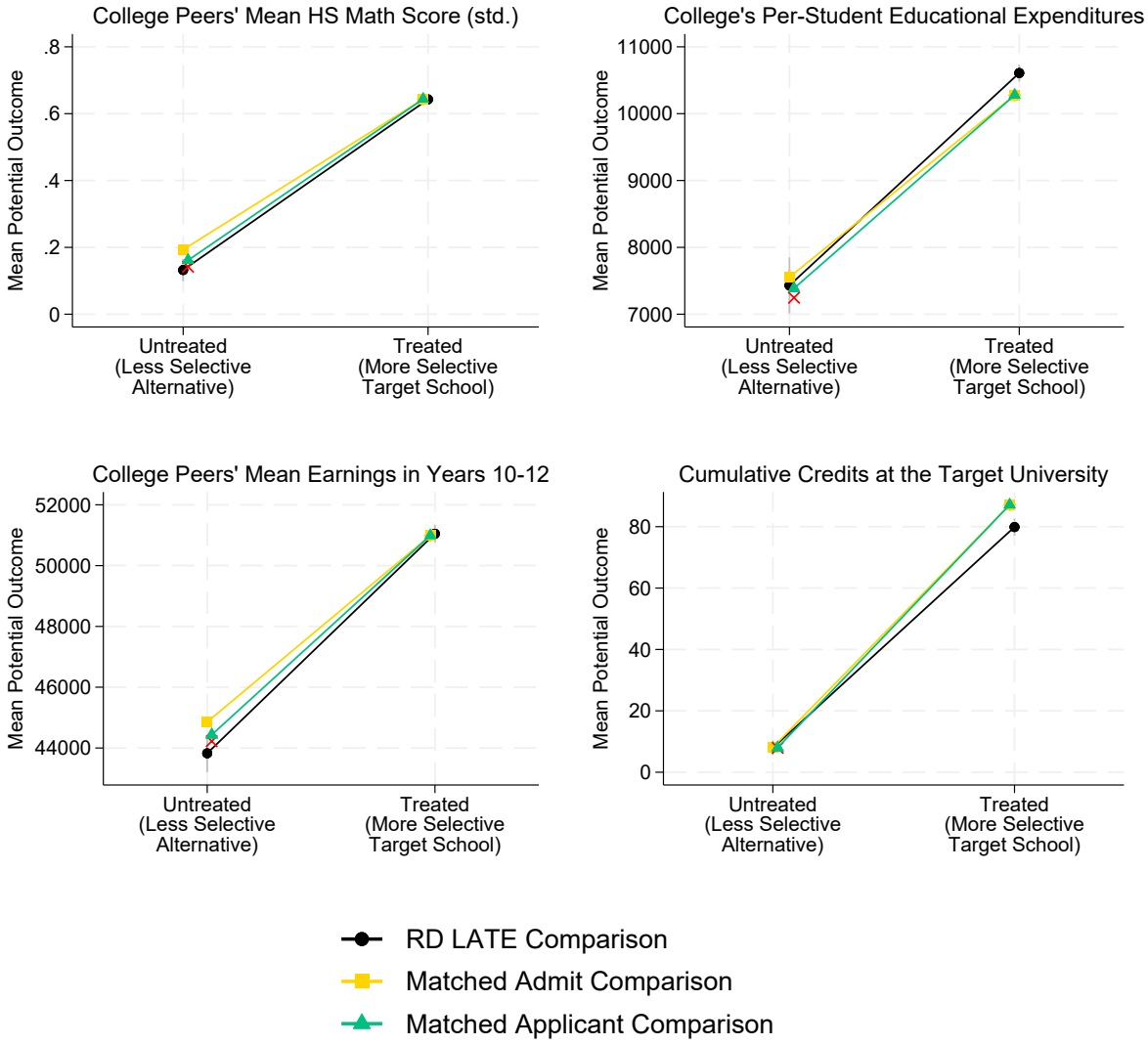
Notes: This figure plots the mean potential outcomes of each type of marginal applicant involved in the RD LATE and Matched Admit comparisons. The middle two points on the horizontal axis distinguish between marginal students' own outcomes in the treated state on the right (enrolling at the target school) versus the untreated state on the left (enrolling at the student's less selective alternative to the target school). The outer two points of the horizontal axis show the mean outcomes of marginal students' observationally identical peers in the treated state on the right versus in the untreated state on the left. Observationally identical peers' outcomes are the predicted values from a regression of earnings on all college treatment indicators interacted with all observable 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. Vertical lines are 95% confidence intervals around each point estimate.

they were admitted to a more selective one might be the students with attributes that lead them to benefit more from attending a less selective school" (p. 1524). Our results also show, however, that such "attributes" are not well captured by any of the pre-college observables typically available in administrative data, complicating potential efforts to tailor college VAM schedules to different types of prospective students. Nevertheless, turning to potential mechanisms, we show below that completed majors and industries of employment may be identifiable mechanisms driving this phenomenon, as Matched Admit decliners end up in more lucrative majors and industries than their peers.

First, though, we rule out the explanation that the Dale and Krueger (2002)-style Matched Admit comparison simply involves a weaker treatment "dosage" of college quality relative to the other methods. In the first three panels of Figure 15, we plot treated-untreated contrasts as in Figures 14 and A.12, but we replace each student's earnings outcome with quality measures of their initial college (including none). The contrasts are large and, importantly, very similar across methods: all three of the RD LATE, Matched Admit, and Matched Applicant comparisons involve very similar levels and changes in institutional selectivity (as measured by peer average high school test scores), per-student educational spending, and peer mean earnings. The final panel of Figure 15 shows similar treated-untreated contrasts in cumulative credits ever completed specifically at the target school, with untreated students completing close to zero credits there across all three methods. This rules out the possibility that the Matched Admit earnings effect is diluted by initially untreated students later transferring into the more selective target school. Thus, the very small estimated earnings effect in the Dale and Krueger (2002)-style Matched Admit comparison is not attributable to weaker treatment dosage along these dimensions.

In Figure 16, we turn to degrees, majors, and industries as potential mechanisms. The top left panel reproduces Figure 14's pattern of self-selection on earnings outcomes. The remaining panels of Figure 16 all repeat the same exercise but replace each student's actual earnings outcome with degree, major, and industry "premia" to explore the extent to which students' actual earnings patterns are explained by average earnings in the degrees they complete, the majors they study, and the industries that employ them. We estimate these observational premia as fixed effects from regressions of the earnings of all high school graduates on mutually exclusive and exhaustive degree/major/industry categories, with controls for high school cohort graduation year. Because these fixed effects are measured relative to an omitted category, the levels on the vertical axes of the remaining panels of Figure 16 are different from the actual earnings in the top left panel, but the scales (in dollars) are comparable. In Figure A.13, we show that all of the results in Figure 16 are similar when we add the set of Typical Controls to the regressions that estimate the degree/major/industry

Figure 15: Similar Treatment Dosage across Methods



Notes: This figure plots the treated-untreated contrasts in college quality and cumulative exposure to the target school experienced by the marginal students compared in the RD LATE, Matched Applicant, and Matched Admit methods. The horizontal axis distinguishes between the treated state (enrolling at the target school) versus the untreated state (enrolling at the less selective alternative). Solid lines represent the treated vs. untreated comparisons involved in each method. Vertical lines are 95% confidence intervals around each point estimate.

premia fixed effects. Figure A.14 adds the Matched Applicant comparison to Figure 16 for completeness.

The top right panel of Figure 16 begins by replacing each student's actual earnings outcome with the premium associated with their highest completed postsecondary degree: none, certificate, associate's degree, bachelor's degree, or graduate degree.[16] These simple degree premia can explain some of the earnings gains in both the RD LATE comparison and the Matched Admit comparison, but they do so in a largely parallel way, thus failing to explain why the actual Matched Admit earnings gain is so much smaller than the RD LATE.

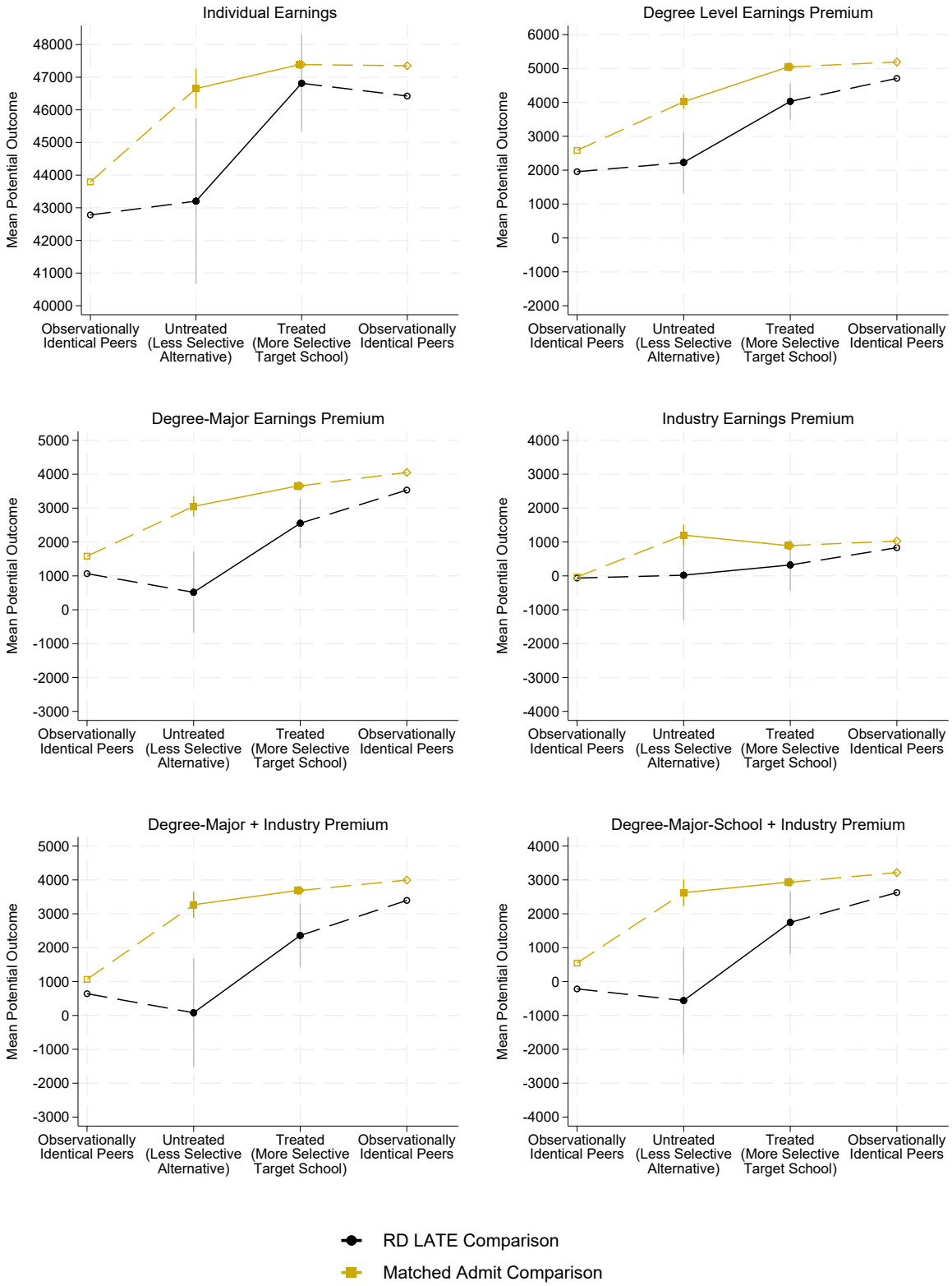
The middle left panel of Figure 16 adds majors. The outcome is now the premium associated with the student's 4-digit CIP major code interacted with their highest degree level, yielding roughly 800 categories (including a null major for non-completers). Here the pattern starts to resemble the actual earnings pattern in the top left, with a larger gap between the Matched Admit mean Y_0 and the RD Complier mean Y_0, and a larger RD LATE relative to the Matched Admit treated-untreated comparison.

The middle right panel of Figure 16 solely considers industry of employment. The outcome is the premium associated with the 6-digit NAICS industry in which the student earned the most income in years 10-12 after high school graduation, comprising roughly 1,200 categories. In this panel, the Matched Admit mean Y_0 stands out: Matched Admit decliners (Admitted Never-Takers) tend to work in higher-paying industries than all of the other types of marginal students.

In the bottom left panel of Figure 16, we consider majors and industries together. We estimate major and industry premia jointly from a regression of all high school graduates' earnings on the roughly 800 degree-major categories and the roughly 1,200 industry categories (additively), then replace each student's actual outcome with the sum of their estimated major premium and industry premium. The resulting figure in the bottom left further resembles the actual earnings patterns in the top left, suggesting that much of the self-selection pattern can be explained by Matched Admit decliners (Admitted Never-Takers) both completing more lucrative majors and working in more lucrative industries relative to their peers. The final panel in the bottom right shows little change when interacting the degree-major categories with indicators for the specific institution from which the highest degree was earned. This last result suggests that institution-specific major premia have little explanatory power to add beyond the main effects of major and industry premia.

[16] We also include a small sixth category of students who appear to have no degree but began college at an out-of-state institution, as these students may have completed a degree that we cannot observe in the THECB data.

Figure 16: Potential Mechanisms: Degree, Major, and Industry



Notes: The top left panel reproduces the selection pattern from Figure 14. The remaining panels replace the student's actual earnings outcome with degree, major, and industry "premia", themselves estimated as fixed effects from regressions of the earnings of all high school graduates on mutually exclusive and exhaustive degree/major/industry categories, with controls for high school cohort graduation year.

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 postsecondary landscape, and rich longitudinal 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 discontinuities, 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 raw mean outcomes, which are the most popular and prominent college value-added measures, consistently overpredict causal impacts on educational attainment and earnings by roughly a factor of two. Regression-adjusted VAMs for typically available controls like test scores and demographics also overpredict educational impacts, but accurately predict earnings gains. The more stringent VAMs that compare matched applicants and matched admits predict educational impacts well, but underpredict earnings gains, especially when comparing students choosing among the same realized admission offers.

To understand this divergence across methods and outcomes and learn about student selection behavior, we theoretically and empirically characterized the types of students involved in the observational and quasi-experimental comparisons. We found that students in the Dale and Krueger (2002)-style matched admit method who voluntarily decline a more selective admission offer tend to have unusually high earnings at their less selective choice, both relative to involuntarily rejected admission cutoff compliers and relative to the decliners' observationally identical peers at the same college, consistent with Roy (1951) self-selection on earnings potential that is not easily predictable by typical pre-college observables. We identified majors and industries as likely mediators, as the selectivity "decliners" end up completing more lucrative majors and working in higher-paying industries than their peers at their less selective choice. Together, these results offer new evidence on the prospects of credible college value-added measures and on the nature of student 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.

Abdulkadiroglu, Atila, Parag A Pathak, Jonathan Schellenberg, and Christopher R Walters (2020). "Do Parents Value School Effectiveness?" *American Economic Review* 110.5, pp. 1502–1539.

Abdulkadiroglu, Atila, Parag A. Pathak, and Christopher R. Walters (2026). "Who Gets What in Education: Can School Matching Improve Student Achievement?", NBER Working Paper No. 34936.

Angrist, Joshua, Peter Hull, Parag A. Pathak, and Christopher Walters (2024). "Credible School Value-Added with Undersubscribed School Lotteries". *The Review of Economics and Statistics* 106.1, pp. 1–19.

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.

Antman, Francisca M, Brian Duncan, and Michael Lovenheim (2024). "The Long-Run Impacts of Banning Affirmative Action in US Higher Education". *Oxford Review of Economic Policy* 40.3, pp. 607–628.

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.

Bau, Natalie (2022). "Estimating an Equilibrium Model of Horizontal Competition in Education". *Journal of Political Economy* 130.7, pp. 1717–1764.

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

— (2024). *Top Percent Policies and the Return to Postsecondary Selectivity*. Working Paper.

Bleemer, Zachary and Aashish Mehta (2022). "Will Studying Economics Make You Rich? A Regression Discontinuity Analysis of the Returns to College Major". *American Economic Journal: Applied Economics* 14.2, pp. 1–22.

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.

Bruhn, Jesse (2023). "The Consequences of Sorting for Measuring Educational Quality", Working paper.

Bruhn, Jesse M., Christopher Campos, Eric Chyn, and Ahn Tran (2025). "Who Benefits from Remote Schooling? Self-selection and Match Effects", NBER Working Paper No. 31542.

Calonico, Sebastian, Matias D Cattaneo, and Rocio Titiunik (2014). "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs". *Econometrica* 82.6, pp. 2295–2326.

Campos, Christopher, Pablo Muñoz, Alonso Bucarey, and Dante Contreras (2026). "College Major Choice, Payoffs, and Gender Gaps", NBER Working Paper No. 34736.

Card, David, Jesse Rothstein, and Moises Yi (2025). "Location, Location, Location". *American Economic Journal: Applied Economics* 17.1, pp. 297–336.

Carneiro, Pedro, James J. Heckman, and Edward J. Vytlacil (2011). "Estimating Marginal Returns to Education". *American Economic Review* 101.5275, pp. 2754–2781.

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, and Jonah E. Rockoff (2014). "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates". *American Economic Review* 104.9, pp. 2593–2632.

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.

Cohodes, Sarah R. and Joshua S. Goodman (2014). "Merit Aid, College Quality, and College Completion: Massachusetts' Adams Scholarship as an in-Kind Subsidy". *American Economic Journal: Applied Economics* 6.4, pp. 251–285.

Cornelissen, Thomas, Christian Dustmann, Anna Raute, and Uta Schönberg (2018). "Who Benefits from Universal Child Care? Estimating Marginal Returns to Early Child Care Attendance". *Journal of Political Economy* 126.6, pp. 2356–2409.

Cotton, Christopher S., Brent R. Hickman, John A. List, Joseph Price, and Sutanuka Roy (2026). "Why Don't Struggling Students Do Their Homework? Disentangling Motivation and Study Productivity as Drivers of Human Capital Formation". *Journal of Political Economy* 134.1, pp. 86–149.

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.

Dynarski, Susan, Brian Jacob, and Daniel Kreisman (2018). "How Important Are Fixed Effects and Time Trends in Estimating Returns to Schooling? Evidence from a Replica-

tion of Jacobson, Lalonde, and Sullivan, 2005". *Journal of Applied Econometrics* 33.7, pp. 1098–1108.

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.

Heckman, James J., John Eric Humphries, and Gregory Veramendi (2018). "Returns to Education: The Causal Effects of Education on Earnings, Health, and Smoking". *Journal of Political Economy* 126.S1, S197–S246.

Heckman, James J., Sergio Urzua, and Edward Vytlacil (2006). "Understanding Instrumental Variables in Models with Essential Heterogeneity". *The Review of Economics and Statistics* 88.3, pp. 389–432.

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.

Kirkeboen, Lars, Edwin Leuven, Magne Mogstad, and Jack Mountjoy (2026). "College as a Marriage Market", NBER Working Paper No. 28688.

Kirkeboen, Lars J., Edwin Leuven, and Magne Mogstad (2016). "Field of Study, Earnings, and Self-Selection". *The Quarterly Journal of Economics* 131.3, pp. 1057–1111.

Kline, Patrick and Christopher R. Walters (2016). "Evaluating Public Programs with Close Substitutes: The Case of Head Start". *The Quarterly Journal of Economics* 131.4, pp. 1795–1848.

Kolesár, Michal (2013). "Estimation in an Instrumental Variables Model with Treatment Effect Heterogeneity", Working Paper.

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

Kozakowski, Whitney (2023). "Are Four-Year Public Colleges Engines for Economic Mobility? Evidence from Statewide Admissions Thresholds". *Annenberg Institute EdWorkingPaper No. 23-727*.

Kulkarni, Siddarth and Jonathan Rothwell (2015). *Beyond College Rankings: A Value-Added Approach to Assessing Two- and Four-Year Schools*. Tech. rep. Washington, D.C.: The Bookings Institution.

LaLonde, Robert J (1986). "Evaluating the Econometric Evaluations of Training Programs with Experimental Data". *The American Economic Review* 76.4, pp. 604–620.

Londoño-Vélez, Juliana, Catherine Rodriguez, Fabio Sanchez, and Luis E. Álvarez-Arango (2023). "Financial Aid and Social Mobility: Evidence from Colombia's Ser Pilo Paga", NBER Working Paper No. 31737.

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

Mabel, Zachary, C. J. Libassi, and Michael Hurwitz (2020). "The Value of Using Early-Career Earnings Data in the College Scorecard to Guide College Choices". *Economics of Education Review* 75, p. 101958.

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*. National Bureau of Economics Working Paper No. 29276.

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

— (2024). *Integrated Postsecondary Education Data System*.

Riehl, Evan, Juan E. Saavedra, and Miguel Urquiola (2019). "Learning and Earning: An Approximation to College Value Added in Two Dimensions". *Productivity in Higher Education*. Ed. by Caroline M. Hoxby and Kevin Stange. University of Chicago Press, pp. 105–132.

Roy, A D (1951). "Some Thoughts on the Distribution of Earnings". *Oxford Economic Papers* 3.2, pp. 135–146.

Silliman, Mikko and Hanna Virtanen (2022). "Labor Market Returns to Vocational Secondary Education". *American Economic Journal: Applied Economics* 14.1, pp. 197–224.

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.

Sorkin, Isaac (2018). "Ranking Firms Using Revealed Preference". *The Quarterly Journal of Economics* 133.3, pp. 1331–1393.

Staiger, Douglas O., Thomas J. Kane, and Brian D. Johnson (2025). *Why Does Value-Added Work? Implications of a Dynamic Model of Student Achievement*. National Bureau of Economic Research Working Paper No. 34119.

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

The Economist (2015). "Our First-Ever College Rankings". *The Economist*.

U.S. Department of Education (2024). *College Scorecard*.

UTD-ERC (2026). *UT-Dallas Education Research Center Data Holdings*.

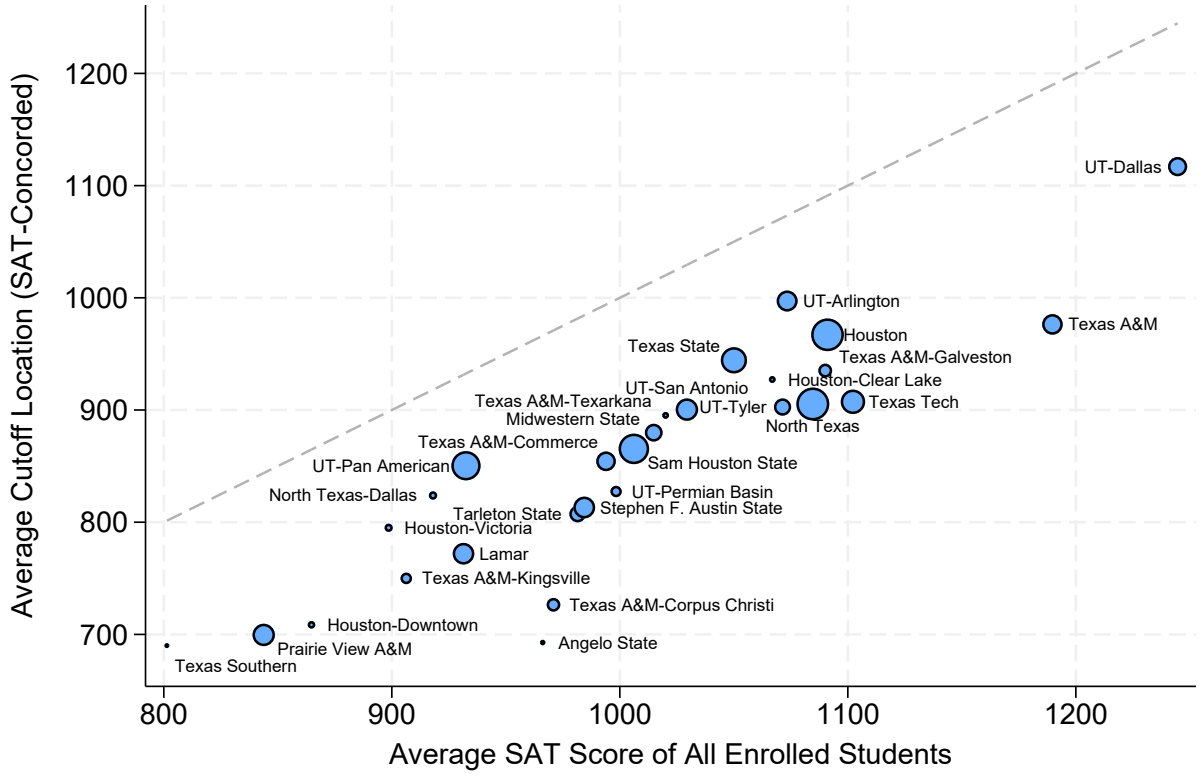
Walters, Christopher R (2018). "The Demand for Effective Charter Schools". *Journal of Political Economy* 126.6, pp. 2179–2223.

Willis, Robert J and Sherwin Rosen (1979). "Education and Self-Selection". *Journal of Political Economy* 87.5, Part 2, S7–S36.

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

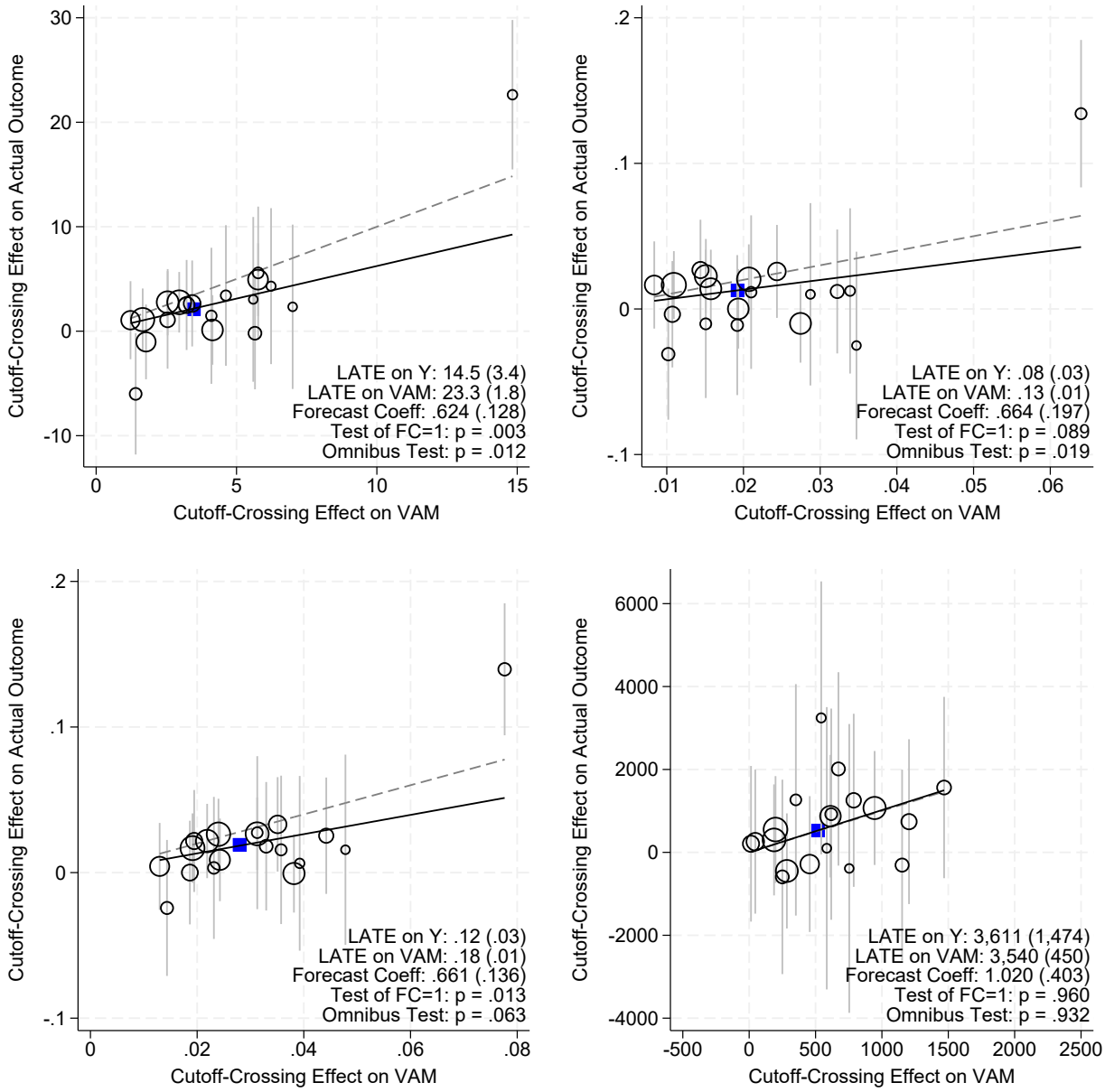
A Appendix Tables and Figures

Figure A.1: Describing the RD Target Schools



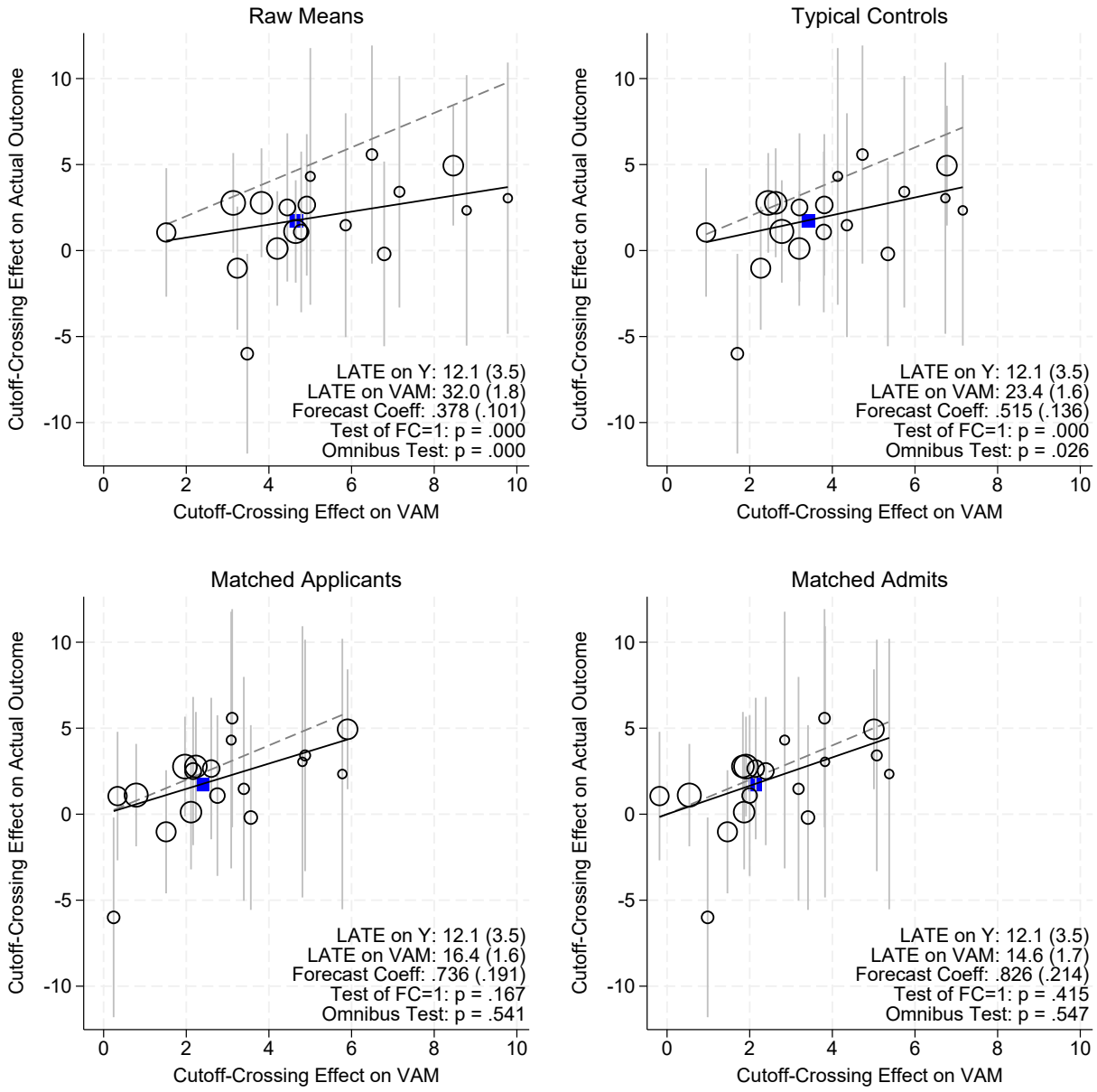
Notes: Each circle represents one of the 28 target schools with admission discontinuities. The horizontal axis plots the average SAT score of all enrolled students. The vertical axis plots the average SAT-concorded cutoff location. Circle sizes are proportional to each school's complier-weighted share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

Figure A.2: VAM Prediction Tests: Typical Controls Interacted with Institution



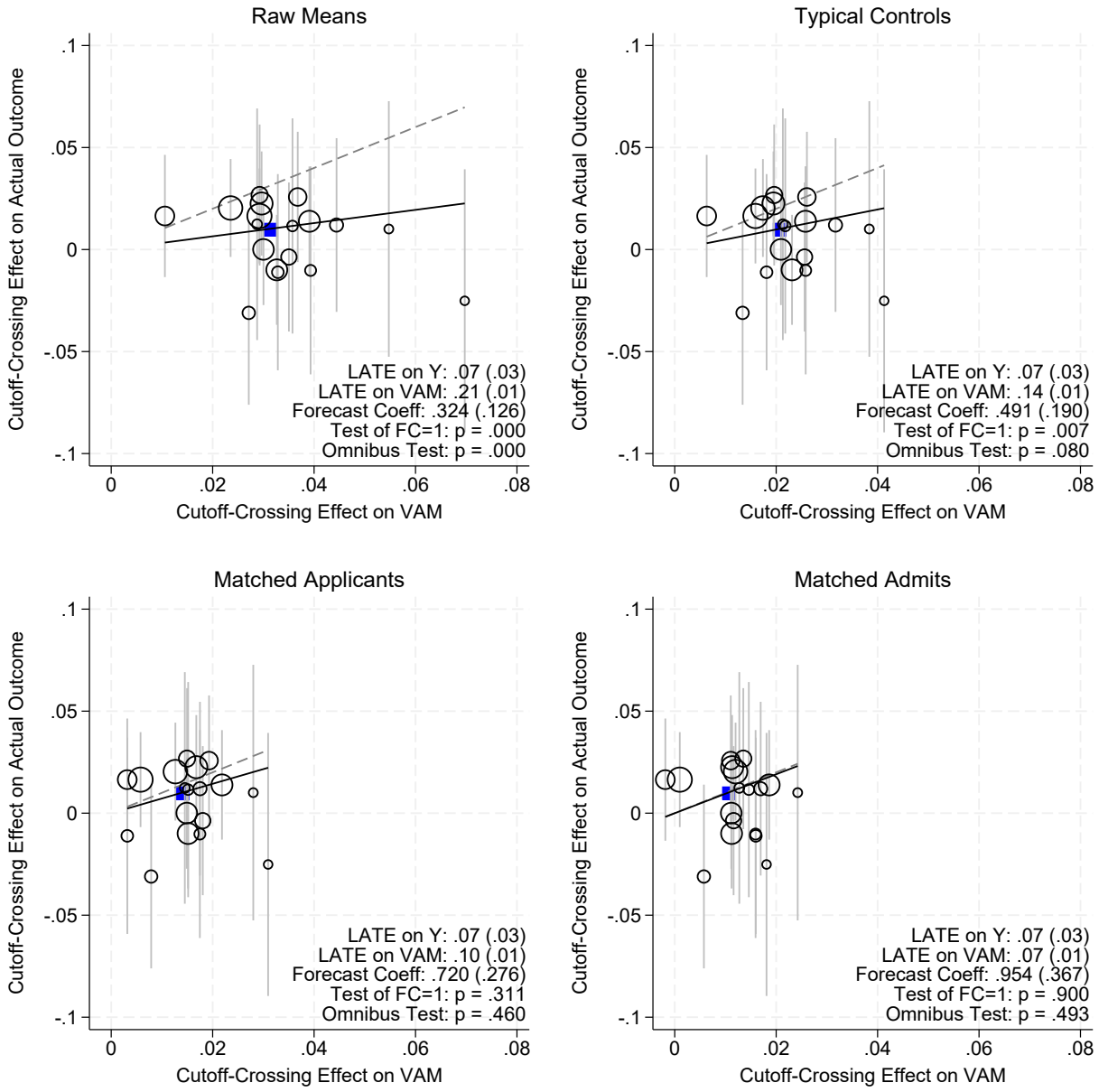
Notes: This figure tests an Interacted Typical Controls VAM model where the covariates in the Typical Controls model are all interacted with all the college treatment options. N applicants for educational outcomes = 234,271. N applicants with earnings = 188,275.

Figure A.3: VAM Prediction Tests: Cumulative Credits, Excluding Prairie View A&M



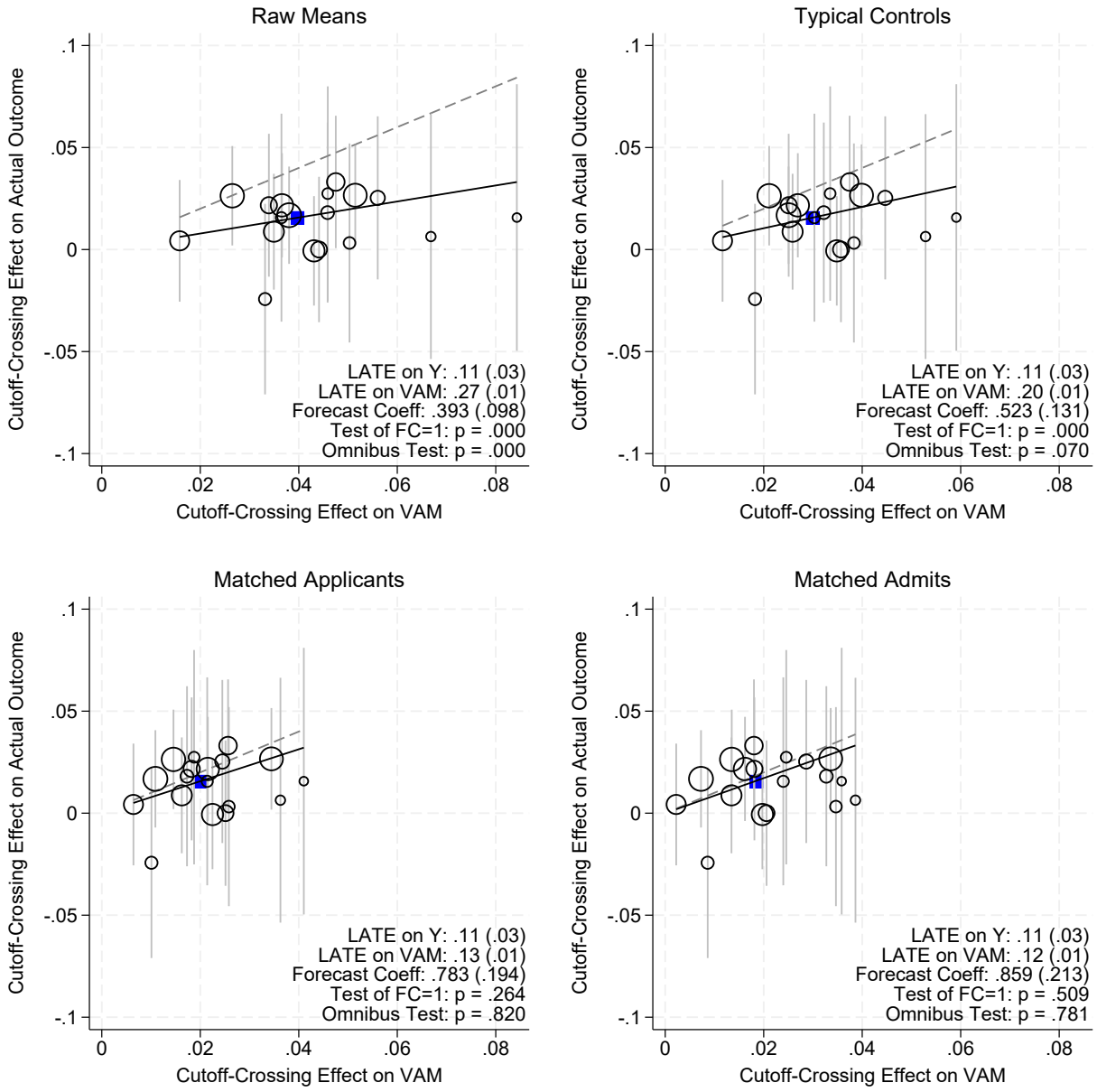
Notes: This figure is the same as Figure 8 but excludes applications targeting Prairie View A&M. N applicants = 229,645.

Figure A.4: VAM Prediction Tests: Any Degree, Excluding Prairie View A&M



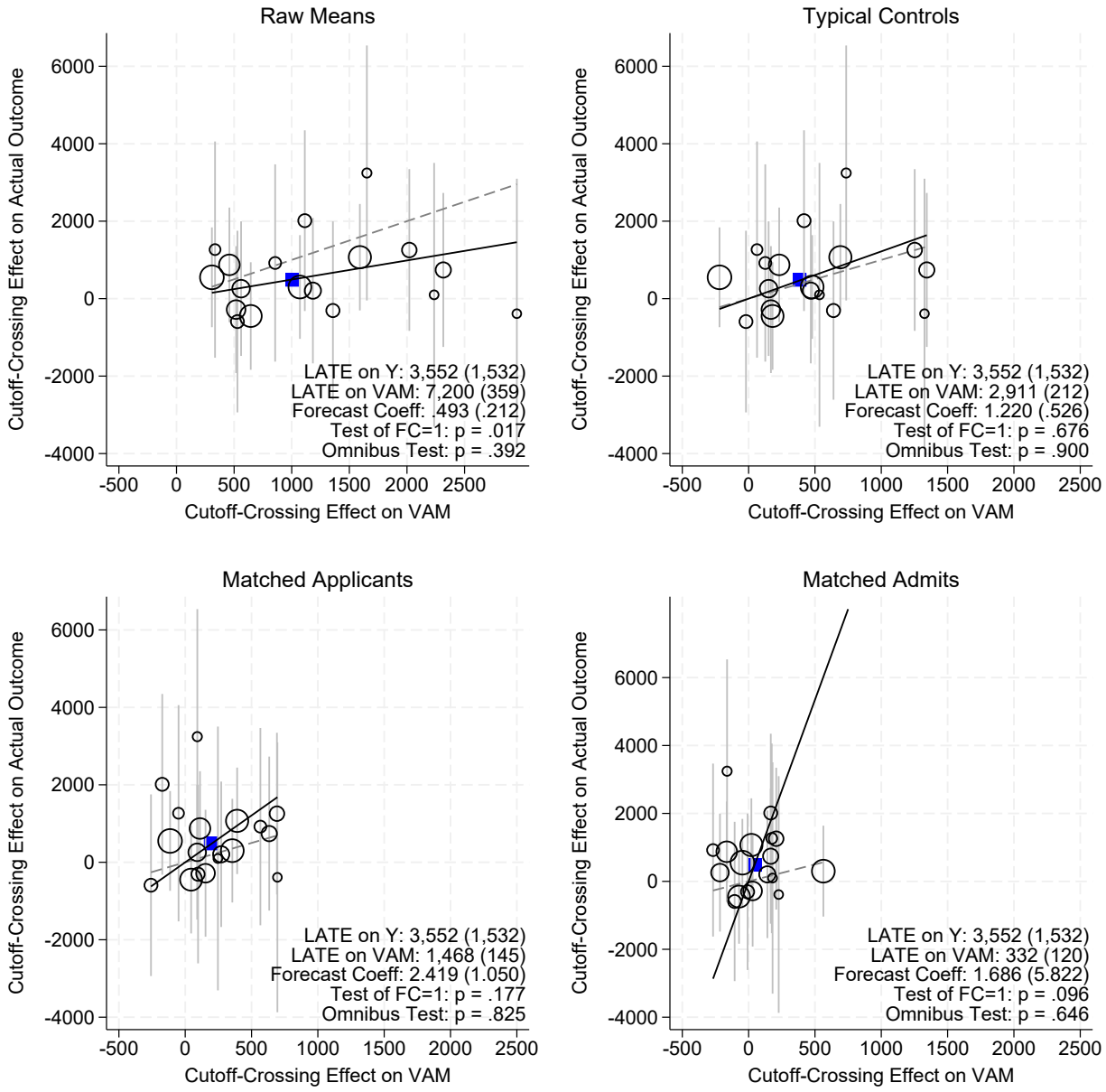
Notes: This figure is the same as Figure 9 but excludes applications targeting Prairie View A&M. N applicants = 229,645.

Figure A.5: VAM Prediction Tests: Bachelor's Degree, Excluding Prairie View A&M



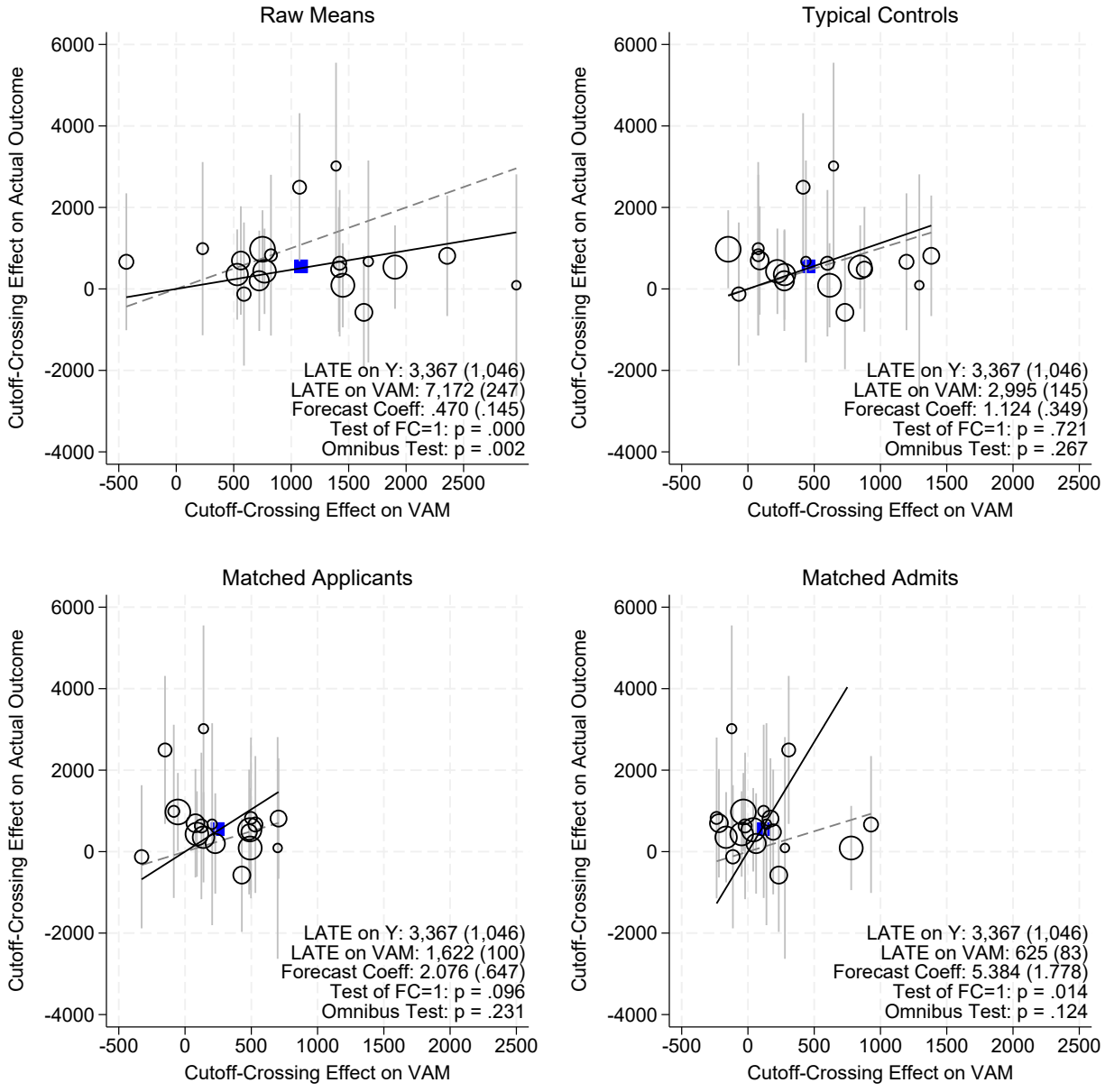
Notes: This figure is the same as Figure 10 but excludes applications targeting Prairie View A&M. N applicants = 229,645.

Figure A.6: VAM Prediction Tests: Earnings, Excluding Prairie View A&M



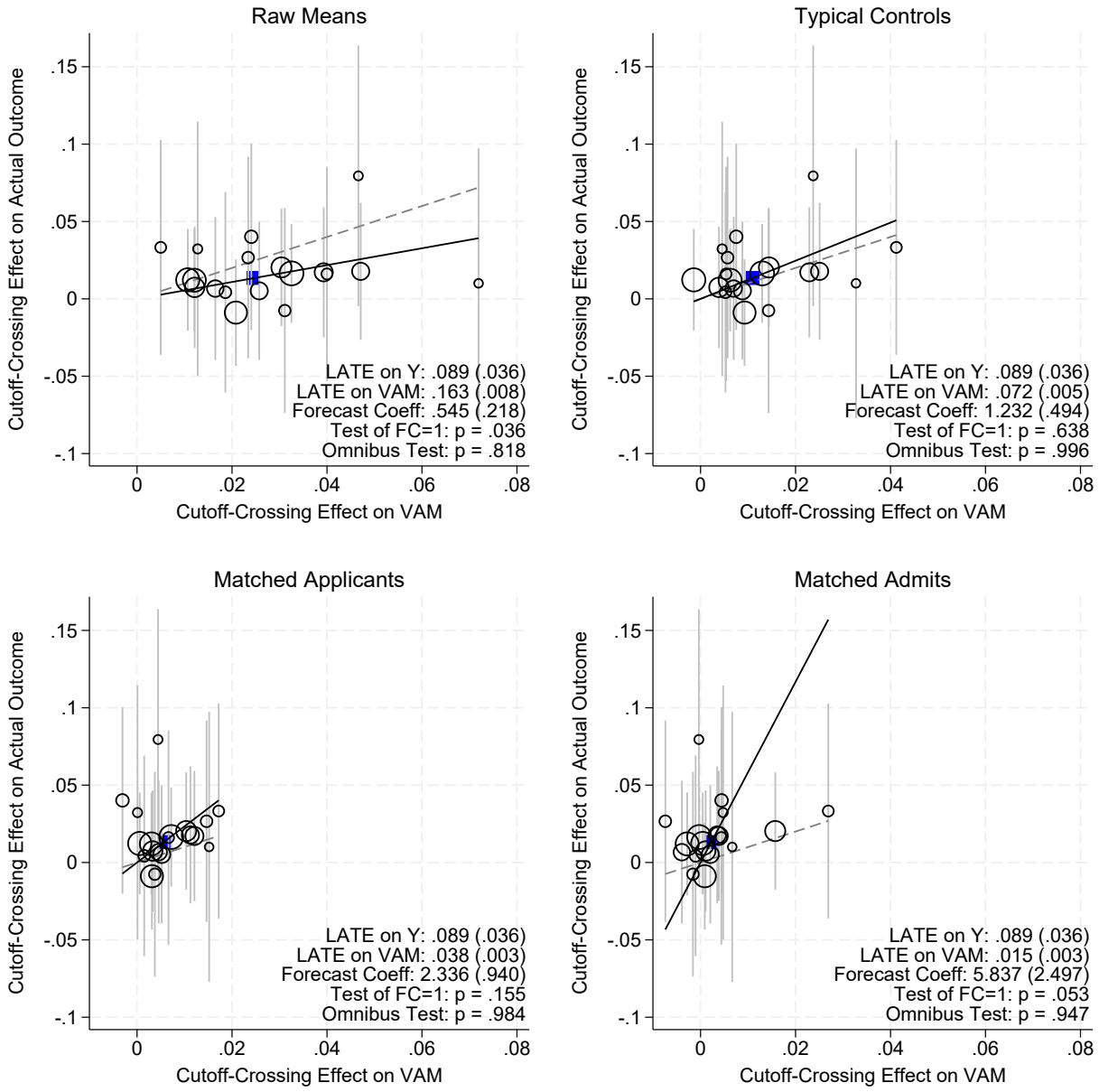
Notes: This figure is the same as Figure 11 but excludes applications targeting Prairie View A&M. N applicants = 184,438.

Figure A.7: VAM Prediction Tests: Earnings, RD Bandwidth=6



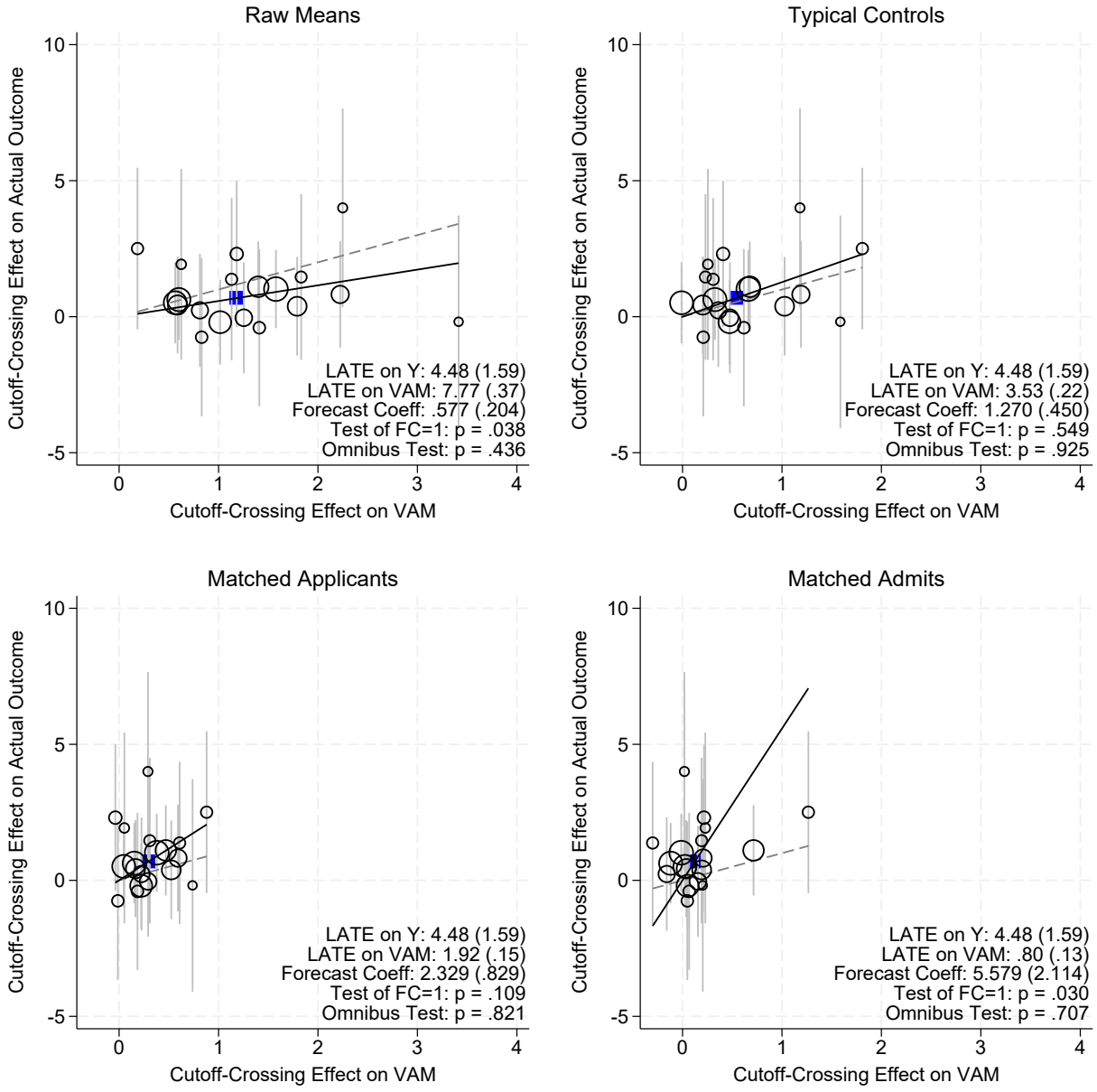
Notes: See the notes to Figure 8. N applicants with earnings = 301,766.

Figure A.8: VAM Prediction Tests: Log Earnings



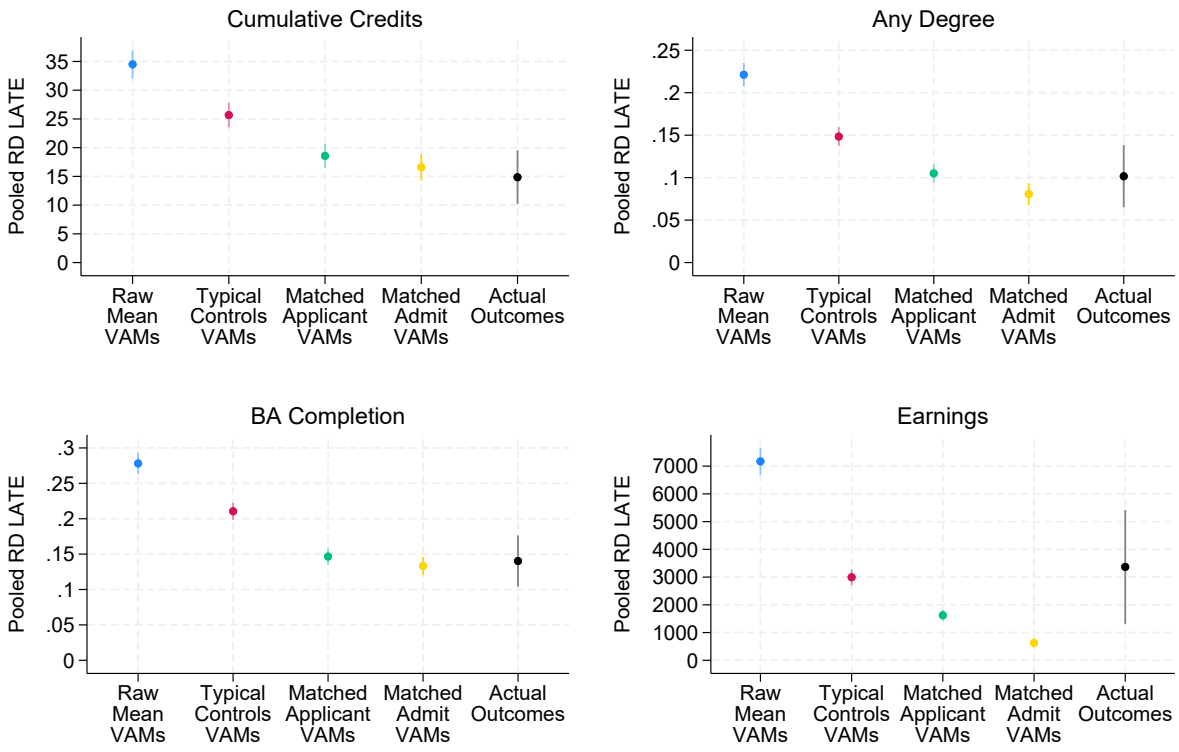
Notes: This figure follows the same format as Figure 11 but measures individual earnings in logs rather than dollars. N applicants with earnings = 188,275.

Figure A.9: VAM Prediction Tests: Earnings Rank



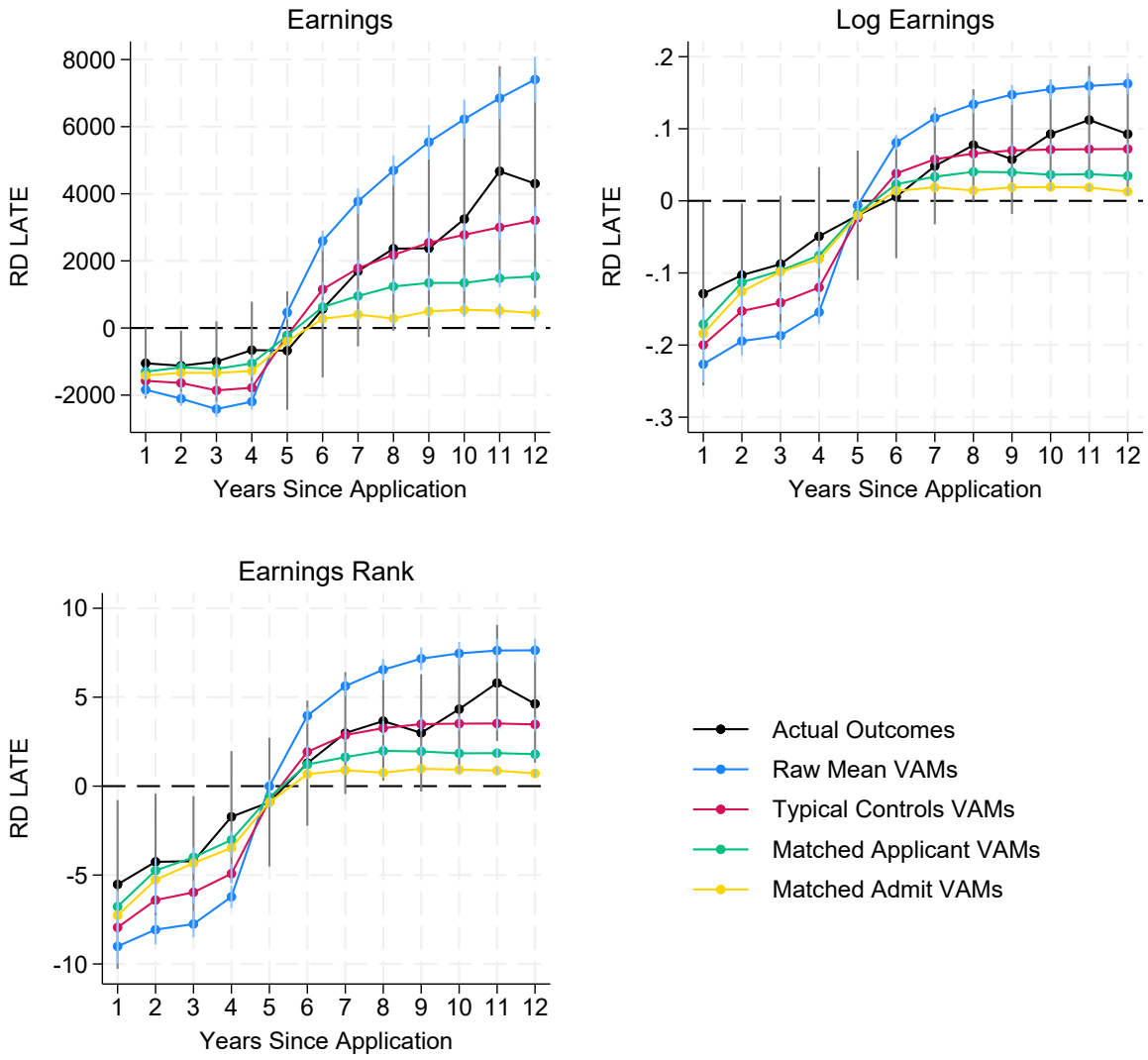
Notes: This figure follows the same format as Figure 11 but measures each individual's earnings as their rank within their statewide high school graduating cohort rather than dollars. N applicants with earnings = 188,275.

Figure A.10: VAM Prediction Tests: Summarizing with Pooled LATEs, RD BW=6



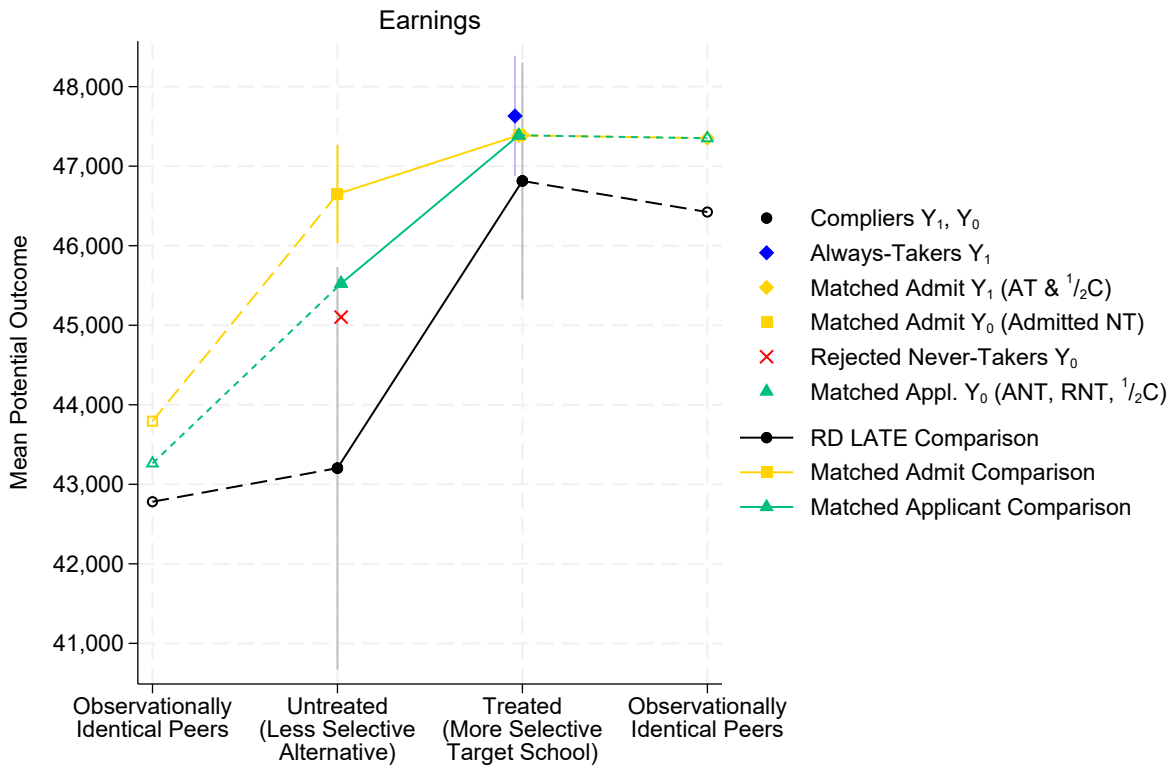
Notes: This figure is the same as Figure 12 but doubles the RD bandwidth from 3 concorded ACT points to 6. Each panel summarizes the VAM prediction test results for a different outcome. Each point is an RD LATE of enrolling in a target university among cutoff compliers, pooled across all target universities with admission cutoffs. The far right point in each panel uses the applicant's actual outcome, while the other points replace the applicant's actual outcome with the observational VAM of their initial college (including no college).

Figure A.11: VAM Prediction Tests: Summarizing with Dynamic LATEs



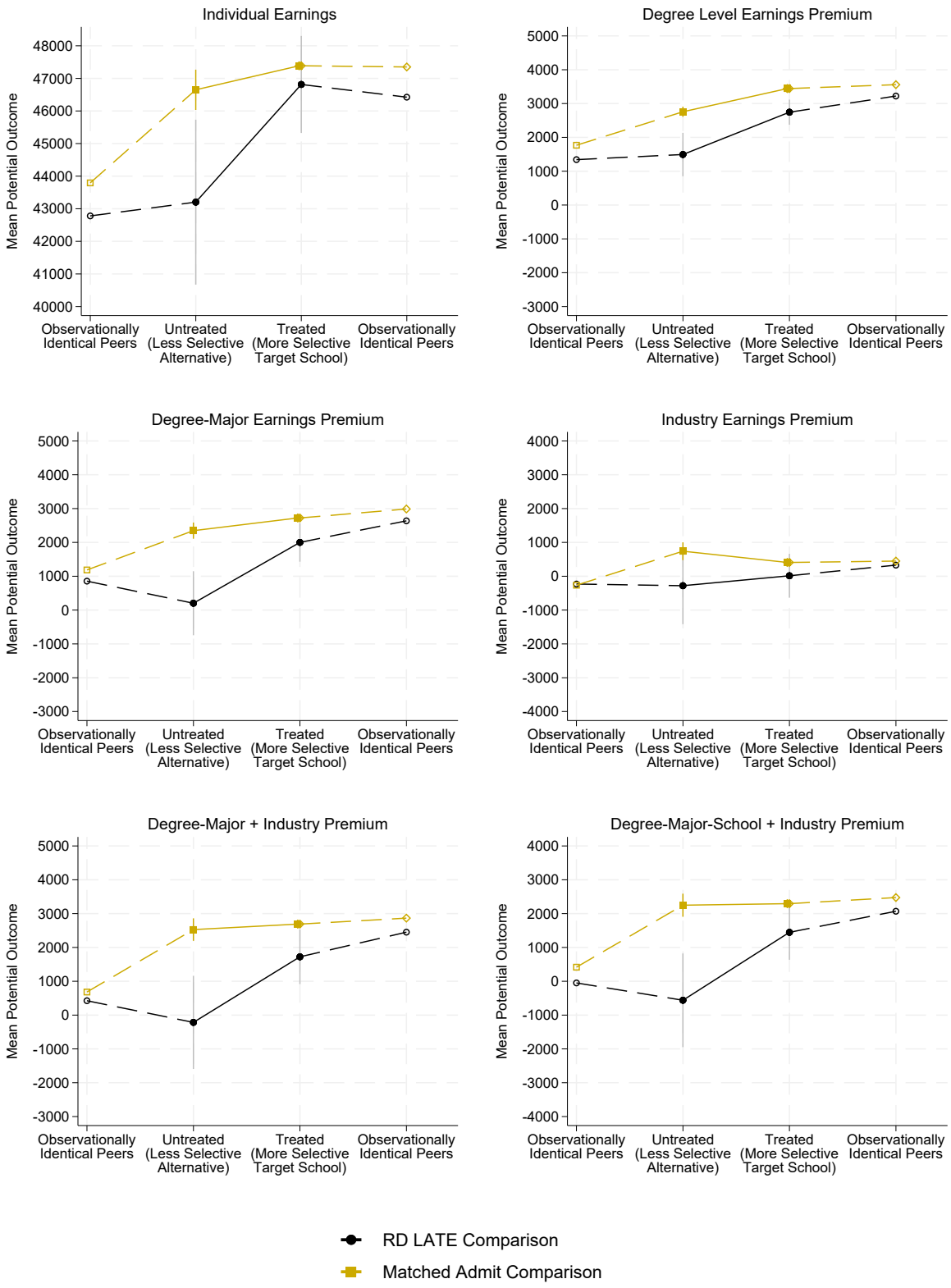
Notes: This figure expands the exercise in the bottom right panel of Figure 12 to estimate earnings LATEs separately for each year since application. Each point is an RD LATE of enrolling in a target university among cutoff compliers, pooled across all target universities with admission cutoffs, estimated separately for each outcome year since application. The black series in each panel uses the applicant's actual outcome, while the other series replace the applicant's actual outcome with the observational VAM of their initial college (including no college).

Figure A.12: Unpacking the Methods: Adding the Matched Applicant Comparison



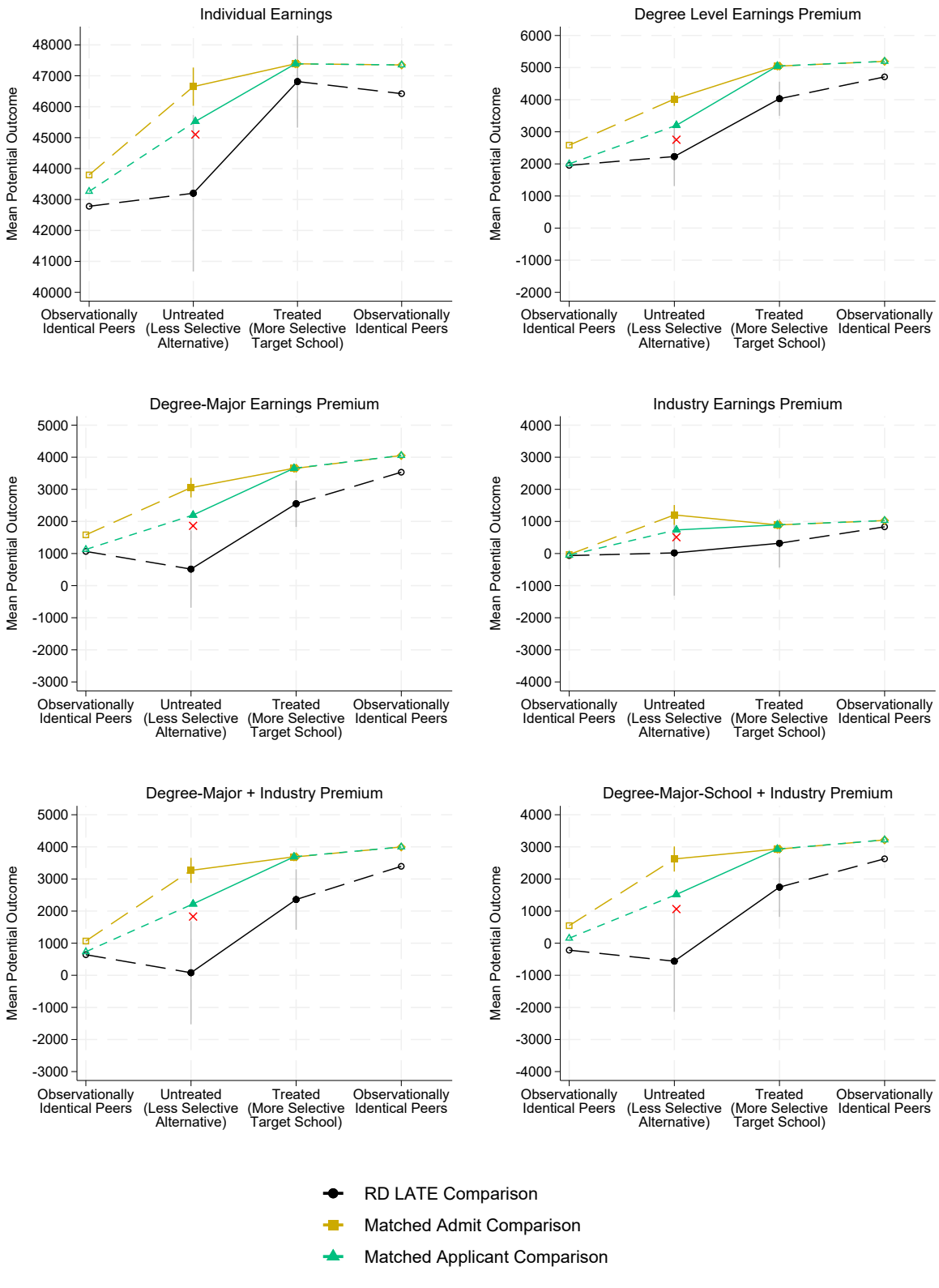
Notes: This figure is the same as Figure 14 but adds the Matched Applicant components. See the notes to Figure 14.

Figure A.13: Degree, Major, and Industry Premia, Estimated with Typical Controls



Notes: This figure is the same as Figure 16 but adds the Typical Controls to the fixed effect regressions that estimate the degree, major, and industry premia.

Figure A.14: Degree, Major, and Industry Premia, Including Matched Applicants



Notes: This figure is the same as Figure 16 but adds the Matched Applicant comparison.