Blueprint Labs

Discussion Paper #2021.17

The Returns to College(s): Relative Value-Added and Match Effects in Higher Education

Jack Mountjoy
Brent R. Hickman

September 2021

MIT Department of Economics
77 Massachusetts Avenue, Bldg. E53-390
Cambridge, MA 02139

National Bureau of Economic Research
1050 Massachusetts Avenue, 3rd Floor
Cambridge, MA 02138

NBER WORKING PAPER SERIES

THE RETURNS TO COLLEGE(S): RELATIVE VALUE-ADDED AND MATCH EFFECTS IN HIGHER EDUCATION

Jack Mountjoy
Brent R. Hickman

Working Paper 29276
http://www.nber.org/papers/w29276

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 2021

For helpful comments, we are grateful to Josh Angrist, Raj Chetty, and John Friedman, along with Dan Black, Zach Bleemer, Kirill Borusyak, David Deming, Michael Dinerstein, Peter Ganong, Nathan Hendren, Peter Hull, Adam Kapor, Ezra Karger, Larry Katz, David Lee, Thomas Lemieux, Alex Mas, Jordan Matsudaira, Arnaud Maurel, Casey Mulligan, Derek Neal, Chris Neilson, Matt Notowidigdo, Jordan Richmond, Jonah Rockoff, Raffaele Saggio, Constantine Yannelis, Owen Zidar, Seth Zimmerman, and seminar participants at Princeton, the Federal Reserve Banks of New York and Chicago, the University of Oslo, Statistics Norway, Duke, Chicago Booth, Chicago Economics, NBER Labor Studies, Washington University in St. Louis, Harvard Opportunity Insights, Columbia Teachers College, NBER Summer Institute, LSE, Penn State, UCL, LMU Munich, and FGV EPGE. We also thank Rodney Andrews, Janie Jury, Mark Lu, Greg Phelan, John Thompson, and especially Greg Branch at the UT-Dallas Education Research Center for expert guidance on the administrative data. Nidhaanjit Jain provided excellent research assistance. We are grateful for financial support from the Industrial Relations Section at Princeton University and the Robert H. Topel Faculty Research Fund at the University of Chicago Booth School of Business. 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. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

The Returns to College(s): Relative Value-Added and Match Effects in Higher Education
Jack Mountjoy and Brent R. Hickman
NBER Working Paper No. 29276
September 2021
JEL No. C21,H75,I23,I24,I26,J24,J31,J62

ABSTRACT

Students who attend different colleges in the U.S. end up with vastly different economic outcomes. We study the role of relative value-added across colleges within student choice sets in producing these outcome disparities. Linking high school, college, and earnings registries spanning the state of Texas, we identify relative college value-added by comparing the outcomes of students who apply to and are admitted by the same set of institutions, as this approach strikingly balances observable student potential across college treatments and renders our extensive set of covariates irrelevant as controls. Methodologically, we develop a framework for identifying and interpreting value-added under varying assumptions about match effects and sorting gains. Empirically, we estimate a relatively tight, though non-degenerate, distribution of relative value-added across the wide diversity of Texas public universities. Selectivity poorly predicts value-added within student choice sets, with only a fleeting selectivity earnings premium fading to zero after a few years in the labor market. Non-peer college inputs like instructional spending more strongly predict value-added, especially conditional on selectivity. Colleges that boost BA completion, especially in STEM majors, also tend to boost earnings. Finally, we probe the potential for (mis)match effects by allowing value-added schedules to vary by student characteristics.

Jack Mountjoy
Booth School of Business
University of Chicago
5807 South Woodlawn Avenue
Chicago, IL 60637
and NBER
jack.mountjoy@chicagobooth.edu

Brent R. Hickman
Olin Business School
Washington University in Saint Louis
One Brookings Drive, Campus Box #1133
Saint Louis, MO 63130
USA
hickmanbr@wustl.edu

1 Introduction

Enrolling in college tends to pay off for students on the margin of non-attendance.[1] But for the already college-bound, the relevant choice is not whether to enroll, but where. Comparing typical student outcomes across institutions suggests this choice may be highly consequential. Graduation rates and mean earnings at the top of the college distribution are vastly different from those at the bottom: relative to students who attend Alabama State, those who attend MIT are nearly four times as likely to graduate (93% vs. 25%) and earn nearly four times as much ($104,700 vs. $27,700). These divergent outcomes may reflect the impacts of differing college inputs—MIT's per-student spending on instruction, academic support, and student services in 2018 was nearly $130,000, while Alabama State's analogous per-student total was just $15,000—but little is known about how effectively different colleges use these inputs to add value to their students' outcomes.[2]

Interpreting differences in student outcomes across colleges is further complicated by the strong sorting of students with vastly different levels of academic preparation and socioeconomic backgrounds into different institutions. To continue the example above, the average incoming student at MIT scored 1530 (out of 1600) on the SAT, while average scores at Alabama State are around 860. Moreover, 74% of students at Alabama State have family incomes low enough to qualify for Pell grants, compared to only 17% at MIT. Different universities 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, 2019). The selective nature of enrollment leaves us with a difficult question: to what extent do differences in graduation rates and earnings across colleges reveal active institutional impacts on student outcomes, rather than passive reflections of durable pre-college inequalities? Understanding how these outcome differences arise is an important consideration for college-bound students deciding where to enroll. With state legislators increasingly exploring funding models that directly tie public college appropriations to student outcomes, it is also an important consideration for policy-makers deciding where and how much to invest.[3]

In this paper, we study the role of relative value-added across colleges within student choice sets in producing these college-level outcome disparities. We allow each college to have its own unique treatment effect on each outcome, which enables us to explore several interrelated questions. Do some colleges boost student outcomes more than others? Are there observable institutional characteristics, like selectivity or per-student spending, that predict such value-added? Do colleges that boost specific outcomes, like degree completion, also boost others, like earnings? And does value-added vary across students from different backgrounds, producing (mis)match effects?

In answering these questions we must confront three significant challenges. First, estimating unique treatment effects for each college is a data-intensive exercise. In order to learn something meaningful, we need to observe a wide enough swath of colleges to span the quality spectrum, and a large enough sample of students within each college to estimate value-added with meaningful precision. Second, this exercise requires a series of individual-level data linkages, from student characteristics determined prior to college entry, to enrollment spells potentially spanning multiple institutions, to policy-relevant outcomes like degree completions, major choices, and earnings. Third, the college market in the U.S. is a highly decentralized endogenous selection process, both on the part of students deciding where to apply and colleges deciding whom to admit, so identifying causal impacts of colleges requires a research design that credibly reckons

[1] See, for example, Card (2001); Lemieux and Card (2001); Carneiro et al. (2011); Angrist and Chen (2011); Zimmerman (2014); Bhuller et al. (2017); Heckman et al. (2018); Smith et al. (2020); Mountjoy (2021).

[2] Sources: College Scorecard (`http://collegescorecard.ed.gov`) and IPEDS (`https://nces.ed.gov/ipeds`).

[3] At least 30 states base postsecondary appropriations on student outcomes in some way, and at least 11 others are in the process of designing such formulas for implementation (National Conference of State Legislatures, 2018; Ward and Ost, 2021).

with the strong sorting of students with different potential outcomes into different institutions.

We address these challenges by linking multiple administrative registries that span the state of Texas, a setting that offers a vast diversity of colleges and students who attend them. Our data include the entire population of Texas public high school students—nearly 10% of all public school students in the United States—and their college choices across the Texas postsecondary landscape. This solves the first challenge of acquiring large sample sizes of enrollees across a highly heterogeneous array of colleges. We overcome the second challenge by linking the high school records of these students, including rich pre-college measures of skills and backgrounds, to statewide higher education registries that track each student's college enrollments, degree completions, and major choices, as well as to administrative earnings registries that continue to track these students well into their early careers. To address the third challenge of selection bias, a unique strength of our dataset is the ability to further link these students to administrative records of all applications and admissions decisions at all thirty Texas public universities. Acquiring admissions data from one institution is a difficult task; observing it across the entirety of a very large state's public university landscape is quite rare, so we focus our analysis on students who begin college at one of these thirty Texas public universities to take advantage of this unique collection of data.

Specifically, we employ the application and admissions records to implement a research design that compares the outcomes of students who enroll in different colleges but applied to and were admitted by the same set of institutions.[4] The key idea behind this "matched applicant" approach, pioneered by Dale and Krueger (2002) to study linear returns to college selectivity, 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.[5]

Our careful adaptation of the matched applicant research design to the task of identifying relative value-added of individual colleges involves several methodological innovations worth highlighting. First, we begin by grounding our identification analysis in a potential outcomes framework with unrestricted heterogeneity in the value-added of each college across different students, generalizing the constant treatment effects approach typically employed in the value-added literature. This framework allows us to draw a clear distinction between selection bias, in which different colleges enroll students with vertically differentiated average potential, and the more subtle question of sorting on match effects, in which students may choose colleges based on their own idiosyncratic deviations from an average value-added schedule—i.e., essential heterogeneity (Heckman et al., 2006) with multiple unordered treatments.

Second, we present evidence that our straightforward implementation of the matched applicant design appears to address selection bias on vertical potential fairly compellingly. The scope of the problem is enormous: we first document that average *predicted* student outcomes—from a model trained on our rich array of pre-college student covariates—vary nearly as much across colleges as raw, *observed* outcomes,

[4]Many of the colleges in our data and across the U.S. practice "holistic admissions" with no sharp quantitative admissions cutoffs. This hampers a regression discontinuity (RD) research design, which has been increasingly employed to study admission impacts in countries with centralized college admissions systems (e.g. Saavedra, 2009; Hastings et al., 2014; Kirkeboen et al., 2016; Canaan and Mouganie, 2018; Anelli, 2019; Jia and Li, 2021; Sekhri, 2020). Because the U.S. has no such centralization, the handful of RD studies in the U.S. focus on binary treatment comparisons of getting marginally admitted to a specific institution (Hoekstra, 2009; Zimmerman, 2014), or a set of cutoff-sharing institutions (Goodman et al., 2017; Smith et al., 2020), relative to a composite counterfactual of various fallback options. Such a composite counterfactual pools other 4-year institutions, 2-year colleges, and non-attendance together into a single comparison group, hampering the ability of the RD approach to identify college-by-college value-added in the decentralized U.S. higher education system.

[5]Subsequent variants of the matched applicant approach include Arcidiacono (2005), Long (2008), Fryer and Greenstone (2010), Broecke (2012), Smith (2013), Cunha and Miller (2014), Arcidiacono et al. (2016b), Ge et al. (2018), Eller (2019), Abdulkadiroglu et al. (2020), and Angrist and Frandsen (2020).

suggesting massive selection bias in uncontrolled college comparisons. We then show that controlling solely for fixed effects for each distinct portfolio of college admissions nearly eliminates these differences in predicted outcomes across colleges. Students who apply to and are admitted by the same set of institutions thus appear to have very similar vertical potential regardless of where they actually enroll.

Due to this striking balance result, including or excluding typical controls like test scores and demographics makes little difference in the value-added estimates that result from our baseline specification, which controls solely for admission portfolio fixed effects. Even the inclusion of less commonly available covariates—i.e., more detailed measures of academic preparation like advanced high school coursework, behavioral measures of non-cognitive skills like attendance and disciplinary records, and fixed effects for every high school to flexibly account for the environmental influences of schools, neighborhoods, and local labor markets—fails to reveal any additional selection bias in the baseline estimates, despite the relevance of all of these covariates in predicting student outcomes, even within admission portfolios. The converse is not true: if we begin with the typical selection-on-observables approach by controlling for commonly available covariates, and even the less commonly available ones, the subsequent addition of admission portfolio controls substantially reduces the dispersion in value-added estimates across colleges. Admission portfolios thus appear to account for significant residual selection bias lurking in typical selection-on-observables estimates of college value-added. We also show that enrolling in a less selective college when admitted to a more selective alternative—an important source of identifying variation in the matched applicant approach—is very common in our sample of public university students, helping to assuage concerns about rare and odd behavior driving the research design.[6]

Third, with selection on vertical potential addressed, we explore the more subtle question of "horizontal" sorting into colleges on the basis of student-specific match effects. We do so in order to generalize beyond the common but strong assumption of homogeneous treatment effects underpinning the vast majority of the value-added literature.[7] We make explicit how this assumption dramatically simplifies both the definition and identification of value-added by removing any possibility of match effects and sorting gains, but we demonstrate that it is much stronger than needed, being sufficient but not necessary for recovering causal comparisons. Importantly, even in the presence of unrestricted heterogeneity in the value-added of each college across students, and unrestricted sorting on such match effects, one can identify treatment-effect-on-the-treated comparisons by differencing mean student outcomes across colleges within admission portfolios. From that fully unrestricted foundation, we then build a menu of progressively stronger, nested restrictions that deliver more ambitious causal parameters with greater statistical precision. Specifically, if the students who share each admission portfolio reap sorting gains that are *on average* symmetric across the colleges in that portfolio—a stronger special case of which is zero average sorting gains, which we might expect if students are unaware of or uninfluenced by their idiosyncratic deviation from the mean treatment effect at a given college—then admission portfolio fixed effects absorb these portfolio-specific sorting gains to deliver average treatment effects when regressing outcomes on college treatment indicators and admission

[6] Recent work by Fillmore (2021) shows that price discrimination through financial aid drives significant within-student variation in net utility across colleges in student admission portfolios; other non-selectivity factors influencing college enrollment decisions include geographic preferences (Fu et al., 2021), amenities (Jacob et al., 2017), institutional mission (Delavande and Zafar, 2019), athletics (Pope and Pope, 2012), coordination with friends (Fletcher, 2015), and weather (Simonsohn, 2010).

[7] In our context, a treatment effect refers to the causal value-added of initially enrolling in one of our 30 Texas universities, which may or may not vary across students for a given institution. The overwhelming majority of papers in the value-added literature—including studies of colleges (Cunha and Miller, 2014; Hoxby, 2019), teachers (Chetty et al., 2014a), K-12 schools (Angrist et al., 2017, 2020), firms (Abowd et al., 1999), industries (Krueger and Summers, 1988), hospitals (Gowrisankaran and Town, 1999), and regional health care systems (Finkelstein et al., 2016)—focus on identifying models that restrict the value-added of a given treatment to be homogeneous across individuals.

portfolio indicators. Empirically, we find that such a regression delivers similar, but much more precise, value-added estimates compared to the more flexible but more data-demanding within-portfolio pairwise averaging approach, which leaves sorting gains completely unrestricted. More generally, our identification framework highlights how the ambitious goal of identifying a population-level average value-added schedule is not threatened by heterogeneous treatment effects per se, or even by systematic sorting gains derived from such match effects, but rather by *mean asymmetry* in realized sorting gains across colleges. Moreover, we establish that even in the presence of such mean asymmetry one can still identify "local" comparisons of causal value-added reaped by the students who actually apply, get admitted, and choose to enroll at each college. While these discussions are grounded in the college setting, the broader lessons of this framework apply to other value-added environments where an analog of our Assumption 1 addresses selection bias on vertical potential but researchers wonder how "horizontal" treatment effect heterogeneity affects the identification and interpretation of value-added.

Our empirical results reveal several notable patterns in the distribution of relative value-added across colleges in student choice sets. First, in terms of magnitude, we estimate a relatively tight, though non-degenerate, value-added distribution across the wide diversity of Texas public universities: moving up one standard deviation in the value-added distribution corresponds to a 3.7 percentage point gain in the probability of completing a BA, and a $1,300 boost in annual earnings around age 28, roughly a 3% premium. Value-added thus varies non-trivially across colleges, but with magnitudes dwarfed by the variation in raw outcome means. A college's raw BA completion rate correlates fairly strongly with its value-added on BA completion ($\rho = 0.5$), while raw earnings means and earnings value-added have a correlation of only 0.2, even after correcting for minor attenuation from estimation error. These results suggest that both the ordering and the magnitudes of the raw outcome comparisons often highlighted in college guides, the popular press, and state funding formulas are driven largely by selection bias, leaving a smaller and more nuanced role for causal college contributions, at least across the actual choices students face within their admission portfolios.

We then investigate how well various observable college characteristics predict relative value-added. While selectivity is a nearly perfect predictor of raw outcome means across colleges, it is a poor predictor of value-added within student choice sets. Students who attend the more selective colleges in their admission portfolios are only slightly more likely to complete a BA. Moreover, these students are significantly less likely to successfully major in STEM, a result that strongly switches signs when moving from raw correlations to within-portfolio value-added estimates. With respect to earnings, we do find a modest selectivity premium in a student's first few years in the post-college labor market, but this premium fades out quickly to a stable and precise zero thereafter. Students with similar potential who attend less selective colleges therefore appear to catch up to their elite peers, suggesting the labor market initially rewards a selectivity signal but then learns about underlying productivity relatively quickly (e.g. Lange, 2007). Moving beyond selectivity, non-peer college inputs like instructional expenditures per student and the share of faculty who are full-time do covary positively with value-added—especially conditional on selectivity—and deliver a durable earnings premium, suggesting that resources outweigh peers in the production of college value-added. We also explore the potential for college-level intergenerational mobility statistics (e.g. Chetty et al., 2020) to serve as predictors of value-added; we do not find strong correlations between these measures, perhaps because mobility statistics that condition solely on family income remain at risk of largely reflecting selection bias rather than causal college impacts. Taken together, this set of results suggests that various prominent measures of college "quality" are neither interchangeable nor sufficient as barometers of causal value-added.

Turning to mechanisms, we find that the value-added estimates for each college tend to be correlated across different student outcomes, shedding light on the potential channels though which colleges shape the path from enrollment to earnings. Colleges that boost BA completion also tend to boost earnings: a 10 percentage point increase in BA value-added is associated with a roughly $3,000 increase in earnings VA across colleges. This relationship strengthens considerably as we lengthen the degree completion window from 4 years to 6 years to 8 years, indicating that strict measures of on-time graduation may underestimate a college's ultimate value-added in the labor market. Unpacking BA completion by major category, we find that a college's value-added on the probability of completing a STEM degree has a strong correlation with its earnings VA, while non-STEM VA has almost no bivariate relationship with earnings VA. A college's STEM and non-STEM value-added are negatively correlated, however; conditional on STEM VA, the relationship between non-STEM VA and earnings becomes reasonably positive, though still substantially weaker than the STEM-earnings relationship. Value-added on non-degree outcomes like persistence and transfer, as well as on industry of employment, also predict earnings value-added, highlighting the multiplicity of channels through which colleges can influence the outcomes of their students.

Finally, we probe the potential for (mis)match effects by allowing the relative value-added of each college to vary flexibly by student characteristics, including race, gender, family income, and pre-college measures of cognitive and non-cognitive skills. Overall, we find little evidence of substantial heterogeneity in value-added schedules across students: separate estimates by student subgroups tend to cluster around the main pooled estimates, and interaction effects between college indicators and student characteristics have very little explanatory power on outcomes. At first glance, one modest exception is that Black students appear to face small negative impacts of attending more selective colleges, but this minor pattern of "mismatch" is driven by the presence of two large historically Black universities in Texas that have low average incoming SAT scores but yield above-average value-added. These HBCUs are over 95 percent Black, and thus do not tend to enter the choice sets of non-Black students. This racial heterogeneity in choice sets, rather than mismatch per se, drives the result, as Black students face similar returns to selectivity, and similar value-added schedules more generally, across the non-HBCUs compared to their peers from other backgrounds.

Our methods and results advance several strands of the literature on the consequences of college choices. First, a large body of prior work has sought to estimate the return to selectivity, or college "quality" more generally, typically by specifying earnings as a function of the mean SAT score of a student's college peers.[8] While several papers in this strand of literature have estimated significant gains to attending a more selective college, Dale and Krueger (2002, 2014) stand out as an exception after controlling for student application behavior. On one hand, we replicate, validate, and extend this null return to selectivity (after documenting a fleeting initial premium) in our administrative data, which span a much larger, more diverse, more recent, and more precisely measured sample of students and colleges than the College and Beyond survey at the center of Dale and Krueger (2002, 2014)'s analysis. On the other hand, our results on non-peer college inputs build on the work of Black and Smith (2006) and others in exploring other important dimensions of college quality that may not be captured by selectivity. Our work thus sheds light on a long-standing apparent tension in this literature: while college selectivity—which, by definition, largely reflects endogenous selection of students with different abilities into different colleges—fails to predict relative value-added within admission sets,

[8] See Brewer and Ehrenberg (1996) for a review of the early literature. Black and Smith (2006) and Dillon and Smith (2018) employ a more comprehensive definition of college quality as an index of multiple institutional measures. Bodoh-Creed and Hickman (2019) endogenize pre-college human capital investment in a structural analysis of college sorting and student outcomes. Ge et al. (2018) expand the sample and outcomes of Dale and Krueger (2002) to study impacts of selectivity on female labor supply and family outcomes. Other related work examines impacts of attending different categories of institutions, e.g. flagship vs. non-flagship public universities (Andrews et al., 2016, 2019; Black et al., 2020), 4-year vs. 2-year colleges (Reynolds, 2012; Mountjoy, 2021), and for-profit vs. public institutions (Armona et al., 2018; Cellini and Turner, 2019).

non-peer college inputs—which are correlated with selectivity, but not perfectly so—do appear to durably boost student outcomes, or at least correlate with factors that do.

More importantly, our college-by-college value-added approach allows us to move beyond traditional analyses of vertical differentiation in higher education that assume college quality can be well proxied by a single observable characteristic. Rather, by flexibly allowing each college to impart its own unique impact on student outcomes, we are able to characterize the distribution of causal effectiveness across colleges without pre-specifying a single observable dimension of vertical differentiation. Our work thus contributes to a nascent set of papers that have begun to leverage large administrative data sources in the U.S. to examine student outcomes at the granularity of individual colleges, including Cunha and Miller (2014), Arcidiacono et al. (2016b), Hoxby (2019), and Chetty et al. (2020). We advance this recent work on several fronts: we merge in administrative admissions records that permit a precise implementation of the matched applicant approach; we conduct a battery of validation exercises to probe the ability of this approach to disentangle causal college contributions from selection bias; and we derive and interpret value-added estimands under a range of assumptions about treatment effect heterogeneity and sorting on gains. We also estimate value-added on a significantly expanded suite of student outcomes—including persistence, transfer, degree completion and its timing, major choice, out-of-state migration, industry of employment, and earnings dynamics—illustrating the variety of interconnected channels through which colleges can influence student trajectories.[9]

Furthermore, our ability to merge in a rich set of pre-college measures of student skills and backgrounds allows us to explore heterogeneity in the college value-added schedules faced by different types of students, and bring these estimates to bear on related academic and policy debates (e.g. Bowen and Bok, 2000; Sander and Taylor, 2012; Arcidiacono and Lovenheim, 2016; Bleemer, 2020a). For example, critics of race-based affirmative action often argue that it harms its intended beneficiaries through a mismatch effect, whereby minority students who attend selective colleges would have fared better in terms of graduation rates and STEM degree completion by choosing less selective schools instead.[10] We contribute a suite of additional policy-relevant outcomes, including earnings, to inform this and other related questions about heterogeneous impacts, including whether low-income students benefit more or less than their higher-income peers from choosing selective colleges, and whether a student's own abilities interact with her institution to generate complementarities in the production of value-added. Our results suggest little heterogeneity in college value-added across these observable dimensions of student diversity, with the slight apparent mismatch effect for Black students driven by heterogeneity in alternatives (i.e., the availability of HBCUs) rather than significant heterogeneity in the value-added reaped by different students at any given university.[11]

In interpreting our results, it is important to keep in mind that the matched applicant research design may sacrifice some external validity on the altar of internal validity by limiting the identifying variation to enrollment choices within admission portfolios. This variation is particularly relevant for learning about

[9]Cunha and Miller (2014) offer a pioneering proof-of-concept demonstration of the matched applicant approach to estimating college value-added. We advance their exploratory study by deriving and interpreting value-added estimands in the presence of treatment effect heterogeneity; by empirically testing assumptions about selection bias and effect heterogeneity; by investigating predictors of value-added and early career dynamics; by estimating value-added on an expanded suite of outcomes and exploring their interrelationships; and by probing (mis)match effects and the role of HBCUs. We also find that Cunha and Miller (2014)'s grouping of all students who applied to four or more schools into a single admission portfolio, and keeping students with no admissions data in another single portfolio, appears to under-control for selection in their main estimates, as these two coarse portfolios comprise one fourth of their identifying sample (students in admission portfolios with treatment variation) and permit significant residual selection into college treatments, given the uncontrolled vertical heterogeneity of the students within them.

[10]For prominent legal articulations of this view, see the opinion of Clarence Thomas in *Fisher v. University of Texas at Austin*, 570 U.S. 297 (2013), and oral arguments by Antonin Scalia in *Fisher v. University of Texas at Austin*, 579 U.S. (2016).

[11]This result is similar in spirit to Angrist et al. (2019), who explore the roles of match effects versus treatment alternatives in explaining the impacts of selective high school admissions on student outcomes.

the consequences of 4-year college choices made after multiple admission letters arrive. This variation does not necessarily speak to impacts at other important margins, like admitting minimally qualified applicants just above a cutoff (e.g. Hoekstra, 2009; Zimmerman, 2014; Goodman et al., 2017; Sekhri, 2020), or intervening early enough to transform students' college ambitions (Hoxby and Turner, 2013; Dynarski et al., 2020) or pre-college human capital investments (Bodoh-Creed and Hickman, 2019; Akhtari et al., 2020).[12] Research designs that employ variation at other margins may not only involve different student subpopulations with potentially different treatment responses, but also involve categorically different counterfactuals to the treatment of interest, thus identifying different causal comparisons. Notably, in contrast to our relative comparisons of individual 4-year institutions, quasi-experimental designs that exploit discontinuities or policy changes in admission into particular 4-year universities (e.g. Hoekstra, 2009; Zimmerman, 2014; Goodman et al., 2017; Smith et al., 2020; Bleemer, 2020b; Black et al., 2020) combine other 4-year institutions, 2-year colleges, and non-enrollment into a composite counterfactual, reflecting the diverse fallback options of students who are exogenously denied admission. Treatment effects recovered by these designs may therefore reflect significant impacts of attending a 4-year instead of a 2-year institution (e.g. Reynolds, 2012; Mountjoy, 2021), and/or attending any college instead of none at all, in addition to any within-sector differences in the value-added of 4-year colleges, which we aim to isolate.

2 Setting and Data

In this section, we describe our setting and data sources. We also define our analysis sample and variables of interest, and we present summary statistics on students and the colleges they attend.

2.1 Setting, Sample, and Sources

Our data come from linking several administrative registries that span the entire state of Texas (UTD-ERC, 2021). Texas is the second largest U.S. state by population, land area, and GDP, with nearly 30 million residents and an economy that would rank 10th largest in the world as a sovereign nation. This immensity supports a comprehensive higher education system that provides a fruitful setting for studying value-added across a rich diversity of colleges and students.

The analysis sample begins with the population of students who graduated from a Texas public high school between 1999 and 2008.[13] We link several student registries maintained by the Texas Education Agency (TEA) to assemble pre-college data on these students' demographics, standardized test scores, academic preparation, attendance, disciplinary infractions, and high school campus. We then link these high school graduates to administrative application and admissions records from all Texas public 4-year universities, maintained by the Texas Higher Education Coordinating Board (THECB). To take advantage of this uniquely comprehensive repository of admissions data, we focus our analysis sample on students who begin college at one of these 30 Texas public universities.[14] We then follow these students longitudinally through

[12] Our variation is also limited to in-state public institutions. Most students in the U.S. attend such colleges, and the Texas public higher education market is one of the largest and most diverse in the country, but our variation may not speak to comparisons that involve significantly more expensive and well-resourced private colleges (e.g. Cohodes and Goodman, 2014).

[13] The 1999 cohort is the first to have college application and admission data, and we stop at the 2008 cohort to observe at least 10 years of post-high-school earnings for all sample members. 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).

[14] We do not observe applications or admissions records at private Texas colleges. Our 30 public universities enroll roughly 83% of 4-year college-goers in Texas. This set also excludes small and young public institutions with limited or no first-year enrollment during our sample period. We do not observe enrollments or degrees at for-profit colleges, which accounted for less than 5 percent of national enrollments during our sample period (National Center for Education Statistics, 2018).

administrative registries of all college enrollment spells and degree completions (also using THECB data) spanning all public and private non-profit postsecondary institutions in the state, allowing us to observe educational outcomes inclusive of transfer. Finally, we merge in quarterly earnings records for these students from the Texas Workforce Commission (TWC) that cover all Texas employees subject to the state unemployment insurance system.[15] We supplement the student-level data with college-level institutional characteristics from the Integrated Postsecondary Education Data System (IPEDS) and college-level intergenerational mobility statistics from Chetty et al. (2020).

As with any data spanning a particular state, we are unable to track activity outside of Texas. This means we will not observe college enrollments, degree attainments, and earnings for students who outmigrate. Fortunately in our case, Texas has the lowest outmigration rate of any state in the U.S. (Aisch et al., 2014). Auxiliary data from the National Student Clearinghouse, which tracks nationwide college enrollments, show that less than 4 percent of the 2008 and 2009 cohorts of Texas public high school graduates left the state to attend college (Mountjoy, 2021). With respect to earnings, we are unable to distinguish nonemployment within Texas from unobserved employment outside the state, so we condition the earnings outcomes on having at least some positive Texas earnings. We conduct a battery of robustness checks in Section 4.3.5 and fail to find any evidence that selection into the earnings sample biases the value-added estimates.

2.2 Variable Definitions

Pre-College Covariates—Our pre-college student covariates from the TEA high school files fall into four categories: demographics, test scores, other measures of academic preparation, and behavioral measures of attendance and discipline. For demographics, we observe each student's gender, race, and free or reduced-price lunch (FRPL) eligibility as a proxy for low family income.[16] We also observe unique indicators for each high school, which we use in robustness checks to flexibly capture environmental influences of the schools, neighborhoods, and local labor markets that students experience prior to college entry. We measure achievement on mandatory statewide 10th grade math and reading exams by standardizing the raw scores to have mean zero and standard deviation one within each cohort. We also observe SAT scores for the subsample of students who graduate high school in 2003-2008. Other measures of academic preparation include the number of Advanced Placement, International Baccalaureate, and other advanced courses taken, as well as indicators for whether a student was ever at risk of dropping out of high school and (from the THECB admissions data) whether she was in the top 10% of her high school class.[17] Finally, we measure attendance behavior as the share of total school days in 10th-12th grade that a student is not absent, and we measure disciplinary infractions as the average number of days suspended in 10th-12th grade, with both measures standardized to have mean zero and standard deviation one within each cohort.

Application/Admission Portfolios—To compare students who apply to and are admitted by the same institutions, we construct unique indicators for each distinct portfolio of observed college applications and

[15] 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.

[16] The three race categories are white, Black, and Hispanic. Asians are pooled with whites and Native Americans are pooled with Hispanics due to small shares at several colleges.

[17] We do not observe a direct measure of high school grade point average (GPA) beyond this discrete indicator for Top Ten Percent program eligibility, as neither the TEA nor the THECB collect high school GPA on a statewide basis. Under the Top Ten Percent program, students in the top decile of their local high school GPA distribution earn automatic admission to any public Texas university they apply to. We show in Section 4.3 that the matched applicant research design does not appear to be sensitive to the presence of this program, as interacting admission portfolios with Top Ten status barely moves the main estimates. This is likely because Top Ten students still tend to apply to a variety of schools before making a final enrollment decision, and thus assemble admission portfolios that are still informative about their vertical potential.

admissions decisions. With respect to each college there is a ternary outcome—*do not apply*, *apply and get rejected*, or *apply and get admitted*—and there are 30 colleges, so theoretically there are $3^{30} \approx 2.06 \times 10^{14}$ possible portfolios. In practice, we observe 21,331 distinct portfolios in our data, since most students only apply to a small number of schools. We show in Section 4.3 that we can reduce the dimensionality of these portfolios without affecting the value-added estimates by discarding information about rejections; that is, by reducing the college-specific outcome to a binary of *do not get admitted* vs. *apply and get admitted*, where *do not get admitted* could be due either to not applying or applying and getting rejected.[18] There are 8,002 such distinct admissions portfolios in the data. Of these, 4,689 are unique to a single student, and thus do not contribute to identification in our portfolio fixed effects approach; neither do the 30 portfolios, one for each institution, in which students are only admitted to that institution and thus generate no within-portfolio enrollment variation. The remaining 3,283 admission portfolios with at least two students and at least two colleges from which to choose form the basis of our research design, which compares outcomes of students who reveal themselves to be sufficiently qualified to get admitted to the same set of schools but end up making different enrollment choices. We describe these choices in more detail in Table 2 below.

Treatments—We construct mutually exclusive and exhaustive treatment indicators for each of the 30 public universities in Texas, with UT-Austin omitted as the comparison institution against which all the others are measured. We define each student's treatment using their first college enrollment starting in the fall after high school graduation.[19] In our causal framework, it is important to define the treatment as the first college attended, rather than the modal/longest college attended (e.g. Chetty et al., 2020), for two reasons. First, starting college at one institution rather than another likely has a causal impact on how long a student stays enrolled and whether she transfers elsewhere.[20] We find evidence of this in Section 7, implying that persistence and transfer should be considered endogenous outcomes along a causal chain that begins with the initial college treatment. Second, within a college, students who remain at that college as their modal institution may differ systematically from students who transfer elsewhere, and this selection pattern may vary across colleges. Indeed, in Section 4.3.4 below we find that while student covariates are well-balanced across initial college treatments conditional on admission portfolios, redefining the treatment as the modal/longest college attended introduces substantial covariate imbalances. The lesson from these results is that our matched applicant research design is well-suited for studying the consequences of initial college enrollment choices, i.e. the "intention to treat" (ITT) effect of enrolling in each college, with subsequent persistence and transfer behavior analyzed as endogenous outcomes.

Academic Outcomes—We study the relative value-added of colleges on several academic outcomes along the path from a student's first college enrollment to her last. To study endogenous transfer behavior, we construct an indicator for whether a student subsequently enrolls in a different institution from her initial treatment.[21] We measure BA completion as an indicator for obtaining a baccalaureate degree from any public or private institution in the THECB registries, which avoids the penalization of transfer students that commonly occurs in datasets that can only track students within their initial institution of enrollment.

[18] The binary portfolio specification effectively condenses multiple ternary portfolios into a single portfolio. For example, suppose Tom applies to colleges A, B, and C, and is admitted to A and C, while Harry applies only to A and C, and is admitted to both. The ternary specification will consider these as two separate portfolios, while the binary specification considers them as the same one. In Section 4.3 we show that both methods yield nearly identical estimates, suggesting that, conditional on colleges' collective acceptance decisions, applications that lead to rejections contain no additional identifying information.

[19] We follow Andrews et al. (2014) and ignore summer terms when defining the institution of first enrollment, as some students take supplemental summer courses at local campuses before starting their main degree program.

[20] See additional discussions of this "attribution issue" in Cunha and Miller (2014) and Hoxby (2019).

[21] To avoid mis-classifying supplemental coursework and dual enrollment, episodes in which the student eventually returns to the initial institution are not coded as transfers.

We also capture dynamics in time-to-degree by constructing separate indicators for completing a BA within 4, 6, and 8 years. For a more continuous measure of persistence, we also calculate years of college completed.[22] Finally, we observe the majors of students who complete a BA, so we construct binary indicators for completing a STEM (science, technology, engineering, or mathematics) degree within 4, 6, and 8 years.[23]

Labor Market Outcomes—Our main labor market outcome is annualized earnings around ten years out from college entry, which we measure as follows. We begin by summing earnings within each person-quarter across different jobs and deflating by the quarterly U.S. consumer price index (base year 2015). We winsorize at the top and bottom 0.1^{th} percentiles and average the non-missing earnings quarters within person over the range of 8-10 years out from college entry, roughly ages 27-29, which are the oldest ages that all of our analysis cohorts have in common.[24] We arrive at an annualized measure by multiplying this quarterly average by 4. We also construct a panel of person-year earnings to explore earnings trajectories. Since the TWC data contain industry of employment, we also construct an indicator for working in the Texas oil and gas industry to explore this as a potential mediator of colleges' earnings impacts. Finally, to probe the robustness of the earnings value-added estimates to sample selection from outmigration or nonemployment, we construct an indicator for simply appearing in the TWC earnings data during our measurement ages and conduct checks with this sample indicator as an outcome.

2.3 Summary Statistics

Table 1 summarizes this data. Our ten cohorts of Texas public high school graduates who begin college at one of the 30 Texas public universities comprise 422,949 students, of whom 54 percent are female, 24 percent are low-income (eligible for free or reduced-price lunch in high school), 12 percent are Black, and 23 percent are Hispanic. 60 percent of students only apply to one Texas public university, and 69 percent are only admitted to the school they end up attending. These students do not contribute to the identification of college value-added in our matched applicant approach, since there is no variation in treatment within their single-college admission portfolios. The second column of Table 1 therefore describes the identifying subsample of 125,876 students who do face at least two choices in their admission portfolio, and share that portfolio with at least one other student. These students are demographically similar to the full sample of college-goers, but are moderately higher achievers in terms of test scores, attendance, and outcomes, as would be expected from students admitted to more colleges.

In terms of treatments, roughly one quarter of the full sample, and nearly one third of the identifying sample, begin college at one of the two flagships, UT-Austin and Texas A&M, which are among the largest universities in the country. The University of Texas (UT) System included 7 other standalone institutions during our sample period (UT-San Antonio, UT-Pan American, UT-Arlington, UT-El Paso, UT-Dallas, UT-Tyler, and UT-Permian Basin), and the Texas A&M (TAMU) System included 8 other standalone institutions (Tarleton State, TAMU-Corpus Christi, Prairie View A&M, TAMU-Kingsville, West Texas A&M, TAMU-Commerce, TAMU-International, and TAMU-Galveston). Three other public systems also oversee multiple standalone institutions: the Texas State System (Texas State-San Marcos, Sam Houston

[22]Specifically, years of college completed in our college-going sample range from 0 to 4. To be classified as completing 0 years: complete high school but less than one year of college. To complete 1: enroll in college with 2nd year standing, or complete a certificate, or complete the academic core requirement at a community college. 2: enroll in college with 3rd year standing, or complete an associate's degree. 3: enroll in college with 4th year standing. 4: complete a bachelor's degree.

[23]STEM majors are defined by the federal government: `https://studyinthestates.dhs.gov/stem-opt-extension-overview`.

[24]We show below in Section 4.3.5 that the main value-added estimates are highly correlated with estimates from the subsample of older cohorts for whom we can measure earnings at later ages, and in Section 6 we document that the earnings effects of attending a more selective college flatten out by 8-10 years out.

Table 1: Summary Statistics

	Full sample	Identifying subsample		Full sample		Identifying subsample	
	Mean (SD)	Mean (SD)		Share	Count	Share	Count
Covariates			*Treatments*				
Female	.544	.566	TX A&M (TAMU)	.13	54,953	.164	20,670
Low-income (FRPL)	.241	.221	UT-Austin	.124	52,508	.155	19,526
Black	.121	.129	TX Tech	.077	32,371	.081	10,172
Hispanic	.227	.21	UT-San Antonio	.065	27,569	.072	9,053
10th grade test score	0 (1)	.14 (.882)	North TX	.057	24,146	.055	6,864
HS attendance (std.)	0 (1)	.076 (.936)	TX State-San Marcos	.056	23,686	.076	9,577
			Houston	.056	23,528	.055	6,983
Applied to:			Stephen F. Austin	.041	17,372	.04	4,973
1 school	.601	0	Sam Houston State	.037	15,703	.034	4,340
2 schools	.233	.538	UT-Pan American	.035	15,000	.022	2,749
3+ schools	.167	.462	UT-Arlington	.035	14,595	.039	4,866
			UT-El Paso	.034	14,361	.01	1,223
Admitted to:			Angelo State	.025	10,585	.015	1,896
1 school	.691	0	Lamar	.025	10,569	.015	1,912
2 schools	.212	.712	Tarleton State	.023	9,791	.014	1,710
3+ schools	.097	.288	TAMU-Corpus Christi	.02	8,550	.023	2,902
			TX Southern	.018	7,736	.01	1,267
Academic Outcomes			Prairie View A&M	.017	7,351	.017	2,112
Ever transfer	.271	.276	TAMU-Kingsville	.016	6,675	.014	1,748
Years completed	2.89 (1.52)	3.18 (1.36)	West TX A&M	.015	6,498	.007	916
BA within 4 years	.274	.325	UT-Dallas	.015	6,453	.021	2,628
BA within 6 years	.592	.677	Midwestern State	.014	5,873	.008	1,053
BA within 8 years	.652	.733	Houston-Downtown	.012	5,196	.007	893
STEM degree	.13	.154	TAMU-Commerce	.01	4,293	.007	920
			TX Woman's	.009	4,001	.01	1,270
Earnings Outcomes			TAMU-International	.008	3,541	.007	835
Positive earnings	.848	.851	UT-Tyler	.008	3,248	.008	945
Annualized earnings	44,834 (28,485)	47,703 (29,057)	TAMU-Galveston	.007	2,797	.009	1,148
			Sul Ross State	.005	2,037	.003	417
			UT-Permian Basin	.005	1,963	.002	308
Students	422,949	125,876					
Admission portfolios	8,002	3,283					

Notes: The *full sample* is students who graduate from a Texas public high school in 1999-2008 and enroll at one of the Texas public universities listed in the Treatments column. The *identifying subsample* is students in admission portfolios with at least one other student and at least two colleges to choose from. Treatments are defined as the first college attended in the academic year after high school graduation. *FRPL* denotes free or reduced price lunch eligibility in high school. *10th grade test score* is an average of math and reading scores and *High school attendance* is measured as days not absent in 10th-12th grade; both are standardized to have mean 0 and standard deviation 1. Academic outcomes are measured 6 years out from college entry unless otherwise noted. Earnings outcomes are measured 8-10 years from college entry in real 2015 dollars.

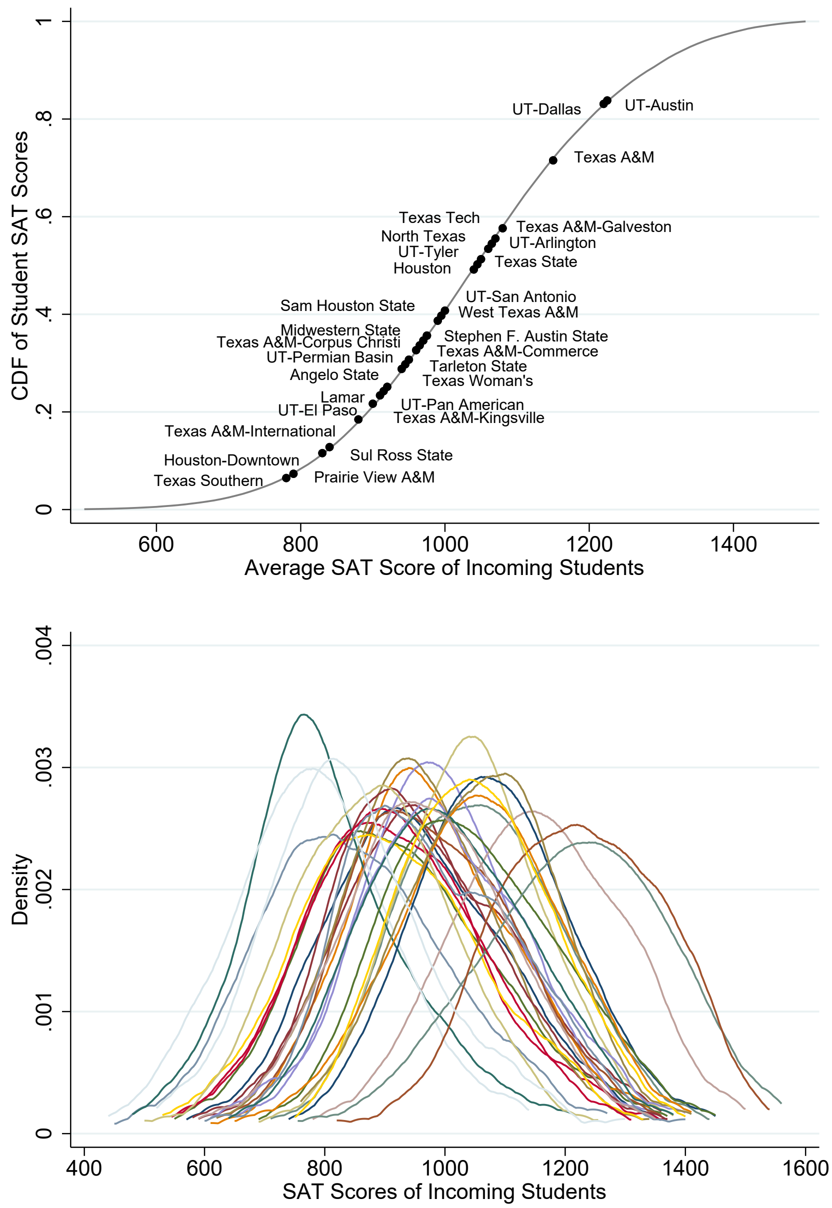
State, Lamar, and Sul Ross State), the Texas Tech System (Texas Tech and Angelo State), and the University of Houston System (Houston and Houston-Downtown). The University of North Texas belongs to its own system, and the remaining universities—Stephen F. Austin State, Midwestern State, Texas Southern, and Texas Woman's—operate independently of any system.

The top panel of Figure 1 illustrates the vast diversity of these institutions in terms of selectivity. Each school's mean SAT score of incoming students is plotted on the student-level CDF of SAT scores, showing that our campus averages span nearly 80 percentiles of the student-level score distribution.[25] Despite these large cross-campus differences in means, the bottom panel of Figure 1 shows that the majority of the student-level variation in SAT scores is actually within campuses; a variance decomposition attributes 75% within campuses and 25% across campuses. Such overlapping support provides the kind of variation necessary for identifying (mis)match effects of a given institution across the heterogeneous students who attend it.

In terms of outcomes, Table 1 shows that beginning college is far from synonymous with finishing a degree, especially on time: only 27% of students complete a BA within 4 years of entry (33% for the identifying

[25] Coincidentally, while we do not observe schools at the very top of the selectivity distribution, this missing complement of our data is precisely the range of selectivity that Dale and Krueger (2002, 2014) focus on in their main dataset, the College and Beyond survey. Our results thus complement and extend their work by studying the lower 80 percent of the selectivity distribution not captured in the highly selective schools of the C&B survey.

Figure 1: Selectivity Across Colleges and SAT Score Variation Within Colleges



Notes: The top panel positions the average SAT score of incoming students at each college in our sample within the overall CDF of student-level SAT scores. The bottom panel plots the distribution of SAT scores within each college.

sample with 2+ admissions), with completion rates climbing to 59% (68%) after 6 years and 65% (73%) after 8 years, in line with national statistics. On average, enrollees complete 2.9 (3.2) years of college, and 27% (28%) transfer to another Texas institution at least once, including to private colleges and 2-year community colleges.[26] Thirteen (fifteen) percent of students successfully complete a STEM degree, such that .13/.59 = 22% (.15/.68 = 22%) of all 6-year BA completions are in STEM fields. Finally, 85% of our students (both in the full sample and identifying sample) appear in the earnings data 8-10 years out from college entry, with real annualized earnings averaging around \$45,000 (\$48,000) with a standard deviation of nearly \$30,000.

[26] Auxiliary data from the National Student Clearinghouse, available for students who began at Texas public colleges in 2011 and 2012, show that only 2.5% transfer out-of-state, and less than 1% earn an out-of-state degree, so our within-Texas estimates are unlikely to be meaningfully biased by out-of-state transfers.

3 Identifying and Interpreting Relative College Value-Added

To set the stage for identifying relative value-added of individual colleges, let Y_{ij} denote the potential outcome of student i if she were to attend college $j \in \mathcal{J}$. We can decompose Y_{ij} into two components:

$$Y_{ij} = \underbrace{\alpha_i}_{\substack{\text{Student effect} \\ \text{(Ability, ambition, advantage)}}} + \underbrace{\nu_{ij}}_{\substack{\text{School effect} \\ \text{(Value-added)}}} \tag{1}$$

where $\alpha_i \equiv \frac{1}{|\mathcal{J}|} \sum_{k \in \mathcal{J}} Y_{ik}$ and $\nu_{ij} \equiv Y_{ij} - \alpha_i$. The student effect α_i is student i's average potential outcome across all schools; it vertically differentiates students, reflecting traits like ability, ambition, and advantage that generally drive student i's outcome regardless of the specific school she actually attends.

Our primary interests are the school effects ν_{ij}, which capture each school j's causal value-added on student i's outcome relative to the average school. Note that the vast majority of papers in the value-added literature—whether concerning colleges (Cunha and Miller, 2014; Hoxby, 2019), teachers (Chetty et al., 2014a), K-12 schools (Angrist et al., 2017, 2020), firms (Abowd et al., 1999), industries (Krueger and Summers, 1988), hospitals (Gowrisankaran and Town, 1999), regional health care systems (Finkelstein et al., 2016), etc.—focus on identifying models that restrict the value-added of a given treatment j to be homogeneous across individuals, i.e. $\nu_{ij} = \nu_j$.[27] One of our contributions in this section is to depart from the value-added canon and refrain from imposing homogeneity at the outset, which allows us to clarify exactly what non-experimental comparisons identify in a world of unrestricted heterogeneity. These derivations, in turn, highlight why homogeneity is such a powerful and pervasive assumption in the value-added literature. But they also reveal weaker conditions under which policy-relevant value-added parameters can be identified, even in the presence of student-college match effects, sorting on gains, and limited treatment support. We therefore conclude this section by compiling these results into a framework for interpreting value-added estimates under various assumptions about treatment effect heterogeneity and sorting behavior, setting the stage for the baseline estimates in Section 4.

3.1 Observed Outcomes and Selection Bias

The most prominent outcome-based comparisons of colleges—pervasive in popular media, college guides, government statistics, and in many states' postsecondary funding formulas—simply compare the raw mean outcomes of students who attend each school. Figure 2 conducts such an exercise across our thirty Texas public universities for two key outcomes: BA completion (top panel) and earnings (bottom panel). The leftmost points on each plot (solid diamonds) indicate the mean outcomes of students who attend each school, measured relative to UT-Austin (signified by the vertical line at zero). We see enormous disparities in these outcomes across campuses. Students who begin college at Texas Southern University are 67 percentage points less likely to complete a BA within 6 years compared to students who start at UT-Austin, who have an 82% completion rate. The average Texas Southern student goes on to earn \$27,000 less each year 8-10 years after college entry compared to the average UT-Austin student, who earns \$55,975.

These large differences in raw outcomes, of course, need not reflect differences in causal college value-added if different types of students systematically attend different schools. Letting $S_i = j$ indicate that

[27] See Bonhomme et al. (2019), Lamadon et al. (2019), Hull (2020), Abdulkadiroglu et al. (2020), and Goldsmith-Pinkham et al. (2021) for recent approaches relaxing value-added homogeneity.

student i actually attended school j, we can use (1) to write student i's observed outcome Y_i as

$$Y_i = \alpha_i + \sum_{j \in \mathcal{J}} \nu_{ij} \, \mathbb{1}[S_i = j], \tag{2}$$

and decompose the raw mean outcome comparison of school j to school k as follows:

$$\underbrace{\mathbb{E}[Y_i|S_i = j] - \mathbb{E}[Y_i|S_i = k]}_{\text{Raw outcome comparison of } j \text{ vs. } k} = \underbrace{\underbrace{\mathbb{E}[\nu_{ij}|S_i = j]}_{ATT_j} - \underbrace{\mathbb{E}[\nu_{ik}|S_i = k]}_{ATT_k}}_{\substack{\text{Relative value-added of } j \text{ vs. } k \\ \text{among their respective students}}} + \underbrace{\mathbb{E}[\alpha_i|S_i = j] - \mathbb{E}[\alpha_i|S_i = k]}_{\text{Selection bias}}. \tag{3}$$

Comparing the mean outcomes of students who attend college j to those who attend college k thus conflates causal and non-causal components, and (3) clarifies exactly what these components are. $ATT_j \equiv \mathbb{E}[\nu_{ij}|S_i = j]$ is the average treatment effect on the treated at school j: the causal value-added of attending j, relative to the average school, among the students who actually attend j. Likewise, $ATT_k \equiv \mathbb{E}[\nu_{ik}|S_i = k]$ is the average value-added of k, relative to the average school, among students who actually attend k. Their difference, $ATT_j - ATT_k$, therefore answers the following question: how effective is college j at boosting its own students' outcomes, relative to the effectiveness of college k at boosting *its* own students' outcomes?

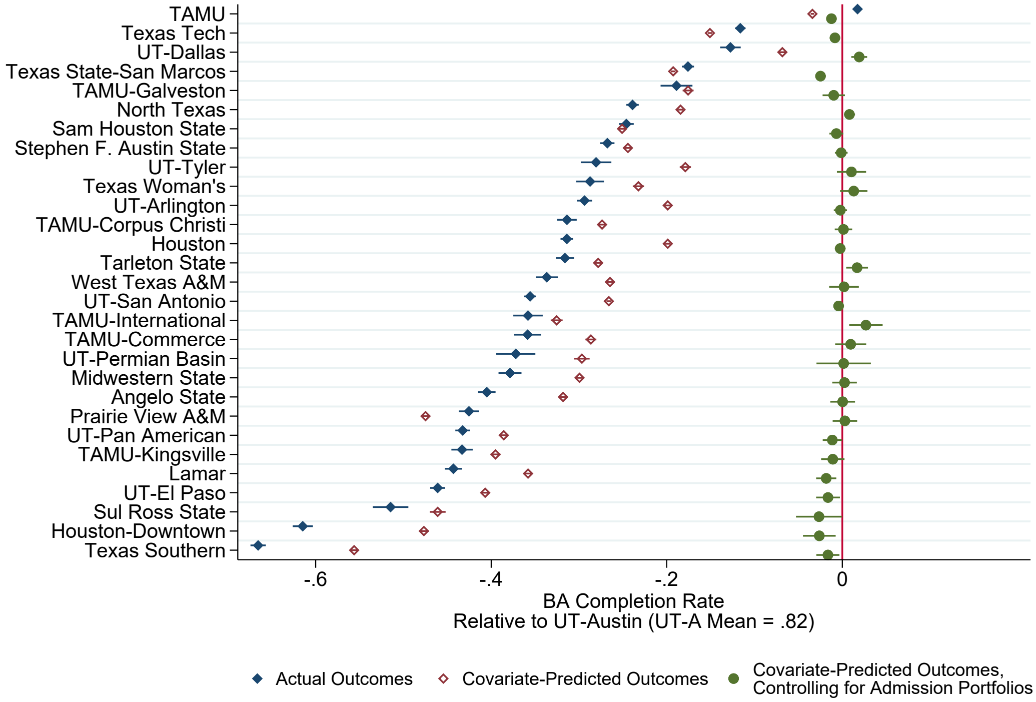
We explore this value-added contrast in more detail later in this section, but first we tackle the key threat to identifying it from observed outcome comparisons: selection bias. This is the remaining component of the decomposition in (3), and typically the principal identification concern in the value-added literature. Students who choose school j may have systematically higher or lower levels of ability, ambition, and advantage relative to students who choose school k, captured by their different mean values of α_i. Since α_i is, by definition, unaffected by the school a student actually attends, selection bias can generate mean outcome differences at j vs. k unrelated to any causal influences of the colleges themselves. To gauge the potential severity of such bias, the second set of points in Figure 2 (hollow diamonds) replace each student's actual outcome with her *predicted* outcome based on covariates measured prior to college entry.[28] The outcomes that each college's incoming students are predicted to have, based entirely on their pre-college characteristics, follow a remarkably similar pattern as the actual outcomes. This clearly illustrates how different colleges enroll systematically different types of students, and suggests that selection bias plays a dominant role in generating the raw outcome differences we observe across colleges.

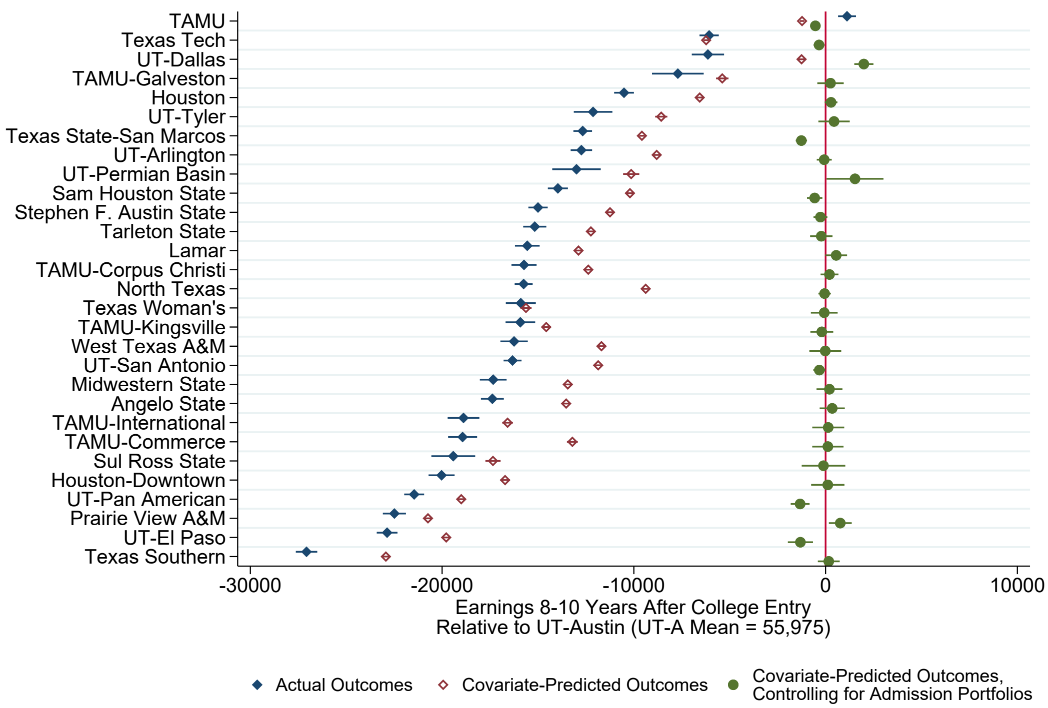
3.2 Addressing Selection Bias: Admission Portfolios as Proxies

The structure of Equation (2) evokes a student fixed effects approach to addressing selection bias: in a regression of student outcomes on college treatment indicators, include unique indicators for each student to absorb the student effect α_i. Unlike many other value-added environments, however, our setting does not feature a panel structure of outcomes driven by contemporaneous treatments that vary within person over time, like the earnings histories of individual workers who switch across firms (e.g. Abowd et al., 1999) or the test score histories of individual students who face a series of different teachers (e.g. Rockoff, 2004). While many college students do transfer across multiple institutions, we cannot logically observe outcomes like BA completion and adult earnings among our traditional-aged college-goers both before and after these

[28] Specifically, we use the predicted values from an OLS regression of each outcome on the following set of pre-college covariates: quartics in 10th grade math and reading scores, attendance, days of disciplinary suspension, and number of advanced high school courses; binary indicators for free/reduced price lunch eligibility, gender, at-risk status, and top decile GPA; and fixed effects for each race, high school cohort, and high school campus.

Figure 2: Raw Outcomes, Predicted Outcomes, and Within-Portfolio Balance Across College Treatments





Notes: Each set of point estimates and robust 95% confidence intervals come from regressions of individual student outcomes on college treatment indicators (and cohort fixed effects), omitting UT-Austin as the reference treatment (signified by the vertical line at zero). The UT-Austin outcome mean appears in parentheses below each plot. *Covariate-Predicted Outcomes* replace actual outcomes with the predicted values from a separate OLS regression of each outcome on the following set of pre-college covariates: quartics in 10th grade math and reading scores, attendance, days of disciplinary suspension, and number of advanced high school courses; binary indicators for free/reduced price lunch eligibility, gender, at-risk status, and top decile GPA; and fixed effects for each race, high school cohort, and high school campus. The rightmost specification regresses these covariate-predicted outcomes on the college treatment indicators and solely controls for each distinct portfolio of college admissions (and cohort fixed effects).

enrollment changes. Hence we cannot include indicators for each student, or more generally exploit within-student variation, to directly control for the individual student effects α_i.[29]

Instead, we propose to proxy for α_i at a coarser, but still rather detailed, level with indicators for each distinct portfolio of college applications and admissions. Intuitively, the idea behind this approach is that when a student assembles a portfolio of applications, she provides a high-dimensional signal of her abilities, ambitions, and advantages, while admissions officers layer on an additional high-dimensional signal of her potential through their collection of acceptance and rejection decisions.[30] This intricate filtering process groups students with similar vertical potential—i.e. values of α_i—into similar portfolios of college applications and admissions, which may substantially reduce the scope for selection bias *within* each portfolio.[31]

Concretely, consider the population projection of the student effect α_i onto a mutually exclusive and exhaustive set of indicators for each admission portfolio A_i, indexed by $a \in \mathcal{A}$:

$$\alpha_i = \sum_{a \in \mathcal{A}} \phi_a \mathbb{1}[A_i = a] + \eta_i, \tag{4}$$

where $\phi_a = \mathbb{E}[\alpha_i | A_i = a]$ and $\mathbb{E}[\eta_i | A_i] = 0$ by construction. The key condition of the proxy approach is that a student's admission portfolio A_i sufficiently proxies for α_i such that the residual within-portfolio variation in vertical student potential, η_i, does not vary systematically with where students ultimately decide to enroll, S_i. This condition is formalized in the following assumption:

Assumption 1 *Admission portfolios proxy for student fixed effects:* $\mathbb{E}[\eta_i | S_i, A_i] = \mathbb{E}[\eta_i | A_i]$.

Assumption 1 allows us to eliminate selection bias by conditioning on observable admission portfolios, which serve as coarse but sufficient proxies for unobservable student fixed effects:

$$\underbrace{\mathbb{E}[\alpha_i | S_i = j, A_i = a] - \mathbb{E}[\alpha_i | S_i = k, A_i = a]}_{\text{Selection bias within portfolio } a} = \mathbb{E}[\phi_a + \eta_i | S_i = j, A_i = a] - \mathbb{E}[\phi_a + \eta_i | S_i = k, A_i = a] = 0,$$

where the first equation plugs in (4) and the second imposes Assumption 1.

3.2.1 Testing Assumption 1

The final set of results in Figure 2 directly test Assumption 1: do students with the same admission portfolio systematically sort into different colleges by vertical potential? To fix ideas, consider a simple balance test where a high school test score is regressed on the college treatment indicators to see whether pre-college ability is balanced across college treatments. Appendix Figure A.2 shows that without any controls, 10th grade math and reading scores are massively unbalanced across colleges, yet strikingly align when controlling solely for admission portfolio fixed effects. This demonstrates a balanced profile of test scores across college treatments within each admission portfolio. However, if standardized test scores are not a sufficient statistic for vertical student potential, this simple exercise ignores the need for balance on a wider array of student

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

[30] See Kapor (2020), for example, for a model of college admissions in which admissions officers observe, and reveal through their decisions, components of a student's "caliber" about which the student herself may have only imperfect information. Figure A.5 below provides evidence that admission decisions deliver additional selection correction beyond applications alone.

[31] A related strategy, employed frequently in the value-added literature, proxies for α_i with lagged values of the outcome of interest, like controlling for a student's prior-year test score when estimating teacher value-added on the current year's score (e.g. Chetty et al., 2014a). See Riehl et al. (2019) for an innovative application of this strategy to colleges in Colombia, where students take entrance and exit exams with common components. Our setting precludes this strategy, again given our logical inability to observe outcomes like BA completion and adult earnings prior to college entry for traditional-aged college-goers.

Table 2: Describing College Choices within Admission Portfolios

	Identifying sample
	Mean (SD)
Selectivity	
Enroll in most selective school in admission portfolio	.545
Selectivity range across schools in admission portfolio (SAT points)	116 (77.8)
Flagship	
Admitted to UT-Austin	.268
Enroll in UT-Austin if admitted	.579
Enroll in UT-Austin if admitted and portfolio does not include Texas A&M	.603
Value-Added	
Enroll in highest earnings value-added school in admission portfolio	.469
Earnings value-added range across schools in admission portfolio (USD)	1,843 (1,228)
Students	125,876
Admission portfolios	3,283

Notes: The *identifying sample* is students in admission portfolios with at least one other student and at least two colleges. Selectivity is defined as the mean SAT score of a college's incoming students. Value-added is estimated using our main specification in Section 4.

attributes. Thus, the rightmost estimates in Figure 2 (solid circles) test for such balance by collecting our rich array of pre-college covariates—going beyond test scores to include high school coursework, attendance, discipline, demographics, and high school fixed effects to flexibly capture academic and socioeconomic pre-college environments—into an index of predicted outcomes. As with the hollow diamonds in Figure 2, we regress these predicted outcomes on college treatments, but now control for admission portfolio fixed effects. The results provide rather stark empirical evidence in support of Assumption 1. While vertical student potential, as measured by covariate-predicted outcomes, varies enormously across campuses in unconditional comparisons (hollow diamonds), controlling solely for admission portfolio fixed effects dramatically reduces these differences to rough balance across all of our treatment colleges (solid circles).[32] Students who share the same portfolio of college admissions thus appear to have similar pre-college potential regardless of where they actually enroll, in line with Assumption 1.

While Figure 2 shows that admission portfolio controls nearly eliminate imbalances in observable student potential across college treatments, a lingering concern with the matched applicant approach is that it relies on "the small share of students who make what is a very odd choice" (Hoxby, 2009): enrolling in a less selective college when admitted to a more selective alternative. Table 2 directly investigates the nature of this identifying variation in our sample, and finds that such choices are actually far more common than the concern supposes. The first section of Table 2 shows that among students in our identifying sample, i.e. those admitted to at least two colleges, only 55% end up enrolling in their most selective option, with the average admission portfolio spanning a selectivity range of 116 SAT points (SD 78). The second section of Table 2 shows that this phenomenon is not confined to students choosing solely among less selective institutions. Of the 27% of students in our identifying sample who are admitted to UT-Austin, the most selective campus in our students' choice sets, only 58% choose to enroll there. And the remainder are not simply sidestepping en masse over to the other flagship: conditional on being admitted to UT-Austin and *not* being admitted to Texas A&M, still only 60% of students choose UT-Austin, meaning 40% forgo it in favor of a less selective, non-flagship school. Very similar patterns also arise among Top Ten Percent students in our sample, who are guaranteed admission to every Texas public university they apply to. Only 30% of them end up enrolling at UT-Austin, and 25% at Texas A&M, leaving fully 45% who decline to exercise their flagship entitlement and instead fan out across the 28 non-flagship campuses in our sample.

[32] The small covariate imbalances that remain do not drive the main value-added estimates, as we show in Figure 5 below.

Thus, nearly half of the students in our identifying sample choose a less selective college than their most selective option, suggesting this identifying variation is not merely an odd choice confined to a small faction of quirky students. As we discuss further below, students choose colleges based on a motley array of considerations, well beyond selectivity alone. Importantly, if the nearly half of students who turn down their most selective option were negatively (or positively) selected on vertical potential, we might expect to see some of this selection show up as covariate imbalances across treatments in our research design. Instead, we see the rather striking balance results of Figure 2. Furthermore, if these students were negatively (or positively) selected on unobservable vertical potential, captured neither by admission portfolios nor our rich set of covariates, we would expect the value-added estimates to be biased towards finding a positive (negative) "return" to attending the more selective colleges in our sample. Instead, in Section 6 we estimate an initially positive earnings return to selectivity that fades to zero after a few years in the labor market—a dynamic that is difficult to dismiss as selection on unobservables, either positive or negative, into more selective colleges. Taken together, these results lend empirical credence to the notion in Assumption 1 that students facing the same choice set of college admissions do not further sort into institutions systematically by vertical traits like ability, ambition, and advantage.

3.3 Value-Added for Whom? Match Effects, Sorting Gains, & Limited Support

With selection bias addressed by Assumption 1, we now turn to a more subtle identification concern that has received far less attention in the value-added literature: "horizontal" student sorting on the basis of heterogeneous treatment effects, i.e. match effects. In our framework, value-added ν_{ij} can vary both across colleges j *and* across individual students i, generalizing beyond the typical assumption of homogeneous treatment effects ($\nu_{ij} = \nu_j$) in the value-added literature. Our analysis in this subsection makes explicit the implicit simplifications afforded by homogeneity in previous work, but more importantly demonstrates that it is much stronger than needed, being sufficient but not necessary for recovering value-added comparisons.

For more concise notation going forward, let $\mathbb{E}[W_i|S_i = j, A_i = a] \equiv \mathbb{E}[W_i|j, a]$ denote the conditional expectation of some variable W_i among students who enroll at school $S_i = j$ and share the admission portfolio $A_i = a$. Returning to Equation (3), we now condition on portfolio $A_i = a$ and apply Assumption 1 to yield

$$\underbrace{\mathbb{E}[Y_i|j,a] - \mathbb{E}[Y_i|k,a]}_{\substack{\text{Outcome comparison of } j \text{ vs. } k \\ \text{within admission portfolio } a}} = \underbrace{\underbrace{\mathbb{E}[\nu_{ij}|j,a]}_{ATT_j^a} - \underbrace{\mathbb{E}[\nu_{ik}|k,a]}_{ATT_k^a}}_{\substack{\text{Relative value-added of } j \text{ vs. } k \\ \text{among their respective students in } a}} + \underbrace{\mathbb{E}[\alpha_i|j,a] - \mathbb{E}[\alpha_i|k,a]}_{\substack{\text{Selection bias within portfolio } a \\ = \mathbf{0} \text{ by Assumption 1}}} \tag{5}$$

Comparing the mean outcomes of students within admission portfolio a who attend college j vs. college k thus identifies $ATT_j^a - ATT_k^a$. This is a well-defined difference of two causal effects reaped by students in portfolio a: each ATT answers the question of how much value that college adds to its own students' outcomes relative to the average college. Each within-portfolio pairwise comparison $\mathbb{E}[Y_i|j, a] - \mathbb{E}[Y_i|k, a]$ available in the data therefore tells us about the relative value-added of j vs. k among their respective students who shared the same admission portfolio a.

From a policy perspective, the ATT comparisons identified in (5) yield valuable *status quo* assessments of relative college value-added. They compare the gains reaped by the students who actually apply and get admitted to a given portfolio of colleges and then choose to enroll in each school. At minimum, under only Assumption 1, our matched applicant approach can identify such status quo assessments. The relevance of these parameters for more ambitious policy questions, however, may not be immediately clear in the presence of match effects, i.e. heterogeneity in the value-added of a given college across different students. Specifically, we can first ask whether the ATT comparison identified in (5) is informative about the causal effect of attending college j instead of college k for the average student in portfolio a, or whether it can only be interpreted as a difference of separate effects that are each "local" to the disjoint subpopulations of students who actually decide to sort into each school. Second, we can ask whether aggregating these within-portfolio comparisons across portfolios can inform how the average college-chooser in the population would fare at each college relative to a common benchmark, in line with the typical (but in fact rather ambitious) aim of a value-added exercise.

To gain clarity on these questions, first consider the causal effect of attending college j instead of college k for the average student facing that choice in admission portfolio a, unconditional on where she actually enrolls: $\mathbb{E}[\nu_{ij} - \nu_{ik}|a] = \mathbb{E}[Y_{ij} - Y_{ik}|a] \equiv ATE^a_{j,k}$. Adding and subtracting this parameter from the right side of Equation (5) yields:

$$\underbrace{\mathbb{E}[Y_i|j,a] - \mathbb{E}[Y_i|k,a]}_{\substack{\text{Outcome comparison of } j \text{ vs. } k \\ \text{within admission portfolio } a}} = \underbrace{\mathbb{E}[\nu_{ij} - \nu_{ik}|a]}_{\substack{ATE^a_{j,k} \\ \text{Value-added of } j \text{ vs. } k \\ \text{for the average student in } a}} + \underbrace{\underbrace{\mathbb{E}[\nu_{ij}|j,a] - \mathbb{E}[\nu_{ij}|a]}_{\substack{\text{Sorting on gains into } j \\ \text{within portfolio } a}} - \underbrace{\{\mathbb{E}[\nu_{ik}|k,a] - \mathbb{E}[\nu_{ik}|a]\}}_{\substack{\text{Sorting on gains into } k \\ \text{within portfolio } a}}}_{\text{Potentially asymmetric sorting on gains into } j \text{ vs. } k \text{ within } a} \quad (6)$$

Equation (6) clarifies that whether the within-portfolio outcome comparison of college j vs. k identifies $ATE^a_{j,k}$ depends on the nature of sorting on match effects among students who share the same admission portfolio. The first sorting gains term, $\mathbb{E}[\nu_{ij}|j,a] - \mathbb{E}[\nu_{ij}|a]$, will be nonzero if the students in a who actually choose school j reap systematically different value-added from attending j compared to what their average peer in portfolio a would gain from attending j. Likewise, the second sorting gains term, $\mathbb{E}[\nu_{ik}|k,a] - \mathbb{E}[\nu_{ik}|a]$, will be nonzero if the students in a who actually choose school k reap systematically different value-added from attending k compared to what their average peer in portfolio a would gain from attending k. Importantly, the differencing of these sorting gains terms in (6) clarifies that *asymmetry* in sorting gains across colleges is the relevant determinant of whether the treatment-effect-on-the-treated contrast $ATT^a_j - ATT^a_k$ identified in (5) also corresponds to the more ambitious parameter $ATE^a_{j,k}$ in (6). That is, the existence of match effects, and even the sorting of students into colleges on the basis of those match effects, need not drive a significant wedge between $ATT^a_j - ATT^a_k$ and $ATE^a_{j,k}$ if the sorting gains generated by such behavior are roughly symmetric across the colleges in a portfolio. In that case, the two sorting gains terms in (6) are each non-zero but roughly cancel each other out. We discuss this particular case further in the next subsection.

Second, to generalize from within-portfolio comparisons to broader population parameters, consider the average treatment effect of attending college j instead of college k across the entire population of college-choosers: $\mathbb{E}[\nu_{ij} - \nu_{ik}|\mathcal{A}_{2+}] = \mathbb{E}[Y_{ij} - Y_{ik}|\mathcal{A}_{2+}] \equiv ATE^{\mathcal{A}_{2+}}_{j,k}$, where $\mathcal{A}_{2+}$ denotes the population of college-choosers as all students with admission portfolios that offer choices among at least two colleges.[33] If every portfolio a featured full treatment support, i.e. offered admission at all colleges in $\mathcal{J}$ and saw at least some students enrolling in each college, we could simply take a population-weighted average of $ATE^a_{j,k}$ across all portfolios $a \in \mathcal{A}_{2+}$ to yield $ATE^{\mathcal{A}_{2+}}_{j,k}$. Ordering each college j by $ATE^{\mathcal{A}_{2+}}_{j,0}$ for some common benchmark school $k = 0$ would then accomplish the aim of a causal value-added schedule for the average college-chooser.

[33]Given that within-portfolio outcome contrasts $\mathbb{E}[Y_i|j,a] - \mathbb{E}[Y_i|k,a]$ are the building blocks of our identification strategy, and that these contrasts are only possible in admission portfolios with at least two colleges, we limit our broadest population of interest to "college-choosers" who are admitted to at least two colleges. As discussed in the Introduction, this population may have different treatment responses, and different treatment alternatives, compared to populations studied in other research designs, e.g. marginally admitted applicants (Hoekstra, 2009; Zimmerman, 2014; Goodman et al., 2017; Bleemer, 2020b) or disadvantaged high school students encouraged to apply more ambitiously (Hoxby and Turner, 2013; Dynarski et al., 2020).

In reality, admission portfolios violate the full support condition by construction, since they are defined by the particular colleges they include and exclude as treatment options available to the students within them. This opens up the possibility that the average of $ATE^a_{j,k}$ across the subset of portfolios that actually contain j and k as treatment options may deviate from the population parameter $ATE^{\mathcal{A}_{2+}}_{j,k}$ required for the average college-chooser value-added schedule, since the students who sort into portfolios with j and k may have different treatment effects at those schools compared to the average college-chooser (e.g. Heckman et al., 1998; Miller et al., 2019). To gain clarity on the consequences of such deviations, add and subtract the population parameter $ATE^{\mathcal{A}_{2+}}_{j,k}$ from the portfolio-specific $ATE^a_{j,k}$ in (6) to yield:

$$\underbrace{\mathbb{E}[\nu_{ij}-\nu_{ik}|a]}_{\substack{ATE^a_{j,k} \\ \text{Value-added of } j \text{ vs. } k \\ \text{for avg. student in } a}} = \underbrace{\mathbb{E}[\nu_{ij}-\nu_{ik}|\mathcal{A}_{2+}]}_{\substack{ATE^{\mathcal{A}_{2+}}_{j,k} \\ \text{Value-added of } j \text{ vs. } k \\ \text{for avg. college-chooser}}} + \underbrace{\underbrace{\mathbb{E}[\nu_{ij}|a]-\mathbb{E}[\nu_{ij}|\mathcal{A}_{2+}]}_{\substack{\text{Sorting on gains into } a \\ \text{w.r.t. gains at } j}} - \underbrace{\{\mathbb{E}[\nu_{ik}|a]-\mathbb{E}[\nu_{ik}|\mathcal{A}_{2+}]\}}_{\substack{\text{Sorting on gains into } a \\ \text{w.r.t. gains at } k}}}_{\text{Potentially asymmetric sorting on gains into portfolio } a \text{ w.r.t. } j \text{ vs. } k} \tag{7}$$

Like Equation (6), Equation (7) clarifies that asymmetry in sorting on match effects, rather than match effects and sorting *per se*, is what governs whether the portfolio-specific treatment effect $ATE^a_{j,k}$ deviates from the broader population parameter $ATE^{\mathcal{A}_{2+}}_{j,k}$. The first sorting gains term, $\mathbb{E}[\nu_{ij}|a]-\mathbb{E}[\nu_{ij}|\mathcal{A}_{2+}]$, will be nonzero if the students who sort into admission portfolio a reap systematically different value-added from attending college j compared to what the average college-chooser would gain from attending j. Likewise, the second sorting gains term, $\mathbb{E}[\nu_{ik}|a]-\mathbb{E}[\nu_{ik}|\mathcal{A}_{2+}]$, will be nonzero if those same students who sort into a also reap systematically different value-added from attending college k compared to what the average college-chooser would gain from attending k. As in (6), the differencing of these sorting gains terms in (7) shows that match effects, and sorting into admission portfolios on the basis of those match effects, need not drive a significant wedge between the portfolio-specific parameter $ATE^a_{j,k}$ and the population parameter $ATE^{\mathcal{A}_{2+}}_{j,k}$ if such sorting is roughly symmetric across the college options within a portfolio.

3.3.1 Interpreting Value-Added Under Various Restrictions on Heterogeneity

As noted at the outset of this section, the vast majority of papers in the value-added literature avoid the foregoing concerns about match effects, sorting on gains, and limited treatment support by writing down models that restrict value-added to be homogeneous across individuals. In our notation, this amounts to simply removing the i subscript from the value-added of each college: ν_{ij} becomes ν_j. But this small change has sweeping implications. If the value-added of attending a given college does not vary across students, then it does not vary across different conditioning sets, which collapses all of the various treatment effect concepts in Equations (5), (6), and (7) into a single identified comparison:

$$\underbrace{\mathbb{E}[Y_i|j,a]-\mathbb{E}[Y_i|k,a]}_{\substack{\text{Outcome comparison of} \\ j \text{ vs. } k \text{ within admission} \\ \text{portfolio } a}} = \underbrace{\underbrace{\mathbb{E}[\nu_{ij}|j,a]}_{ATT^a_j} - \underbrace{\mathbb{E}[\nu_{ik}|k,a]}_{ATT^a_k}}_{\substack{\text{Relative value-added of} \\ j \text{ vs. } k \text{ among their} \\ \text{respective students in } a}} = \underbrace{\mathbb{E}[\nu_{ij}-\nu_{ik}|a]}_{\substack{ATE^a_{j,k} \\ \text{Value-added of} \\ j \text{ vs. } k \text{ for the} \\ \text{average student in } a}} = \underbrace{\mathbb{E}[\nu_{ij}-\nu_{ik}|\mathcal{A}_{2+}]}_{\substack{ATE^{\mathcal{A}_{2+}}_{j,k} \\ \text{Value-added of} \\ j \text{ vs. } k \text{ for the} \\ \text{average college-chooser}}} = \underbrace{\nu_j-\nu_k}_{\substack{\text{Value-added of} \\ j \text{ vs. } k \text{ for } \textit{all} \\ \text{college choosers}}} \tag{8}$$

The assumption of value-added homogeneity delivers identification of *the* causal effect of attending college j instead of college k, $\nu_j-\nu_k$, by comparing student outcomes in any portfolio a that includes colleges j and k. Homogeneity therefore implies simple answers, by construction, to the two questions about policy relevance posed in the previous subsection. First, the $ATT^a_j - ATT^a_k$ comparisons identified in (5) are indeed informative about $ATE^a_{j,k}$, since treatment effect homogeneity ensures these are the same parameter (see the second equation in (8)). Second, averaging $ATE^a_{j,k}$ across portfolios is indeed informative about $ATE^{\mathcal{A}_{2+}}_{j,k}$, since treatment effect homogeneity also ensures these are the same parameter (see the third equation in (8)).

In turn, homogeneity implies that assembling a causal value-added schedule relevant for the entire population of college-choosers is as simple as ordering each college j by $\nu_j - \nu_0$ for some benchmark college $k = 0$. This population value-added schedule is identified even in extreme versions of limited treatment support where no students actually face the choice between college j and the benchmark college $k = 0$ in their admission portfolio, as long as a chain of other pairwise comparisons can eventually indirectly link j to 0.[34] For example, $\nu_j - \nu_0 = (\nu_j - \nu_m) + (\nu_m - \nu_0)$ is identified by $(\mathbb{E}[Y_i|j,a] - \mathbb{E}[Y_i|m,a]) + (\mathbb{E}[Y_i|m,a'] - \mathbb{E}[Y_i|0,a'])$ for any portfolio a that contains j and m and any other portfolio a' that contains m and 0, which does not require any portfolio to actually contain both j and 0.

Making these simplifications explicit, after starting from a potential outcomes framework with unrestricted value-added heterogeneity, aids our understanding of why treatment effect homogeneity is such a powerful assumption, and in turn a pervasive feature of the value-added literature. But our derivations in the previous subsection also revealed weaker conditions under which policy-relevant value-added comparisons can be identified, even in the presence of match effects, sorting on gains, and limited treatment support. Thus, we can now compile these results into a framework for interpreting value-added estimates under different assumptions about treatment effect heterogeneity and sorting gains, ranging from completely unrestricted to completely shut down. While our discussions are anchored to the college setting, the broader lessons of this framework may prove useful in other value-added environments where an analog of Assumption 1 addresses selection bias on vertical potential but researchers wonder how "horizontal" heterogeneity in value-added across individuals affects identification and interpretation.

Unrestricted Heterogeneity in Value-Added. Assumption 1 alone, in a world of unrestricted heterogeneity in value-added, delivers the first equation in (8), but guarantees no more. In this most general case, outcome comparisons across colleges within an admission portfolio identify $ATT^a_j - ATT^a_k$, the relative value-added of j among students in a who actually choose j vs. the value-added of k among students in a who actually choose k. On one hand, averaging these treatment-effect-on-the-treated comparisons across all of the portfolios that include j and k yields policy-relevant evaluations of realized college value-added across the landscape of actual college choices. On the other hand, these local status quo evaluations cannot necessarily speak to more ambitious policy questions that involve re-sorting students into counterfactual admission portfolios and enrollment choices. The resulting pairwise estimates also do not necessarily compile into a single value-added schedule for a common student population, since different pairwise estimates may average over different sets of admission portfolios (limited treatment support). And in finite data, some of the more esoteric college pairings may involve too few students to estimate with meaningful statistical precision. These limitations, while natural and conservative, may motivate restrictions on value-added heterogeneity that deliver more ambitious causal parameters with greater precision.

Heterogeneity in Value-Added, but with Symmetric Mean Sorting Gains. Our derivation in (6) revealed that even in a world of heterogeneous value-added, and even if students systematically sort into colleges on the basis of those match effects, we can advance from $ATT^a_j - ATT^a_k$ to $ATE^a_{j,k}$—i.e. to the second equation of (8)—if the sorting gains that students reap from their enrollment choices are symmetric (on average) across the colleges in a portfolio. Likewise, our derivation in (7) revealed that even without

[34] Such a chain relates to the "connected set" concept in the worker-firm literature (Abowd et al., 2002) and the paired comparison method discussed in Hoxby (2019).

full treatment support within all portfolios, we can further advance from within-portfolio effects $ATE^a_{j,k}$ to average treatment effects across all college-choosers $ATE^{\mathcal{A}_{2+}}_{j,k}$—i.e. to the third equation of (8)—if the sorting gains into any given portfolio are also symmetric (on average) across the colleges in that portfolio. Letting $\mathcal{J}_a \subseteq \mathcal{J}$ denote the set of colleges offering admission in a given portfolio a, we can write these conditions as:

Assumption 2a *Symmetric mean sorting gains:*

1. *Into enrollment choices:* $\mathbb{E}[\nu_{ij}|j,a] - \mathbb{E}[\nu_{ij}|a] = \mathbb{E}[\nu_{ik}|k,a] - \mathbb{E}[\nu_{ik}|a] \equiv \delta_a \quad \forall j,k \in \mathcal{J}_a$

2. *Into admission portfolios:* $\mathbb{E}[\nu_{ij}|a] - \mathbb{E}[\nu_{ij}|\mathcal{A}_{2+}] = \mathbb{E}[\nu_{ik}|a] - \mathbb{E}[\nu_{ik}|\mathcal{A}_{2+}] \equiv \lambda_a \quad \forall j,k \in \mathcal{J}_a$

Assumption 2a allows for heterogeneity in value-added, and two layers of systematic student sorting on the basis of such match effects. First, relative to their peers in the same admission portfolio, students can choose to enroll in a given college because they have particularly good match quality there ($\delta_a > 0$). And second, relative to all college-choosers, students can end up in a given admission portfolio (both through their own application behavior and through the decisions of admissions officers) because they have particularly good match quality at the schools within it ($\lambda_a > 0$). Note that the signs of these sorting gains are unrestricted: λ_a and/or δ_a could be negative if students in portfolio a tend to sort into options with poorer match quality with respect to the outcome Y, perhaps being motivated to do so by other compensating factors. Note also that the signs and magnitudes of λ_a and δ_a can vary arbitrarily across different portfolios a, such that students in some portfolios may be more effective at sorting on gains than students in other portfolios. Finally, note that the signs and magnitudes of each student's idiosyncratic deviations from the mean treatment effect of her peers who enroll at the same school from the same portfolio—$\nu_{il} - \mathbb{E}[\nu_{il}|l,a],\ l = j,k$—can vary widely across students i as well, so long as they are equal *on average* across colleges j and k within portfolio a.

The key restriction in Assumption 2a is within-portfolio mean symmetry in these sorting gains across colleges. If the students in a who choose college j reap sorting gains of average magnitude δ_a, then their peers in a who choose college k must also reap sorting gains of similar magnitude, on average. And if the students who end up in portfolio a tend to be particularly well-matched at j relative to the average college chooser, reaping sorting gains of average magnitude λ_a, then these students must also be similarly well-matched at k, relative to the average college chooser. This is a strong assumption, and difficult to test directly in observational data; however, we discuss testing more restrictive special cases of this assumption below. The takeaway is simply that Assumption 2a is weaker than the constant-effects approach typical to the value-added literature, which shuts down match effects and sorting gains altogether. And suspicions of violations of Assumption 2a, like students only choosing to forego more selective options when their less selective alternatives provide especially high match quality, need not doom the identification of policy-relevant value-added parameters, as our first case above—unrestricted heterogeneity in value-added—does not impose any restrictions on match effects and sorting gains and yet still recovers useful value-added comparisons.

Heterogeneity in Value-Added, but with Zero Mean Sorting Gains. One particularly clear manifestation of symmetric mean sorting gains is zero mean sorting gains. This comprises a special case of Assumption 2a in which the terms δ_a and λ_a, representing realized mean sorting gains from students' purposeful choices, are equal to zero for every portfolio a:

Assumption 2b *Zero mean sorting gains:*

1. *Into enrollment choices:* $\mathbb{E}[\nu_{ij}|j,a] = \mathbb{E}[\nu_{ij}|a] \quad \forall j \in \mathcal{J}_a$

2. *Into admission portfolios:* $\mathbb{E}[\nu_{ij}|a] = \mathbb{E}[\nu_{ij}|\mathcal{A}_{2+}] \quad \forall j \in \mathcal{J}_a$

Assumption 2b, like Assumption 2a, delivers the first, second, and third equations in (8).[35] That is, it allows within-portfolio outcome comparisons of college j vs. k to identify the value-added of j vs. k for the average college-chooser, $ATE^{\mathcal{A}_{2+}}_{j,k}$. Unlike value-added homogeneity, this assumption still allows for match effects to exist between students and colleges. But in contrast to the previous two cases, students are now assumed to be unaware of, or at least not influenced to a meaningful degree, by such match effects, at least at the time they are making their application and initial enrollment choices.

This assumption might be too strong when modeling the choice of whether to enroll in any college as opposed to none (Willis and Rosen, 1979; Carneiro et al., 2011) or which field to specialize in (Kirkeboen et al., 2016). However, the evidence base is very thin on whether first-time college students sort into different *institutions* on the basis of heterogeneous match effects with those institutions, especially within the public university sector. A broader literature consistently documents the difficulties students face in forecasting their own gains from various educational endeavors (e.g. Dominitz and Manski, 1996; Betts, 1996; Jensen, 2010; Stinebrickner and Stinebrickner, 2012, 2014b; Wiswall and Zafar, 2015a,b; Hastings et al., 2016). These difficulties may be especially pronounced at the application and initial enrollment stages, prior to learning about any relevant match effects through direct experience and experimentation (Manski, 1989; Altonji, 1993; Arcidiacono, 2004; Malamud, 2011; Stange, 2012; Stinebrickner and Stinebrickner, 2014a; Arcidiacono et al., 2016a). In our own data, 27% of students end up transferring out of their initial college to another institution, and only 54% ever earn a BA from their initial college, despite just 4.5% of incoming first-years in the U.S. expecting to ever transfer and 98.5% expecting to earn a BA from their initial institution (Stolzenberg et al., 2020). Consistent with Assumption 2b, then, heterogeneous match effects between students and colleges may be real and non-trivial—recall that λ_a and δ_a represent *realized* gains from purposeful sorting choices, rather than potential gains from sorting—but students may only start to develop an accurate assessment of them after actually enrolling. This would hamper their ability to generate sorting gains at the initial decision point of where to *begin* college, which is our treatment of interest.

More starkly, it is possible that relative value-added itself, regardless of any heterogeneity across students, may simply fail to enter into the decision process of where to enroll (Rothstein, 2006; Abdulkadiroglu et al., 2020). Students may have difficulty assessing causal value-added differences across colleges, or may put little weight on them relative to other considerations like peer characteristics (Abdulkadiroglu et al., 2014), tuition costs (Cohodes and Goodman, 2014; Fillmore, 2021), proximity (Fu et al., 2021), on- and off-campus consumption amenities (Jacob et al., 2017), institutional mission (Delavande and Zafar, 2019), athletics (Pope and Pope, 2012), coordination with friends (Fletcher, 2015), and even the weather on campus visit day (Simonsohn, 2010). In our data, the final section of Table 2 shows that only 47% of students enroll at the college in their admission portfolio with the highest earnings value-added (estimated using our main specification described below), which is barely higher than the 46% share we would expect if students chose randomly with respect to value-added,[36] and lower than the 55% share who opt for the highest peer SAT

[35] Note that the subparts of Assumptions 2a and 2b can be mixed and matched to equivalently deliver the first three equations of (8): symmetric sorting into enrollment choices with zero sorting into admission portfolios (2a.1 + 2b.2), or zero sorting into enrollment choices with symmetric sorting into admission portfolios (2b.1 + 2a.2). Note also that 2b.1, as well as Assumption 1, are implied by the conditional independence assumption $\{Y_{ij}\}_{j\in\mathcal{J}} \perp\!\!\!\perp S_i|A_i$, i.e. if potential outcomes are independent of enrollment choices conditional on admission portfolios. Such a CIA therefore delivers $ATE^a_{j,k}$ in the second equation of (8), but, as noted above, averages of $ATE^a_{j,k}$ across the portfolios a that actually contain j and k may not deliver $ATE^{\mathcal{A}_{2+}}_{j,k}$ in the third equation of (8) due to limited treatment support across portfolios—an issue which 2a.2 and 2b.2 address.

[36] If all students were choosing between two colleges in their admission portfolio, clearly 50% would randomly pick the college with higher value-added. But 28.8% of students in our identifying sample are choosing among three, four, five, etc. colleges, so their probability of randomly picking the highest value-added school in their portfolio is 33%, 25%, 20%, etc. Averaging these probabilities across the student-level distribution of portfolio sizes yields the random-choice benchmark probability of 46%.

scores.

To be clear, excluding relative value-added entirely from student decisionmaking is sufficient, but far stronger than necessary, for Assumption 2b to hold. Students can systematically prefer colleges with higher (or lower) value-added and still satisfy Assumption 2b, which specifically concerns whether students go beyond the *average* value-added of a given college to sort on their own idiosyncratic deviation from that average gain, i.e. their match effect.[37] Altogether, while direct evidence on the validity of Assumption 2b is scant, broader evidence suggests that students may have difficulty assessing and/or systematically acting upon causal match effects as they decide where to initially enroll. The next and final case concerns whether value-added schedules actually vary meaningfully across students in the first place.

No Heterogeneity in Value-Added. We opened this subsection with a discussion of the powers of value-added homogeneity, codified in the following assumption:

Assumption 2c *Value-added homogeneity:* $\nu_{ij} = \nu_j \; \forall j \in \mathcal{J}$

To summarize the discussion above, Assumption 2c delivers all of the equations in (8): value-added homogeneity allows within-portfolio outcome comparisons of college j vs. k to identify *the* value-added of j vs. k, which is a constant parameter across all students. In a world of homogeneous treatment effects, there is no asymmetric sorting on gains (satisfying Assumption 2a), because there is no sorting on gains (satisfying Assumption 2b), because there are no idiosyncratic gains on which to sort (Assumption 2c).

Assumption 2c is testable in our data, as we can stratify value-added estimates across two important domains of student heterogeneity: our detailed set of observable student covariates, and the large set of student admission portfolios. We devote Section 8 entirely to the first question of whether value-added schedules vary with student covariates, including those often hypothesized to be drivers of (mis)match effects—race, cognitive skills, and family income—as well as gender and non-cognitive skills. Second, in Section 4.2 below we test whether value-added differences vary across the admission portfolios that help identify them. In short, we find little evidence of systematic value-added heterogeneity across these rich dimensions of student diversity. These two sets of diagnostics, of course, are ultimately limited to observable student heterogeneity, and thus cannot rule out match effects along unobserved dimensions, and sorting on those idiosyncratic gains. But our covariates and admission portfolios do capture a significant amount of student diversity, including across policy-relevant stratifications.

Lessons from This Framework. Assumption 1 alone, in a world of unrestricted effect heterogeneity and sorting gains, allows within-portfolio outcome differences across colleges to identify comparisons of the value-added realized by students who actually apply, get admitted, and end up enrolling at each college. Asymmetric sorting gains across colleges and limited treatment support within portfolios may prevent the more ambitious aggregation of these comparisons into a value-added schedule for the average college-chooser. Invoking any one of Assumption 2a, 2b, or 2c, or the combination of subparts of 2a and 2b, allows for the identification of this more ambitious population schedule. This framework generalizes the typical constant-effects approach in the value-added literature along two fronts. First, value-added homogeneity (2c) is sufficient, but not necessary, for identifying a value-added schedule for the average college-chooser, since 2b and 2a are weaker yet still sufficient for this goal. Second, no restrictions on value-added heterogeneity are needed to identify local comparisons of realized value-added, which only require Assumption 1.

[37]For example, Campos and Kearns (2021) estimate that families in the Zones of Choice program in Los Angeles prefer high schools with higher average value-added, but not with higher student-specific match quality. This behavior would satisfy Assumption 2b. Beuermann and Jackson (2020) review the literature on preferences for schools and their value-added, which rarely considers heterogeneity in value-added and sorting gains of the type necessary to violate Assumption 2a or 2b.

4 Estimating and Validating Relative College Value-Added

We now apply this framework to estimate relative college value-added. We first present baseline regression estimates that, under Assumption 1 and any one of Assumption 2a, 2b, or 2c, correspond to an average value-added schedule across all college-choosers with at least two admissions. We then show that taking Equation (5) directly to the data by manually averaging within-portfolio pairwise college comparisons, which only imposes Assumption 1 and thus leaves effect heterogeneity and sorting gains unrestricted, delivers similar estimates as the regression approach for the more common college match-ups in our data but uninformatively imprecise estimates for the less common pairs. This motivates employing the regression approach for the remainder of the paper, further bolstered by a battery of validation exercises and specification checks.

4.1 Baseline Regression Estimates

To build up to the baseline estimates, Figure 3, like Figure 2, begins on the left with raw outcome means (in diamonds) at each college in our sample, measured relative to UT-Austin (signified by the vertical line at zero). The schools are sorted from top to bottom by these disparate raw means, which help contextualize the ordinal and cardinal content of the subsequent estimates. Appendix Tables B.1 and B.2 report the numerical estimates corresponding to each specification in Figure 3.

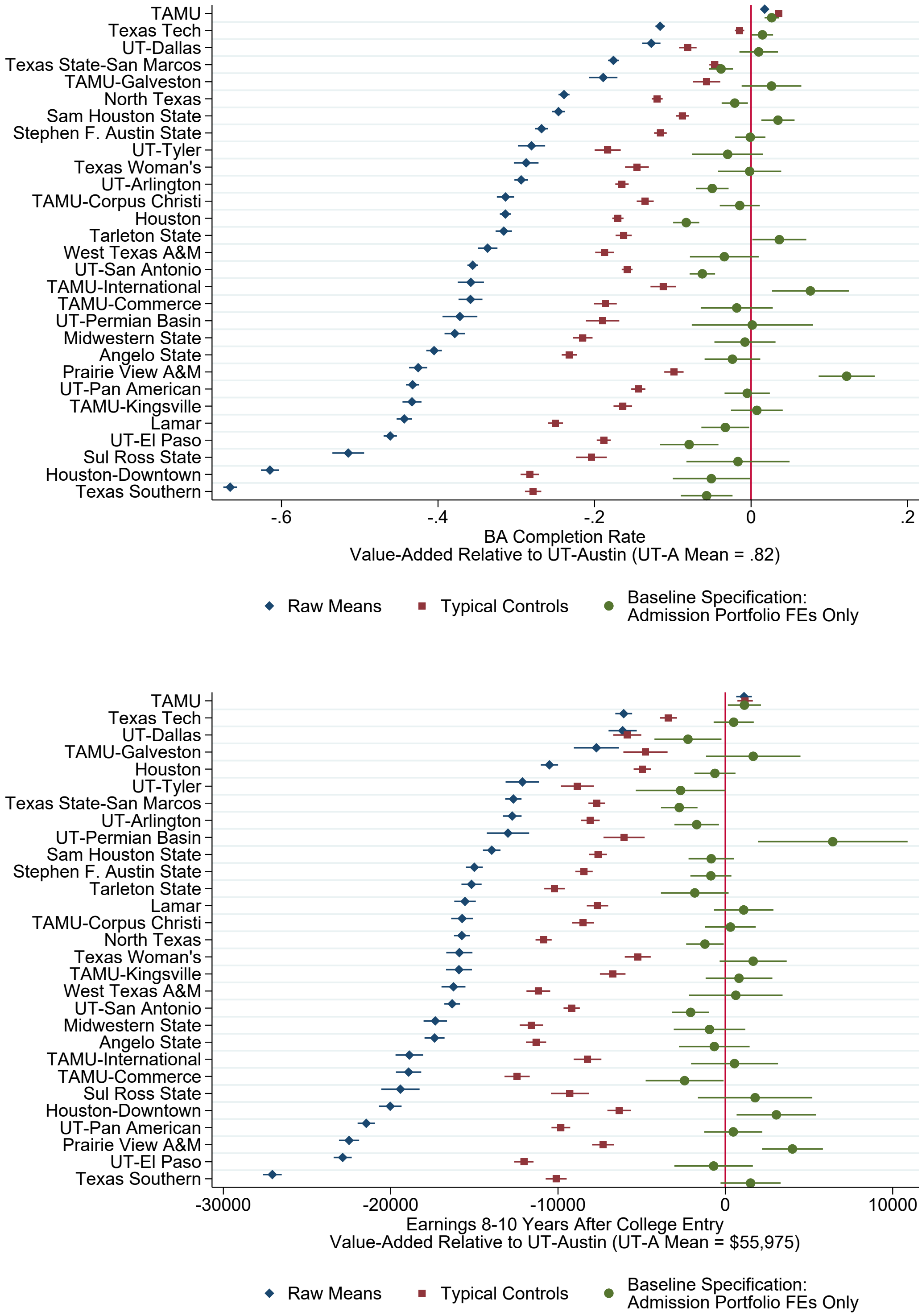
The estimates in the middle column of Figure 3 (squares) add a rich set of student covariate controls, akin to a typical specification in the literature on the return to college quality. Specifically, we regress student outcomes on college treatment indicators (omitting UT-Austin as the comparison case) plus controls for demographics (gender, race), family income (proxied by free/reduced price lunch status), high school academic preparation (10^{th} grade test scores, advanced coursework, and an indicator for graduating in the top GPA decile), and behavioral measures of non-cognitive skills (high school attendance record, disciplinary infractions, and an indicator for ever being at-risk of dropping out). All of these covariates are highly predictive of college choices and outcomes. Controlling for them attenuates the mean outcome differences between most colleges and UT-Austin by roughly half, but interpreting the remaining differences as causal would still attribute substantial returns to attending the more selective colleges at the top of the distribution relative to those at the bottom.

Our baseline specification includes college treatment indicators (omitting UT-Austin) and controls solely for admission portfolio fixed effects. We can derive this specification by imposing Assumption 1 and any one of Assumption 2a, 2b, or 2c, which allows us to write expected student outcomes Y_i, conditional on choosing college $S_i = j$ from admission portfolio $A_i = a$, as

$$\begin{aligned}
\mathbb{E}[Y_i|j,a] = \mathbb{E}[\alpha_i + \nu_{ij}|j,a] &= \underbrace{\mathbb{E}[\alpha_i|j,a]}_{\phi_a} + \underbrace{\mathbb{E}[\nu_{ij}|j,a] - \mathbb{E}[\nu_{ij}|a]}_{\delta_a} + \underbrace{\mathbb{E}[\nu_{ij}|a] - \mathbb{E}[\nu_{ij}|\mathcal{A}_{2+}]}_{\lambda_a} + \mathbb{E}[\nu_{ij}|\mathcal{A}_{2+}] \\
&= k + \sum_{a\neq 0} p_a \, \mathbb{1}[A_i = a] + \sum_{j \neq 0} v_j \, \mathbb{1}[S_i = j], \qquad (9)
\end{aligned}$$

where $k = \phi_0 + \delta_0 + \lambda_0 + \mathbb{E}[\nu_{i0}|\mathcal{A}_{2+}]$, $p_a = \phi_a + \delta_a + \lambda_a - \phi_0 - \delta_0 - \lambda_0$, and $v_j = \mathbb{E}[\nu_{ij} - \nu_{i0}|\mathcal{A}_{2+}]$. Under Assumption 2a, the sorting gains terms δ_a and λ_a can be nonzero for a given admission portfolio a, but are absorbed into the portfolio fixed effect p_a; under Assumption 2b or 2c, $\delta_a = \lambda_a = 0$. Our baseline specification thus follows (9) and regresses individual student outcomes Y_i on college treatment indicators $\mathbb{1}[S_i = j]$ (omitting UT-Austin as the $j = 0$ reference case) and admission portfolio indicators $\mathbb{1}[A_i = a]$ (omitting the reference case $a = 0$).

Figure 3: Baseline Value-Added Estimates and Comparison to Other Common Approaches



Notes: Each set of point estimates and robust 95% confidence intervals come from regressions of individual student outcomes on college treatment indicators, omitting UT-Austin as the reference treatment (signified by the vertical line at zero). The UT-Austin outcome mean appears in parentheses below each plot. All specifications control for cohort fixed effects. The *Raw Means* specification controls for nothing else. The *Typical Controls* specification adds controls for demographics (gender, race, FRPL), high school academic preparation (10th grade test scores, advanced coursework, and top high school GPA decile indicator), and behavioral measures of non-cognitive skills (high-school attendance, disciplinary infractions, and an indicator for ever being at risk of dropping out). The *Baseline Specification* controls solely for college admission portfolio fixed effects (and cohort fixed effects). See Appendix Tables B.1 and B.2 for the corresponding numerical estimates.

Estimates of the v_j relative value-added parameters from (9) appear farthest to the right in Figure 3 (in circles). Remarkably, they orbit loosely around zero: students who face the same choice sets of college admissions tend to have relatively similar graduation and earnings outcomes regardless of where they end up enrolling. As we quantify in Section 5, however, this distribution of relative value-added across colleges is not degenerate, even after accounting for estimation error. But it is substantially less dispersed than the raw means, and it does not correspond well with some prominent college orderings like selectivity, as we explore in Section 6. Before pursuing those analyses, however, we next compare these baseline regression estimates to the more heterogeneity-agnostic but more data-demanding pairwise averaging approach.

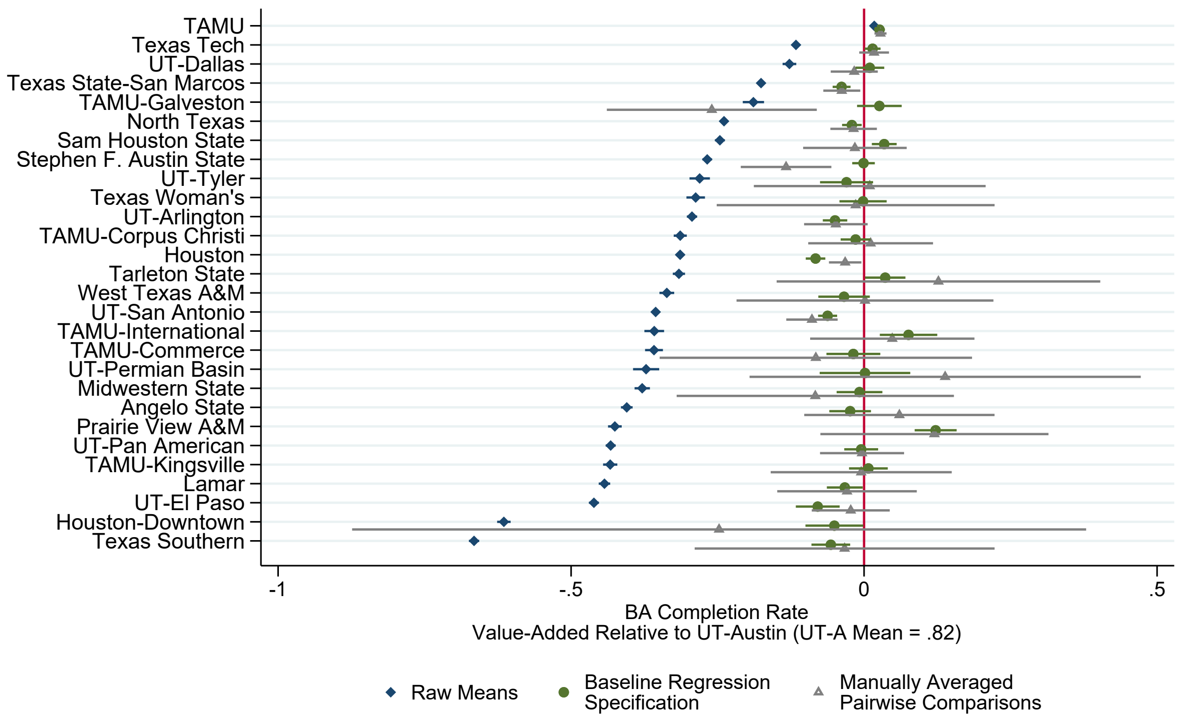
4.2 Unrestricted Heterogeneity with Average Pairwise Comparisons

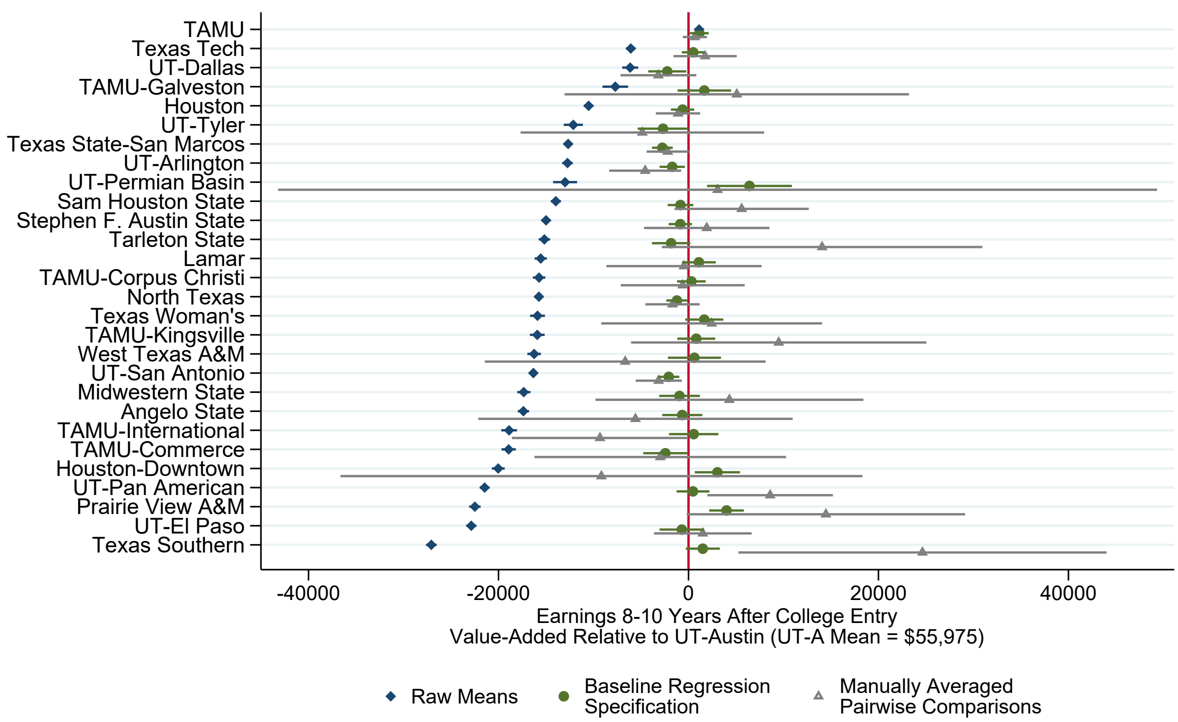
Imposing any one of Assumption 2a, 2b, or 2c allows us to interpret the baseline regression estimates from (9) as a value-added schedule for the average student choosing among at least two admissions. We now compare those baseline regression estimates to the more general, but more data-demanding, approach of taking Equation (5) directly to the data, which does not impose Assumption 2a, 2b, or 2c. That is, by manually averaging within-portfolio pairwise college comparisons across portfolios, we can leave effect heterogeneity and sorting gains unrestricted while still identifying useful value-added comparisons. Concretely, for a given j, k college pair, we first estimate the average outcome difference $\mathbb{E}[Y_i|j,a] - \mathbb{E}[Y_i|k,a]$ within each admission portfolio a that contains j and k as available options. We then average these within-portfolio comparisons across all the different portfolios that contain j and k, weighted by the number of students in each portfolio who enroll at j or k. By Equation (5), which only imposes Assumption 1, this delivers estimates of $\mathbb{E}_{a\in\mathcal{A}_{j,k}}\left[ATT_j^a - ATT_k^a\right]$, the relative value-added of j vs. k among their respective students who shared the same admission portfolio a, averaged across the set of admission portfolios $a \in \mathcal{A}_{j,k}$ that contain colleges j and k as options.

Figure 4 presents these estimates (in triangles), with UT-Austin fixed as the common comparison school $k = 0$. Comparing these average pairwise estimates to the baseline regression estimates (in circles) yields two main takeaways. First, the two approaches deliver very similar point estimates for the larger schools in our data, e.g. Texas A&M, Texas Tech, UT-San Antonio, North Texas, Texas State, etc. These schools regularly face off against UT-Austin within student admission portfolios, and thus produce fairly precise estimates in the pairwise approach that relies exclusively on these direct, within-portfolio comparisons. On the other hand, the smaller schools in our data—TAMU-Galveston, UT-Permian Basin, Houston-Downtown, etc.—simply do not have enough students in these direct pairwise comparisons to deliver statistically informative estimates, transparently highlighting a tradeoff between statistical precision and unrestricted heterogeneity.

In light of this tradeoff, the baseline regression approach offers a reasonable way forward. It delivers value-added estimates for the more popular colleges in our sample that are very similar to the pairwise averages, which do not impose Assumption 2a, 2b, or 2c. For the smaller colleges, the regression approach generates substantially more precise estimates by incorporating additional variation from chains of pairwise comparisons that indirectly link these colleges to UT-Austin, akin to the "connected set" of comparisons central to the firm effects literature (Abowd et al., 2002). These chained comparisons are interpretable under Assumption 2a, 2b, or 2c, since imposing any one of those assumptions allows pairwise college comparisons made within one portfolio to also be informative about expected impacts for other students in different portfolios. We find little evidence against this restriction across the observed portfolios in our sample: for the earnings outcome, an F-test fails to reject that each school's value-added relative to UT-Austin is homogeneous across the different admission portfolios that identify it. For BA completion, the same

Figure 4: Unrestricted Heterogeneity in Value-Added with Average Pairwise Comparisons





Notes: The *Raw Means* specification regresses individual student outcomes on college treatment indicators, omitting UT-Austin as the reference treatment (signified by the vertical line at zero). The UT-Austin outcome mean appears in parentheses below each plot. The *Baseline Regression Specification* controls solely for college admission portfolio fixed effects (and cohort fixed effects). The *Manually Averaged Pairwise Comparisons* first take the average outcome difference $\mathbb{E}[Y_i|j,a] - \mathbb{E}[Y_i|k,a]$ within each admission portfolio a that contains college j and k (fixed as UT-Austin) as available options, then averages these within-portfolio comparisons across all the different portfolios that contain j and k, weighted by the number of students in each portfolio who enroll at j or k. All comparisons involve at least ten students. Estimates for Sul Ross State are omitted due to extreme imprecision. Confidence intervals for the average pairwise comparisons are computed from 500 bootstraps.

F-test does reject omnibus homogeneity across all portfolios, but this is driven by just two colleges with significant value-added heterogeneity at the 5% level: Texas A&M and UT-Pan American. And the top panel of Figure 4 shows that, for these two schools, the cross-portfolio averages that emerge from the strict pairwise estimation approach are nearly identical to the baseline regression estimates for BA completion, indicating that this sporadic heterogeneity across portfolios is not large and systematic enough to drive the two approaches toward different conclusions.[38]

4.3 Validation Exercises and Specification Checks

Having established the baseline value-added estimates, we now subject them to a battery of validation exercises and specification checks to probe their robustness. We first validate the baseline research design with excluded covariate controls, then explore richer specifications of admission portfolios, simpler specifications, alternative treatment definitions, and alternative earnings measures.

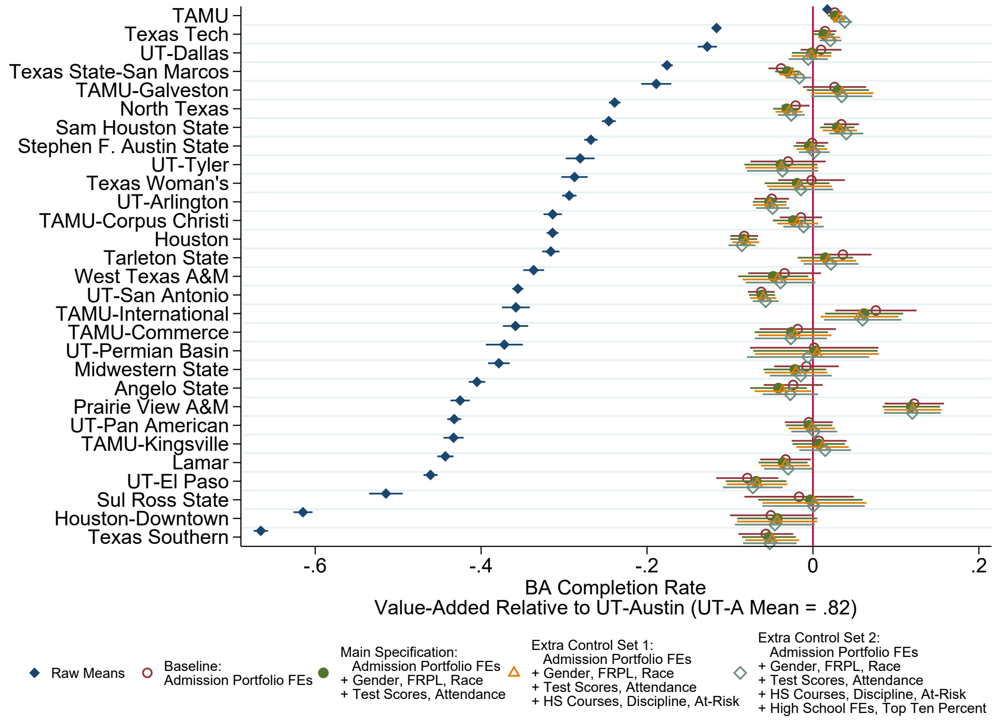
4.3.1 Validating the Research Design with Excluded Controls

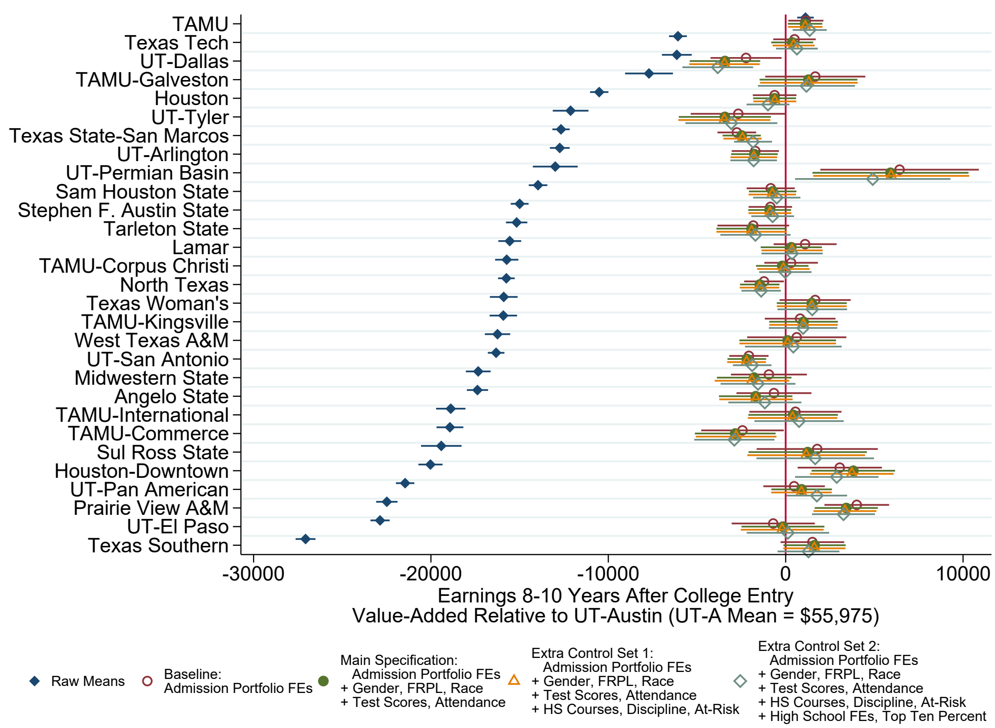
Figure 2 established that our baseline research design, which solely controls for admission portfolio fixed effects, nearly eliminates imbalances in observable pre-college student potential across the college treatments in our sample. Some schools exhibit minor deviations from perfect balance, however, so to move from our baseline specification to our main specification, we add a core set of covariate controls, which are also the dimensions of student heterogeneity along which we investigate match effects in Section 8: categorical indicators for gender, race, and low family income (FRPL), and continuous measures of cognitive skills (10th grade test scores) and non-cognitive skills (high school attendance).

Figure 5 shows that adding these core controls barely moves the value-added estimates, suggesting that the minor deviations from perfect balance in Figure 2 do not translate into meaningful bias. The additional results in Figure 5 demonstrate robustness of the value-added estimates to further expansions of the control set. *Extra Control Set 1* includes additional pre-college measures of cognitive and non-cognitive skills: advanced high school coursework, disciplinary infractions, and an indicator for ever being at risk of dropping out. *Extra Control Set 2* further adds the indicator for graduating in the top decile of high school GPA, as well as fixed effects for each of the 1,457 different high schools from which our students graduate, flexibly accounting for the environmental influences of the schools, neighborhoods, and local labor markets in which students are embedded prior to college entry. All of these additional controls are strong predictors of student outcomes, even within admission portfolios: moving from the baseline specification to the specification with all extra controls increases the adjusted R^2 proportionally by about 60% (from .145 to .235 for BA completion, and from .089 to .142 for earnings). Meanwhile, the value-added estimates barely change. While this exercise cannot completely rule out bias from remaining unobservable confounders, the stability of the value-added estimates in Figure 5 at least fails to invalidate our research design across the expansive swath of potential confounders that we do observe.

[38]Related evidence against empirically consequential heterogeneity across portfolios comes from robustness across alternative specifications that employ different weighting schemes when averaging pairwise estimates across portfolios. For example, running separate binary treatment regressions for each college relative to UT-Austin, controlling for portfolio fixed effects, weights within-portfolio binary treatment comparisons with greater treatment variance more heavily (Angrist, 1998) compared to the manually averaged estimates in Figure 4. With substantial treatment effect heterogeneity across portfolios, these different weighting schemes could produce contrasting results (e.g. Gibbons et al., 2019; Goldsmith-Pinkham et al., 2021). Instead, they produce very similar estimates in our sample.

Figure 5: Validating the Research Design: Robustness to Omitted Variables





Notes: Each set of point estimates and robust 95% confidence intervals come from regressions of individual student outcomes on college treatment indicators, omitting UT-Austin as the reference treatment (signified by the vertical line at zero). The UT-Austin outcome mean appears in parentheses below each plot. All specifications control for cohort fixed effects. The *Raw Means* specification controls for nothing else. The *Baseline Specification* controls solely for college admission portfolio fixed effects (and cohort fixed effects). The *Main Specification* adds our core set of covariate controls: gender, FRPL, race, 10th grade test scores, and high school attendance. *Extra Control Set 1* adds controls for advanced high school coursework, disciplinary infractions, and an indicator for ever being at risk of dropping out. *Extra Control Set 2* adds fixed effects for every high school and an indicator for being in the top decile of high school GPA. See Appendix Tables B.1 and B.2 for the corresponding numerical estimates.

4.3.2 Richer Specifications of Admission Portfolios

Figure A.3 probes robustness to more demanding specifications of the admission portfolio controls. First, limiting the regression sample to the subset of students who are actually admitted to at least two colleges delivers essentially identical results as the full sample, since our value-added parameters are identified from variation in enrollment choices among this subset. Students with only one admission necessarily have no scope for variation in enrollment choices and thus do not contribute to identifying value-added in this approach.

The next set of estimates in Figure A.3 add information about rejections (applications that do not result in admissions) in the portfolio control. Recall that our main specification uses portfolios based on simplified binary entries for each college: *do not get admitted* or *apply and get admitted*, where *do not get admitted* could be due either to not applying or applying and getting rejected. The estimates in triangles in Figure A.3 use portfolios with all three possible entries for each college: *do not apply*, *apply and get rejected*, or *apply and get admitted*. The resulting estimates are nearly identical to those from our main specification. Student applications that lead to rejections therefore do not provide additional selection correction beyond portfolios of successful admissions.

As a final check for robustness to richer specifications, the last two sets of estimates in Figure A.3 interact our main admission portfolios with additional covariates. The first specification interacts them with the indicator for being in the top decile of high school GPA, to address the concern that a given portfolio may convey different information if the student is eligible for the Texas Top Ten Percent automatic admissions program. The second specification interacts the admission portfolios with our core set of student covariates—gender, race, low-income (FRPL), high school test scores, and high school attendance—to address the concern that a given portfolio may convey different information for students with different demographic backgrounds and skills. In both cases, the value-added estimates are virtually unaffected.

4.3.3 Simpler Specifications of Admission Portfolios

In the other direction, a natural question prompted by the use of exhaustive fixed effects for every distinct admission portfolio is whether substantially simpler specifications deliver similar value-added estimates. Appendix Figure A.4 demonstrates viable simplifications in two directions, both of which maintain the restriction to the subsample of students who are admitted to at least two colleges in order to facilitate comparisons with our baseline results, which are identified from this subsample of college choosers. First, we can dramatically reduce the number of fixed effects with additive, rather than combinatorial, indicators for applying and getting admitted to each college. Specifically, the estimates in hollow triangles in Figure A.4 come from a specification that controls for 30 additive indicators for applying to each school, plus 30 additive indicators for being admitted to each school, which reduces the number of portfolio fixed effects from over 3,000 to just 60. Estimated on our identifying sample of students with at least two admissions, this drastically simpler specification delivers roughly similar estimates as our main specification. Second, we can even further reduce the dimensionality of the admission controls by eliminating fixed effects altogether, and simply control linearly for two covariates: the average peer SAT score across the colleges a student is admitted to, and the number of applications she sent. This specification is reminiscent of Dale and Krueger (2002), and yields a similar pattern of estimates (in squares) as our main specification, though with meaningful deviations for some schools. Overall, these two drastically simplified specifications in Figure A.4 help to assuage concerns that the basic patterns of relative value-added that emerge from our main specification are a problematic artifact of including high-dimensional fixed effect controls.

4.3.4 Insufficient Specifications

We now turn to three alternative specifications that appear to systematically under-control for selection bias, despite falling under the broad umbrella of exploiting application and/or admission behavior as controls. First, in Appendix Figure A.5, we run the same simplified additive specification as in Figure A.4—controlling for 30 additive indicators for applying to each school, plus 30 additive indicators for being admitted to each school—but now instead of limiting the sample to students with at least two admissions, we include the full sample of all college-goers, i.e. adding students who were only admitted to one college, the one in which they enrolled. The resulting estimates (in x's) imply substantially larger value-added differences between nearly every school and UT-Austin. Intuitively, unlike the main specification where the combinatorial admission portfolio controls automatically limit the identifying variation to students with at least two admissions, the additive specification allows single-admission students to contribute identifying outcome variation when included in the sample. These single-admission students tend to be negatively selected on ability and outcomes relative to the "college-choosers" who enjoy at least two admissions (Table 1), leading to unbalanced comparisons of students across colleges when single-admission students are included in the additive specification.

Second, the estimates in hollow circles in Figure A.5 come from a specification that includes fixed effects for each distinct combinatorial portfolio of *applications* only, ignoring any information about admissions decisions. These estimates get a bit closer to the main results, but they still appear to systematically under-control for selection, leaving larger outcome gaps between the selective flagships and the rest of the schools. Admission decisions thus appear to reveal additional information about student potential beyond application behavior. This suggests that research designs relying solely on application controls (e.g. SAT/ACT score sending behavior) may not fully correct for selection, and thus may overestimate the relative value-added of more selective institutions.

Third, in Figures A.6 and A.7 we explore the consequences of redefining the treatment variable as the *modal* college attended over a student's undergraduate career, as in Chetty et al. (2020), rather than the initial college attended. The first two sets of estimates in Figure A.6 show that raw mean outcomes across colleges are similar regardless of this treatment distinction. The third and fourth sets of estimates, however, show that the modal treatment definition leads to systematically larger value-added differences compared to the initial treatment definition. Figure A.7 shows why: the modal treatment definition generates systematic imbalances in vertical student potential (as measured by covariate-predicted outcomes) across colleges in our baseline specification, in contrast to the rough balance achieved by the initial treatment definition. As discussed in Section 2.2, this lack of balance in the modal approach could be due to at least two factors: initial institutions have different causal impacts on how long a student stays enrolled and whether she transfers elsewhere (as we find in Section 7 below), implying that the modal college definition problematically conditions on endogenous outcomes; and within a college, students who remain likely differ systematically from students who transfer elsewhere, with this selection pattern likely varying across colleges. Thus, our research design appears appropriate for studying the consequences of initial college enrollment choices, i.e. the intention-to-treat (ITT) effect of enrolling in each college, rather than cumulative exposure over time to a given college. The latter, as these results indicate, is best analyzed as an endogenous outcome, rather than as an unbalanced treatment in our matched applicant approach.

4.3.5 Alternative Earnings Measures and Missing Earnings

Our last set of checks probe the robustness of the earnings results to alternative earnings definitions and missing earnings. The top two panels in Appendix Figure A.8 scatter the main value-added estimates on earnings levels against VA estimates on log earnings (left) and earnings rank within our sample of college-goers (right). Both alternate earnings definitions produce VA estimates with correlations above 0.90 with the main estimates. The results are also unaffected by the age at which we measure earnings (third plot of Figure A.8): the main estimates of VA on earnings at ages 27-29 feature a correlation of 0.84 with less precise estimates from the subsample of older cohorts for whom we can extend earnings ages to 30-32.

Regarding missing earnings, the fourth plot in Figure A.8 shows a near-perfect correlation (0.98) between the raw earnings means at each college in our sample with the raw earnings means for those same colleges calculated by Chetty et al. (2020) using nationwide tax records, suggesting that our Texas-workers-only sample accurately captures earnings patterns across schools.[39] We also compute the correlation between a school's value-added on earnings and its value-added on appearing in the earnings sample (fifth plot in Figure A.8) and find that it is roughly zero. This helps further assuage concerns about bias from systematic selection out of the earnings sample. As a final check, the last plot in Figure A.8 shows that the main VA estimates on BA completion, which are not conditioned on appearing in the earnings sample, are essentially identical (correlation 0.99) to estimates from the subsample of students who do appear in the earnings sample.

5 Describing the Distribution of Value-Added Across Colleges

The results in the previous section presented visual evidence that relative value-added varies much less across colleges than raw outcome means, with the two measures being imperfectly correlated. We now quantify these distributional magnitudes, taking into account finite-sample estimation error. We begin by decomposing the value-added estimate $\hat{v}_j$ for each college j (relative to UT-Austin) into true (signal) value-added v_j plus orthogonal estimation error (noise) e_j: $\hat{v}_j = v_j + e_j$, where $\mathbb{E}[e_j|v_j] = 0$. The variance of the estimates across colleges, $\mathbb{V}[\hat{v}_j] \equiv \sigma^2_{\hat{v}}$, is thus the sum of the signal variance $\mathbb{V}[v_j] \equiv \sigma^2_v$ and the noise variance $\mathbb{V}[e_j] \equiv \sigma^2_e$. Subtracting the noise variance from the variance of the estimates identifies the underlying signal standard deviation of true value-added:

$$\sigma_v = \sqrt{\sigma^2_{\hat{v}} - \sigma^2_e}. \tag{10}$$

We estimate $\sigma^2_{\hat{v}}$ as the sample variance of the VA estimates across colleges, and we estimate σ^2_e as the average of the squared standard errors of those VA estimates, with both calculations weighted by enrollment to reflect variability in the value-added experienced by enrolled students.

Table 3 presents estimates of the components of Equation (10). Panel A quantifies the dispersion in raw outcome means across colleges, which is precisely measured and economically large: one standard deviation (SD) spans 17.9 percentage points of BA completion, and \$8,066 of annual earnings 8-10 years out from college entry. In contrast, Panel B shows substantially less dispersion in relative value-added, which is also rather precisely measured: one SD in the signal distribution of value-added spans 3.7 percentage points of BA completion and \$1,327 of annual earnings, or roughly a 3% increase relative to sample mean earnings of \$44,834.[40] As a gauge of the market-wide quality gradient within student admission portfolios, these

[39] Our mean earnings levels tend to be proportionally lower across all schools given that we measure earnings at slightly younger ages compared to most of the cohorts in Chetty et al. (2020).

[40] Since students complete an average of 2.89 years of college, a one SD increase in earnings VA roughly boosts earnings by

Table 3: Distributional Magnitudes of Relative Value-Added, Accounting for Estimation Error

	BA Completion	Earnings
Panel A: Raw Outcome Means		
Standard deviation of estimates across colleges	.179	8,070
Standard deviation of signal component	.179	8,066
Standard deviation of noise component	.004	276
Panel B: Relative Value-Added Estimates		
Standard deviation of estimates across colleges	.039	1,526
Standard deviation of signal component	.037	1,327
Standard deviation of noise component	.012	753
Panel C: Relationships between Raw Outcome Means and Value-Added		
Signal SD of value-added ÷ signal SD of raw outcome means	.207	.165
Correlation of VA estimate with raw outcome mean (uncorrected for noise)	.471	.174
Correlation of signal VA with raw outcome mean (corrected for noise)	.496	.200
Regression of school's value-added estimate on its raw outcome mean (SE)	.103 (.036)	.033 (.035)

Notes: Panels A and B decompose the distribution of raw outcome means and relative VA estimates, respectively, into signal and noise components per Equation (10). Panel C describes the joint relationship between these two distributions. Footnote 41 describes the procedure for correcting correlations for estimation error. All calculations weight colleges by student enrollment.

numbers imply (under a normal distribution of value-added) a gain of roughly 12.3 percentage points of BA completion and \$4,440 of annual income between the 10^{th} to the 90^{th} percentiles of the college distribution.

The signal SD of relative value-added for each outcome is around one-fifth of the SD in raw outcome means, as calculated in the first row of Panel C. As for their joint distribution, the third row of Panel C shows signal correlations of .496 for BA completion and .200 for earnings, which reflect only modest corrections for attenuation bias from estimation error.[41] Finally, to get a sense of the magnitudes of these relationships in outcome units, the fourth row of Panel C presents coefficients from a bivariate regression of a school's VA estimate on its raw outcome mean. For BA completion, a 10 percentage point increase in the raw graduation rate predicts an increase in value-added of just 1 percentage point; for earnings, a \$10,000 increase in the raw earnings mean predicts an increase in value-added of just \$330. All of the above results help quantify the visual intuition from the plots in Section 4: while economically meaningful differences in value-added do exist across colleges within student choice sets, comparisons of raw outcome means are not strongly informative about either the ordinal rankings or the cardinal magnitudes of these differences in causal effectiveness.

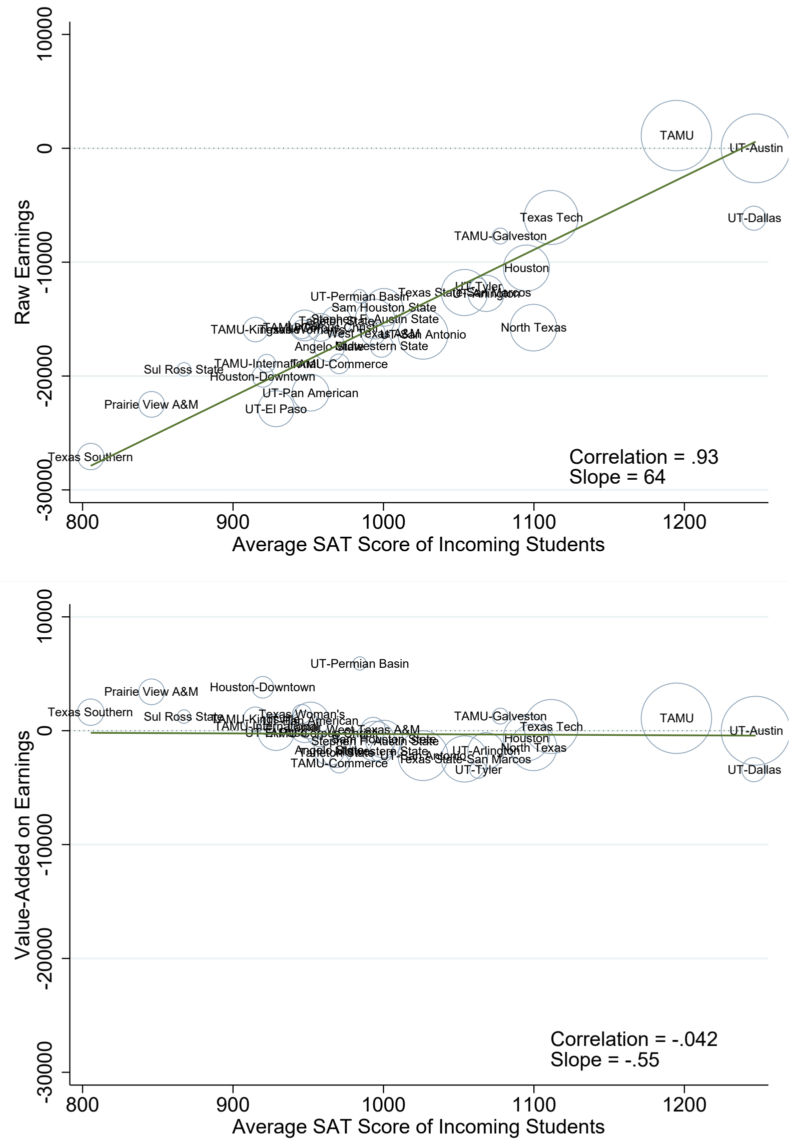
6 Institutional Predictors of Relative Value-Added

One takeaway from the preceding results is that a college's raw outcome mean is an imperfect predictor of its relative value-added within admission sets, especially for earnings. We now ask whether other prominent college characteristics—including selectivity, non-peer inputs like instructional spending and the faculty/student ratio, and intergenerational mobility statistics—are useful predictors of value-added.

just over 1% per year enrolled. This is similar to Chetty et al. (2014b)'s estimate of a 1.3% earnings boost of having a teacher with one SD better test score VA for one year of elementary or middle school.

[41] We can correct for estimation error in the correlation between the VA estimates $\hat{v}_j$ and precisely measured institutional characteristics x_j (in this case school j's raw outcome mean) by multiplying their raw correlation by $\frac{\sigma_{\hat{v}}}{\sigma_v}$, since $Corr(v,x) = \frac{Cov(v,x)}{\sigma_v \sigma_x} = \frac{Cov(\hat{v}-e,x)}{\sigma_v \sigma_x} = \frac{Cov(\hat{v},x)}{\sigma_v \sigma_x} = Corr(\hat{v},x)\frac{\sigma_{\hat{v}}}{\sigma_v}$.

Figure 6: Predicting Raw Mean Earnings vs. Value-Added with College Selectivity



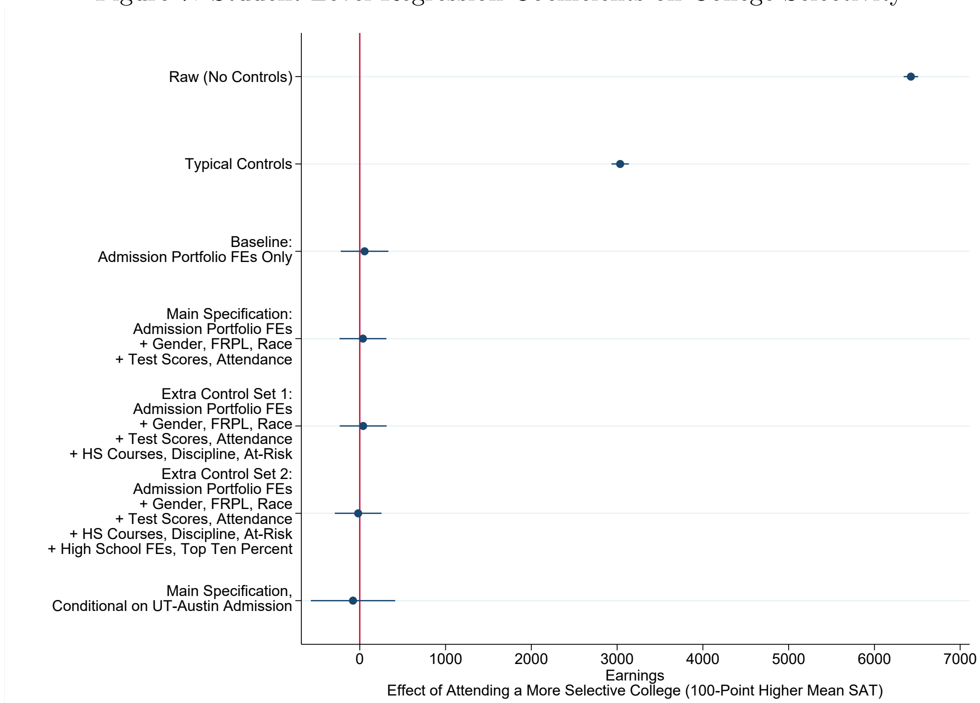
Notes: The top panel plots raw mean earnings at each college, relative to UT-Austin, against the average SAT score of incoming students at each college. The bottom panel replaces the vertical axis with the main value-added estimates from Section 4. Correlations, regression slopes, and circle sizes are weighted by student enrollment.

6.1 Selectivity vs. Value-Added

Selectivity, typically measured as the mean SAT score of a college's incoming students, is a popular measure of vertical stratification in college guides, the popular press, and the academic literature on the returns to attending different types of colleges. The rich variation in selectivity across our sample colleges (recall Figure 1) provides a fruitful setting for exploring the relationship between selectivity and the relative value-added of colleges in student choice sets. To set the stage, the top panel of Figure 6 shows that a college's mean incoming SAT score is a very strong predictor of the *raw* mean earnings of its students, with a correlation of 0.93 and a regression slope of $6,400 in annual earnings for each 100-point increase. In stark contrast, the bottom panel of Figure 6 shows that selectivity is an uninformative predictor of relative earnings value-added across the colleges in student choice sets.[42]

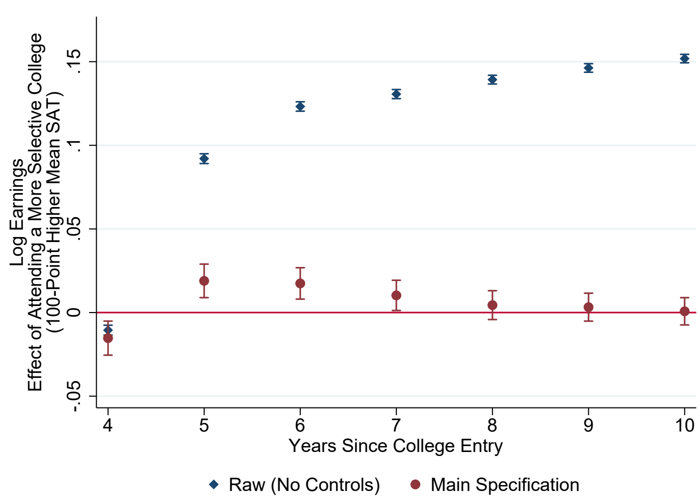
[42]Correcting for estimation error in value-added has little impact on this correlation, consistent with the results in Table 3. Replacing each college's mean SAT with its rejection rate, as another measure of selectivity, delivers similar results.

Figure 7: Student-Level Regression Coefficients on College Selectivity



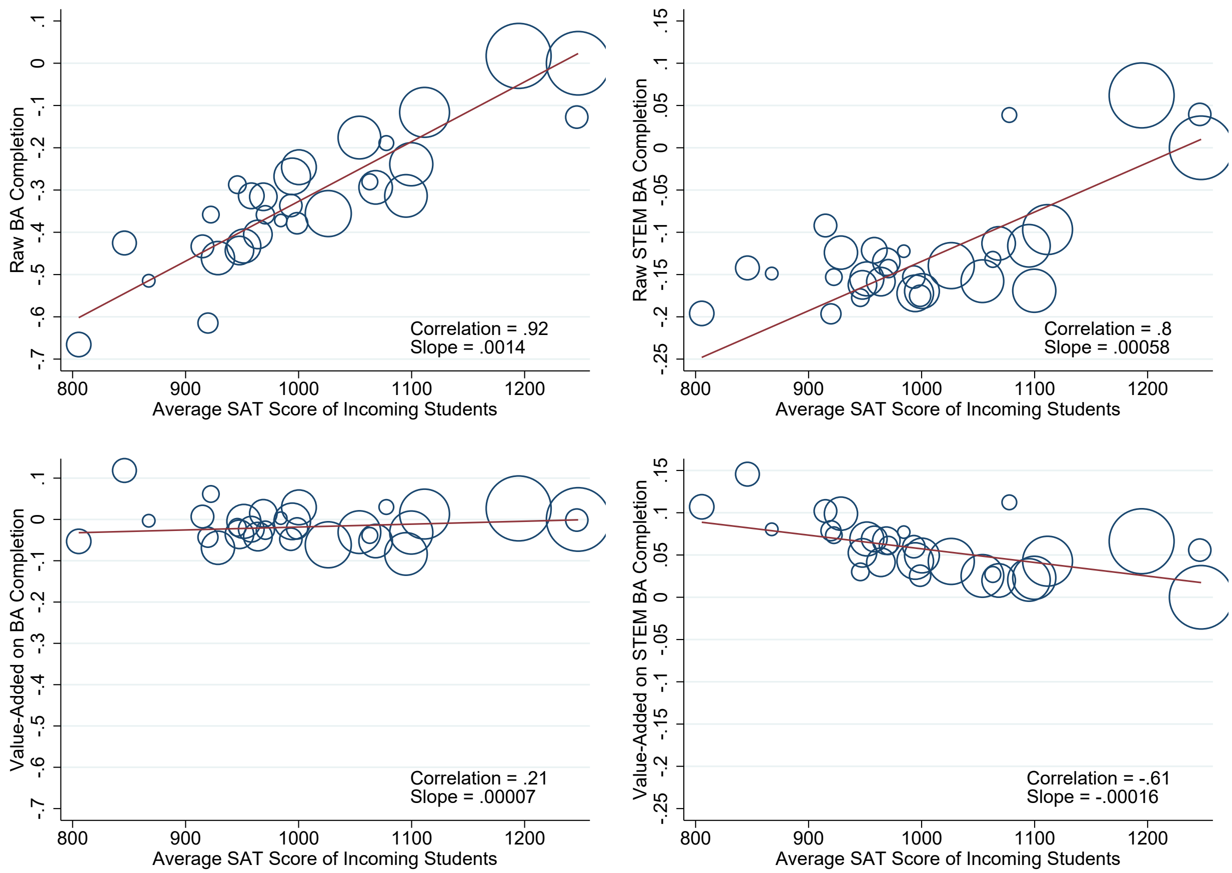
Notes: Each point estimate and robust 95% confidence interval comes from a regression of individual student outcomes on the mean incoming SAT score of the student's college. The coefficients are scaled to correspond to a 100-point increase in mean SAT scores. All specifications control for cohort fixed effects. The *Raw Specification* controls for nothing else. The *Typical Controls* specification adds controls for demographics (gender, race, FRPL), high school academic preparation (10th grade test scores, advanced coursework, and top high school GPA decile indicator), and behavioral measures of non-cognitive skills (high school attendance, disciplinary infractions, and an indicator for ever being at risk of dropping out). The *Baseline Specification* controls solely for college admission portfolio fixed effects (and cohort fixed effects). The *Main Specification* adds our core set of covariate controls: gender, FRPL, race, 10th grade test scores, and high school attendance. *Extra Control Set 1* adds controls for advanced high school coursework, disciplinary infractions, and an indicator for ever being at risk of dropping out. *Extra Control Set 2* adds fixed effects for every high school and an indicator for being in the top decile of high school GPA. The final estimate comes from running the Main Specification on the subsample of students who receive admission to UT-Austin. See Appendix Table B.3 for the corresponding numerical estimates.

Figure 8: Early Career Dynamics of the Return to College Selectivity



Notes: Each point estimate and robust 95% confidence interval comes from a separate regression of individual log student earnings, measured at a given number of years since college entry, on the mean incoming SAT score of the student's college. The coefficients are scaled to correspond to a 100-point increase in mean SAT scores. All specifications control for cohort fixed effects. The *Raw Specification* controls for nothing else. The *Main Specification* controls for college admission portfolio fixed effects and our core set of covariate controls: gender, FRPL, race, 10th grade test scores, and high school attendance.

Figure 9: Predicting Raw Degree Outcomes vs. Value-Added with Selectivity



Notes: The top two panels plot raw BA completion rates (left) and raw STEM BA completion rates (right), relative to UT-Austin, against the average incoming SAT score at each college. STEM BA completion is not conditioned on completing a BA. The bottom two panels replace the vertical axes with value-added estimates from our main specification, which regresses individual student outcomes on college treatment indicators (with UT-Austin as the reference treatment at zero), controlling for admission portfolio fixed effects and our core set of covariate controls: gender, FRPL, race, 10th grade test scores, and high school attendance. Correlations, slopes, and circles are weighted by student enrollment.

Figure 7 presents further analyses from student-level regressions of earnings on the mean SAT score of a student's college, akin to canonical specifications in the literature on the returns to college selectivity. The figure plots the regression coefficient on selectivity across several specifications. The top two omit our admission portfolio controls and suggest large returns to college selectivity, even after controlling for typical covariates like test scores and demographics. The remaining specifications control for admission portfolio fixed effects and yield selectivity effects close to (and statistically indistinguishable from) zero. The final specification in Figure 7 is a robustness check: the zero return to selectivity still holds even within the subset of students who were admitted to UT-Austin, indicating that the previous results are not driven by students choosing solely among less-selective institutions.

Interestingly, when plotting the early-career dynamics of the selectivity effect in Figure 8, we find that there does exist an earnings premium of about 2% per 100 SAT points in years 5-6 since college entry (i.e. in the first two years after projected graduation), but this premium fades out quickly to generate the stable null return in Figure 7, which averages over years 8-10. This fleeting selectivity premium could be generated by employer learning (e.g. Farber and Gibbons, 1996; Altonji and Pierret, 2001) or other early labor market frictions that give students from more selective colleges an initial advantage (e.g. campus recruiting, alumni

Table 4: Correlations of Non-Peer College Inputs with Value-Added

	BA Completion	Earnings
Non-Peer College Inputs: Correlation with Causal Value-Added		
Instructional expenditures per student	.342	.314
Academic support expenditures per student	.159	.285
Student services expenditures per student	.296	.076
Share of faculty who are full-time	.372	.447
Share of faculty who are tenured or on tenure-track	.269	.409
Average faculty salary	.083	.087
Faculty/student ratio	.433	.434
Share of degrees in STEM fields	.333	.419

Notes: This table presents correlations between college-level measures of non-peer inputs and the main value-added estimates from Section 4. Non-peer input measures come from IPEDS and are averaged across available years from 2000 to 2017. Correlations are weighted by student enrollment.

networks) which fades quickly as students with similar potential from less selective colleges catch up.[43]

Figure 9 explores the relationships between selectivity and value-added on two additional outcomes: completing any BA, and completing a BA in a STEM field. The two left panels show that, similar to earnings, raw BA completion rates exhibit a very strong correlation with selectivity (top left), but this relationship weakens dramatically when replacing raw outcomes with value-added (bottom left). A 100-point increase in incoming SAT scores predicts a 14 percentage point increase in raw BA completion, but only 0.7 percentage points in value-added. The two right panels of Figure 9 show that while raw STEM BA completion rates (top) are also strongly positively correlated with selectivity (driven especially by the large selective flagships), the sign for value-added actually reverses (bottom) to a correlation of -0.61. STEM VA increases roughly linearly at a rate of 1.6 percentage points for every 100-SAT-point *decline* away from UT-Austin, the most selective college in our sample.[44]

6.2 Non-Peer Inputs vs. Value-Added

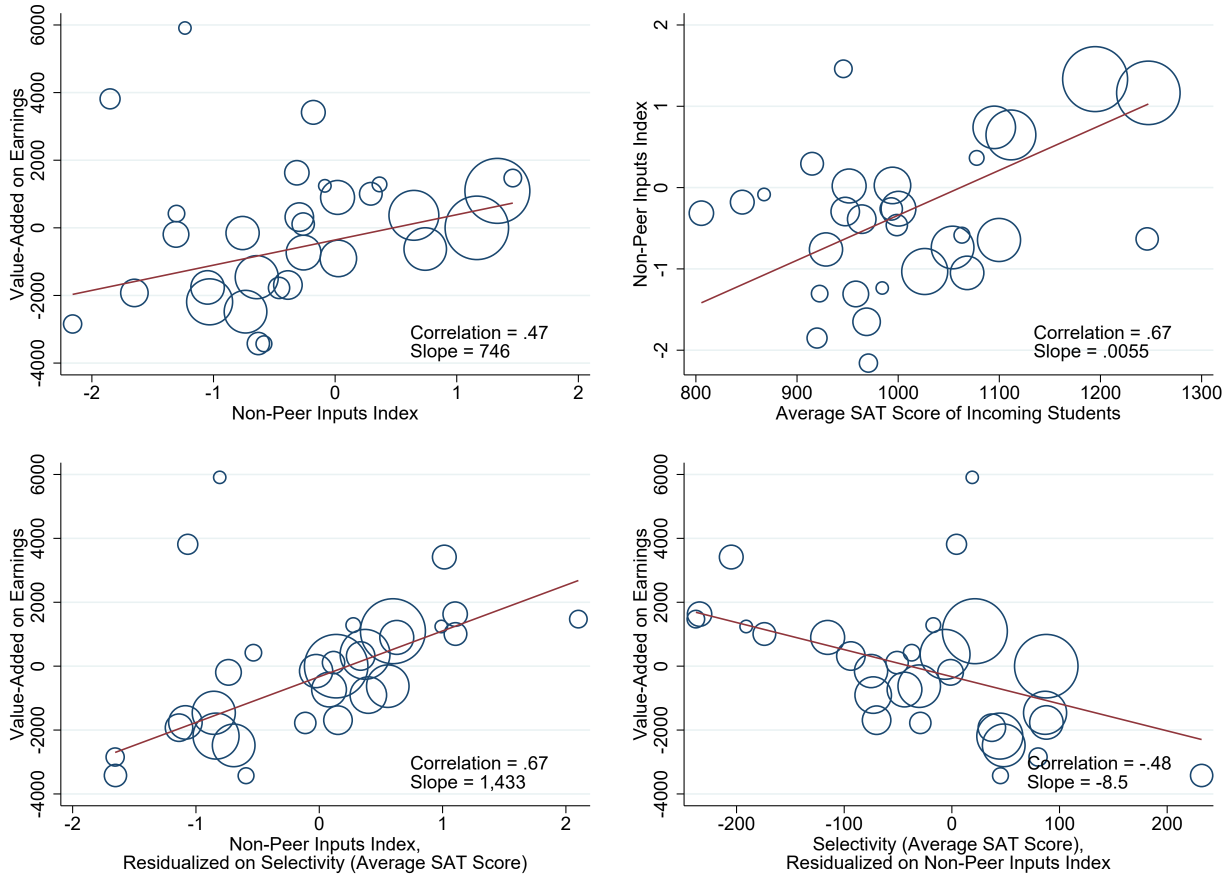
The weak correlation between selectivity and value-added suggests that peer quality may not be an essential input in causal college effectiveness. We now explore whether non-peer college inputs, like instructional spending and faculty characteristics, do a better job predicting value-added. Table 4 presents correlations to this end. Among core postsecondary educational expenses tallied by IPEDS, instructional expenditures per student modestly correlate with value-added on both BA completion (.342) and earnings (.314), while expenditures on academic support and student services offer a mixture of somewhat weaker correlations. Among faculty characteristics, we tend to see stronger predictions for earnings value-added: the share of faculty who are full-time, the share of faculty who are tenured or on the tenure-track, and the faculty-to-student ratio feature healthy correlations with earnings VA in the range of .41-.45. Average faculty salary, on the other hand, is not a strong predictor of value-added on either outcome. Finally, as a crude measure of differences in curricular inputs across campuses, we see that the share of degrees granted in STEM fields also covaries positively with value-added, especially on earnings, a pattern we will revisit in Section 7 when studying relationships between value-added on STEM degrees and earnings.

Since many of the input measures in Table 4 are correlated with each other, the top left panel of Figure 10 plots earnings VA against a single index of non-peer inputs, constructed as the predicted factor from

[43] Bordón and Braga (2020) find a similar pattern of fading selectivity returns in Chile using a regression discontinuity design.

[44] Texas A&M, which has a stronger emphasis on STEM fields, performs above trend as a selective college with relatively high VA on STEM completion. Prairie View A&M, an HBCU that also emphasizes STEM fields, exhibits the highest STEM VA estimate. See Weinberger (2018) for historical context on the role of HBCUs in producing STEM degrees.

Figure 10: Peer vs. Non-Peer Inputs and Value-Added

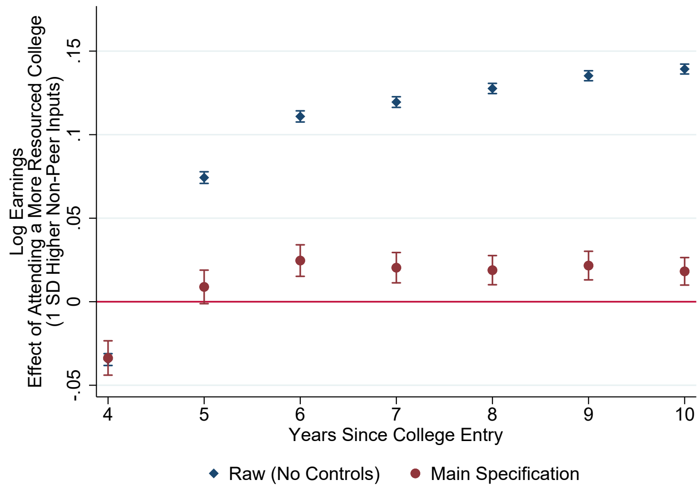


Notes: The top left panel plots earnings value-added against an index of non-peer inputs, constructed as the predicted factor from an enrollment-weighted one-factor model of instructional expenditures, full-time faculty share, tenure-track faculty share, and faculty-student ratio. The top right panel plots this non-peer inputs index against the average incoming SAT score at each college. The bottom left panel plots earnings VA against the residuals of a college-level regression of the non-peer inputs index on average SAT scores. Likewise, the bottom right panel plots earnings VA against the residuals of a college-level regression of average SAT scores on the non-peer inputs index. All correlations, regressions, and circles are weighted by student enrollment.

an enrollment-weighted one-factor model of instructional expenditures, full-time faculty share, tenure-track faculty share, and faculty-student ratio. Reflecting a distillation and tightening of the component correlations from Table 4, we see an overall correlation of 0.47 between earnings VA and this non-peer inputs index, and the slope implies that a one SD increase in non-peer inputs across colleges predicts $746 in additional earnings VA. Furthermore, in terms of early-career dynamics, Figure 11 shows an enduring return of about 2% per year to attending a college with one SD higher non-peer inputs, in contrast to the fleeting selectivity premium documented in Figure 8. One interpretation of this contrast is that the selectivity of a student's college may primarily serve as a signaling device, with early returns to that signal but rapid employer learning about true productivity (Lange, 2007; Aryal et al., 2021), which is itself more durably influenced by non-peer college resources (Bound and Turner, 2007; Bound et al., 2010; Deming and Walters, 2018).

The top right panel of Figure 10 shows that non-peer inputs and peer inputs (i.e. selectivity) are positively correlated, but not perfectly so: there are plenty of off-diagonal colleges investing more in non-peer inputs, and others investing less, than their selectivity level would predict. We therefore explore this conditional variation in the bottom two panels of Figure 10. The bottom left plot shows that the positive relationship between value-added and non-peer inputs strengthens appreciably when controlling for selectivity: the partial correlation jumps to 0.67, and the slope nearly doubles to $1,433 in extra predicted value-added for each

Figure 11: Early Career Dynamics of the Return to Non-Peer College Inputs



Notes: Each point estimate and robust 95% confidence interval comes from a separate regression of individual log student earnings, measured at a given number of years since college entry, on the index of non-peer inputs of a student's college, constructed as the predicted factor from an enrollment-weighted one-factor model of instructional expenditures, full-time faculty share, tenure-track faculty share, and faculty-student ratio. The coefficients are scaled to correspond to a one standard deviation increase in non-peer inputs. All specifications control for cohort fixed effects. The *Raw Specification* controls for nothing else. The *Main Specification* controls for college admission portfolio fixed effects and our core set of covariate controls: gender, FRPL, race, 10th grade test scores, and high school attendance.

standard deviation increase in non-peer inputs, residualized on selectivity. The bottom right panel shows the joint implication of this result and those from the previous subsection: selectivity, residualized on non-peer inputs, is actually a negative predictor of earnings VA, perhaps reflecting a within-campus competition channel that becomes more apparent when comparing colleges with similar non-peer resources but different peer composition.[45] These results challenge popular notions of college "quality" as positively-weighted indices of peer and non-peer inputs: although they broadly move together, peer and non-peer inputs offer contrasting partial correlations with relative college value-added.

More broadly, these results help to resolve a long-standing controversy in the literature on whether meaningful economic returns accrue to students who attend higher "quality" colleges. A majority of papers in this vein—e.g., Brewer et al. (1999), Black and Smith (2006), Long (2008), and Dillon and Smith (2018)—have found consistent evidence for strong returns, while Dale and Krueger (2002, 2014) stand out as a notable exception. Our approach highlights not only the econometric importance of addressing selection bias on vertical potential and thinking carefully about sorting on horizontal match effects, but also the importance of carefully defining college "quality."[46] While we extend Dale and Krueger (2002, 2014)'s main conclusion that college selectivity *per se* does not predict relative value-added within student choice sets (beyond the short-lived initial premium we document in Figure 8), this does not imply that college-level value-added differentials—a more flexible notion of quality—are absent. Rather, we document a non-trivial distribution of relative value-added across colleges, and simply show that selectivity poorly summarizes this distribution, with non-peer inputs performing better, especially conditional on selectivity. Moreover, by letting each college have its own unique impact on student outcomes, our value-added approach lets the data flexibly determine the ordering of the quality space, rather than imposing an ex-ante ordering based on a single-dimensional college observable or index of observables.

[45] Consistent with this result, de Roux and Riehl (2020) find a negative return to peer quality among Colombian college students in the same institution and major but exogenously assigned to vertically differentiated peer cohorts.

[46] In a similar spirit, Black and Smith (2006) emphasize the need for multiple proxies to mitigate measurement error in univariate constructions of college quality.

6.3 Intergenerational Mobility Statistics vs. Value-Added

Finally, we explore the potential for college-level intergenerational mobility statistics to serve as proxies for value-added. College mobility measures—the likelihood that students from disadvantaged backgrounds at a given college end up with better outcomes as adults—have gained prominence in debates over whether higher education mitigates or exacerbates economic inequality. A key question is to what extent these measures reflect causal college effectiveness in boosting the outcomes of disadvantaged students, versus differential ability across colleges in selecting disadvantaged students who would do well regardless of where they attend.

Chetty et al. (2020) estimate a comprehensive set of college mobility statistics using nationwide administrative tax records linking parental income, child income, and child college attendance. They measure mobility at college j as the probability that a child who attends j makes it into the top 20% of the national earnings distribution, conditional on that child having parents in the bottom 20% of the earnings distribution.[47] Our data unfortunately do not include a continuous measure of parental income, nor do they include nationwide earnings, so we cannot reproduce Chetty et al. (2020)'s exact mobility measures with our microdata. The top left panel of Figure 12, however, demonstrates that we can get rather close: instead of conditioning on students from the bottom family income quintile, we can condition on eligibility for free or reduced price lunch (FRPL) in high school, and instead of measuring the probability of making it into the top quintile of national earnings, we can measure the probability of making it into the top tercile of earnings among our Texas college-goers. This analogous mobility measure in our data has a correlation of 0.96 with Chetty et al. (2020)'s measure from national tax data, suggesting that both samples are capturing the same patterns in raw intergenerational mobility across colleges.[48]

Having established common measurement of the raw mobility patterns, the top right panel of Figure 12 investigates to what degree a college's raw mobility measure predicts its relative value-added for disadvantaged students. We first estimate each college's value-added on our analog mobility measure by running our main specification on the subsample of low-income students, and specifying the outcome as an indicator for whether the student makes it into the top tercile of earners. We plot these causal mobility estimates against the Chetty et al. (2020) raw mobility measures in the top right panel of Figure 12, adding back the UT-Austin mean to make the levels comparable. The correlation is roughly zero: a college's raw mobility statistic is an uninformative predictor of its relative value-added for disadvantaged students, at least those choosing among multiple admissions offers. This outcome measure may miss causal impacts across less ambitious leaps in the income distribution, so in the bottom panel of Figure 12 we swap in our main measure of value-added on average earnings levels, which exhibits a modest positive correlation of 0.28 with Chetty et al. (2020)'s raw mobility statistic.[49]

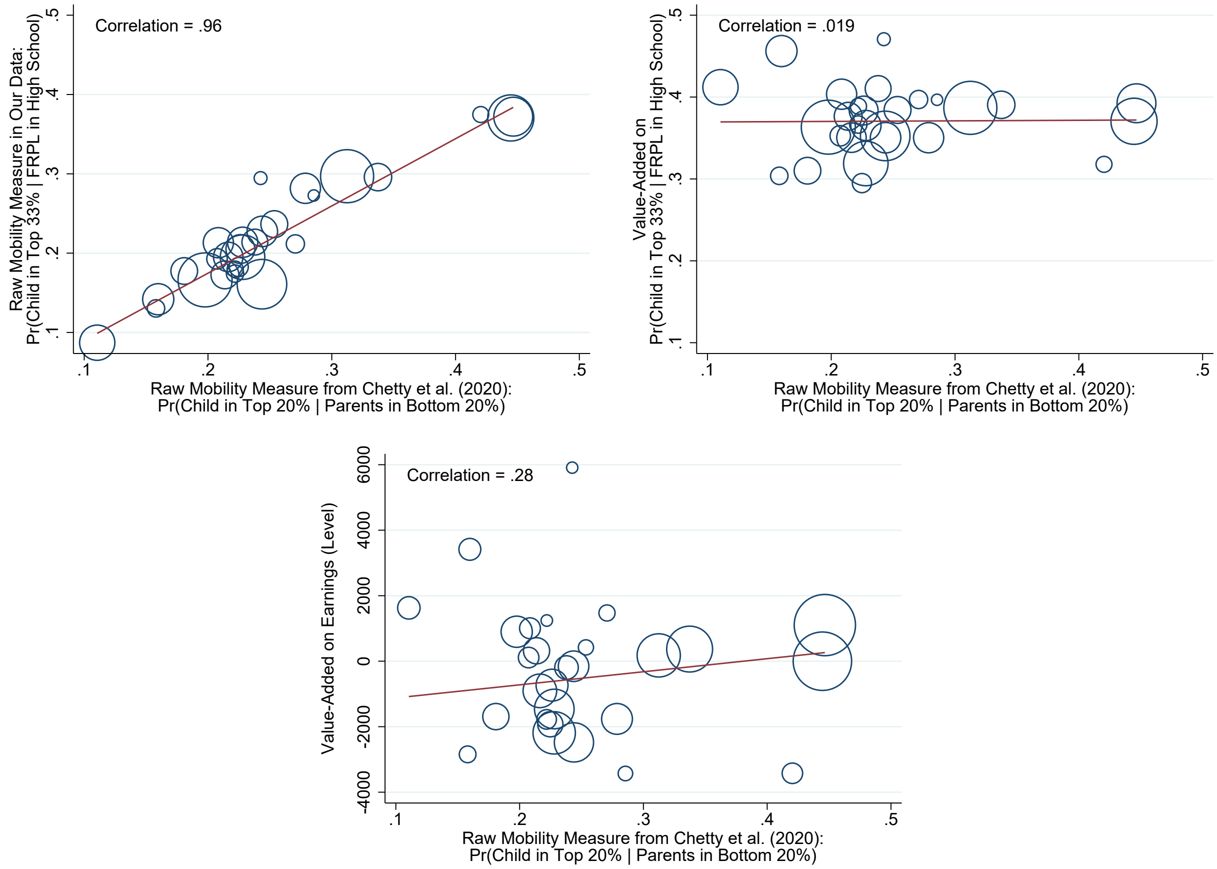
The weak relationship between a college's raw mobility statistic and its relative value-added is consistent with the results from Figure 3: the "typical controls" approach, which controls for family income like the raw mobility statistics, appears to substantially under-correct for selection bias. Because the raw intergenerational mobility statistics control solely for family income, they likely still reflect other systematic differences

[47] Chetty et al. (2020) denote this term the "success rate," which they then multiply by an "access" measure (the fraction of students at each college who actually come from families in the bottom quintile) to generate their "mobility rate" metric. We focus on the success rate, since this is the component of the mobility rate that captures variation in the outcomes of observably similar students across colleges and is thus readily comparable to value-added.

[48] Chetty et al. (2020) condense some multi-campus university systems into single observational units; e.g., merging TAMU-Galveston with TAMU-College Station, and Houston-Downtown with the larger University of Houston campus. For comparability in this subsection we combine the two schools in each pair as an enrollment-weighted average of individual estimates.

[49] This main measure of value-added on earnings levels is estimated on our full sample of students; using VA estimates from the subsample of low-income students yields similar results, given our findings below in Section 8 that value-added schedules do not appear to vary systematically by low-income status.

Figure 12: Intergenerational Mobility Statistics vs. Value-Added



Notes: The horizontal axis in each plot is the college-level intergenerational mobility statistic reported by Chetty et al. (2020): the probability that a student at a given college reaches the top quintile of earnings conditional on having parents in the bottom quintile of earnings, estimated using national tax records. The vertical axis in the top left panel is the closest analog to this measure in our Texas sample: the raw probability that a student reaches the top income third of Texas college-goers conditional on being eligible for free or reduced price lunch in high school. The vertical axis in the top right panel swaps out the raw mobility measure for its causal analog: a college's value-added on that probability, estimated using our main specification but limited to the subsample of FRPL students. The bottom panel replaces the vertical axis with our main earnings VA measure from Section 4. Correlations and circles are weighted by student enrollment.

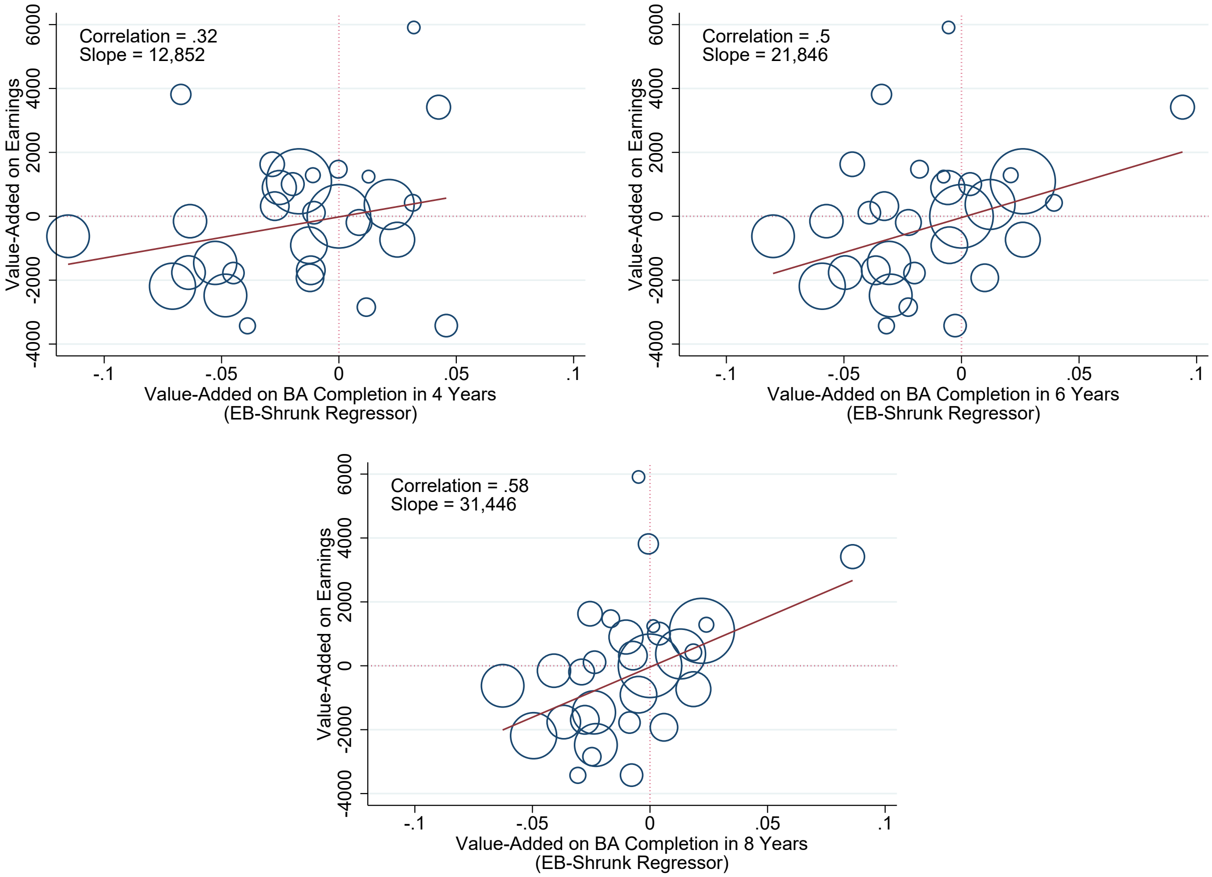
in student characteristics across campuses, like levels of academic preparation, rather than causal college impacts.[50]

7 Potential Mechanisms: Relationships Across Student Outcomes

The preceding sections have focused on earnings and degree completion as separate outcomes on which colleges may add value. We now explore the interrelationships of value-added on these and other key outcomes to illustrate some of the potential mechanisms through which colleges shape the path from enrollment to earnings. To address attenuation bias in these relationships from estimation error in value-added, we construct Empirical Bayes (EB) forecasts (e.g. Robbins, 1956) that shrink each college's value-added estimate

[50]Chetty et al. (2020) also estimate college mobility statistics that control for SAT/ACT scores but, due to data agreements, are unable to present college-level estimates that would allow individual comparisons. Chetty et al. (2020) further implement a specification in the spirit of Dale and Krueger (2002) that controls for SAT/ACT score sending behavior as a proxy for college applications, which they find only modestly attenuates outcome differences across colleges. This contrasts with our results, as well as those of Dale and Krueger (2002, 2014), possibly because their specification involves features we flag in Section 4.3.4 as liable to under-correct for selection bias, including defining the treatment as the modal rather than initial college, having application proxies only, and including students who only apply to the college they end up attending.

Figure 13: Value-Added on Earnings vs. Value-Added on BA Completion in 4, 6, 8 Years



Notes: The vertical axis of each plot is the main earnings VA estimate from Section 4. The horizontal axes are the Empirical Bayes shrunk forecasts (see Appendix A.1) of value-added on BA completion within a given number of years, which correct the regression slopes for mild attenuation bias from estimation error in the regressor. Correlations, regression slopes, and circles are weighted by student enrollment.

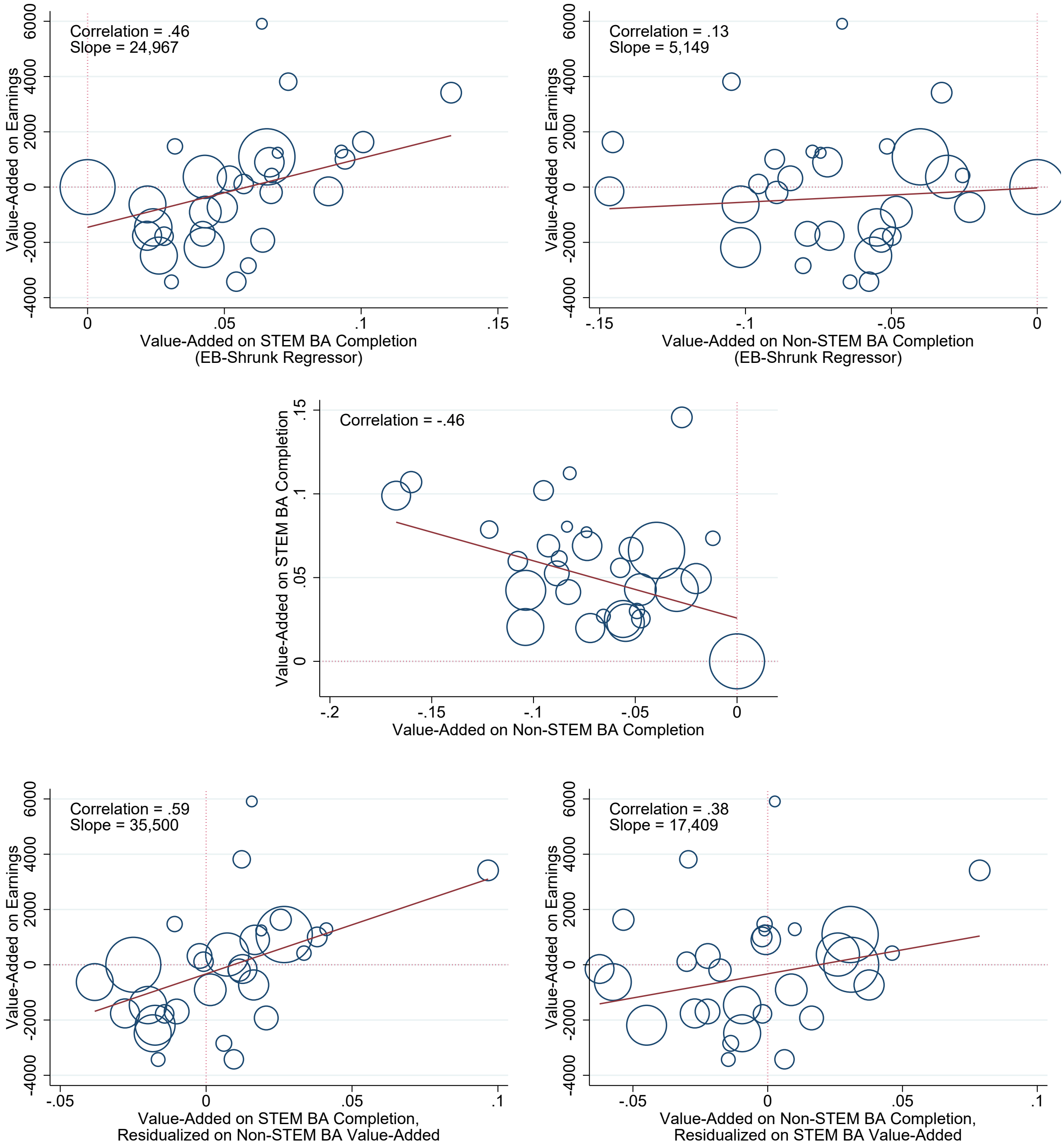
towards the mean in proportion to its imprecision.[51] This issue turns out to be of minimal quantitative importance in our case, but we use EB-shrunk regressors to mitigate concerns about measurement error that generally arise when conducting such analyses. See Appendix A.1 for further details on how we construct the EB forecasts.

7.1 Value-Added on BA Completion vs. Value-Added on Earnings

We start in Figure 13 with the relationship between a college's value-added on BA completion (horizontal axis) and its value-added on earnings (vertical axis). The top left panel shows a moderate positive correlation of 0.32 when measuring BA completion according to the "on-time" criterion of finishing within 4 years of entry. The top right panel extends the completion window to 6 years, and the bottom panel to 8 years. Each extension both tightens the distribution of BA VA and strengthens its relationship with earnings VA, with the slopes increasing from \$1,285 to \$2,185 to \$3,145 in additional earnings VA associated with a 10 percentage point increase in BA VA. Colleges are therefore more dispersed in their effectiveness at boosting on-time completion relative to eventual completion, and strict measures of on-time graduation may underestimate a college's ultimate value-added in the labor market.

[51] Appendix Figure A.1 plots these EB forecasts against the VA estimates. Since the VA estimates for most colleges are relatively precise and/or already close to the mean, their shrunk EB forecasts are rather similar.

Figure 14: Value-Added on Earnings vs. Value-Added on STEM and Non-STEM BA Completion



Notes: Empirical Bayes shrunk value-added estimates are used as regressors to correct for mild attenuation bias from estimation error in value-added (see Appendix A.1). The top two panels plot earnings VA against EB-shrunk STEM BA (left) and non-STEM BA (right) completion VA. The middle panel plots value-added on STEM BA completion against value-added on non-STEM BA completion. The bottom left panel plots earnings VA against the residual from a college-level regression of EB-shrunk STEM BA VA on EB-shrunk non-STEM BA VA. The bottom right panel plots earnings VA against the residual from a college-level regression of EB-shrunk non-STEM BA VA on EB-shrunk STEM BA VA. Correlations, regressions, and circles are weighted by student enrollment.

7.2 Unpacking BA Completion: STEM vs. Non-STEM Degrees

We next unpack BA completion by STEM versus non-STEM degrees to explore whether value-added on completing different majors has different predictions for earnings effects. The top two panels of Figure 14 shows that a college's value-added on completing a STEM degree (left) has a strong correlation with its earnings VA, while value-added on non-STEM completion (right), in contrast, has almost no bivariate relationship with earnings VA. The middle panel shows that STEM and non-STEM VA are negatively correlated, however, suggesting that colleges may face tradeoffs across fields in boosting degree completion. Exemplifying this to the extreme, UT-Austin at the (0,0) origin has the lowest value-added on STEM completion, as we also saw in Section 6, but simultaneously has the *highest* VA on non-STEM degree completion. Simple bivariate correlations like the top right panel of Figure 14 may thus underestimate the earnings gains of producing more non-STEM degrees, since this is correlated with fewer STEM degrees in the cross-section. The bottom right panel of Figure 14, consistent with this hypothesis, shows that the relationship between earnings VA and non-STEM VA becomes significantly more positive when controlling for STEM VA. On the other hand, STEM VA becomes an even stronger predictor of earnings VA when controlling for non-STEM VA (bottom left panel of Figure 14). The partial slope coefficients imply that a 10 percentage point controlled increase in STEM VA predicts roughly twice as much earnings VA ($3,550) as the same controlled increase in non-STEM VA ($1,741).

7.3 Other Potential Mechanisms: Persistence, Transfer, and Industry

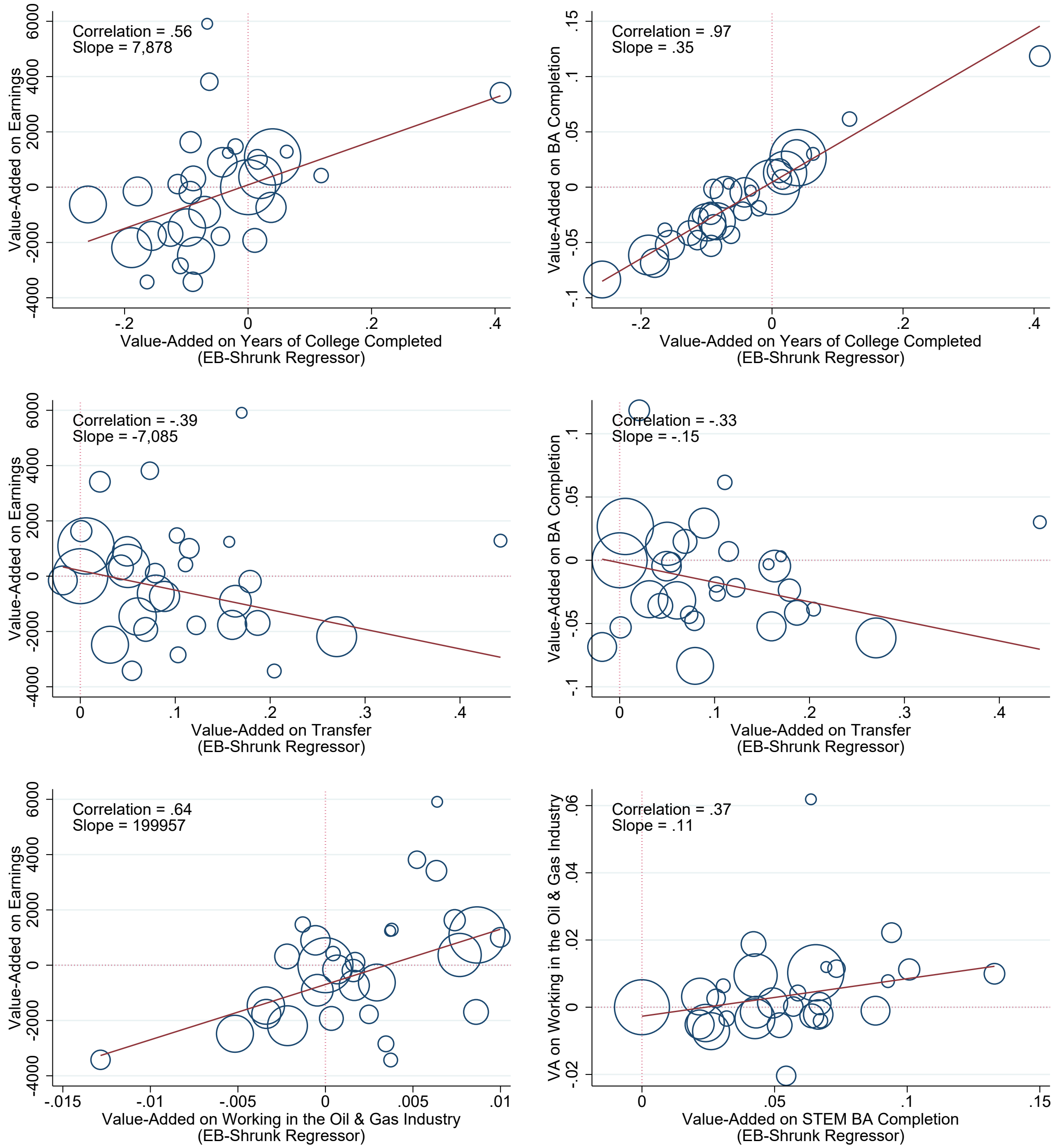
Figure 15 highlights the predictive power of other intermediate outcomes on earnings value-added. The top left panel shows that value-added on years of college completed, a more continuous measure of educational attainment, exhibits a strong correlation with earnings VA: an extra year of attainment VA is associated with an additional $7,878 in earnings VA.[52] The top right panel shows that the colleges that boost persistence are nearly always the colleges that boost BA completion, which is not strictly mechanical given that over a third of college-goers make some progress but never complete a degree. The middle two panels illustrate transfer as a potentially detrimental mediator: colleges that induce more students to transfer to other institutions (horizontal axis) also tend to have lower value added on earnings (left) and degree completion (right). Finally, the bottom two panels illustrate the potential for college-industry linkages to help explain differences in earnings effects. As a test case, the bottom left plot shows a strong positive correlation between a college's value-added on working in the oil and gas industry and its earnings VA. The bottom right plot, which shows the correlation between VA on oil and gas employment and VA on STEM degrees, suggests that majors may play a role in these linkages.

8 Match Effects: Student Heterogeneity in Relative Value-Added

Our final set of results probe the potential for (mis)match effects by allowing the impacts of attending different colleges to vary flexibly across our suite of observable student characteristics. We first visualize separate value-added estimates at each college for different student subpopulations to gauge the overall quantitative importance of match effects, then analyze student heterogeneity in the returns to selectivity to speak directly to empirical questions that arise in debates over the role of mismatch in higher education.

[52] Under a strong exclusion restriction that different colleges only affect earnings through their differential effects on persistence, this slope could be interpreted as an IV estimate of the return to a year of college, with college indicators acting as instruments, the horizontal axis as the first stage, and the vertical axis as the reduced form. $7,878 is a gain of 17.6% above the sample mean of $44,834, which is on the high end of the distribution of estimated returns (e.g. Card, 2001; Oreopoulos and Petronijevic, 2013), suggesting violations of exclusion by colleges adding value to earnings through channels other than persistence.

Figure 15: Other Potential Mechanisms: Persistence, Transfer, and Industry of Employment

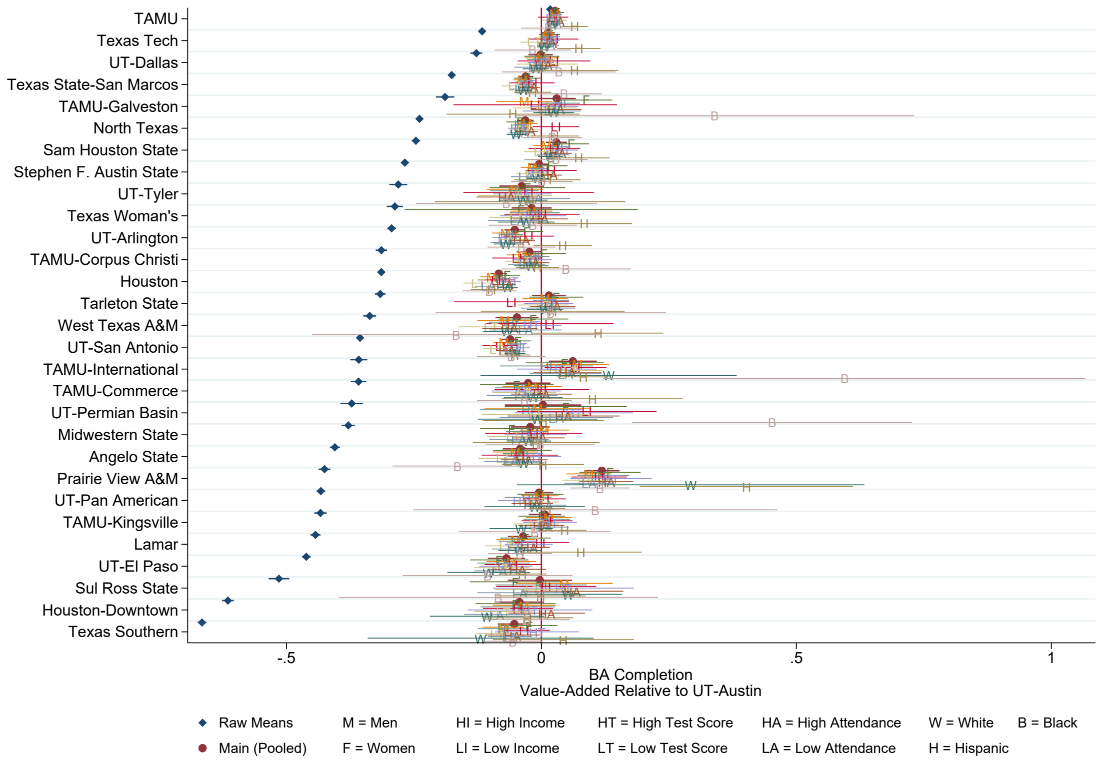


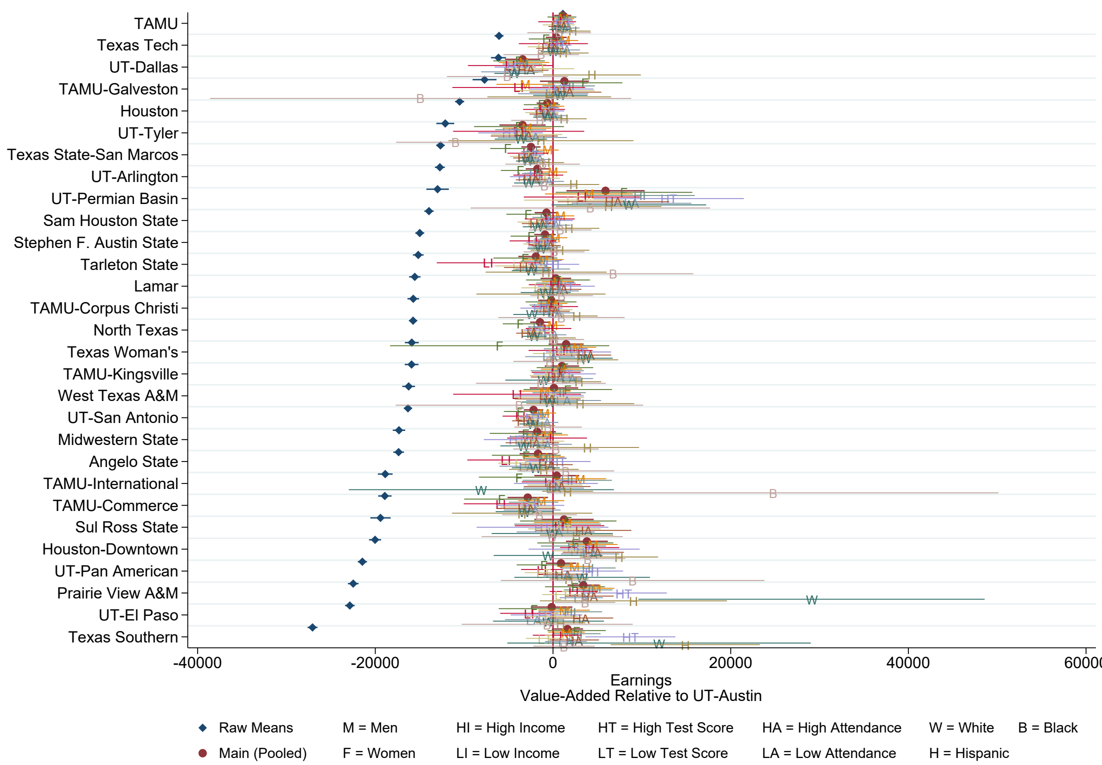
Notes: Empirical Bayes shrunk value-added estimates are used as regressors to correct for mild attenuation bias from estimation error in value-added (see Appendix A.1). Each panel plots the main VA estimate on one outcome against the EB-shrunk VA estimate on another outcome. Years of college completed, transfer, and working in the oil and gas industry are defined in Section 2. Correlations, regressions, and circles are weighted by student enrollment.

8.1 Value-Added Estimates Across Student Subpopulations

Figure 16 plots the distribution of value-added estimates across different student subpopulations. Following the graphical format from Section 4, we plot the raw outcome means at each college for comparison, and the main VA estimates (which pool across all student subgroups) appear in solid circles. Below each college's main estimate is the set of subgroup VA estimates that come from estimating our main specification within

Figure 16: Value-Added Estimates Across Student Subpopulations





Notes: Each set of point estimates and robust 95% confidence intervals come from regressions of individual student outcomes on college treatment indicators, omitting UT-Austin as the reference treatment (signified by the vertical line at zero). All specifications control for cohort fixed effects. The Raw Means specification controls for nothing else. The Main Specification adds our core set of covariate controls: gender, FRPL, race, 10th grade test scores, and high school attendance. Each remaining specification estimates the Main Specification in a specified subpopulation of students: men vs. women, high income (non-FRPL) vs. low income (FRPL), high 10th grade test score (above median) vs. low 10th grade test score (below median), high high school attendance (above median) vs. low high school attendance (below median), and white vs. Hispanic vs. Black.

each split of the sample: men vs. women, high income (non-FRPL) vs. low income (FRPL), high test score (above median) vs. low test score (below median), high attendance (above median) vs. low attendance (below median), and white vs. Hispanic vs. Black.

Two patterns emerge immediately from this exercise. First, a handful of estimates are extremely imprecise, with large confidence intervals and outlier point estimates. These come from college-subpopulation cells with small numbers of students, e.g., Black students at TAMU-International, which is 95% Hispanic, and white students at Prairie View A&M, which is 96% Black.[53] Second, and more substantively, we see that the more precisely estimated subgroup-specific value-added estimates tend to cluster around the main pooled estimate for each college, with deviations rarely statistically distinguishable from each other.

To quantify the role of student-college match effects in explaining student outcomes, we also run a pooled specification in the full sample, reported in Tables B.4 and B.5, that interacts each college treatment (and each admission portfolio) with each student covariate: gender, low-income (FRPL), race, standardized test score, and standardized attendance. An F-test shows that these treatment-covariate interactions are jointly statistically significant at the 1% level, suggesting match effects may be present, but the question remains as to whether they are economically meaningful. To gauge the magnitude of this heterogeneity in driving student outcomes, we can compare the R^2 of this fully interacted specification to a specification without treatment-covariate interactions (but keeping admission portfolio-covariate interactions to maintain a common control set). This comparison attributes only a trivial contribution of match effects to explaining student outcomes, with the R^2 increasing from .2423 to .2428 for BA completion (.2088 to .2090 in adjusted R^2), and from .1611 to .1616 for earnings (.1206 to .1207 in adjusted R^2).

These results suggest limited scope for student heterogeneity in relative college value-added, at least within actual student choice sets and across the dimensions of diversity that we can observe in our data. Methodologically, this may also help to allay concerns that pooling the estimation of college value-added across different types of students introduces biases from match effects. We provide further evidence of this in Figure A.9, which shows that the main estimates from Section 4 are quite similar to estimates from the specification with treatment-covariate interactions evaluated for a student with average covariate values.[54]

8.2 Testing for Mismatch: Heterogeneity in the Returns to Selectivity

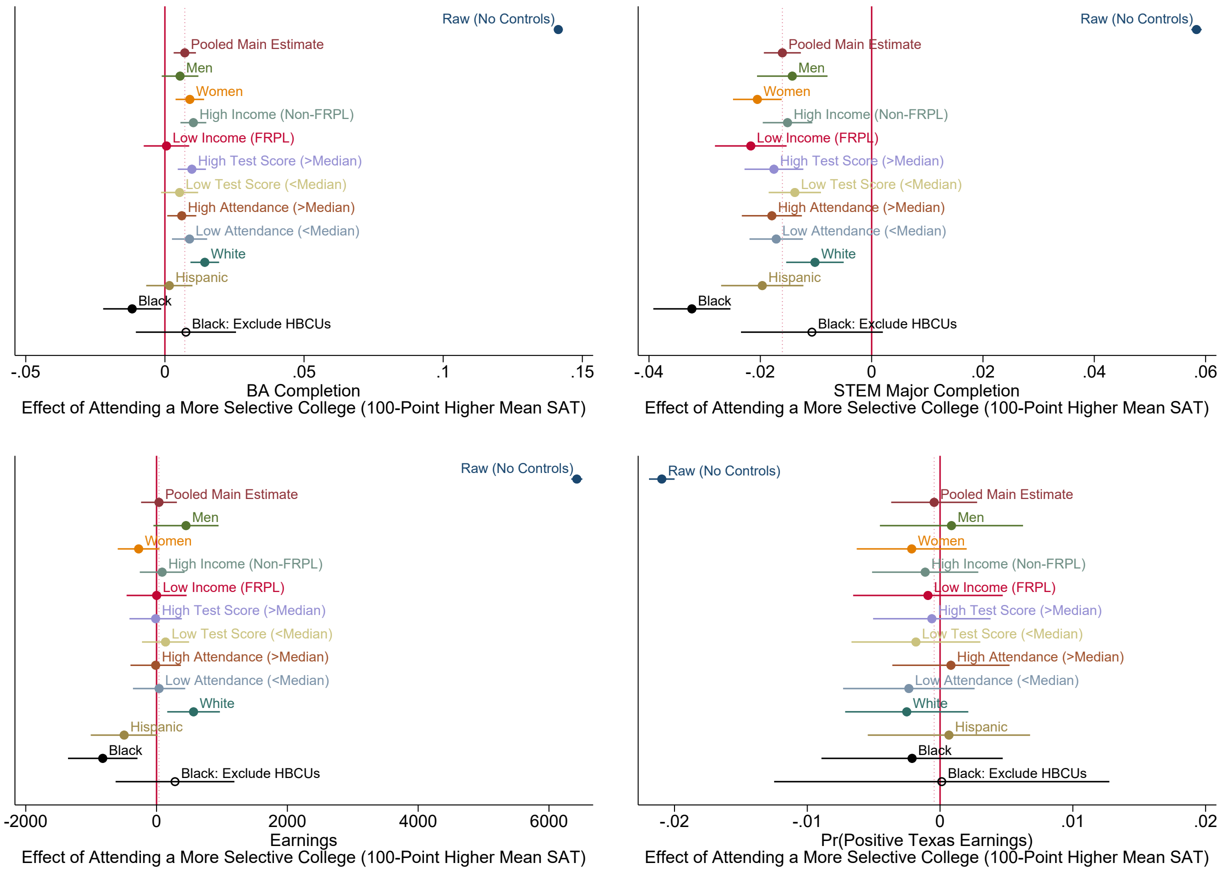
Finally, Figure 17 directly examines whether the returns to college selectivity within student admission sets vary systematically across different types of students. Each estimate is the coefficient from a student-level regression of a student's outcome on the mean SAT score of her college peers. For comparison, the estimate at the top of each plot includes no controls, and the next is the overall pooled estimate from our main specification, which controls for admission portfolio fixed effects and our core set of student covariates. We then run this specification separately for each student subpopulation, as in the previous subsection: men vs. women, high income vs. low income, high test score vs. low test score, high attendance vs. low attendance, and white vs. Hispanic vs. Black.

The results in Figure 17 show that for each outcome, the estimated impacts of attending a more selective college are rather similar across student subpopulations. We find little evidence of significant heterogeneity

[53] All of these cells have at least 10 students in them.

[54] We drop the five colleges that enroll almost exclusively one particular demographic group—Blacks at Prairie View A&M and Texas Southern, Hispanics at TAMU-International and UT-Pan American, and women at Texas Woman's University—since these schools have extremely imprecise estimates for the average student. The two schools with non-trivial deviations between the two estimates—TAMU-Galveston and UT-Permian Basin—are two of our smallest schools, and thus their VA estimates also become noisier when interacted with covariates.

Figure 17: Testing for Mismatch: Heterogeneity in the Returns to Selectivity



Notes: Each point estimate and robust 95% confidence interval comes from a regression of individual student outcomes on the mean incoming SAT score of the student's college. The coefficients are scaled to correspond to a 100-point increase in mean SAT scores. All specifications control for cohort fixed effects. The Raw specification controls for nothing else. The Pooled Main Estimate adds college admission portfolio fixed effects and our core set of covariate controls: gender, FRPL, race, 10th grade test scores, and high school attendance. Each remaining specification estimates the main specification in a specified subpopulation of students: men vs. women, high income (non-FRPL) vs. low income (FRPL), high 10th grade test score (above median) vs. low 10th grade test score (below median), high high school attendance (above median) vs. low high school attendance (below median), and white vs. Hispanic vs. Black. The final specification in each plot omits the two historically Black universities, Texas Southern and Prairie View A&M, from the set of college treatments in the subsample of Black students.

in the returns to selectivity by gender, family income as measured by FRPL eligibility, cognitive skills as measured by high school test scores, and non-cognitive skills as measured by high school attendance.[55] Black students are the main exception to this trend: they experience economically small but statistically significant negative impacts of selectivity, with a 100-point increase in peer mean SAT scores predicting a 1 percentage point reduction in BA completion, a 3 percentage point reduction in STEM completion (compared to a 2 percentage point reduction for the average student), and a $1,000 reduction in earnings.

The estimates at the bottom of each plot demonstrate that this modest pattern of "mismatch" for Black students is driven by the two large historically Black universities (HBCUs) in Texas: Prairie View A&M and Texas Southern. The selectivity coefficients for Black students increase to become indistinguishable from the pooled estimate when we exclude these two schools from the sample. These two HBCUs have the lowest mean SAT scores among Texas public universities, but they generate above-average value-added for their students—especially Prairie View A&M—as shown in Figure 6. Their presence thus exerts downward

[55] We also find no gender difference in the effect of selectivity on appearing in the Texas earnings data, as shown in the bottom right panel of Figure 17. This contrasts with some of the results in Ge et al. (2018), though the difference may be driven by the fact that the women in their sample, the College and Beyond survey, attended college in the 1970s, compared to our cohorts enrolling 30 years later in the 2000s.

pressure on the slope between selectivity and value-added for Black students, as Black students comprise over 95% of their enrollment, and 28% of Black students in our sample attend one of them. Hence, the slight apparent "mismatch" effect for Black students is actually driven by their distinct college choice set relative to other students, rather than by significant heterogeneity in the effects of attending a given selective college. That is, across the non-HBCUs, i.e. the schools that non-Black students attend, Black students experience very similar impacts of selectivity as their peers from other backgrounds.

9 Conclusion

Average student outcomes vary enormously across colleges in the United States. We find that these outcome differences attenuate dramatically when comparing students who applied to and gained admission to the same set of schools, suggesting that causal value-added differences across the colleges within student choice sets play a smaller role than selection bias in producing these observed outcome disparities. The distribution of relative value-added is not degenerate, however, and while it is poorly summarized by selectivity, we find that non-peer college inputs like instructional spending do covary positively with value-added, especially conditional on selectivity. Colleges that boost earnings also tend to boost intermediate outcomes like persistence and BA completion, shedding light on the potential mechanisms through which colleges shape the path from enrollment to earnings. Methodologically, we develop a framework for identifying and interpreting value-added under varying assumptions about match effects and sorting gains, and we empirically examine whether college value-added varies systematically across observable student characteristics. We fail to find evidence of substantial match effects across observable dimensions of student diversity, with a small selectivity mismatch effect for Black students driven by a different choice set (the availability of HBCUs) rather than heterogeneity in value-added schedules across the same schools relative to peers from other backgrounds.

While these results offer new insights into the consequences of college choices, the limitations of our approach point to several fruitful avenues for future work. First, our data span a very large state with a rich diversity of students and a wide array of colleges, but our identifying sample is ultimately limited to public universities in Texas. Determining whether the patterns we find extend to other settings and to other postsecondary sectors, including non-profit and for-profit private colleges, will be valuable for understanding variation in college value-added across other segments of the higher education landscape. Second, our identification strategy isolates variation among students who actually applied to and were admitted to the same colleges, leaving open the possibility that college impacts could differ for students outside of these comparison sets, like disadvantaged students induced to attend highly selective colleges through interventions that substantially change their application and enrollment behavior. Third, our investigation of match effects between colleges and students is limited to observable dimensions of heterogeneity; questions remain as to what extent students sort and gain heterogeneously along unobservable dimensions. Fourth, our research design focuses on institutions as the treatments of interest, but our results on STEM vs. non-STEM degrees, and a growing literature on the causal returns to different college majors, suggest curriculum choices within an institution may be more consequential than the institution itself in shaping a student's labor market outcomes (Altonji et al., 2012; Webber, 2014; Kirkeboen et al., 2016; Andrews et al., 2017; Aucejo et al., 2020; Bleemer, 2020c). Finally, while our data linkages permit the measurement of college value-added on multiple policy-relevant outcomes, a wealth of other important outcomes remain to explore, including marriage and family formation, occupational choices, employer characteristics, and geographic mobility.

References

Abdulkadiroglu, A., J. Angrist, and P. Pathak (2014): "The Elite Illusion: Achievement Effects at Boston and New York Exam Schools," *Econometrica*, 82, 137–196.

Abdulkadiroglu, A., P. A. Pathak, J. Schellenberg, and C. R. Walters (2020): "Do Parents Value School Effectiveness?" *American Economic Review*, 110, 1502–1539.

Abowd, J. M., R. H. Creecy, and F. Kramarz (2002): "Computing Person and Firm Effects Using Linked Longitudinal Employer-Employee Data," Tech. rep., Center for Economic Studies, US Census Bureau.

Abowd, J. M., F. Kramarz, and D. N. Margolis (1999): "High Wage Workers and High Wage Firms," *Econometrica*, 67, 251–333.

Aisch, G., R. Gebeloff, and K. Quealy (2014): "Where We Came From and Where We Went, State by State," *New York Times*, 19 August.

Akhtari, M., N. Bau, and J.-W. Laliberte (2020): "Affirmative Action and Pre-College Human Capital," NBER Working Paper No. 27779.

Altonji, J. G. (1993): "The Demand for and Return to Education When Education Outcomes are Uncertain," *Journal of Labor Economics*, 11, 48–83.

Altonji, J. G., E. Blom, and C. Meghir (2012): "Heterogeneity in Human Capital Investments: High School Curriculum, College Major, and Careers," *Annual Review of Economics*, 4, 185–223.

Altonji, J. G. and C. R. Pierret (2001): "Employer Learning and Statistical Discrimination," *The Quarterly Journal of Economics*, 116, 313–350.

Andrews, R. J., S. A. Imberman, and M. F. Lovenheim (2017): "Risky Business? The Effect of Majoring in Business on Earnings and Educational Attainment," NBER Working Paper No. 23575.

——— (2019): "Recruiting and supporting low-income, high-achieving students at flagship universities," *Economics of Education Review*, 101923.

Andrews, R. J., J. Li, and M. F. Lovenheim (2014): "Heterogeneous Paths through College: Detailed Patterns and Relationships with Graduation and Earnings," *Economics of Education Review*, 42, 93–108.

——— (2016): "Quantile Treatment Effects of College Quality on Earnings," *Journal of Human Resources*, 51, 200–238.

Anelli, M. (2019): "Returns to Elite University Education: A Quasi-experimental Analysis," *Journal of European Economic Association*, forthcoming.

Angrist, J. and B. Frandsen (2020): "Machine Labor," NBER Working Paper No. 26584.

Angrist, J., P. Hull, P. A. Pathak, and C. R. Walters (2020): "Simple and Credible Value-Added Estimation Using Centralized School Assignment," NBER Working Paper No. 28241.

Angrist, J. D. (1998): "Estimating the Labor Market Impact of Voluntary Military Service Using Social Security Data on Military Applicants," *Econometrica*, 66, 249–288.

Angrist, J. D. and S. H. Chen (2011): "Schooling and the Vietnam-Era GI Bill: Evidence from the Draft Lottery," *American Economic Journal: Applied Economics*, 3, 96–118.

Angrist, J. D., P. D. Hull, P. A. Pathak, and C. R. Walters (2017): "Leveraging Lotteries for School Value-Added: Testing and Estimation," *The Quarterly Journal of Economics*, 132, 871–919.

Angrist, J. D., P. A. Pathak, and R. A. Zarate (2019): "Choice and Consequence: Assessing Mismatch at Chicago Exam Schools," NBER Working Paper No. 26137.

Arcidiacono, P. (2004): “Ability sorting and the returns to college major,” *Journal of Econometrics*, 121, 343–375.

——— (2005): “Affirmative Action in Higher Education: How Do Admission and Financial Aid Rules Affect Future Earnings?” *Econometrica*, 73, 1477–1524.

Arcidiacono, P., E. Aucejo, A. Maurel, and T. Ransom (2016a): “College Attrition and the Dynamics of Information Revelation,” NBER Working Paper No. 22325.

Arcidiacono, P., E. M. Aucejo, and V. J. Hotz (2016b): “University Differences in the Graduation of Minorities in STEM Fields: Evidence from California,” *American Economic Review*, 106, 525–562.

Arcidiacono, P. and M. Lovenheim (2016): “Affirmative Action and the Quality-Fit Trade-off,” *Journal of Economic Literature*.

Armona, L., R. Chakrabarti, and M. F. Lovenheim (2018): “How Does For-Profit College Attendance Affect Student Loans, Defaults and Labor Market Outcomes?” NBER Working Paper No. 25042.

Aryal, G., M. Bhuller, and F. Lange (2021): “Signaling and Employer Learning with Instruments,” NBER Working Paper No. 25885.

Aucejo, E., C. Hupkau, and J. Ruiz-Valenzuela (2020): “Where versus What: College Value-Added and Returns to Field of Study in Further Education,” Working paper.

Betts, J. R. (1996): “What Do Students Know about Wages? Evidence from a Survey of Undergraduates,” *The Journal of Human Resources*, 31, 27–56.

Beuermann, D. W. and C. K. Jackson (2020): “The Short and Long-Run Effects of Attending The Schools that Parents Prefer,” *Journal of Human Resources*.

Bhuller, M., M. Mogstad, and K. G. Salvanes (2017): “Life-Cycle Earnings, Education Premiums, and Internal Rates of Return,” *Journal of Labor Economics*, 35, 993–1030.

Black, D. A. and J. A. Smith (2006): “Estimating the Returns to College Quality with Multiple Proxies for Quality,” *Journal of Labor Economics*, 24, 701–728.

Black, S. E., J. T. Denning, and J. Rothstein (2020): “Winners and Losers? The Effect of Gaining and Losing Access to Selective Colleges on Education and Labor Market Outcomes,” NBER Working Paper No. 26821.

Bleemer, Z. (2020a): “Affirmative Action, Mismatch, and Economic Mobility after California’s Proposition 209,” Working paper.

——— (2020b): “Top Percent Policies and the Return to Postsecondary Selectivity,” Working paper.

——— (2020c): “Will Studying Economics Make You Rich? A Regression Discontinuity Analysis of the Returns to College Major,” Working paper.

Bodoh-Creed, A. L. and B. R. Hickman (2019): “Pre-College Human Capital Investment and Affirmative Action: A Structural Policy Analysis of US College Admissions,” Working paper, Washington University in St. Louis.

Bonhomme, S., T. Lamadon, and E. Manresa (2019): “A Distributional Framework for Matched Employer Employee Data,” *Econometrica*, 87, 699–739.

Bordón, P. and B. Braga (2020): “Employer learning, statistical discrimination and university prestige,” *Economics of Education Review*, 77, 101995.

Bound, J., M. F. Lovenheim, and S. Turner (2010): “Why Have College Completion Rates Declined? An Analysis of Changing Student Preparation and Collegiate Resources,” *American Economic Journal: Applied Economics*, 2, 129–157.

Bound, J. and S. Turner (2007): "Cohort crowding: How resources affect collegiate attainment," *Journal of Public Economics*, 91, 877–899.

Bowen, W. G. and D. C. Bok (2000): *The Shape of the River: Long-term Consequences of Considering Race in College and University Admissions*, Book collections on Project MUSE, Princeton University Press.

Brewer, D. J. and R. G. Ehrenberg (1996): "Does It Pay to Attend an Elite Private College? Evidence from the Senior High School Class of 1980," *Research in Labor Economics*, 15, 239–271.

Brewer, D. J., E. R. Eide, and R. G. Ehrenberg (1999): "Does It Pay to Attend an Elite Private College? Cross-Cohort Evidence on the Effects of College Type on Earnings," *The Journal of Human Resources*.

Broecke, S. (2012): "University selectivity and earnings: Evidence from UK data on applications and admissions to university," *Economics of Education Review*, 31, 96–107.

Campos, C. and C. Kearns (2021): "The Impact of Neighborhood School Choice: Evidence from Los Angeles' Zones of Choice," Working paper.

Canaan, S. and P. Mouganie (2018): "Returns to Education Quality for Low-Skilled Students: Evidence from a Discontinuity," *Journal of Labor Economics*, 36, 395–436.

Card, D. (2001): "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems," *Econometrica*.

Carneiro, P., J. J. Heckman, and E. J. Vytlacil (2011): "Estimating Marginal Returns to Education," *American Economic Review*, 101, 2754–2781.

Cellini, S. R. and N. Turner (2019): "Gainfully Employed?: Assessing the Employment and Earnings of For-Profit College Students Using Administrative Data," *Journal of Human Resources*, 54, 342–370.

Chetty, R., J. N. Friedman, and J. E. Rockoff (2014a): "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates," *American Economic Review*, 104, 2593–2632.

——— (2014b): "Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood," *American Economic Review*, 104, 2633–2679.

Chetty, R., J. N. Friedman, E. Saez, N. Turner, and D. Yagan (2020): "Income Segregation and Intergenerational Mobility Across Colleges in the United States," *The Quarterly Journal of Economics*, 135, 1567–1633.

Cohodes, S. R. and J. S. Goodman (2014): "Merit Aid, College Quality, and College Completion: Massachusetts' Adams Scholarship as an In-Kind Subsidy," *American Economic Journal: Applied Economics*, 6, 251–285.

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

Dale, S. B. and A. 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, 1491–1527.

——— (2014): "Estimating the Effects of College Characteristics over the Career Using Administrative Earnings Data," *Journal of Human Resources*, 49, 323–358.

de Roux, N. and E. Riehl (2020): "Isolating Peer Effects in the Returns to College Selectivity," Working paper.

Delavande, A. and B. Zafar (2019): "University Choice: The Role of Expected Earnings, Nonpecuniary Outcomes, and Financial Constraints," *Journal of Political Economy*, 127, 2343–2393.

Deming, D. J. and C. R. Walters (2018): "The Impact of State Budget Cuts on U.S. Postsecondary Attainment," Working paper.

Dillon, E. W. and J. Smith (2018): "The Consequences of Academic Match between Students and Colleges," NBER Working Paper No. 25069.

Dominitz, J. and C. Manski (1996): "Eliciting Student Expectations of the Returns to Schooling," *Journal of Human Resources*, 31, 1–26.

Dynarski, S., B. Jacob, and D. Kreisman (2018): "How important are fixed effects and time trends in estimating returns to schooling? Evidence from a replication of Jacobson, Lalonde, and Sullivan, 2005," *Journal of Applied Econometrics*, 33, 1098–1108.

Dynarski, S., C. Libassi, K. Michelmore, and S. Owen (2020): "Closing the Gap: The Effect of a Targeted, Tuition-Free Promise on College Choices of High-Achieving, Low-Income Students," NBER Working Paper No. 25349.

Eller, C. C. (2019): "Superficially Coupled Systems: The Organizational Production of Inequality in Higher Education," Annenberg EdWorkingPaper No. 19–70.

Farber, H. S. and R. Gibbons (1996): "Learning and Wage Dynamics," *The Quarterly Journal of Economics*, 111, 1007–1047.

Fillmore, I. (2021): "Price Discrimination and Public Policy in the U.S. College Market," Working paper.

Finkelstein, A., M. Gentzkow, and H. Williams (2016): "Sources of Geographic Variation in Health Care: Evidence From Patient Migration," *The Quarterly Journal of Economics*, 131, 1681–1726.

Fletcher, J. M. (2015): "Social interactions and college enrollment: A combined school fixed effects/instrumental variables approach," *Social Science Research*, 52, 494–507.

Fryer, R. G. and M. Greenstone (2010): "The Changing Consequences of Attending Historically Black Colleges and Universities," *American Economic Journal: Applied Economics*, 2, 116–148.

Fu, C., J. Guo, A. J. Smith, and A. T. Sorensen (2021): "Students' Heterogeneous Preferences and the Uneven Spatial Distribution of Colleges," NBER Working Paper No. 28343.

Ge, S., E. Isaac, and A. Miller (2018): "Elite Schools and Opting-In: Effects of College Selectivity on Career and Family Outcomes," NBER Working Paper No. 25315.

Gibbons, C. E., J. C. Suárez Serrato, and M. B. Urbancic (2019): "Broken or Fixed Effects?" *Journal of Econometric Methods*, 8.

Goldsmith-Pinkham, P., P. Hull, and M. Kolesar (2021): "On Estimating Multiple Treatment Effects with Regression," Working paper.

Goodman, J., M. Hurwitz, and J. Smith (2017): "Access to Four-Year Public Colleges and Degree Completion," *Journal of Labor Economics*, 35, 829–867.

Gowrisankaran, G. and R. J. Town (1999): "Estimating the quality of care in hospitals using instrumental variables," *Journal of Health Economics*, 18, 747–767.

Hastings, J. S., C. A. Neilson, A. Ramirez, and S. D. Zimmerman (2016): "(Un)informed college and major choice: Evidence from linked survey and administrative data," *Economics of Education Review*, 51, 136–151.

Hastings, J. S., C. A. Neilson, and S. D. Zimmerman (2014): "Are Some Degrees Worth More than Others? Evidence from College Admission Cutoffs in Chile," NBER Working Paper No. 19241.

Heckman, J., H. Ichimura, J. Smith, and P. Todd (1998): "Characterizing Selection Bias Using Experimental Data," *Econometrica*, 66, 1017–1098.

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

Heckman, J. J., S. Urzua, and E. Vytlacil (2006): "Understanding Instrumental Variables in Models with Essential Heterogeneity," *The Review of Economics and Statistics*, 88, 389–432.

Hoekstra, M. (2009): "The Effect of Attending the Flagship State University on Earnings: A Discontinuity-Based Approach," *Review of Economics and Statistics*, 91, 717–724.

Hoxby, C. and S. Turner (2013): "Expanding College Opportunities for High-Achieving, Low Income Students," Tech. Rep. 12-014.

Hoxby, C. M. (2009): "The Changing Selectivity of American Colleges," *The Journal of Economic Perspectives*, 23, 95–118.

——— (2019): "The Productivity of US Postsecondary Institutions," in *Productivity in Higher Education*, ed. by C. M. Hoxby and K. Stange, University of Chicago Press, 31–66.

Hull, P. (2020): "Estimating Hospital Quality with Quasi-experimental Data," Working paper.

Jacob, B. and L. Lefgren (2008): "Can Principals Identify Effective Teachers? Evidence on Subjective Performance Evaluation in Education," *Journal of Labor Economics*, 26, 101–136.

Jacob, B., B. McCall, and K. Stange (2017): "College as Country Club: Do Colleges Cater to Students' Preferences for Consumption?" *Journal of Labor Economics*, 36, 309–348.

Jensen, R. (2010): "The (Perceived) Returns to Education and the Demand for Schooling," *The Quarterly Journal of Economics*, 125, 515–548.

Jia, R. and H. Li (2021): "Just above the Exam Cutoff Score: Elite College Admission and Wages in China," NBER Working Paper No. 28450.

Kapor, A. (2020): "Distributional Effects of Race-Blind Affirmative Action," Working paper.

Kirkeboen, L. J., E. Leuven, and M. Mogstad (2016): "Field of Study, Earnings, and Self-Selection," *The Quarterly Journal of Economics*, 131, 1057–1111.

Krueger, A. B. and L. H. Summers (1988): "Efficiency Wages and the Inter-Industry Wage Structure," *Econometrica*, 56, 259.

Lamadon, T., M. Mogstad, and B. Setzler (2019): "Imperfect Competition, Compensating Differentials, and Rent Sharing in the U.S. Labor Market," Working paper.

Lange, F. (2007): "The Speed of Employer Learning," *Journal of Labor Economics*, 25, 1–35.

Lemieux, T. and D. Card (2001): "Education, Earnings, and the Canadian G.I. Bill," *Canadian Journal of Economics/Revue canadienne d'économique*, 34, 313–344.

Long, M. C. (2008): "College Quality and Early Adult Outcomes," *Economics of Education Review*, 27, 588–602.

Malamud, O. (2011): "Discovering One's Talent: Learning from Academic Specialization," *ILR Review*, 64, 375–405.

Manski, C. F. (1989): "Schooling as experimentation: a reappraisal of the postsecondary dropout phenomenon," *Economics of Education Review*, 8, 305–312.

Miller, D. L., N. Shenav, and M. Z. Grosz (2019): "Selection into Identification in Fixed Effects Models, with Application to Head Start," NBER Working Paper No. 26174.

Morris, C. N. (1983): "Parametric Empirical Bayes Inference: Theory and Applications," *Journal of the American Statistical Association*, 78, 47–55.

Mountjoy, J. (2021): "Community Colleges and Upward Mobility," Working paper.

National Center for Education Statistics (2018): "Digest of Education Statistics," Tech. rep.

National Conference of State Legislatures (2018): "Outcomes-Based Funding as an Evolving State Appropriation Model," Tech. rep.

Oreopoulos, P. and U. Petronijevic (2013): "Making College Worth It: A Review of the Returns to Higher Education," *The Future of Children*, 23, 41–65.

Pope, D. G. and J. C. Pope (2012): "Understanding College Application Decisions: Why College Sports Success Matters," *Journal of Sports Economics*, 15, 107–131.

Reynolds, C. L. (2012): "Where to Attend? Estimating the Effects of Beginning College at a Two-Year Institution," *Economics of Education Review*, 31, 345–362.

Riehl, E., J. E. Saavedra, and M. Urquiola (2019): "Learning and Earning: An Approximation to College Value Added in Two Dimensions," in *Productivity in Higher Education*, ed. by C. M. Hoxby and K. Stange, University of Chicago Press, 105–132.

Robbins, H. (1956): "An Empirical Bayes Approach to Statistics," in *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 1*, ed. by J. Neyman, Berkeley: University of California Press, 157–163.

Rockoff, J. E. (2004): "The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data," *The American Economic Review*, 94, 247–252.

Rothstein, J. M. (2006): "Good Principals or Good Peers? Parental Valuation of School Characteristics, Tiebout Equilibrium, and the Incentive Effects of Competition among Jurisdictions," *American Economic Review*, 96, 1333–1350.

Saavedra, J. (2009): "The Learning and Early Labor Market Effects of College Quality: A Regression Discontinuity Analysis," Working paper.

Sander, R. H. and S. Taylor (2012): *Mismatch: How Affirmative Action Hurts Students It's Intended to Help, and Why Universities Won't Admit It*, Basic Books.

Sekhri, S. (2020): "Prestige Matters: Wage Premium and Value Addition in Elite Colleges," *American Economic Journal: Applied Economics*, 12, 207–225.

Simonsohn, U. (2010): "Weather To Go To College," *The Economic Journal*, 120, 270–280.

Smith, J. (2013): "Ova and out: Using twins to estimate the educational returns to attending a selective college," *Economics of Education Review*, 36, 166–180.

Smith, J., J. Goodman, and M. Hurwitz (2020): "The Economic Impact of Access to Public Four-Year Colleges," NBER Working Paper No. 27177.

Stange, K. M. (2012): "An Empirical Investigation of the Option Value of College Enrollment," *American Economic Journal: Applied Economics*, 4, 49–84.

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

Stinebrickner, R. and T. Stinebrickner (2014a): "Academic Performance and College Dropout: Using Longitudinal Expectations Data to Estimate a Learning Model," *Journal of Labor Economics*, 32, 601–644.

Stinebrickner, R. and T. R. Stinebrickner (2014b): "A Major in Science? Initial Beliefs and Final Outcomes for College Major and Dropout," *The Review of Economic Studies*, 81, 426–472.

Stinebrickner, T. and R. Stinebrickner (2012): "Learning about Academic Ability and the College Dropout Decision," *Journal of Labor Economics*, 30, 707–748.

Stolzenberg, E. B., M. C. Aragon, E. Romo, V. Couch, D. McLennan, M. K. Eagan, and N. Kang (2020): *The American Freshman: National Norms Fall 2019*, Los Angeles: Higher Education Research Institute, UCLA.

UTD-ERC (2021): "UT Dallas Education Research Center Data Holdings," https://tsp.utdallas.edu.

Ward, J. and B. Ost (2021): "The Effect of Large-scale Performance-Based Funding in Higher Education," *Education Finance and Policy*, 16, 92–124.

Webber, D. A. (2014): "The lifetime earnings premia of different majors: Correcting for selection based on cognitive, noncognitive, and unobserved factors," *Labour Economics*, 28, 14–23.

Weinberger, C. J. (2018): "Engineering Educational Opportunity: Impacts of 1970s and 1980s Policies to Increase the Share of Black College Graduates with a Major in Engineering or Computer Science," in *U.S. Engineering in a Global Economy*, University of Chicago Press, 87–128.

Willis, R. J. and S. Rosen (1979): "Education and Self-Selection," *Journal of Political Economy*, 87, S7–S36.

Wiswall, M. and B. Zafar (2015a): "Determinants of College Major Choice: Identification using an Information Experiment," *The Review of Economic Studies*, 82, 791–824.

——— (2015b): "How Do College Students Respond to Public Information about Earnings?" *Journal of Human Capital*, 9, 117–169.

Zimmerman, S. D. (2014): "The Returns to College Admission for Academically Marginal Students," *Journal of Labor Economics*, 32, 711–754.

A Appendix

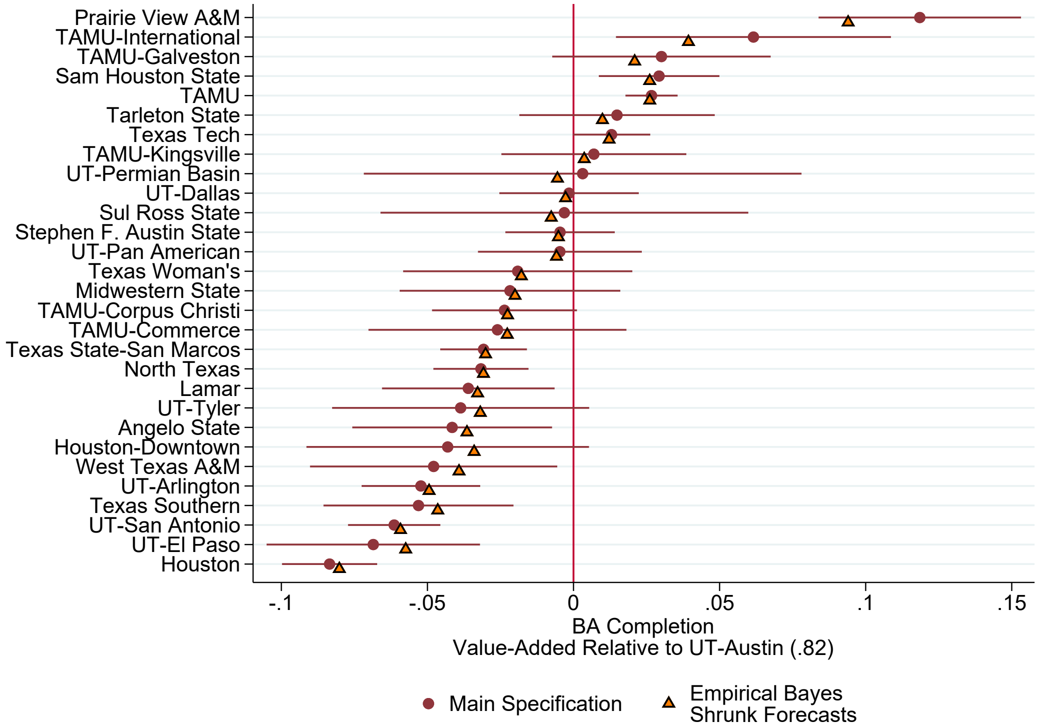
A.1 Empirical Bayes Shrunk Forecasts

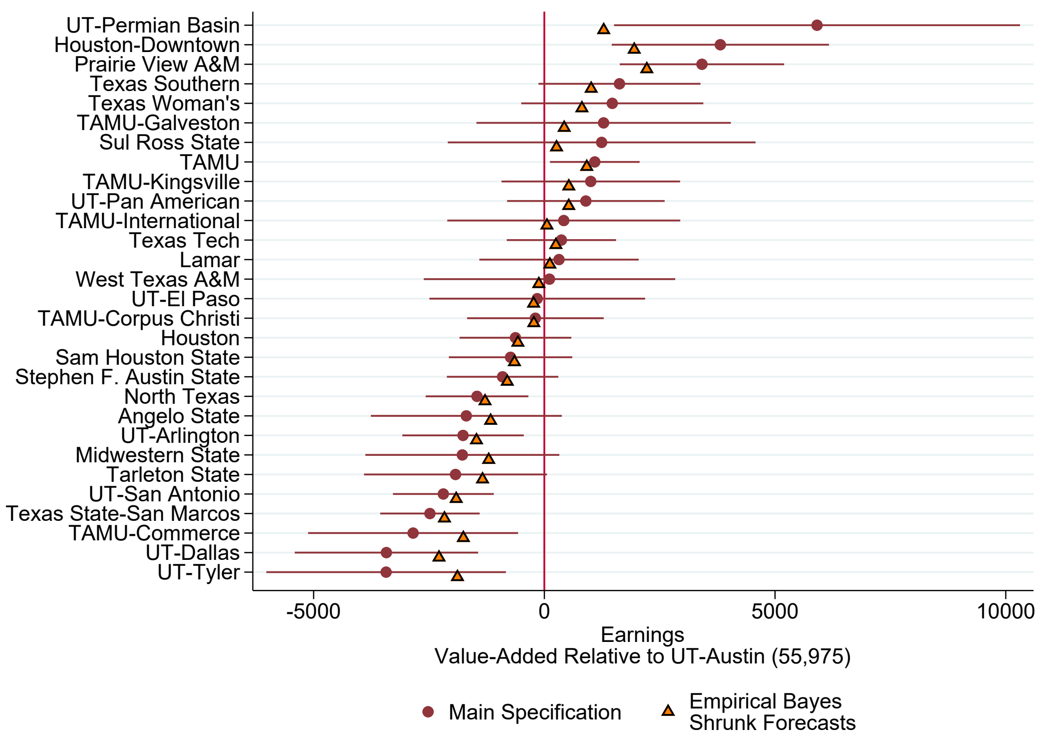
As a means of accounting for estimation error in value-added, we construct Empirical Bayes forecasts (e.g., Robbins, 1956) that shrink each college's value-added estimate towards the mean in proportion to its imprecision. These EB estimates are interesting in themselves as efficient predictions of true college value-added given our finite data, but they deliver additional utility as error-correcting regressors (e.g., Jacob and Lefgren, 2008) in the analysis in Section 7.

Following the standard approach in the value-added literature, we can derive the EB shrinkage factor either by assuming a normally distributed prior over true value-added v_j and estimation error e_j (e.g., Morris, 1983; Abdulkadiroglu et al., 2020), or by restricting the forecast function of v_j given $\hat{v}_j$ to linear projection (e.g., Chetty et al., 2014a). Both premises deliver a shrinkage factor of the form $\frac{\sigma_v^2}{\sigma_v^2+\sigma_{e_j}^2}$, where $\sigma_{e_j}^2$ is the college j-specific estimation error variance. We estimate σ_v^2 using the empirical counterparts of Equation (10), and $\sigma_{e_j}^2$ is estimated as the square of the standard error of the VA estimate $\hat{v}_j$ for college j.

Figure A.1 presents the Empirical Bayes forecasts and compares them to the unshrunk value-added estimates. Since the VA estimates for most of our colleges are measured relatively precisely and/or are already close to zero, the EB shrinkage procedure does not have much impact on them; as expected, the greatest shrinkage is exerted on the handful of small schools with noisier and more far-flung estimates at the top and bottom of the distribution.

Figure A.1: Empirical Bayes Shrunk Forecasts of College Value-Added

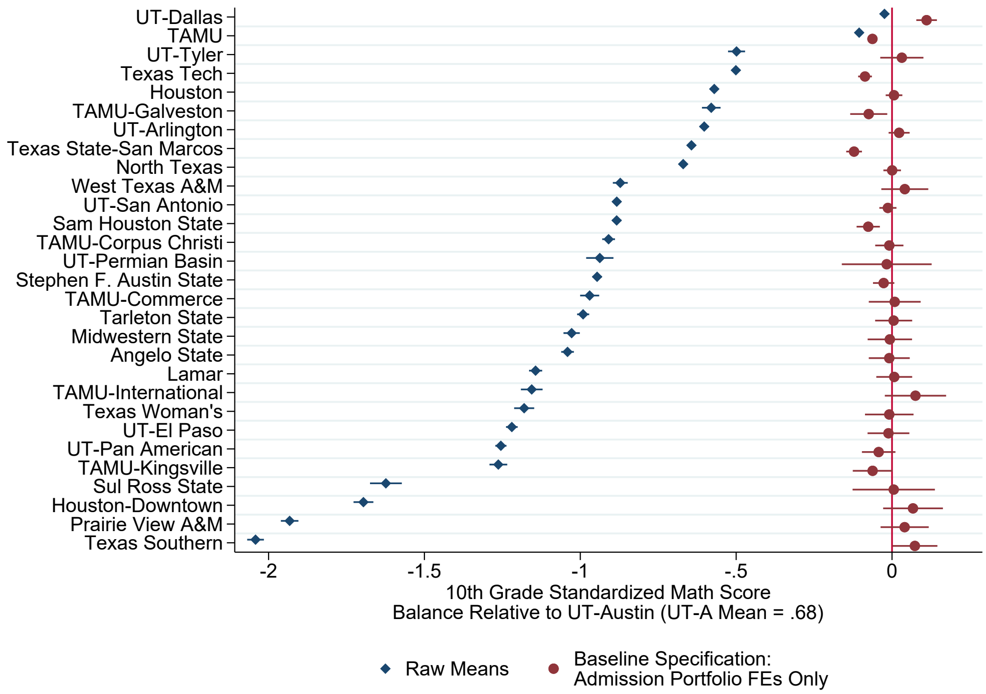


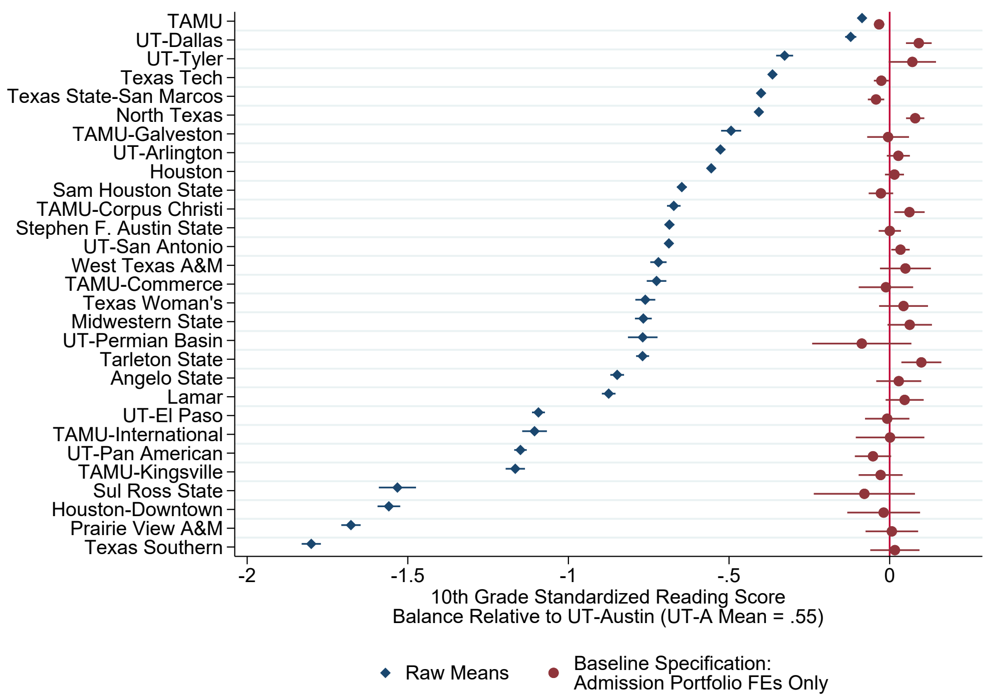


Notes: The Main Specification point estimates and robust 95% confidence intervals come from regressions of individual student outcomes on college treatment indicators (with UT-Austin as the reference treatment at zero), controlling for admission portfolio fixed effects and our core set of covariate controls: gender, FRPL, race, 10th grade test scores, and high school attendance. The Empirical Bayes Shrunk Forecasts shrink the Main Specification estimates towards their mean by the shrinkage factor $\frac{\sigma_v^2}{\sigma_v^2+\sigma_{e_j}^2}$, as described in Section A.1.

A.2 Additional Figures and Tables

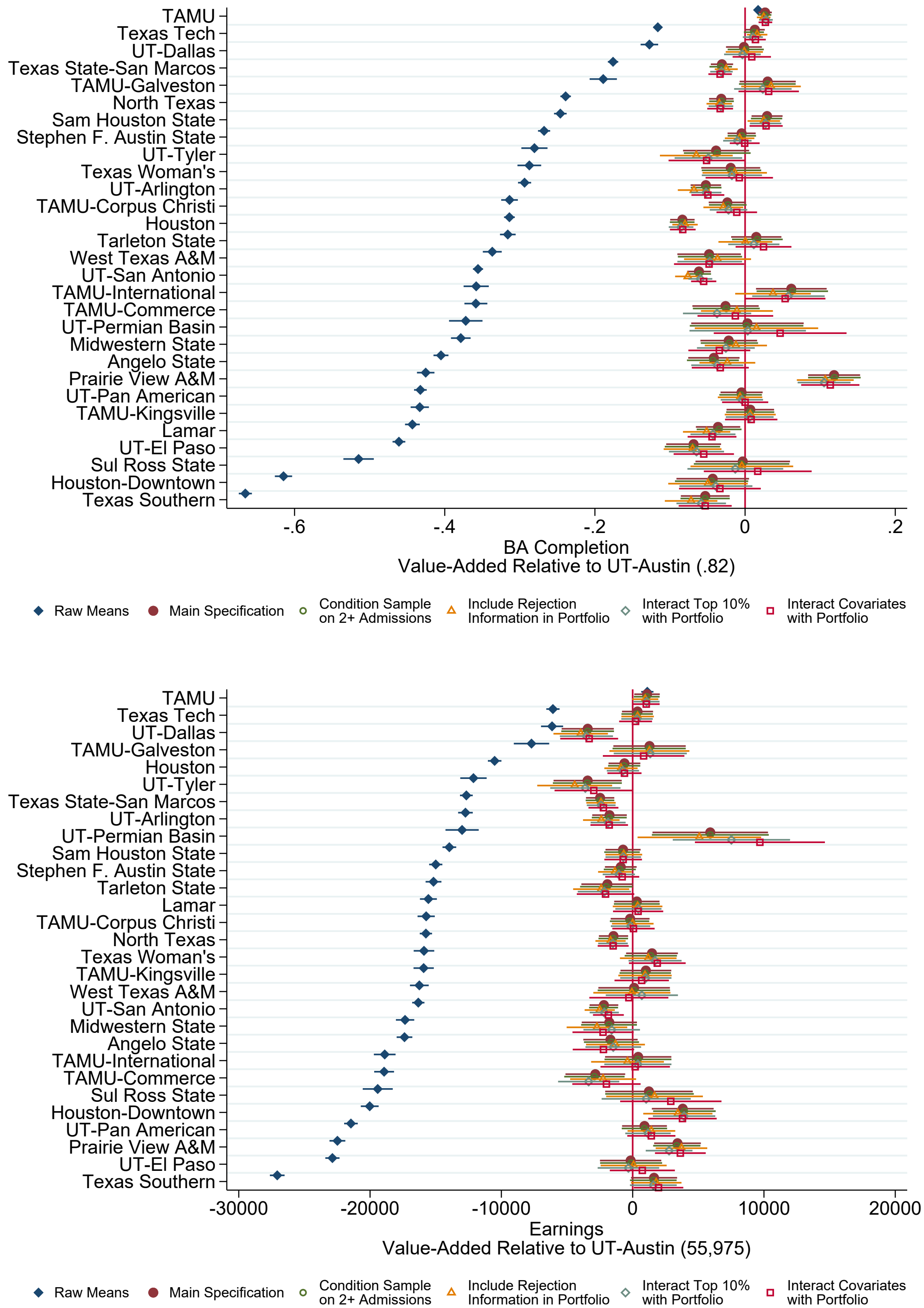
Figure A.2: Validating the Matched Applicant Approach: Ability Balance Across College Treatments





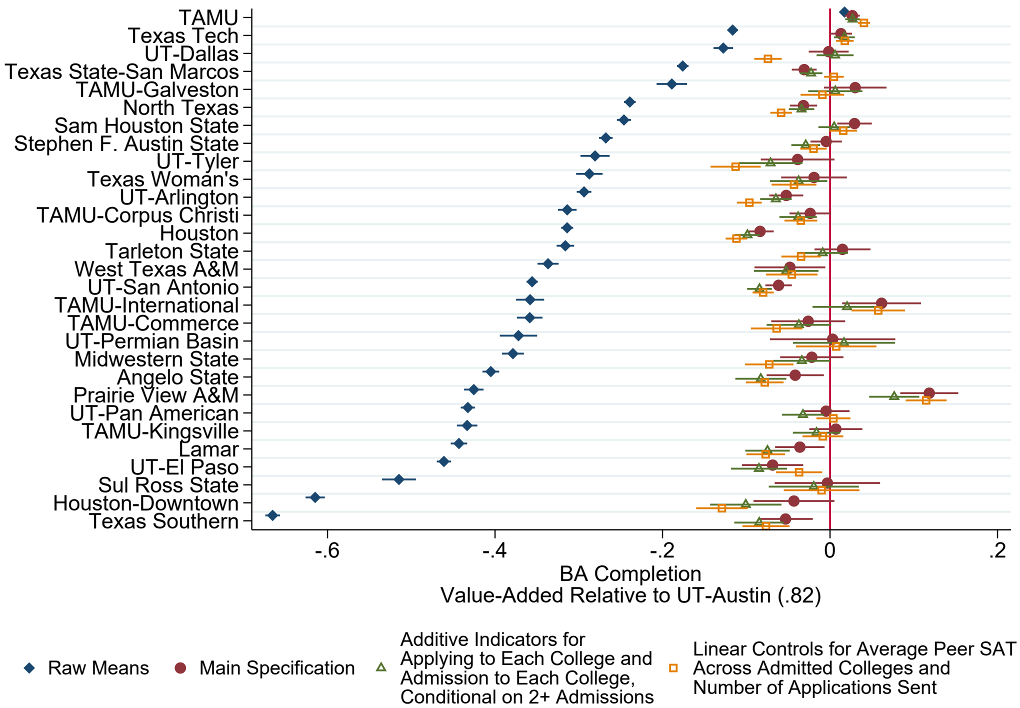
Notes: Each set of point estimates and robust 95% confidence intervals come from regressions of individual student standardized test scores on college treatment indicators, omitting UT-Austin as the reference treatment (signified by the vertical line at zero). The UT-Austin mean appears in parentheses below each plot. The *Raw Means* specification controls only for cohort fixed effects. The *Baseline Specification* controls solely for college admission portfolio fixed effects (and cohort fixed effects).

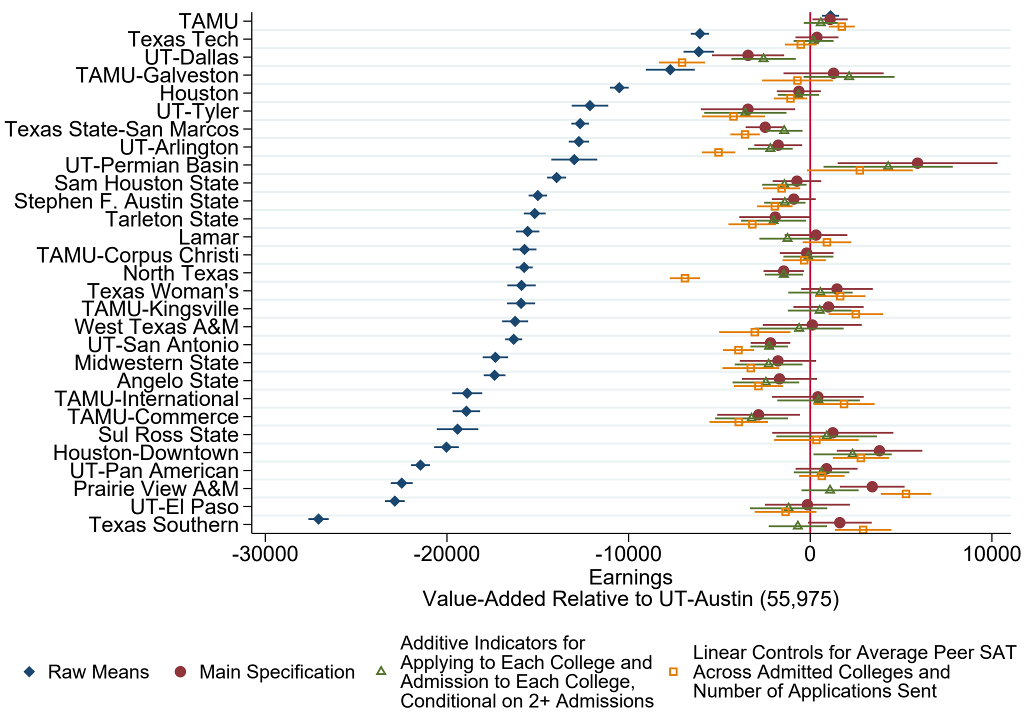
Figure A.3: Robustness to Richer Specifications of Admission Portfolios



Notes: Each set of point estimates and robust 95% confidence intervals come from regressions of individual student outcomes on college treatment indicators, omitting UT-Austin as the reference treatment (signified by the vertical line at zero). The UT-Austin outcome mean appears in parentheses below each plot. All specifications control for cohort fixed effects. The *Raw Means* specification controls for nothing else. The *Main Specification* controls for college admission portfolio fixed effects and our core set of covariate controls: gender, FRPL, race, 10th grade test scores, and high school attendance. The *Condition Sample* specification runs the Main Specification on the subsample of students who are admitted to at least two colleges. The *Include Rejection Information* specification replaces the admission portfolio fixed effects with fixed effects for every distinct portfolio of applications and admissions, thereby including additional information about applications that do not result in admissions, i.e. rejections. The *Interact Top 10%* specification interacts the admission portfolio indicators with the indicator for being in top decile of high school GPA. The *Interact Covariates* specification interacts the admission portfolio indicators with our core covariate controls: gender, FRPL, race, 10th grade test scores, and high school attendance.

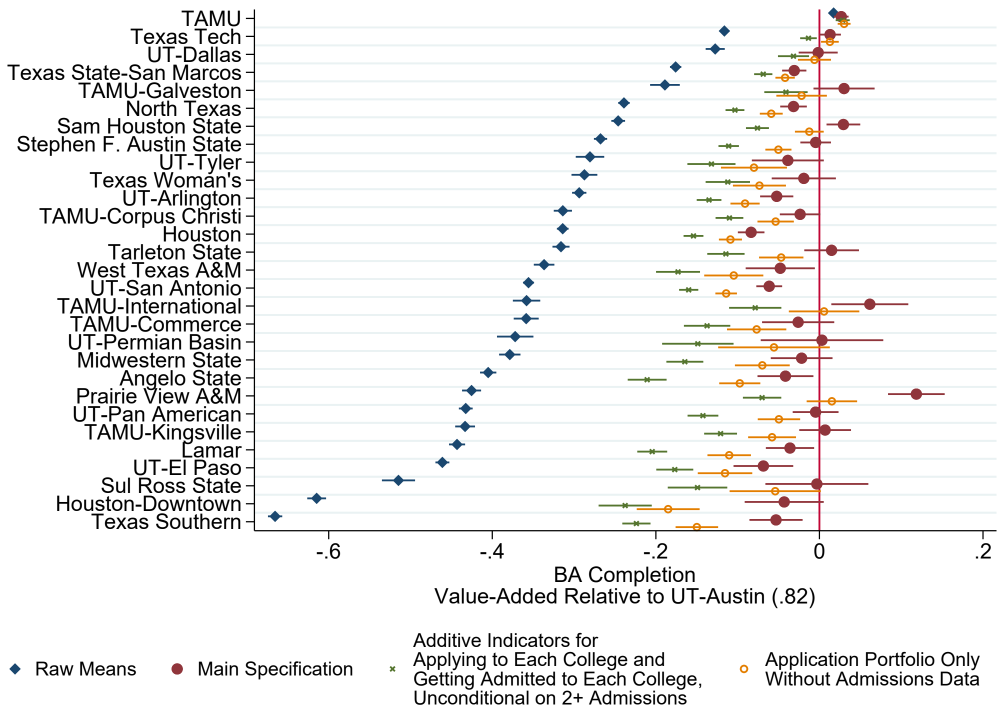
Figure A.4: Simpler Specifications of Admission Portfolios

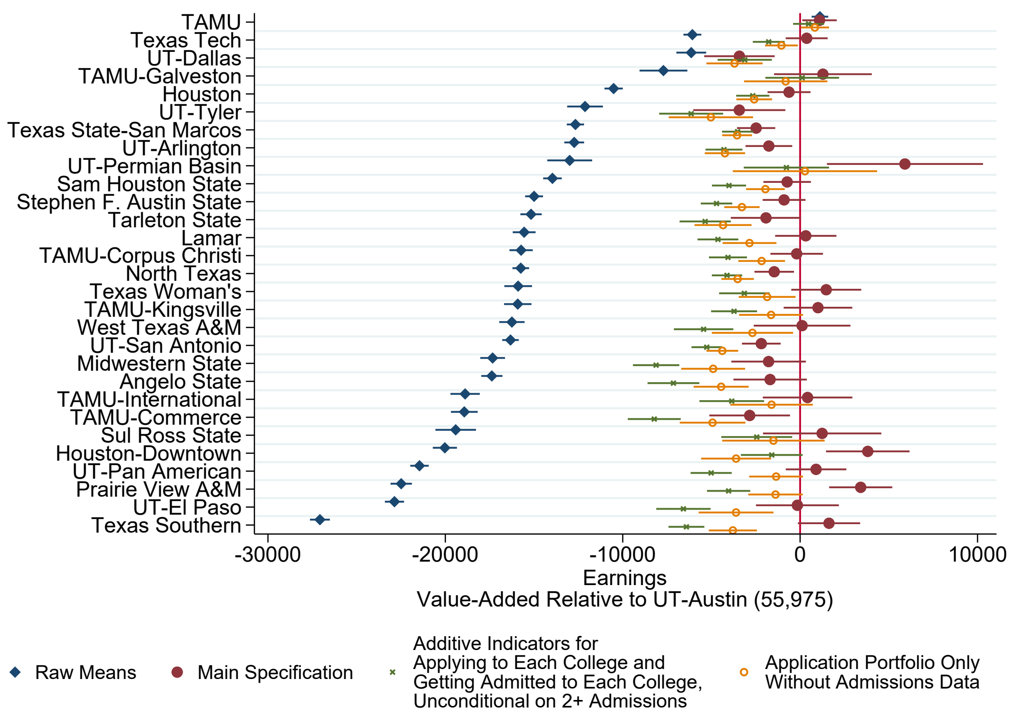




Notes: Each set of point estimates and robust 95% confidence intervals come from regressions of individual student outcomes on college treatment indicators, omitting UT-Austin as the reference treatment (signified by the vertical line at zero). The UT-Austin outcome mean appears in parentheses below each plot. All specifications control for cohort fixed effects. The *Raw Means* specification controls for nothing else. The *Main Specification* controls for college admission portfolio fixed effects and our core set of covariate controls: gender, FRPL, race, 10th grade test scores, and high school attendance. The *Additive Indicators* specification replaces the full set of college admission portfolio fixed effects from the Main Specification with 30 dummies indicating an application to each college, and 30 dummies indicating an admission to each college, and is estimated on the sample of students with at least two admissions. The *Linear Controls* specification replaces the college admission portfolio fixed effects from the Main Specification with just two linear control variables—the average peer SAT score across all the colleges a student is admitted to, and the number of applications she sent—and is estimated on the sample of students with at least two admissions.

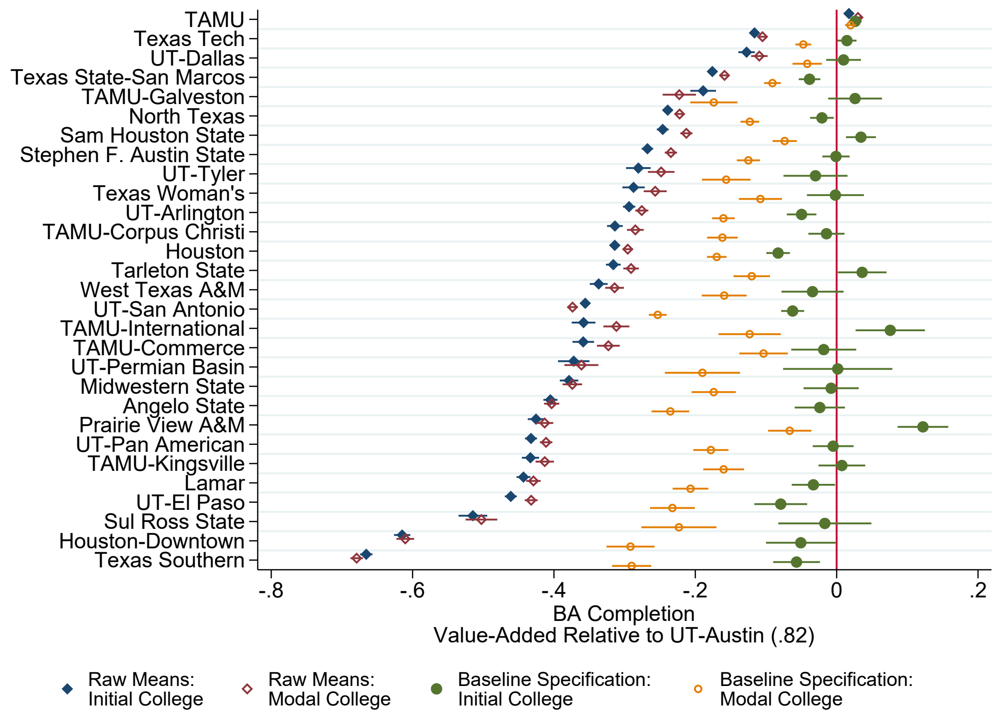
Figure A.5: Insufficient Specifications

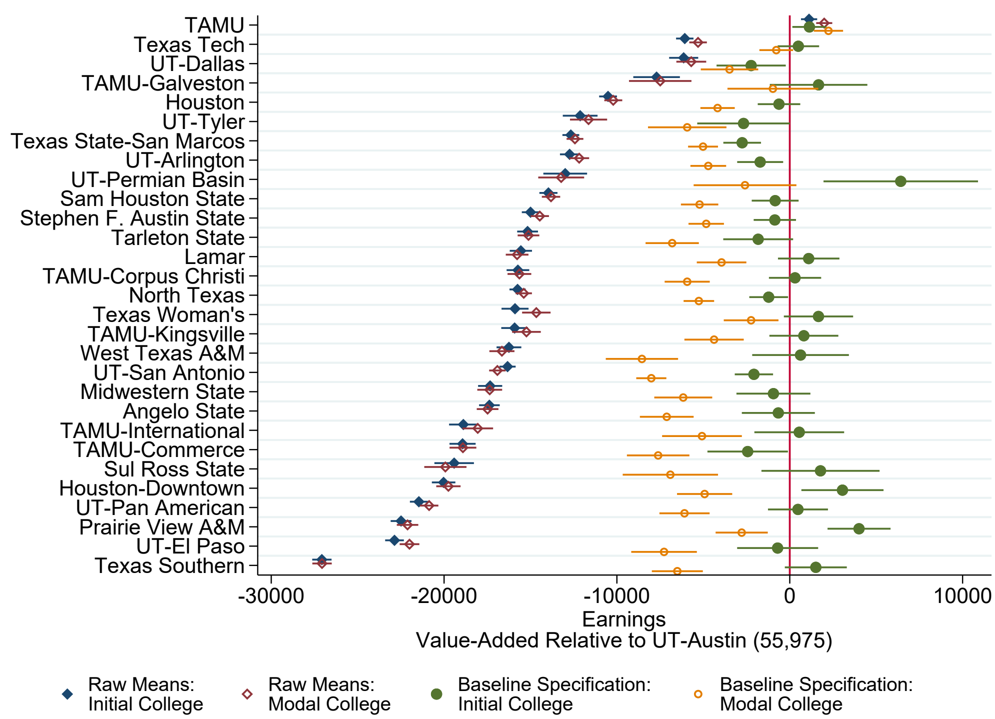




Notes: Each set of point estimates and robust 95% confidence intervals come from regressions of individual student outcomes on college treatment indicators, omitting UT-Austin as the reference treatment (signified by the vertical line at zero). The UT-Austin outcome mean appears in parentheses below each plot. All specifications control for cohort fixed effects. The *Raw Means* specification controls for nothing else. The *Main Specification* controls for college admission portfolio fixed effects and our core set of covariate controls: gender, FRPL, race, 10th grade test scores, and high school attendance. The *Additive Indicators* specification replaces the full set of college admission portfolio fixed effects from the Main Specification with 30 dummies indicating an application to each college, and 30 dummies indicating an admission to each college, and is estimated on the full sample, including students who were only admitted to one college. The *Application Portfolio Only* specification replaces the college admission portfolio fixed effects from the Main Specification with fixed effects solely for each combination of applications, ignoring information on admissions.

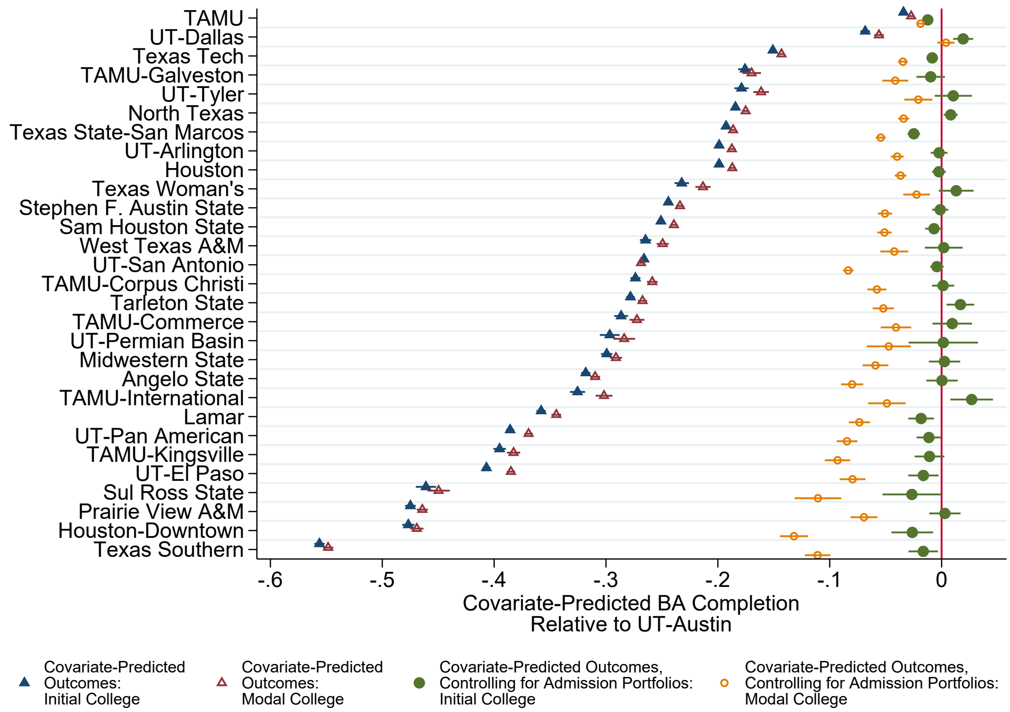
Figure A.6: Redefining the Treatment as the Modal College Attended: Value-Added Estimates

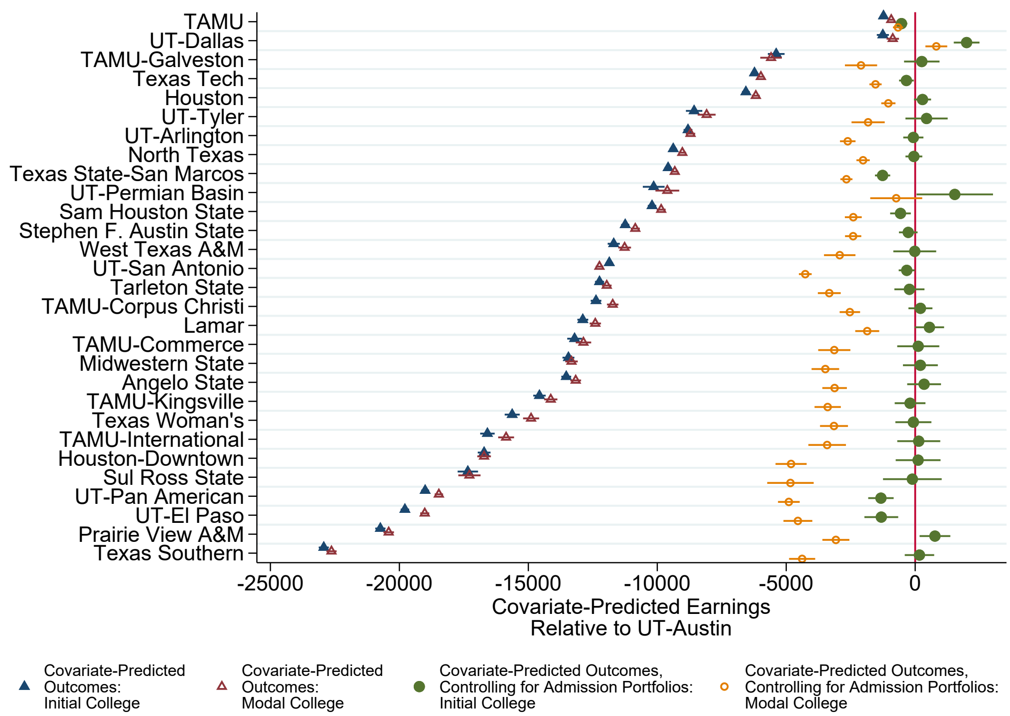




Notes: Each set of point estimates and robust 95% confidence intervals come from regressions of individual student outcomes on college treatment indicators, omitting UT-Austin as the reference treatment (signified by the vertical line at zero). The UT-Austin outcome mean appears in parentheses below each plot. All specifications control for cohort fixed effects. The *Raw Means* specification controls for nothing else. The *Baseline Specification* controls solely for college admission portfolio fixed effects (and cohort fixed effects). The *Initial College* treatment is the first undergraduate institution a student attends, starting in the fall after high school graduation. The *Modal College* treatment is the undergraduate institution that a student attends for the largest number of semesters, starting in the fall after high school graduation.

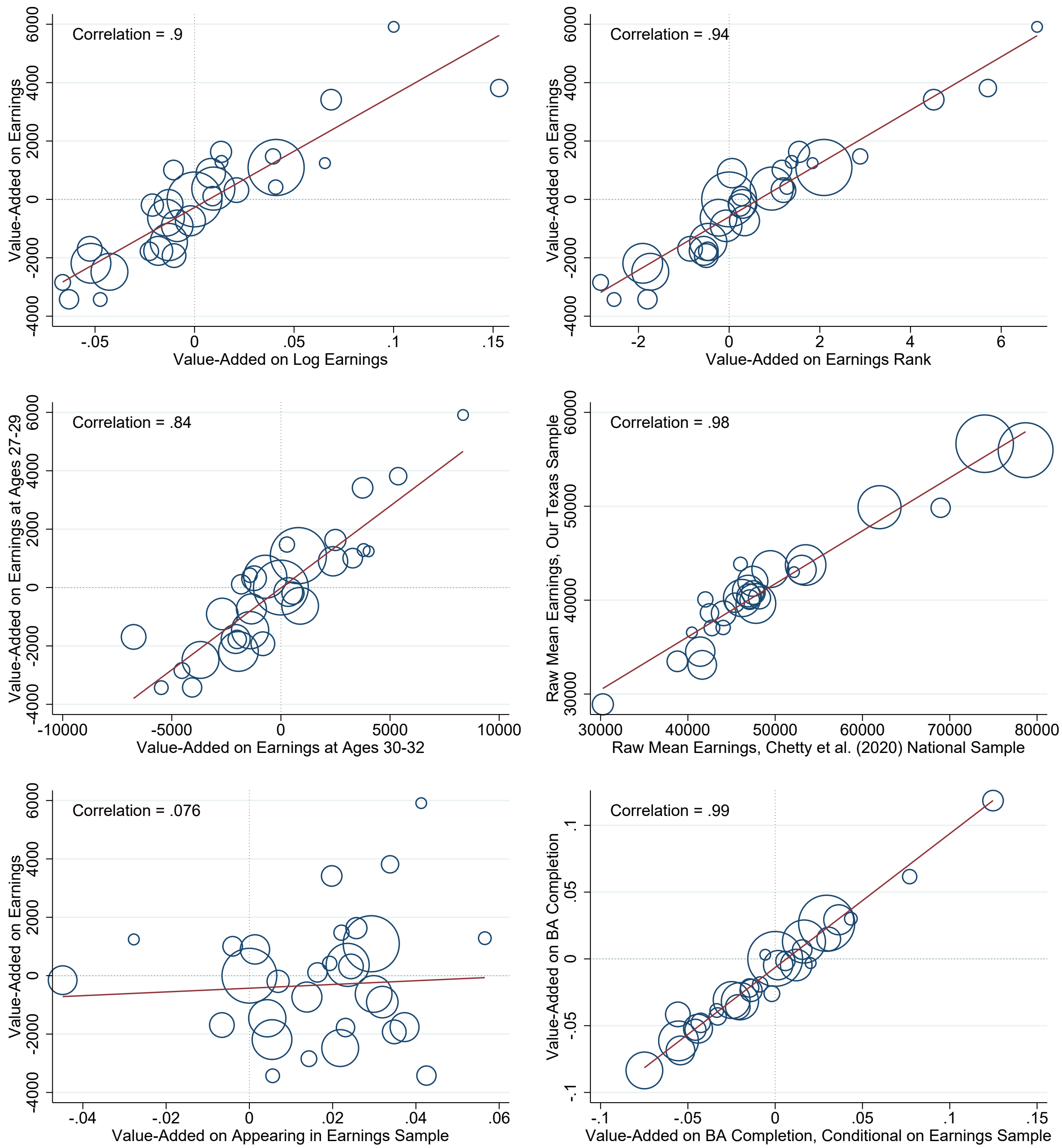
Figure A.7: Redefining the Treatment as the Modal College Attended: Lack of Balance





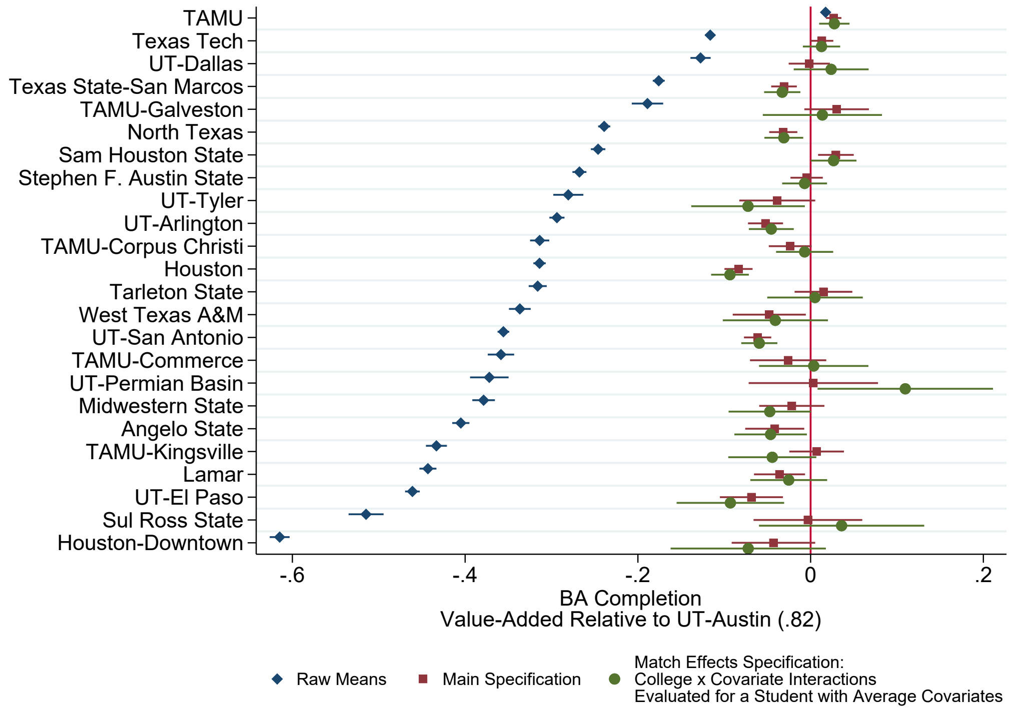
Notes: Each set of point estimates and robust 95% confidence intervals come from regressions of covariate-predicted student outcomes on college treatment indicators, omitting UT-Austin as the reference treatment (signified by the vertical line at zero). *Covariate-Predicted Outcomes* replace actual outcomes with the predicted values from a separate OLS regression of each outcome on the following set of pre-college covariates: quartics in 10th grade math and reading scores, attendance, days of disciplinary suspension, and number of advanced high school courses; binary indicators for free/reduced price lunch eligibility, gender, at-risk status, and top decile GPA; and fixed effects for each race, high school cohort, and high school campus. The *Initial College* treatment is the first undergraduate institution a student attends, starting in the fall after high school graduation. The *Modal College* treatment is the undergraduate institution that a student attends for the largest number of semesters, starting in the fall after high school graduation. The two rightmost estimates come from our baseline specification, which controls for fixed effects for each distinct portfolio of college admissions (and cohort fixed effects).

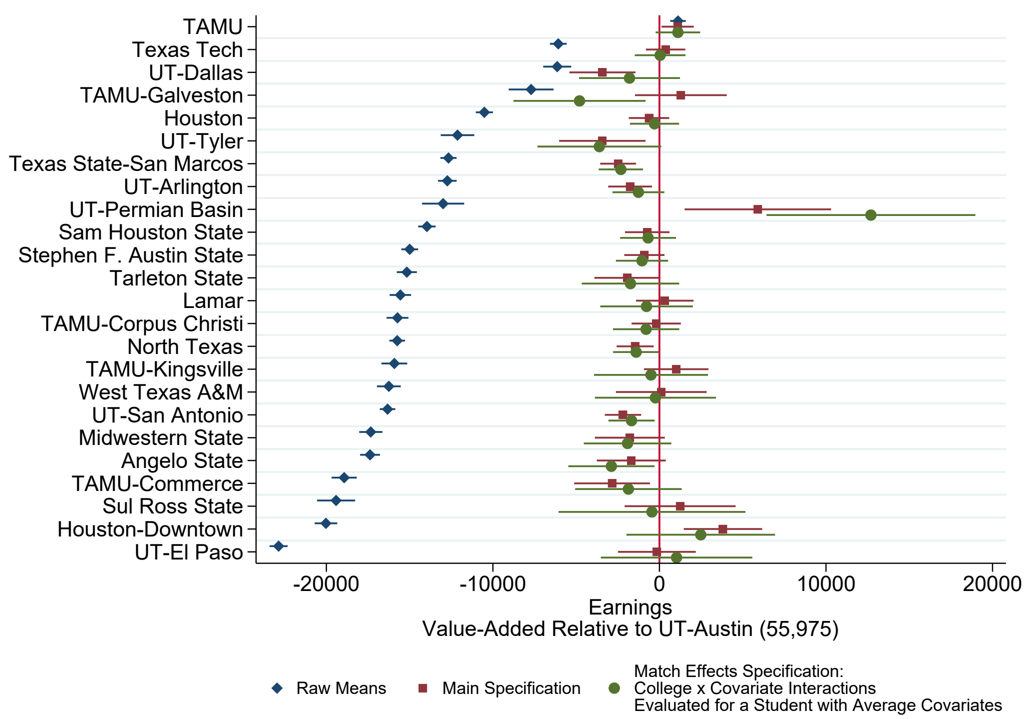
Figure A.8: Robustness to Alternative Earnings Definitions and Missing Earnings



Notes: Each circle plots college value-added estimates from our main specification on different earnings outcomes, which regresses individual student outcomes on college treatment indicators (with UT-Austin as the reference treatment at zero), controlling for admission portfolio fixed effects and our core set of covariate controls: gender, FRPL, race, 10th grade test scores, and high school attendance. The one exception is the fourth panel, which simply plots raw mean earnings at each college in our Texas sample against raw mean earnings from Chetty et al. (2020)'s national sample. All correlations and circles are weighted by student enrollment.

Figure A.9: Allowing for Match Effects Delivers Similar Estimates for the Average Student





Notes: Each set of point estimates and robust 95% confidence intervals come from regressions of individual student outcomes on college treatment indicators, omitting UT-Austin as the reference treatment (signified by the vertical line at zero). The UT-Austin outcome mean appears in parentheses below each plot. All specifications control for cohort fixed effects. The *Raw Means* specification controls for nothing else. The *Main Specification* adds college admission portfolio fixed effects and our core set of covariate controls: gender, FRPL, race, 10th grade test scores, and high school attendance. The *Match Effects Specification* adds interactions between the college treatment indicators and the core covariate controls, and the VA estimates are evaluated for a student with average values of the interacted covariates. See Tables B.4 and B.5 for the corresponding numerical estimates.

B Online Appendix

Table B.1: Value-Added Estimates: BA Completion

	Raw Means	Typical Controls	Baseline	Main Spec	Extra Controls 1	Extra Controls 2
Angelo State	-0.405	-0.232	-0.0238	-0.0415	-0.0364	-0.0272
	(0.00508)	(0.00494)	(0.0181)	(0.0174)	(0.0174)	(0.0170)
TAMU-Commerce	-0.359	-0.186	-0.0183	-0.0260	-0.0216	-0.0266
	(0.00779)	(0.00745)	(0.0235)	(0.0225)	(0.0225)	(0.0222)
Lamar	-0.443	-0.250	-0.0329	-0.0360	-0.0334	-0.0299
	(0.00502)	(0.00494)	(0.0156)	(0.0151)	(0.0150)	(0.0148)
Midwestern State	-0.379	-0.215	-0.00782	-0.0217	-0.0204	-0.0146
	(0.00668)	(0.00640)	(0.0199)	(0.0193)	(0.0192)	(0.0191)
North Texas	-0.239	-0.120	-0.0208	-0.0317	-0.0288	-0.0260
	(0.00359)	(0.00358)	(0.00855)	(0.00831)	(0.00830)	(0.00827)
UT-Pan American	-0.432	-0.144	-0.00494	-0.00467	-0.00133	0.00149
	(0.00431)	(0.00461)	(0.0148)	(0.0143)	(0.0142)	(0.0141)
Sam Houston State	-0.246	-0.0878	0.0344	0.0293	0.0325	0.0403
	(0.00428)	(0.00429)	(0.0108)	(0.0105)	(0.0105)	(0.0104)
Texas State-San Marcos	-0.176	-0.0464	-0.0384	-0.0308	-0.0254	-0.0163
	(0.00353)	(0.00358)	(0.00778)	(0.00757)	(0.00757)	(0.00749)
Stephen F. Austin State	-0.268	-0.116	-0.000980	-0.00461	-0.00124	0.00162
	(0.00413)	(0.00412)	(0.00986)	(0.00955)	(0.00953)	(0.00946)
Sul Ross State	-0.515	-0.204	-0.0167	-0.00313	0.00178	0.000748
	(0.0103)	(0.0100)	(0.0336)	(0.0321)	(0.0320)	(0.0315)
Prairie View A&M	-0.425	-0.0986	0.122	0.119	0.120	0.120
	(0.00594)	(0.00630)	(0.0183)	(0.0177)	(0.0176)	(0.0175)
Tarleton State	-0.316	-0.163	0.0361	0.0149	0.0187	0.0218
	(0.00533)	(0.00527)	(0.0176)	(0.0171)	(0.0170)	(0.0169)
TAMU	0.0174	0.0355	0.0264	0.0267	0.0316	0.0383
	(0.00230)	(0.00226)	(0.00466)	(0.00455)	(0.00456)	(0.00455)
TAMU-Kingsville	-0.433	-0.164	0.00740	0.00697	0.0117	0.0148
	(0.00619)	(0.00605)	(0.0169)	(0.0161)	(0.0161)	(0.0159)
Texas Southern	-0.666	-0.278	-0.0568	-0.0531	-0.0491	-0.0519
	(0.00446)	(0.00532)	(0.0169)	(0.0166)	(0.0165)	(0.0165)
Texas Tech	-0.116	-0.0147	0.0145	0.0130	0.0196	0.0211
	(0.00304)	(0.00307)	(0.00693)	(0.00675)	(0.00676)	(0.00675)
Texas Woman's	-0.287	-0.146	-0.00175	-0.0191	-0.0163	-0.0144
	(0.00806)	(0.00775)	(0.0206)	(0.0200)	(0.0199)	(0.0198)
Houston	-0.314	-0.170	-0.0829	-0.0835	-0.0810	-0.0856
	(0.00366)	(0.00367)	(0.00852)	(0.00831)	(0.00830)	(0.00829)
UT-Arlington	-0.294	-0.165	-0.0496	-0.0522	-0.0522	-0.0486
	(0.00446)	(0.00436)	(0.0107)	(0.0104)	(0.0103)	(0.0103)
UT-El Paso	-0.461	-0.188	-0.0791	-0.0685	-0.0673	-0.0723
	(0.00434)	(0.00456)	(0.0191)	(0.0186)	(0.0186)	(0.0184)
West Texas A&M	-0.337	-0.187	-0.0342	-0.0479	-0.0425	-0.0390
	(0.00643)	(0.00615)	(0.0224)	(0.0216)	(0.0215)	(0.0215)
TAMU-International	-0.358	-0.112	0.0759	0.0616	0.0561	0.0599
	(0.00855)	(0.00829)	(0.0250)	(0.0240)	(0.0239)	(0.0239)
UT-Dallas	-0.127	-0.0808	0.00976	-0.00153	-0.00173	-0.00568
	(0.00596)	(0.00572)	(0.0126)	(0.0122)	(0.0122)	(0.0121)
UT-Permian Basin	-0.372	-0.190	0.00154	0.00314	0.00487	-0.00588
	(0.0114)	(0.0109)	(0.0394)	(0.0382)	(0.0381)	(0.0376)
UT-San Antonio	-0.356	-0.158	-0.0623	-0.0614	-0.0599	-0.0569
	(0.00344)	(0.00355)	(0.00831)	(0.00806)	(0.00805)	(0.00797)
TAMU-Galveston	-0.189	-0.0570	0.0261	0.0301	0.0354	0.0347
	(0.00927)	(0.00898)	(0.0194)	(0.0191)	(0.0191)	(0.0188)
TAMU-Corpus Christi	-0.314	-0.135	-0.0144	-0.0236	-0.0183	-0.0113
	(0.00567)	(0.00554)	(0.0130)	(0.0127)	(0.0126)	(0.0124)
UT-Tyler	-0.281	-0.183	-0.0299	-0.0386	-0.0377	-0.0366
	(0.00889)	(0.00851)	(0.0231)	(0.0224)	(0.0223)	(0.0219)
Houston-Downtown	-0.615	-0.283	-0.0507	-0.0431	-0.0431	-0.0458
	(0.00587)	(0.00608)	(0.0252)	(0.0247)	(0.0246)	(0.0246)
R-Squared	0.1281	0.2130	0.1517	0.2166	0.2219	0.2443
N	422949	422949	418260	418260	418260	418158

Notes: These estimates correspond to Figures 3 and 5. Each column presents point estimates and robust standard errors from a regression of individual student outcomes on college treatment indicators, omitting UT-Austin as the reference treatment. The UT-Austin outcome mean is 0.82. All specifications control for cohort fixed effects. The Raw Means specification controls for nothing else. The Typical Controls specification adds controls for demographics (gender, race, FRPL), high school academic preparation (10th grade test scores, advanced coursework, and top high school GPA decile indicator), and behavioral measures of non-cognitive skills (high school attendance, disciplinary infractions, and an indicator for ever being at risk of dropping out). The Baseline Specification controls solely for college admission portfolio fixed effects (and cohort fixed effects). The Main Specification adds our core set of covariate controls: gender, FRPL, race, 10th grade test scores, and high school attendance. Extra Control Set 1 adds controls for advanced high school coursework, disciplinary infractions, and an indicator for ever being at risk of dropping out. Extra Control Set 2 adds fixed effects for every high school and an indicator for being in the top decile of high school GPA.

Table B.2: Value-Added Estimates: Earnings

	Raw Means	Typical Controls	Baseline	Main Spec	Extra Controls 1	Extra Controls 2
Angelo State	-17377.9	-11311.4	-660.5	-1690.7	-1676.9	-1174.4
	(305.2)	(308.6)	(1077.8)	(1055.9)	(1055.8)	(1051.0)
TAMU-Commerce	-18930.1	-12443.7	-2434.8	-2843.6	-2792.1	-2893.5
	(385.9)	(385.1)	(1187.6)	(1160.7)	(1159.4)	(1155.1)
Lamar	-15556.7	-7640.6	1096.3	318.3	364.0	363.6
	(328.7)	(326.6)	(906.6)	(880.8)	(879.4)	(881.8)
Midwestern State	-17334.2	-11588.1	-944.8	-1776.5	-1899.8	-1559.8
	(356.2)	(359.1)	(1091.5)	(1072.3)	(1071.4)	(1076.6)
North Texas	-15746.9	-10860.4	-1219.9	-1459.1	-1475.9	-1380.8
	(239.5)	(246.1)	(572.3)	(568.1)	(568.2)	(566.9)
UT-Pan American	-21456.3	-9830.5	475.1	899.9	892.4	1759.0
	(265.1)	(285.8)	(885.1)	(871.1)	(870.9)	(869.9)
Sam Houston State	-13961.6	-7611.6	-843.4	-731.8	-752.1	-504.8
	(267.1)	(273.8)	(691.7)	(682.5)	(682.0)	(680.4)
Texas State-San Marcos	-12666.4	-7685.5	-2753.2	-2479.1	-2433.6	-1840.5
	(247.8)	(254.0)	(556.2)	(549.8)	(549.9)	(546.4)
Stephen F. Austin State	-14997.2	-8445.0	-863.5	-904.6	-886.6	-733.2
	(257.3)	(264.9)	(624.8)	(616.3)	(616.3)	(614.5)
Sul Ross State	-19416.5	-9292.3	1779.1	1242.2	1156.8	1665.9
	(583.4)	(578.1)	(1743.4)	(1701.7)	(1692.7)	(1690.3)
Prairie View A&M	-22487.4	-7304.9	4008.7	3415.8	3341.4	3260.3
	(307.6)	(335.3)	(929.9)	(908.4)	(905.7)	(904.5)
Tarleton State	-15168.0	-10211.6	-1826.9	-1924.9	-1920.5	-1703.8
	(307.6)	(312.2)	(1031.7)	(1011.4)	(1011.8)	(1009.6)
TAMU	1117.2	1180.9	1138.1	1094.1	1104.8	1354.5
	(238.7)	(234.8)	(504.3)	(495.8)	(495.9)	(491.2)
TAMU-Kingsville	-15918.0	-6731.9	815.9	1006.9	1007.7	983.3
	(394.5)	(390.7)	(1017.4)	(987.2)	(985.0)	(984.3)
Texas Southern	-27069.7	-10109.9	1507.7	1628.6	1614.4	1286.7
	(285.7)	(320.5)	(915.3)	(895.8)	(894.5)	(889.8)
Texas Tech	-6073.4	-3402.8	498.9	370.1	446.0	632.8
	(257.1)	(260.9)	(611.8)	(604.0)	(604.2)	(601.5)
Texas Woman's	-15894.7	-5233.9	1662.8	1473.1	1468.0	1499.9
	(399.1)	(395.5)	(1022.4)	(1006.5)	(1003.6)	(1001.7)
Houston	-10512.3	-4957.5	-622.5	-625.9	-613.0	-995.0
	(263.3)	(266.5)	(627.3)	(618.3)	(618.3)	(617.1)
UT-Arlington	-12736.1	-8072.7	-1711.4	-1761.7	-1796.2	-1805.0
	(285.2)	(287.8)	(678.8)	(671.6)	(671.4)	(669.6)
UT-El Paso	-22868.2	-12036.9	-701.2	-153.3	-194.0	128.0
	(276.2)	(293.4)	(1195.9)	(1192.5)	(1191.7)	(1185.1)
West Texas A&M	-16246.5	-11175.5	624.3	111.8	115.4	432.7
	(365.0)	(362.0)	(1427.1)	(1390.2)	(1388.6)	(1391.0)
TAMU-International	-18883.7	-8235.8	548.9	421.6	388.4	756.4
	(422.1)	(423.1)	(1324.9)	(1288.0)	(1287.0)	(1284.7)
UT-Dallas	-6137.0	-5857.7	-2233.9	-3421.9	-3448.7	-3827.1
	(429.5)	(426.6)	(1021.9)	(1014.0)	(1014.1)	(1016.8)
UT-Permian Basin	-12990.2	-6051.4	6421.5	5910.2	5949.4	4911.1
	(645.6)	(627.8)	(2280.6)	(2244.6)	(2243.6)	(2235.9)
UT-San Antonio	-16326.1	-9183.9	-2073.5	-2188.2	-2204.9	-1889.2
	(238.7)	(247.8)	(565.2)	(557.0)	(556.7)	(554.6)
TAMU-Galveston	-7705.0	-4772.6	1667.5	1285.1	1308.9	1172.0
	(687.9)	(674.0)	(1439.6)	(1406.3)	(1405.0)	(1394.9)
TAMU-Corpus Christi	-15727.8	-8499.8	305.9	-193.9	-129.6	-19.19
	(334.4)	(334.3)	(770.6)	(755.3)	(754.7)	(753.4)
UT-Tyler	-12125.6	-8842.5	-2673.0	-3428.8	-3458.6	-3062.9
	(513.4)	(501.7)	(1367.7)	(1324.5)	(1323.5)	(1323.2)
Houston-Downtown	-20027.0	-6341.4	3048.9	3814.5	3731.8	2882.2
	(346.8)	(361.1)	(1212.2)	(1201.1)	(1201.5)	(1200.2)
R-Squared	0.0777	0.1236	0.0973	0.1322	0.1335	0.1527
N	358652	358652	354397	354397	354397	354299

Notes: These estimates correspond to Figures 3 and 5. Each column presents point estimates and robust standard errors from a regression of individual student outcomes on college treatment indicators, omitting UT-Austin as the reference treatment. The UT-Austin outcome mean is 55,975. All specifications control for cohort fixed effects. The Raw Means specification controls for nothing else. The Typical Controls specification adds controls for demographics (gender, race, FRPL), high school academic preparation (10th grade test scores, advanced coursework, and top high school GPA decile indicator), and behavioral measures of non-cognitive skills (high school attendance, disciplinary infractions, and an indicator for ever being at risk of dropping out). The Baseline Specification controls solely for college admission portfolio fixed effects (and cohort fixed effects). The Main Specification adds our core set of covariate controls: gender, FRPL, race, 10th grade test scores, and high school attendance. Extra Control Set 1 adds controls for advanced high school coursework, disciplinary infractions, and an indicator for ever being at risk of dropping out. Extra Control Set 2 adds fixed effects for every high school and an indicator for being in the top decile of high school GPA.

Table B.3: Student-Level Regression Estimates of the Return to Attending a More Selective College

	Raw	Typical	Baseline	Main Spec	Extra 1	Extra 2	UTA Admits
College Mean SAT Score (x100)	6425.5	3035.8	55.81	36.98	39.31	-18.83	-78.95
	(43.18)	(51.69)	(141.8)	(139.6)	(139.5)	(139.0)	(250.9)
R-Squared	0.0673	0.1155	0.0970	0.1319	0.1331	0.1524	0.0717
N	358652	358652	354397	354397	354397	354299	52,459

Notes: These estimates correspond to Figure 7. The mean of the dependent variable (annualized earnings) is $44,834. Each point estimate and robust 95 percent confidence interval comes from a regression of individual student earnings on the mean incoming SAT score of the student's college. The coefficients are scaled to correspond to a 100-point increase in mean SAT scores. All specifications control for cohort fixed effects. The Raw specification controls for nothing else. The Typical Controls specification adds controls for demographics (gender, race, FRPL), high school academic preparation (10th grade test scores, advanced coursework, and top high school GPA decile indicator), and behavioral measures of non-cognitive skills (high school attendance, disciplinary infractions, and an indicator for ever being at risk of dropping out). The Baseline Specification controls solely for college admission portfolio fixed effects (and cohort fixed effects). The Main Specification adds our core set of covariate controls: gender, FRPL, race, 10th grade test scores, and high school attendance. Extra Control Set 1 adds controls for advanced high school coursework, disciplinary infractions, and an indicator for ever being at risk of dropping out. Extra Control Set 2 adds fixed effects for every high school and an indicator for being in the top decile of high school GPA. The final estimate comes from running the Main Specification on the subsample of students who receive admission to UT-Austin.

Table B.4: Match Effects Specification with Treatment-Covariate Interactions: BA Completion

	Main Spec	Add Portfolio-Covariate Interactions	Add Treatment-Covariate Interactions
Angelo State	-0.0415	-0.0330	-0.0463
	(0.0174)	(0.0194)	(0.0214)
TAMU-Commerce	-0.0260	-0.0130	0.00361
	(0.0225)	(0.0257)	(0.0323)
Lamar	-0.0360	-0.0440	-0.0253
	(0.0151)	(0.0166)	(0.0227)
Midwestern State	-0.0217	-0.0343	-0.0473
	(0.0193)	(0.0210)	(0.0243)
North Texas	-0.0317	-0.0332	-0.0310
	(0.00831)	(0.00877)	(0.0115)
UT-Pan American	-0.00467	0.000110	-0.00602
	(0.0143)	(0.0157)	(0.0413)
Sam Houston State	0.0293	0.0280	0.0266
	(0.0105)	(0.0113)	(0.0135)
Texas State-San Marcos	-0.0308	-0.0333	-0.0328
	(0.00757)	(0.00797)	(0.0107)
Stephen F. Austin State	-0.00461	-0.000478	-0.00700
	(0.00955)	(0.0101)	(0.0133)
Sul Ross State	-0.00313	0.0169	0.0359
	(0.0321)	(0.0366)	(0.0488)
Prairie View A&M	0.119	0.113	0.315
	(0.0177)	(0.0198)	(0.110)
Tarleton State	0.0149	0.0246	0.00511
	(0.0171)	(0.0190)	(0.0282)
TAMU	0.0267	0.0271	0.0275
	(0.00455)	(0.00465)	(0.00898)
TAMU-Kingsville	0.00697	0.00815	-0.0444
	(0.0161)	(0.0178)	(0.0260)
Texas Southern	-0.0531	-0.0531	-0.0987
	(0.0166)	(0.0180)	(0.0891)
Texas Tech	0.0130	0.0138	0.0126
	(0.00675)	(0.00703)	(0.0110)
Texas Woman's	-0.0191	-0.00771	-0.0277
	(0.0200)	(0.0229)	(0.0553)
Houston	-0.0835	-0.0830	-0.0934
	(0.00831)	(0.00872)	(0.0111)
UT-Arlington	-0.0522	-0.0496	-0.0455
	(0.0104)	(0.0112)	(0.0133)
UT-El Paso	-0.0685	-0.0550	-0.0929
	(0.0186)	(0.0205)	(0.0318)
West Texas A&M	-0.0479	-0.0479	-0.0408
	(0.0216)	(0.0238)	(0.0311)
TAMU-International	0.0616	0.0533	0.311
	(0.0240)	(0.0274)	(0.102)
UT-Dallas	-0.00153	0.00893	0.0238
	(0.0122)	(0.0130)	(0.0221)
UT-Permian Basin	0.00314	0.0467	0.110
	(0.0382)	(0.0451)	(0.0518)
UT-San Antonio	-0.0614	-0.0551	-0.0594
	(0.00806)	(0.00856)	(0.0107)
TAMU-Galveston	0.0301	0.0315	0.0136
	(0.0191)	(0.0205)	(0.0352)
TAMU-Corpus Christi	-0.0236	-0.0111	-0.00692
	(0.0127)	(0.0138)	(0.0169)
UT-Tyler	-0.0386	-0.0513	-0.0724
	(0.0224)	(0.0258)	(0.0336)
Houston-Downtown	-0.0431	-0.0335	-0.0722
	(0.0247)	(0.0278)	(0.0459)
R-Squared	0.2166	0.2423	0.2428
N	418260	418260	418260

Notes: These estimates correspond to Figure A.9. Each column presents point estimates and robust standard errors from a regression of individual student outcomes on college treatment indicators, omitting UT-Austin as the reference treatment. The UT-Austin outcome mean is 0.82. All specifications control for cohort fixed effects. The Main Specification controls for admission portfolio fixed effects and our core set of covariates: gender, FRPL, race, 10th grade test scores, and high school attendance. The Portfolio-Covariate Interactions specification adds interactions between the admission portfolios and the core covariates. The Treatment-Covariate Interactions specification further adds interactions between the college treatments and the core covariates.

Table B.5: Match Effects Specification with Treatment-Covariate Interactions: Earnings

	Main Spec	Add Portfolio-Covariate Interactions	Add Treatment-Covariate Interactions
Angelo State	-1690.7	-2225.8	-2886.0
	(1055.9)	(1193.7)	(1320.7)
TAMU-Commerce	-2843.6	-1996.9	-1862.4
	(1160.7)	(1327.4)	(1633.1)
Lamar	318.3	415.7	-774.5
	(880.8)	(975.1)	(1419.3)
Midwestern State	-1776.5	-2255.7	-1914.3
	(1072.3)	(1187.9)	(1342.3)
North Texas	-1459.1	-1484.6	-1399.2
	(568.1)	(602.1)	(705.4)
UT-Pan American	899.9	1423.1	4942.8
	(871.1)	(935.6)	(2991.3)
Sam Houston State	-731.8	-724.3	-684.7
	(682.5)	(730.7)	(855.3)
Texas State-San Marcos	-2479.1	-2225.9	-2315.5
	(549.8)	(581.8)	(679.6)
Stephen F. Austin State	-904.6	-796.8	-1045.3
	(616.3)	(660.7)	(799.9)
Sul Ross State	1242.2	2905.5	-448.9
	(1701.7)	(1969.3)	(2862.5)
Prairie View A&M	3415.8	3632.0	23111.7
	(908.4)	(987.0)	(6883.0)
Tarleton State	-1924.9	-2064.7	-1740.7
	(1011.4)	(1116.8)	(1493.4)
TAMU	1094.1	1035.4	1109.0
	(495.8)	(521.3)	(682.3)
TAMU-Kingsville	1006.9	687.9	-507.4
	(987.2)	(1056.8)	(1746.1)
Texas Southern	1628.6	1971.5	17293.8
	(895.8)	(962.1)	(6287.3)
Texas Tech	370.1	221.5	43.23
	(604.0)	(639.5)	(777.9)
Texas Woman's	1473.1	1875.7	-499.6
	(1006.5)	(1103.1)	(3125.5)
Houston	-625.9	-623.3	-297.1
	(618.3)	(661.9)	(753.4)
UT-Arlington	-1761.7	-1784.5	-1264.2
	(671.6)	(729.8)	(794.7)
UT-El Paso	-153.3	736.2	1029.4
	(1192.5)	(1264.2)	(2320.6)
West Texas A&M	111.8	-290.0	-242.3
	(1390.2)	(1535.9)	(1855.2)
TAMU-International	421.6	194.0	-3758.1
	(1288.0)	(1349.3)	(5927.2)
UT-Dallas	-3421.9	-3312.8	-1798.8
	(1014.0)	(1125.0)	(1547.7)
UT-Permian Basin	5910.2	9691.9	12700.4
	(2244.6)	(2525.8)	(3199.5)
UT-San Antonio	-2188.2	-1848.5	-1674.5
	(557.0)	(599.2)	(707.3)
TAMU-Galveston	1285.1	835.5	-4799.3
	(1406.3)	(1584.2)	(2024.5)
TAMU-Corpus Christi	-193.9	62.33	-800.0
	(755.3)	(822.0)	(1015.6)
UT-Tyler	-3428.8	-2961.8	-3610.3
	(1324.5)	(1524.5)	(1896.9)
Houston-Downtown	3814.5	3794.1	2481.4
	(1201.1)	(1330.3)	(2277.0)
R-Squared	0.1322	0.1611	0.1616
N	354397	354397	354397

Notes: These estimates correspond to Figure A.9. Each column presents point estimates and robust standard errors from a regression of individual student outcomes on college treatment indicators, omitting UT-Austin as the reference treatment. The UT-Austin outcome mean is 55,975. All specifications control for cohort fixed effects. The Main Specification controls for admission portfolio fixed effects and our core set of covariates: gender, FRPL, race, 10th grade test scores, and high school attendance. The Portfolio-Covariate Interactions specification adds interactions between the admission portfolios and the core covariates. The Treatment-Covariate Interactions specification further adds interactions between the college treatments and the core covariates.