Three Essays on the College Admissions Process

by

Daniel M. Leeds

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy (Economics) in the University of Michigan 2013

Doctoral Committee:

Professor Jeffrey A. Smith, Co-Chair

Professor Brian P. McCall, Co-Chair

Professor Charles C. Brown

Assistant Professor Kevin M. Stange

To my family and to Rachel,
for their constant love and support.

I would like to thank Dan Ackerberg, Emily Beam, Charlie Brown, Emily Collins, Lindsay Daugherty, Steve DesJardins, John DiNardo, Sue Dynarski, Rachel Ezzell, Rob Garlick, Jesse Gregory, Brad Hershbein, Chris House, Bob Hunt, Josh Hyman, Brian Jacob, Hyunjee Kim, Eva Leeds, Michael Leeds, Paco Martorell, Brian McCall, Isaac McFarlin, Edie Ostapik, Jeff Smith, Kevin Stange, Jacob Vigdor, Elias Walsh, Ophira Vishkin, Laura Zimmermann, participants at Michigan's Labor Lunch and Causal Inference in Education Research Seminars and at the MEA conference, and any others whom I have not mentioned above for their advice and assistance, without which this would have been impossible.

Table of Contents

List of Tables

List of Figures

Introduction

Over the past several decades, the high returns to college completion have become a stylized fact influencing both conventional wisdom and national policy. At the same time, there may be significant heterogeneity in these returns both by student and by school. This, combined with the increasingly nationwide and international college application process, has helped to create industries and policies designed to match students to the colleges of their choice. Preparatory industries exist solely to improve students' standardized test scores, automatic admissions programs are designed to extend college access to diverse populations, and merit-based scholarships serve to funnel students toward particular college choices. Each of these programs presents certain sets of students with unique sets of incentives. In this dissertation, I study the impact that each has on students' preparation for college and their eventual enrollment choice.

In my first chapter, "The Impact of Admissions Policies and Test Prep on SAT Scores," I study how the rules that college use in student evaluation impact student behavior and test scores. Students have differential access to preparatory services and to testing based on their socioeconomic status. Students are able to take the SAT as often as they wish, and many colleges encourage them to submit their most favorable sets of scores. This allows affluent students to test repeatedly in hopes of obtaining high scores, and has led to calls for more equitable student evaluation via explicit or implicit restrictions on testing. However, such policy changes may incentivize different levels of test preparation by SES level and thereby perpetuate inequitable outcomes. I study the impact of several counterfactual admissions policies by creating a model of test prep and retesting, calibrating parameter values, and running simulations. I find that while policies geared towards reducing the amount of testing bring students' scores closer in line to their ability levels, they end up reducing college access and increasing low-income students' costs.

In my second chapter, "Does Student Effort Respond to Incentives? Evidence from Automatic College Admissions," co-authored with Lindsay Daugherty, Paco Martorell, and Isaac McFarlin, Jr., I study how students respond to incentives while in high school. Many high school seniors are susceptible to "senioritis," the practice of shirking once they have been admitted to a satisfactory college. Guaranteed college admissions programs, which admit many students to similar sets of schools early in the school year, provide a quasi-experimental setting in which to observe the impact of senioritis. These programs may incentivize students to shirk or to work harder in hopes of attending an elite college, making their net impact ambiguous. In 1996, Texas passed the "Top Ten Percent" law, which guarantees admission to any in-state public university for students in the first decile of their high school class at the end of their junior year. We use regression discontinuity methods and data from a large urban school district to show that students barely qualifying for automatic admission are substantially more likely to enroll at a flagship institution than those barely missing the cutoff. Qualifying students act as if they are "overinsured," as their GPA falls in their senior year. As shirking in high school may result in worse college outcomes or lower wages, it is in our interest to reduce incentives to shirk. Shirking might be countered by explicitly conditioning acceptances on senior year performance.

In my third chapter, "Iowa's National Scholars Award and the Efficiency of Merit Aid: A Regression Discontinuity Analysis," co-authored with Stephen L. DesJardins, I study the impact that rule-based merit aid has on the probability of student enrollment at the University of Iowa. Between the rising number of students applying to college and the rising costs of college, financial aid officers face large constraints in their ability to attract students to a particular college. Finding ways to effectively distribute financial aid money is therefore extremely important to colleges hoping to maintain an academically talented and diverse student body. Iowa's National Scholars Award, given to students from out of state, uses a prescribed formula to determine which students are eligible for an award worth nearly $5,000. We are therefore able to use a regression discontinuity model to analyze the impact of award receipt on several subgroups of out-of-state students, using in-state (and therefore ineligible) students as an important falsification test. We find a large impact on students' odds of enrollment at the cutoff for award eligibility. Preliminary analysis of an earlier, tiered version of the current single-valued award suggests that the NSA could be effectively targeted towards very high-achieving candidates.

These chapters, taken together, analyze several different points in the college application process, from entrance exam-taking to the choice of where to enroll.

There is much additional work to be done in studying the transition to college, and my work here represents a brief foray into a small subset of the issues worth studying. It is my hope that future work – both my own and others' – will build on the material presented here and will represent an important step forward in the literature of the economics of education.

Chapter 1

The Impact of Admissions Policies and Test Prep on SAT Scores

Abstract

A common critique of standardized college entrance exams is that they favor affluent students. Most colleges evaluate students only on their highest exam scores, and affluent students can take advantage of this policy by repeatedly retesting. Therefore, the logic goes, retesting should be disincentivized, perhaps by evaluating students on a different set of scores. However, affluent students under such a policy may use test prep services as an alternative way to improve their test scores, making the net impact on test scores ambiguous. This paper therefore examines the effects that different score evaluation policies may have on SAT scores and on college access for students from varying socioeconomic backgrounds. By modeling students' decisions to retest and to use test prep services, I provide the broadest examination of the college admissions process to date. I find that when colleges incentivize additional testing, such as by using students' highest SAT scores in determining acceptances, the gap between students' submitted SAT scores and their true ability increases, as does the gap between more affluent and less affluent students' scores. However, more restrictive score evaluation policies may actually increase the costs that poorer students face without increasing their odds of applying to college. As a result, the optimal score evaluation policy remains unclear.

I. Introduction

Despite disputes over the existence or nature of returns to attending college, economists agree that there are high returns to college selectivity for low-income students. These students, however, often face major hurdles. Bailey and Dynarski (2011), for instance, find that low-income students are much less likely than high-income students to attend or graduate from selective colleges. It is therefore crucial to examine barriers to entry at these institutions that disproportionately affect low-income and low-SES students.

The SAT has for generations been one of the highest-profile gatekeepers to selective colleges. While two-year and nonselective four-year colleges may admit students who test poorly or do not test at all, state flagships and selective private colleges often restrict their admissions to high scorers.[1] As a result, the process by which students obtain these scores should matter to policymakers. Organizations such as FairTest ("The National Center for Fair and Open Testing") insist that admissions exams are heavily biased against low-income, minority, and female test-takers. The National Association for College Admission Counseling has responded that "[a] substantial body of literature indicates that test bias has been largely mitigated in today's admission tests due to extensive research and development of question items on both the SAT and ACT." The College Board, which administers the SAT, says that score differences among student subgroups "[reflect] the unfortunate reality that

[1] While there are exceptions, such as the University of Texas's requirement to accept any student finishing in the top ten percent of her graduating high school class regardless of SAT scores, many students admitted in this manner would score reasonably well regardless.

there is a great disparity in educational opportunities for students across the United States" and "should be viewed as a call to action to ensure access and equity for all students."

Altering the way in which SAT scores are evaluated could be a low-cost way of ensuring equality in college access for equally able students from different backgrounds; however, such policies could also exacerbate differences in college access. In this paper, I combine the economic literature on school selectivity, score returns to test preparation, and optimal retesting to determine the impact of hypothetical changes in score evaluation policies on testing, college-going behavior, and SAT scores. I particular, I wish to examine whether admissions policies commonly perceived as "fairer" have harmful unintended consequences, particularly for low-income students. If there are benefits to attending a selective college, these policies might contribute to intergenerational wage stagnation or reduce social mobility.

The argument for more restrictive test policies is straightforward. While testing incurs substantial monetary and non-monetary costs, students from families of high socioeconomic status face lower costs and thus are able to retest more often than their low-SES counterparts. Permissive admissions regimes, which allow students to submit their highest scores (either by exam or by subsection), encourage such retesting and allow high-SES students to generate many score draws. Retesting is also associated with an upward drift in SAT scores, possibly because students have gained more knowledge or become more comfortable in test settings in the interim. As a result, high-SES students will submit higher scores than low-SES students conditional on their ability levels. Therefore, using students' average scores or explicitly limiting how often students may retest should remove some of the advantages associated with higher socioeconomic status and allow colleges to more accurately select students for admission.

While this argument is appealing, it accounts for only one way to improve test scores. It is unlikely that high-SES students will quietly abandon hope of score gains if retesting is disincentivized. Instead, they will seek alternate sources of score improvement, most likely through the use of preparatory services. As these services are typically very expensive or effort-intensive, it is more likely that high-SES students rather than low-SES students will use them.[2] As a result, if retesting is disincentivized, high-SES students may instead seek score gains through test prep. It is possible (if prep is both very effective and very costly) that score gaps between high-SES and low-SES students conditional on their ability levels will be maintained or grow, but that overall utility will fall. Even if score gaps are held constant, higher levels of prep usage may widen SES-specific score distributions conditional on underlying ability. If so, SAT scores will contain less information about student ability levels than under current policies.

To investigate the impact of admissions policies on student outcomes, I develop a model of retesting and test prep usage. Students may vary in actual ability, perceived ability, and SES level. Low-SES students face higher costs of testing, test prep usage, and applying to college than do high-SES students. At each node of the implied game

[2] If, for instance, students from Low-SES backgrounds are more likely to work to support themselves or their families, they will be less able to engage in testing or in test prep.

tree, students may choose to take the SAT, apply to a nonselective university, apply to a selective university (if they have previously tested), or use test prep services. Prep services provide students with an extra ability signal, which may discourage some from taking additional tests. Students are aware of admissions policies–here, the combination of SAT scores that colleges will use to evaluate them–and use this knowledge in determining their testing and prep decisions.

I limit my analysis to state flagship universities for several reasons. First, as these institutions receive an extremely high number of applications, they are more likely than smaller colleges to use broad admissions rubrics, sacrificing some intensity of evaluation in favor of administrative efficiency.[3] As a result, Second, the economic literature suggests that there are large returns to college selectivity. At one extreme, Hoekstra (2009) uses a regression discontinuity analysis of students' GPAs and SAT scores to show that attending a state flagship over other in-state institutions may increase earnings by up to 20 percent for white men. Black and Smith (2006) use a set of five variables to proxy for college quality, finding that a one standard deviation increase in college quality improves wages by approximately four percent. Even Dale and Kruger (2002 and 2011), who find minimal returns to college selectivity among students attending at least somewhat selective schools, find that students from low-income households earn more as they attend more selective universities. Thus, even under conservative assumptions about the returns to college selectivity,

[3]For instance, the University of Iowa evaluates students based on a linear combination of their high school GPA, high school class rank, standardized test scores, and courses taken. Students who score above a fixed cutoff are guaranteed admission to the College of Liberal Arts. The Colleges of Engineering and of Business use similar cutoff rules. Remaining students are evaluated on an individual basis.

social welfare could be improved if more low-income but high-ability students were admitted to selective universities.

College choice also matters due to undermatch–highly qualified students' attending weak universities. While it seems more likely that overmatched students would suffer from attending highly competitive schools, Bowen, Chingos, and McPherson (2009) find that it is instead undermatched students who graduate at surprisingly low rates. Matching low-income students to the proper quality school is therefore important not simply for the effect that their diploma will have on their wages, but for making sure that they obtain the diploma in the first place.

I fit my model to data from the Education Longitudinal Study of 2002 and its two follow-ups ("ELS:2002"). I also use publicly available data on state flagship universities' entering freshman classes as a proxy for college selectivity within each state. I then match simulated data on students' testing behavior and SAT scores to actual data, assuming that colleges look only at students' highest scores by section. This allows me to determine which combination of parameter values best predicts student behavior. I then use these parameter values to simulate student outcomes under counterfactual admissions regimes.

Contrary to the conclusions in Vigdor and Clotfelter (2003), it is not clear that using students' highest SAT scores in college applications is a bad thing. On the one hand, this policy results in a large gap between high-SES and low-SES students' scores conditional on ability, providing high-SES students with a leg up in college admissions. On the other hand, score evaluation policies that reduce this gap discourage some students from testing or applying to selective universities. These policies also lead students to substitute away from retesting towards more expensive test prep programs; as a result, low-SES students may end up paying more but having lower access to college. This suggests that changing policies to reduce score gaps may be suboptimal policy.

The rest of the chapter proceeds as follows. In Section II, I discuss the use of standardized tests and test prep and highlight a selection of the literature on test prep and retesting. In Section III, I present my model. In Section IV, I discuss the data available from ELS:2002. In Section V, I discuss how I apply my model to the data and how my simulations are set up. In Section VI, I present the results of these simulations and discuss how they should be interpreted. Section VII concludes.

II. Background and Literature

1. Test Scores and College Admissions

In applying to most four-year colleges, domestic students must submit their scores on the SAT or ACT exams. I focus on the SAT in my analysis to take advantage of its greater variation in composite scores.[4] Both exams are typically taken during

[4] A similar study could be done using ACT scores, or combining the two using official concordance tables. I prefer, however, to sacrifice some power in order to increase the amount of variation in my data. Given the different scoring scales, concordance tables map ranges of SAT scores into a single ACT score, but can only map each ACT score back to a single SAT score. As a result, the amount of variation in converted ACT scores will be somewhat lower than would be observed if the same set of students took the SAT and were equally proficient at both exams.

the junior year of high school, though some students choose to (re)take them early in their senior year. Approximately half of all SAT-takers take two or more exams, while ten percent take three or more. I focus on the Math and Critical Reading (or, less formally, Verbal) sections within the SAT as my data contain only these two sections.

Score-sending policies have varied over time. Initially, students sent all of their SAT scores to each college they applied to. Colleges were free to do with these scores as they wished; quite often, they would evaluate students on their highest composite score or highest Math and Verbal subsections.[5] Of course, if a student took the SAT five times in hopes of getting a lucky score, all five scores would be visible; this presumably discouraged excessive test-taking. Beginning with the high school class of 2010, however, the College Board instituted a "ScoreChoice" program, which allows colleges to request specific combinations of scores. Students in turn are "encouraged" to submit scores to each college in line with its policy; the default policy is to request all scores. Of the 971 institutions with score policies listed by the College Board, approximately 55 percent request students' highest scores by section, another 13 percent request students' highest composite score at a single sitting, 20 percent request all scores, and the remainder require that students contact them ("SAT®Score-Use Practices by Participating Institution").[6] Within each group, there are both elite schools and nonselective schools. By requesting only high scores, schools are better

[5]The Princeton Review's website states that most colleges "consider only your highest scores (by section or by test date) when making admissions decisions."

[6]As I am concerned chiefly with the admissions policies at selective academic institutions within the United States, I omit 86 foreign institutions, 67 bible colleges and other seminaries, and 310 Community and Junior Colleges.

able to advertise the test scores of their students, may attract some students nervous about sending all of their scores, and may be able to reduce some administrative costs. Students, meanwhile, are judged exclusively on favorable scores.[7] The College Board suggests that this will "reduce student stress and improve the test-day experience." Detractors counter that it allows affluent students to test repeatedly without revealing this; as the exam costs $50, repeated testing may be prohibitively expensive for budget-constrained households. While my data were collected before ScoreChoice was implemented, its usage suggests that colleges are collecting less information than before and impacts potential policy options. For instance, colleges may have difficulty incorporating upward score drift from retesting into applicant evaluation.

2. Score Gains from Retesting and Prep

While the gains from test prep have been the subject of much debate, the general consensus is that prep does improve scores moderately. Evidence that companies have exaggerated these gains has hardly discouraged students from using prep services–a 1979 Federal Trade Commission investigation found that the Stanley H. Kaplan Co. produced average gains of only 25 points rather than the 100 that Kaplan claimed, but is credited with boosting Kaplan's business ("Test-Prep Pioneer Stanley H. Kaplan Dies at 90"). There are several possible reasons that students continue to use

[7] Colleges may choose how to interpret the scores that they request; for instance, a school could request all scores but evaluate students chiefly on their highest scores.

test prep services despite these findings–students may believe that they will experience above-average score gains, may be striving for only incremental improvement, or may not know of or trust these results. Regardless of the reasons, test prep remains a multi-million dollar industry.

Most studies of the effectiveness of test prep have samples too small for accurate inference, heavily selected samples with little external validity, or uninformative research designs (such as lacking a control group). While three recent papers use more advanced techniques to study the gains from SAT coaching, each has its flaws.

Powers and Rock (1999) run five separate analyses, using a self-designed survey in addition to College Board data on students' test scores and School Descriptive Questionnaire responses.[8] These analyses include using an instrumental variable selection model, a Heckman correction, and nearest-neighbor propensity score matching. However, they never make clear their exclusion restrictions. In their IV and Heckman estimation, they use variables for ethnicity, previous SAT scores, father's education, high school GPA, grades in math (when estimating scores on the Math section) or social science (when estimating Verbal scores), and the difference between students' prior scores and the mean SAT scores of applicants to their first-choice college. None of these appear to be ideal instruments, as each could plausibly affect both a student's odds of using test prep and her test scores independent of test prep usage. Similarly, these variables taken together do not appear to satisfy the conditional

[8] The Student Descriptive Questionnaire asks students about demographics, parental education and income, and college aspirations, among other topics.

independence assumption required for propensity score matching–motivation, for instance, is not fully captured in the above list of variables and certainly affects both treatment status and untreated outcomes.

Briggs (2001) runs a linear regression of test scores on prep usage using data from National Education Longitudinal Study of 1988, controlling for similar background characteristics as those in Powers and Rock and restricting his sample to students who have taken the PSAT. However, the OLS estimates presented almost certainly suffer from selection bias and omitted variable bias, and he does not report any exclusion restrictions when presenting a Heckman selection model.

Domingue and Briggs (2009) appears to best account for the requirements of its statistical techniques. In addition to replicating Briggs (2001) with new data, the authors use propensity score matching to estimate the gains from coaching. Domingue and Briggs do a better job than Powers and Rock of satisfying the conditional independence assumption, however, by creating several variables designed to measure sources of motivation. These include identifiers for students who outperform their predicted PSAT scores but have fairly low grades (as these students may perceive the SAT as the easiest path to college admission) or students who significantly underperform their expected PSAT scores (who may be more likely to turn to coaching due to test-day anxiety). Using this methodology, Domingue and Briggs estimate that test prep adds 6 to 9 points to Verbal scores and 11 to 15 to Math scores.

While it may contribute to one's score, test prep is not likely to be a major source of long-term skills. Papers have argued as much beginning with Campbell (1976), continuing with Holmstrom and Milgrom (1991) and most recently with Neal (2011). These papers all argue some variant of a principal-agent model–if the student (the principal) is paying test prep services (the agent) for improvements in test scores, an easy-to-measure product, then test prep services will have little incentive to produce long-term skills, a harder-to-measure product. This is not to say that test prep teaches no long-term skills, merely that they will be developed incidentally rather than as their own end. Test prep may signal that students using such services are dedicated to graduating, but I am not aware of any literature on this. I therefore treat test prep services as adding to students' scores but not to their underlying ability. Colleges in my model (and likely in the real world) will be unable to distinguish test prep from other sources of noise in test scores.

Alternatively, students may improve their test scores through retesting. Nathan and Camara (1998) find that students who take a second test improve their scores by an average of 10 to 15 points per section, with gains slightly larger for males on the Math subsection. This result should be interpreted as treatment on the treated, since students self-select into retesting. Though the direction and magnitude of any composition bias are unclear, my prior is that these gains are overstated, partly because the authors do not examine whether students use test prep and partly because students will be more likely to retest the more they expect their scores to improve.[9] There is substantial variation in score gains on students' second exams,

[9] These gains will be understated if students with large expected score gains fail to retest. For instance, intelligent and motivated students might have large expected gains from retesting but also score highly enough on their first test to guarantee admission to their university of choice, making retesting unnecessary.

with approximately ten percent of scores falling by at least 50 points and over twice as many improving by over 50 points.

3. The Effects of Admissions Policies

Vigdor and Clotfelter (2003) (hereinafter "VC") examine the interaction between admissions policies and SAT retesting. VC construct a model of how students retake the SAT, calibrate parameter values to fit the data, and then analyze a range of counterfactual admissions policies. In doing so, they use two groups of students–one with high costs of test-taking and one with low costs of test-taking, both of which have the same expected gains from attending college.

VC evaluate admissions policies on the metrics of "Accuracy," "Precision," "Bias," and "Cost." Accuracy is defined as the mean difference between the percent of SAT questions that students answer correctly and their true ability levels (which VC treat as the probability that they get any particular question correct). Precision is the standard deviation of this difference. Bias (not to be confused with the common statistical usage) measures the difference in Accuracy between high-cost and low-cost students. Cost measures the number of exams taken.

VC find that using students' high scores in college admissions is socially inefficient. In particular, this policy performs poorly in Accuracy, Bias, and Cost. In other words, students' scores are heavily inflated relative to their true ability levels, this inflation occurs primarily among students with low costs of testing, and students are taking too many tests. While it performs somewhat better on Precision than on the other three metrics, several other admissions policies outperform it on that metric as well.

Unfortunately, VC's model does not allow extra studying or prep to affect test scores. While VC do account both for upward drift in scores and for lucky score draws, students' choices are reduced merely to whether to take an(other) exam. I believe, and present evidence below, that this simplification qualitatively affects their model's results.

III. Model

1. Overview of Testing Process

I examine students' optimal paths within a two-test framework, partly for tractability and partly because it is difficult to determine in the data whether students have tested more than twice; however, the vast majority of SAT-takers should be captured in such a model.[10] For comparison, a two-period version of VC's model, which does not include test prep, is shown in Figure 1, while my model is depicted in Figure 2. Students are forward-looking and maximize their utility (rather than their SAT scores or probability of admission to a selective school). Use of prep services does not obligate students to test.

[10] VC state that approximately 90 percent of college applicants in 1997 took fewer than three SATs.

Students are initially assigned an ability level and SES level. Ability is split into two components, $\boldsymbol{\alpha^*} = \left(\alpha^{M*}, \alpha^{V*}\right)$, which dictate students' performance on the Math and Verbal sections of the SAT respectively. Similarly to VC, though I refer to $\boldsymbol{\alpha^*}$ as "ability," I use the term loosely and do not take a stance on what specific characteristics the SAT measures.[11] Colleges value ability both in its own right and as a predictor of graduation, future success, and ability to donate as an alumnus.

Socioeconomic status, $\ell \in \{L, H\}$, reflects the costs of testing, prep usage, and applying to college and the bonuses associated with retesting and with test prep. Since the SAT and ACT both charge a flat testing fee with fee waivers for particularly low-income students, non-monetary costs (such as psychic or travel costs) are included in all costs.[12] This is broadly consistent with Avery and Kane (2004), who show that students from poorer backgrounds are less likely to show up on test day even if they have already registered and paid exam fees.

To prevent students from following a purely deterministic testing process or targeting unrealistically high or low scores, students do not know their true ability and must infer it from their test scores and other signals received during the testing process.[13] Rather than observing their true ability levels, students see their score on the

[11] The SAT itself does not take a particularly strong stance on this question. Initially, SAT was an acronym for "Scholastic Aptitude Test," suggesting that the exam revealed some innate measure of intelligence. In 1990, it was changed to "Scholastic Achievement Test," implying that the exam measured or predicted success in academic settings. Since 1993, the SAT has been an empty acronym; the full exam title is currently the "SAT Reasoning Test." Nevertheless, the exam measures some quality that colleges find attractive, even if that quality can formally be described only as "the ability to perform well on the SAT."

[12] Neither VC nor I address fee waivers, which allow extremely low-income students to avoid paying the SAT's testing fee. However, as monetary costs are only one component of the costs that I analyze, this should not threaten my results.

[13] Students are aware of all parameters other than their true abilities.

PSAT, a test used in part as practice for taking the SAT. These scores are written as $x_0^M = \alpha^{M*} + \epsilon_0^M$ and $x_0^V = \alpha^{V*} + \epsilon_0^V$.[14] The PSAT does not perfectly reflect a student's true test-taking ability–it has a slightly different format and is much lower-stakes–so these scores consist of students' ability levels plus a mean-zero, normally distributed noise term, $\epsilon_0^j \sim N\left(0, \sigma_0^2\right)$. Students realize that this is the case, and understand that their SAT scores may be significantly higher or lower than their PSAT scores would suggest. As students take the SAT, they continue to update their beliefs, realizing that some portion of their score reflects luck or exam-day idiosyncrasies. Similarly, students who use test prep services must take simulated SATs, which provide yet another source of information about their ability levels.

Students in this model eventually apply to either a selective university ("SU") or a nonselective university ("NSU"). The nonselective school functions as an outside option–application is costless, admission is guaranteed, and attending gives zero utility. The selective school admits students on the basis of their SAT scores and other factors. In the primary specification I treat all other factors as unobserved. Students anticipate enrolling in the nonselective school if they are rejected from the selective one.

Students who apply to the selective university are evaluated differently under each admissions regime. As students are forward-looking, the admissions regime may influence decisions other than whether to retest. For instance, students will be more likely to use test prep services the lower the expected gains from retesting.

[14]By omitting student-level subscripts I am sacrificing some notational rigor for the sake of clarity.

2. Paths through the Testing Process

At each point in the testing process, students face a discrete set of decisions. They may choose (depending on where they are in the game tree) to take the SAT, to use test prep, to apply to the selective university, or to apply to the non-selective university.

Students who test receive a score composed of three terms: their ability, any "bonuses" from their test history, and a mean-zero noise term. Students therefore will receive scores $x_t^j = \alpha^{j*} + Y_\ell + \epsilon_E^j$. The first term represents a student's true ability (determined by nature) on section $j \in \{M, V\}$.[15] The second term, $Y_\ell \in \left\{0, Y_{1,\ell}^P, Y_{2,\ell}, Y_{2,\ell}^P\right\}$, reflects the choices that a student has made to that point in the testing process. Students who have neither tested nor used a prep service do not receive any bonus from their prior actions (or lack thereof). Students from socioeconomic group ℓ who use test prep prior to their first exam receive a boost to their score equal to $Y_{1,\ell}^P$, while students who retest without ever using test prep receive $Y_{2,\ell}$. Students who retest and who have ever used test prep receive bonus $Y_{2,\ell}^P$. While the precise values of $Y_{1,\ell}^P$, $Y_{2,\ell}$, and $Y_{2,\ell}^P$ are an empirical question, we should

[15] VC treat ability as the probability that a student will answer any question correctly within a given SAT section. I do not use this construction for several reasons. First, the odds of getting questions right are not constant, as they may vary both by question type within each section and with a student's mental or emotional state as she takes the exam. Second, it ignores the possibility of leaving a question blank, which on the SAT hurts one's score less than an incorrect answer would. Finally, by treating the SAT as a series of Bernoulli trials, VC ensure that students with very high or low ability have much lower score variation than those with middling ability. While the bounded score range may reduce variance for very high or low ability students, treating questions as binomial trials goes too far in this direction. I instead treat ability as an unbounded constant on the same scale as (bounded) SAT scores.

expect $Y^P_{2,\ell} \geq \max\left\{Y^P_{1,\ell}, Y_{2,\ell}\right\}$. Testing costs $C_{E,\ell}$ in monetary, psychic, and other non-monetary costs. Students may test zero times, once, or twice.

Using test prep has two functions in this model. First, using test prep increases students' expected scores. Students know exactly how effective test prep will be in expectation, but understand that the uncertainty inherent in testing makes it impossible to guarantee any particular gain in their actual score. However, since practice tests are a large component of commercial test prep, prep will also help students learn their true ability. Specifically, test prep provides students with ability signals $y^j = \alpha^{j*} + Y_\ell + \epsilon^j_P$. These ability signals are generated according to the same process as test scores, but have several key differences. Chief among these is that these signals are not usable in applying to college. Second, since students are able to use test prep only once, $Y_\ell \in \{0, Y_{2,\ell}\}$. Finally, obtaining these signals costs $C_{P,\ell}$ rather than the exam cost $C_{E,\ell}$. I assume for simplicity going forward that ϵ^j_P and ϵ^j_E are drawn from the same distribution.[16] Students who use test prep are not obligated to actually take a test–those who receive particularly low ability signals may realize that testing is not for them.

As they do not observe their true test-taking ability, students estimate their ability using signals from their testing and prep history. These ability estimates will generally satisfy $\frac{x^j_0 - \hat{\alpha}}{\sigma^2_0} + \sum_t \frac{\tilde{x}^j_t - \hat{\alpha}}{\sigma^2_E} = 0$ if a student who has tested t times has not used test prep and $\frac{x^j_0 - \hat{\alpha}}{\sigma^2_0} + \sum_t \frac{\tilde{x}^j_t - \hat{\alpha}}{\sigma^2_E} + \frac{\tilde{y}^j - \hat{\alpha}}{\sigma^2_P} = 0$ if she has. In these equations, $\tilde{x}^j_t$

[16]This makes sense if practice tests make up a large part of the prep signal–presumably signals from practice tests given under exam conditions are distributed similarly to those from actual exams.

and $\tilde{y}_k^j$ represent scores and prep signals with any accumulated bonuses netted out. Put differently, a student's ability estimate will be somewhere in between her various scores and ability signals, with the exact location depending on the relative amounts of noise associated with each. Unlike in a Bayesian model, students do not adjust their confidence levels–while they may update their ability estimate from one period to the next, they have perfect faith in their ability estimate at any given point in time. Accordingly, students simply use these estimates in place of their true ability when predicting their test scores; while they do account for noise on test day, they do not account for imprecise estimates of their ability.

At any point in time, students may opt out of the testing process and apply to the non-selective school. Doing so functions as an outside option–it costs nothing to pursue and provides zero utility with certainty. Admission to the selective university provides a higher level of utility B, but applying costs fee $C_{A,\ell}$ and admission is not guaranteed.

3. Applying to College

Applicants to the selective university are admitted with probability $\Phi\left(\frac{x^*-\mu}{\sigma_A}\right)$, where μ represents the degree of the university's selectivity, σ_A represents the weight given to the SAT conditional on that degree of selectivity, and x^* is a function of their SAT scores to date. The functional form $\Phi(\cdot)$ makes an intuitive probability of admission; a fixed score improvement will always improve a student's odds of admission, but will have little impact at very low or very high scores, where she would be accepted or rejected with near certainty.[17] DesJardins, Ahlburg, and McCall (2006) show that students' probabilities of admission to the University of Iowa as a function of their ACT scores and high school class ranks take on a similar shape; while I use SAT scores and do not use class rank in my analysis, this still provides some evidence in favor of the functional form I have chosen.[18] As all of these values are known to the student at the time she applies, her expected utility from applying will be $\Phi\left(\frac{x^*-\mu}{\sigma_A}\right) B - C_{A,\ell}$.

The variable x^*, which reflects the score used in admissions, is the primary factor driving my counterfactual policy simulations. Applicants to the selective university who have taken T tests use score $x^* = f(x_1, \ldots, x_T)$ in their application (abstracting again from the two-test restriction), where $f(\cdot)$ varies by admissions regime. The most commonly used admissions regime is a high section regime. Under this policy, a student who has taken the SAT T times would submit $x^* = \max\left\{x_1^M, \ldots, x_T^M\right\} + \max\left\{x_1^V, \ldots, x_T^V\right\}$. A closely related and commonly used system is a high test regime, in which students submit $x^* = \max\left\{x_1^M + x_1^V, \ldots, x_T^M + x_T^V\right\}$. Both of these policies encourage retesting; the former slightly more so, as students may more easily seek to improve one section score without lowering the other.

[17] For an extreme example, consider 50-point increments from 400 to 450, from 1200 to 1250, and from 1550 to 1600. Any selective university would reject the first candidate both before and after the 50-point improvement, while all but the most selective would accept the third both before and after. The middle candidate, however, would see a nontrivial increase in her odds of admission at many selective universities.

[18] Iowa admitted students and awarded merit aid at this time based on an "Admission Index Score" equal to $2^*ACT_i + \%CR_i$, where ACT_i represents student i's composite ACT score and $\%CR_i$ represents her class rank listed as a percentile.

Alternatively, universities could establish an average score regime, in which students submit $x^* = \sum_{n \le T} \frac{x_n^M + x_n^V}{T}$. If minimizing the amount of testing is of primary importance, schools could implement a first score regime, in which students submit $x^* = x_1^M + x_1^V$.[19] Finally, universities may implement a last score regime, requiring students to submit $x^* = x_T^M + x_T^V$. All of these admissions regimes have two attractive characteristics–they can be easily understood by the vast majority of students intending to attend college, and they allow students to apply after taking a single exam.[20] Conveniently, a student who has tested only once has $x^* = x_1^M + x_1^V$ regardless of regime.

I do not include any explicit measures of how often students have tested when computing their estimated probability of admission. While VC do account for repeated testing in several of their simulations, policy changes such as ScoreChoice that make it easier for students to submit a single set of scores may interfere with many colleges' ability to correct for score drift. As a result, while such policies were undoubtedly relevant for VC's paper, they may be substantially less practicable today.

[19]Though they differ in the specific mechanism, this is functionally equivalent to having a low score regime–in neither case is there anything to be gained from retesting.

[20]Some students may be unable to retest due to the monetary or opportunity costs of testing. Admissions regimes requiring multiple scores may therefore be perceived as biased against such students.

4. Value Functions

Working backwards, it is possible to set up value functions associated with each decision point. In this model, students with two test scores may not retest or use prep, and must instead apply to one of the two universities. That is, they face maximization problem

$$V\left(\boldsymbol{x_1}, \boldsymbol{x_2}, \ell\right) = \max\left\{0, \Phi\left(\frac{x^* - \mu}{\sigma_A}\right) B - C_{A,\ell}\right\}$$

I list only scores and SES level in the value function because these are the only variables that should factor into a rational student's application decision once the testing process has stopped.[21] As a result, this maximization problem will apply to all individuals with two valid test scores, regardless of their prep history or perceived ability.

Prior to this decision, some students will need to determine whether they should retake the SAT. A subgroup of these students will have used test prep previously, reducing their choice to whether to retest, apply to the selective school, or apply to the nonselective school. These students will face a similar optimization problem, but with one added option:

[21]Students in real life are of course affected by unobserved factors that do not affect their odds of admission, but this is beyond the focus of my current work. It is also possible that expectations or ability signals may affect the perceived benefit of admission (i.e. admission to the selective university may be valued more by those who consider it a "reach" and rejection taken harder by those who did not expect it).

$$V\left(\boldsymbol{x_0},\boldsymbol{x_1},\boldsymbol{y},\ell\right)=\max\left\{0,\Phi\left(\frac{x^*-\mu}{\sigma_A}\right)B-C_{A,\ell},E\left[V\left(\boldsymbol{x_1},\boldsymbol{x_2'},\ell\right)\right]-C_{E,\ell}\right\}$$

Once again, they may choose whether to apply to the nonselective university or to the selective one (though the interpretation of x^* is different in this case than in the value function above, since these students have only taken one SAT so far). These students, however, are able to retest. They therefore must weigh the expected benefit of a new SAT score (where $\boldsymbol{x_2'}$ represents the unrealized score) against the cost of retesting and determine whether this gain represents an improvement over applying to either university.

Students who have previously taken the SAT but have not used test prep face all four possible decisions:

$$\begin{aligned}V\left(\boldsymbol{x_0},\boldsymbol{x_1},\ell\right)=\max\{0,&\Phi\left(\frac{x^*-\mu}{\sigma_A}\right)B-C_{A,\ell},\\&E\left[V\left(\boldsymbol{x_1},\boldsymbol{x_2'},\ell\right)\right]-C_{E,\ell},E\left[V\left(\boldsymbol{x_0},\boldsymbol{x_1},\boldsymbol{y'},\ell\right)\right]-C_{P,\ell}\}\end{aligned}$$

Using test prep operates analogously to retesting–in both cases, the student will face an updated value function after receiving a new set of scores but must pay a cost to do so.

Students who have not yet taken the SAT cannot apply to the selective university, as they do not yet have an SAT score. Students who use test prep prior to their first exam must therefore decide whether to test at all or whether to apply to the nonselective school:

$$V\left(\boldsymbol{x_0}, \boldsymbol{y}, \ell\right) = \max\left\{0, E\left[V\left(\boldsymbol{x_0}, \boldsymbol{x_1'}, \boldsymbol{y}, \ell\right)\right] - C_{E,\ell}\right\}$$

Meanwhile, those students starting at the beginning of the game tree must decide whether to take the SAT, use test prep, or forego the testing process entirely:

$$V\left(\boldsymbol{x_0}, \ell\right) = \max\left\{0, E\left[V\left(\boldsymbol{x_0}, \boldsymbol{x_1'}, \ell\right)\right] - C_{E,\ell}, E\left[V\left(\boldsymbol{x_0}, \boldsymbol{y'}, \ell\right)\right] - C_{P,\ell}\right\}$$

Figure 3 depicts these choices as part of the game tree. Note again that all paths must eventually end in an application to one of the two universities–students in this model may not, for instance, decide to completely opt out of postsecondary education following an unexpectedly poor SAT score.

IV. Data

Data for this paper came from the Education Longitudinal Study of 2002, provided by the National Center for Education Statistics. ELS surveyed approximately 16,000 students as they began 10th grade in 2002 and conducted follow-up surveys in 2004 and 2006. Over these four years, ELS asked students questions about their high school, college, and work experiences, among other topics.[22] There is little attrition–fewer than 200 students drop out between the first and second follow-ups. The study's key feature, though, is that it is one of very few data sets that contain data both on multiple SAT scores per student and on test prep usage. In particular, at the 2004 follow-up, when most students were beginning their senior years, it asked them to report their highest and most recent SAT scores by section and whether they had used or planned to use several methods of studying for the SAT or ACT.[23] At the second follow-up in 2006, it asked for their most recent SAT scores by section.

College selectivity data comes from the College Navigator at the Integrated Postsecondary Education Data System ("IPEDS"). IPEDS provides the 25th and 75th percentiles of SAT section scores and of ACT section and composite scores among enrolling freshmen. While these values may not be the same as those for all admitted students, they should nonetheless be broadly indicative of selectivity. Two states–Kansas and North Dakota–have a small enough number of SAT-takers that they do not list SAT scores. I computed the interquartile ranges for these students using their composite ACT scores and the concordance tables provided by ACT, Inc. and the College Board. As New York does not have a true public flagship, I use admissions values for Binghamton University, the public university with the highest enrollee test values.

My model uses two parameters to measure colleges' selectivity–the mean and

[22] I do not use their surveys of parents, teachers, or school administrators.

[23] Specific test prep methods listed are a) high school courses, b) commercial courses, c) private tutoring, d) books, e) videos, or f) software.

standard deviation of admitted students' SAT scores. Using both parameters allows me to gauge both how selective a university is and the amount by which a given shift in an applicant's SAT scores will increase her probability of admission. I rely on several assumptions in computing these parameters. First, I assume that the distribution of admitted students' SAT scores (which is not publicly available) is identical to that of enrolled students' SAT scores.[24] Second, I assume that colleges evaluate applicants on their composite SAT scores rather than their individual section scores, and that enrollees' composite scores are normally distributed. Finally, I assume that the average Composite SAT score is equal to the sum of the average Math SAT score and the average Verbal SAT score, but that Math and Verbal scores are not otherwise correlated among the sample of admitted students. While these assumptions may be strong, the recovered values of μ and σ_A match the conventional wisdom regarding university selectivity, use data available to applicants (who are unlikely to adjust for shifts in composition), and allow me to incorporate state-level variation in flagship selectivity. Formally, I compute the average composite SAT score at a particular institution as $\mu = \frac{M_{25}+M_{75}+V_{25}+V_{75}}{2}$ and the standard deviation of SAT scores as $\sigma = \frac{M_{75}+V_{75}-\mu}{\Phi^{-1}(0.75)}$, where M_p and V_q represent the pth percentile Math score and the qth percentile Verbal score respectively among enrollees at that institution. Values for these parameters are described in more detail in the Appendix.

In creating my sample, I first dropped those ELS respondents who attritted between the first and second follow-ups ($n = 180$). I then dropped those students whose

[24] One example in which this would not hold is if the most highly qualified admits at a particular university are lured away by other schools. In this case, the enrolling class would consist of the remaining, less qualified students, and would not accurately reflect the qualifications of admitted students as a whole.

GPAs could not be computed from transcript data ($n = 1,500$),[25] as well as those with missing data for sex ($n = 710$) or race ($n = 110$). Next, I dropped students who had a most recent score listed in the first follow-up but not in the second ($n < 10$) and those who had either a Math score or a Verbal score–but not both–listed for any exam ($n = 20$). I dropped students from Alaska ($n = 10$) and Washington, DC ($n = 40$); the former because Alaska's state flagship school is nonselective and the latter because Washington, DC does not fit into the state flagship framework.[26]

Finally, I dropped students who did not have PSAT scores listed ($n = 10,770$) or who had taken the ACT ($n = 1,570$). The former group is omitted because the PSAT is a natural signal of students' ability in taking the SAT, making students with a PSAT score a natural sample to examine. The resulting sample loss is troublesome, however, and I discuss below how my sample compares with two subgroups of omitted students. The latter is omitted because students who have taken both the ACT and SAT may be systematically different from those who have taken the SAT alone; those who have taken zero SATs but have taken the ACT will be particularly different from those who have taken neither exam. To abstract from this issue, I focus on students who have not taken the ACT. The final sample consists of 1,580 student-level observations.

As ELS does not provide the number of exams taken by each student, I use the implicit timing of scores in the data to create rough exam histories. Each student

[25] I list the number of observations dropped conditional on all prior drops.

[26] While the University of Maryland is the state flagship university closest to the District of Columbia's city limits, Washingtonians are currently eligible for in-state tuition in a number of different states, further complicating efforts to incorporate them into my analysis.

has up to six listed scores–their highest Math and Verbal scores as of 2004, most recent Math and Verbal scores as of 2004, and most recent Math and Verbal scores as of 2006. Considering one section at a time, I refer to these as "HI1," "MR1," and "MR2" respectively. Given the previous data cleaning, a student will have a valid score for MR2 if and only if she has taken at least one exam. If her (non-missing) MR1 score differs, she must have taken at least one exam by the end of 2004 and at least one between 2004 and 2006; I then increase her exam count to two. Otherwise, I leave her exam count at one. Finally, if her HI1 and MR1 scores are different, she must have taken at least two exams by the end of 2004; I then increase her exam count from one to two or from two to three. I am therefore able to assign each student an "observed" number of exams within each section. I find that approximately 360 students took zero SAT exams, 1,220 took at least one, and 490 took at least two.[27]

The chief drawback of this construction is that it presents an incomplete test history. Students who test repeatedly may have fewer observed scores than actual scores.[28] A specific case of this involves the interaction between HI1 and MR1. To illustrate, consider a student who initially scores 400 on both exam sections, then retests and scores 500 on each. If both exams were taken before the first follow-up survey, her highest section scores (500 and 500) will also be her most recent ones, so taking the data at face value both understates her exam count and overstates her scores. I take a moment-matching approach in calibrating the parameter values in my model, as this allows me to easily compare simulated data to the available data

[27]Given the small number of students who must have taken three or more exams, I focus on the distinction between those students who have taken one exam and those students who took multiple exams.

[28]This goes doubly if these exams are concentrated before the 2004 follow-up.

without additional assumptions about how often students have tested or what their missing scores might have been.

Sample statistics sorted by the number of SAT exams taken are presented below. Table 1.1.A lists student demographics, Table 1.1.B lists degree expectations as of 10th grade, and Table 1.1.C lists several college outcomes. These sub-tables should be interpreted differently: test-taking behavior directly affects outcomes in Table 1.1.C in ways that would be impossible in Table 1.1.A and difficult in Table 1.1.B (as these responses were given before most students would have tested). Nonetheless, it does illustrate the types of students in each category. Those who take multiple exams are more likely to be White or Asian and less likely to be African-American or Hispanic than those who take just one; a similar pattern holds when comparing ever-takers to never-takers. Expectations regarding college-going are relatively stable among students who ever test, but students who test at least twice are somewhat more likely to expect to receive an M.A. or higher degree.[29] Students who test more than once apply to more schools, are accepted at more schools, and attend more selective schools than students who test only once.

Table 1.1.D explores which students are more likely to use various test prep services. A majority of the students who ever take the SAT use some form of test prep–in particular, over two-thirds of testers use some form of book to study for the exam. This allows me to focus exclusively on those students who use more expensive methods of test prep, such as tutoring or commercial courses. Accordingly, my

[29] As in Avery and Kane (2004), a surprisingly large number of students expect to receive a Bachelor's or postgraduate degree but then fail to take the SAT.

findings will represent the gains from these two methods over and above those from using cheaper methods, such as books or software. Unsurprisingly, students who test more are more likely to use test prep. Students who use test prep are likely to be wealthier than those who do not; while a nontrivial number of students in each SES quartile use the two most expensive methods–commercial courses and private tutoring–the share of students in the richest quartile using such services is nearly double that in the next-highest quartile (which is itself higher than the bottom two).

Tables 1.2.A through 1.2.D also examine student demographics, expectations, college outcomes, and prep usage. These tables sort students into three groups. The first consists of those students who never reported taking the PSAT and who did not have a score released by their school. The second consists of students who reported taking the PSAT but do not have an observable PSAT score in the data. The third group is the one I use in my analysis, which consists of students who do have observable PSAT scores, regardless of whether they report taking the PSAT. Students who take the PSAT are more likely to be White or Asian, have higher educational expectations, are more likely to attend selective universities, and are more likely to use any form of test prep than are students as a whole. Crucially, students who report taking the PSAT but do not have a score listed in the data are very similar to those students who do have a score listed, suggesting that my results may be generalizable to this additional group of students. Just as importantly, few students who do not take the PSAT end up attending highly selective colleges. While this may be discouraging from a policy standpoint, it suggests that restricting my data to students who have taken the PSAT will in fact capture most of the students who end up attending highly selective colleges and therefore is not an overaggressive restriction.

V. Simulations

To fit the data to my model, I first define several variables according to information contained in ELS. First, I define HSES students as those in the top SES quartile and LSES students as all others. As Table 1.4 shows a large jump in the number of students using a commercial prep service or tutor at the top SES quartile, this makes a natural break point at which to define HSES status. I define "test prep" within my model as using a commercial service or tutor, as the remaining methods are both relatively widespread and do not contribute statistically significantly to students' scores. Students' highest PSAT scores are used as their initial ability signal. Students use selectivity parameters equal to the mean and standard deviations of freshman SAT scores at their state flagship university, and are counted as applying to a selective university if their highest selectivity application is to a "highly selective" university.[30]

In my simulations, I aim to match several moments of the data. These, along with other illustrative moments, are listed in Table 1.5. I consider two broad categories of moments in my simulations–score-based moments and behaviorally-based moments. In the former category, I include the mean and standard deviation of students' high

[30]Unfortunately, some state flagships do not meet this standard; I use their selectivity data, though, as they represent a combination of quality and affordability that many students may use as a baseline for applications.

Math and Verbal scores and last Math and Verbal scores; in the latter, I have the mean and standard deviation of the number of observed scores by section, as well as the means of prep usage and application to highly selective universities. While I perform my matching on the full sample, I also present these moments (as well as several others) by SES status.

I assign several values, listed in Table 1.6, based on previous paper results and preliminary analysis. First, I normalize B, the benefit of attending the selective university, to 1000. Maximum likelihood estimation (not shown here) of σ in an linear regression of PSAT section scores on high SAT scores gives a value of approximately 5; somewhat surprisingly, this does not depend on depend on students' test prep usage or SES level. As PSAT scores are a proxy for ability, σ reflects the noise from generating both PSAT scores and SAT scores. As the variance of the total measurement error in SAT scores is the sum of the variances of each of its two components, I assign a value of 4 to σ_0 and 3 to σ_E.[31] As PSAT scores represent x_0, I assign true ability levels α_0 by adding a normally distributed, mean zero noise term.

I consider a wide range of bonus and cost values in my simulations, and conduct searches separately within each SES group.[32] Costs vary in increments of one util

[31] I assign a larger value to σ_0 because the PSAT is a slightly different exam and because students may experience slightly more score variation due to inexperience, increased nerves, or other factors that fade with experience.

[32] VC assume that LSES students' costs (relative to the benefit of being admitted) are a fixed multiple of HSES students' costs. With two additional types of costs, this assumption becomes untenable–for instance, LSES students may face very similar test costs but much higher prep costs. As a result, I search over separate ranges of parameters within each SES group.

(normalized so that admission is 1000 utils for each group), while bonuses vary in increments of one SAT point. Among HSES students, I consider over 13,000 distinct combinations of the six parameters (three bonus, three cost) in question. Among LSES students, I consider over 37,000 distinct combinations.[33]

For each combination of values, I evaluate fourteen moments of the simulated data and compute their percentage difference from the corresponding moments of the actual data.[34] To compare specifications, I compute the sum of all squared percentage moment differences; I do similar analysis for the moments of test scores and for the moments of other outcomes separately.

VI. Results

1. Evaluation of Regimes

The metrics used by VC may not be the best way to evaluate admissions regimes. Accuracy represents nothing more than a level shift in students' scores above their abilities. This is a problem only if we observe sizable mass points at perfect section scores. While a valid concern in theory (or when observing applicants to the most selective tier of colleges, as VC do), this is not the case even among the heavily

[33] It is worth noting that because I search more closely where particular parameter combinations appear promising, I do not use a uniform grid of parameters.

[34] Specifically, I consider the mean and variance of students' high Math and Verbal scores, last Math and Verbal scores, and observed Math and Verbal sections. As prep usage and application to the selective university are binary variables, I use only their mean values.

selected sample that I study. Bias, while a useful and interesting parameter, faces a straightforward critique–if HSES students outperform LSES students conditional on ability level, why not simply add a handicap to bring this score difference into alignment? Finally, the cost of testing may be misleading without additional context. If reducing the number of times that students test causes them to spend more on test prep, then reducing the amount of testing may actually increase students' overall costs. Similarly, if incentivizing additional testing increases the probability that qualified applicants will apply to appropriately selective institutions, the increased costs that these students pay may be outweighed by the gain in their expected utility.

I therefore examine a variety of outcomes under my calibrated parameters. First, I consider students' testing patterns–the number of students in each SES group who take any tests, the number who retest, and the number who use test prep. Next, I look at students' college outcomes–the number of students applying to the selective university, the expected number of admitted students from this applicant pool, and the implied acceptance rate.[35] I then examine the total test and prep costs paid by each SES group relative to the number of testers, the number of applicants, and the number of applicants.

With a full set of simulated SAT scores, I am able to examine the impact of students' behavior and the university's policy separately. That is, I am able to simulate students' score draws if they believe that they will be evaluated on one metric and then gauge the expected number of admissions if colleges evaluate them

[35]I assume here that the university weights scores differently when students send multiple SAT scores but keeps the same level of overall selectivity.

on a different metric. This will shed additional light on the relative effects of behavior and policy and on students' testing strategies.

Finally, I examine the difference between students' true abilities and their scores under each regime. In particular, I evaluate the mean of this difference (whether test scores predict student ability on average) as well as the standard deviation (whether any particular student's score is likely to be particularly far from her true ability).

Each set of outcomes is interesting for different reasons. Students' testing behavior will necessarily affect the pool of college applicants, implying that restrictive score evaluation policies may discourage qualified students before they have the opportunity to apply. Students' application outcomes will determine the school that they attend and whether they are undermatched. Finally, the examining the difference between students' scores and ability levels provides a window into whether colleges are admitting the most able group of students.

2. Calibrated Parameters

The best-performing set of parameters is presented in Table 1.7. The first column shows the parameters used in the grid search. The second provides a description of each parameter. The third and fourth columns present calibrated values of each parameter for LSES and HSES students respectively.

Moment matching results for the best-performing specification are presented in Table 1.8. The first row of the table shows the simulated values of the moments being matched. The second shows the percentage deviation of these values from those in the actual data. The table's columns are separated into three groupings-–one indicating how well the simulated distribution of scores matches the actual distribution, another doing the same for student behavior, and a third synthesizing this information.

The values in Table 1.7 show several interesting patterns. Unsurprisingly, HSES students face lower costs of test prep and retesting. However, HSES face higher costs when applying to selective universities. Looking at bonus parameters, LSES students have lower returns than HSES students to retesting, regardless of whether they have used test prep. In fact, HSES students who retest but do not use any test prep can expect to boost their score nearly as much as LSES students who both retest and use test prep. LSES students have surprisingly high returns to test prep on their initial exam, while HSES students have surprisingly low returns. Finally, the bonuses to retesting and to test prep "stack" differently for each SES group. The returns to both test prep and retesting combined are less than the sum of their parts for LSES students. An LSES student using test prep on her first exam can expect her score to rise by approximately 18 points, while a similar student who retests without using test prep can expect her score to rise by approximately 14 points. However, an LSES student who both retests and uses test prep will get a bonus of only 24 points, eight points lower than the sum of the two component bonuses. HSES students, however, have bonuses that stack perfectly–prep alone yields 10 points, retesting alone yields 23 points, and the combination of the two yields 33 points.

These bonuses have several implications. Ceteris paribus, we should expect LSES students to take fewer tests than HSES students do. However, since retesting and prep bonuses do not stack well for LSES students, these two paths to score improvement are much more likely to be treated as substitutes for one another. HSES students, on the other hand, are more likely to treat the two as complements. Put differently, if testing is restricted, LSES students may be more likely to use test prep while HSES students may be less likely to do so. The different costs that the two groups face further confirm this intuition. Given the high costs of test prep and the low returns among HSES students, these students will prefer retesting over test prep whenever possible.

Even the more surprising parameter values make some sense in context. The fact that HSES student face higher application costs may reflect the different types of selective schools to which LSES and HSES students are likely to apply, or may act as a way of correcting for higher SAT scores among HSES students. Similarly, there are three possible explanations for the fact that HSES students have lower bonuses to test prep alone. First, HSES students may take their first SAT exam earlier than LSES students do–returns to early prep may therefore reflect that these students have learned less material to date. Second, the fact that HSES students can retest more easily may lead them to treat their first SAT as a lower-stakes exam. Finally, it may simply reflect the different timing of test prep among LSES and HSES students. If LSES students can afford to either use test prep or retest and HSES tend to consume both items together, these parameters values may reflect these tendencies.

2. Calibrated Parameters

Tables 1.9.A - 1.9.D present an overview of counterfactual simulations. Table 1.9.A shows testing behavior under each set of simulations, Table 1.9.B reflects college outcomes, Table 1.9.C displays the costs associated with each policy, and Table 1.9.D analyzes how accurately scores reflect students' ability levels. Rows in these tables reflect the different possible score evaluation policies in increasing order of restrictiveness–using students' highest scores by section, highest scores by test, last scores, average scores, and first scores. Columns reflect the different parameters of interest by socioeconomic group.

Table 1.9.A backs up the intuition presented in the previous subsection. Since HSES students face lower testing and prep costs, they will be significantly more likely to test, retest, and use test prep. Interestingly, LSES students are more likely to use test prep conditional on testing at all–in part because they are less likely to retest. As score usage policies become more restrictive, testing patterns diverge, largely because these students get most of their score bonuses from retesting rather than from using test prep. Both groups test less and retest less–more HSES students quit the testing process altogether, while the dropoff in retesting is greater among LSES students. LSES students make up for their decreased testing by using more test prep, while HSES students prep less.

Table 1.9.B reflects how these students apply to college. As test policy becomes more restrictive, fewer students apply to college from both SES groups. This decline

is larger among HSES students, in part due to the larger drop in their overall testing rate. However, as these students test at a much higher rate to begin with, there is still a sizable gap in the relative rates of application to selective universities. A similar pattern appears in the relative rates of admission. LSES applicants are admitted a lower rate, though this rate stays relatively stable as test policy becomes more restrictive. HSES admission rates fall more sharply as score evaluation policies become more restrictive, though they remain well above the LSES rate.

Table 1.9.C shows the costs that students pay as part of the testing process.[36] Test costs generally fall as admissions policies become more restrictive. The main exception to this occurs among LSES students under an average score regime. Students under this policy pay more in absolute terms than under the less restrictive last score regime. This exception becomes more pronounced when looking at the costs that students pay per application and become more dramatic still when looking at the cost per expected admission.

Table 1.9.D looks at how students' scores reflect their true ability levels. As expected, when students take more exams and are evaluated on more favorable metrics, their scores will overstate their true ability by larger and larger amounts. Interestingly, LSES students under a first score regime appear to significantly outperform similarly-qualified HSES students. While this is consistent with the parameters used above, it is nonetheless puzzling. It may be the case that the parameters currently

[36] I do not examine the costs associated with applying to the selective university here, as higher application rates would result in higher costs. As I do not wish to imply that higher rates of application are a bad thing, omitting these costs results in a clearer metric.

selected fit current policy rather well but that the model holds certain behaviors constant that would certainly vary under different policies. For instance, the lack of a significant prep bonus among HSES students may reflect that these students take retesting almost for granted. If they are able to test only once, the increased stakes may force them to use more intensive forms of test prep, time their testing more strategically, or put greater effort into their first (and only) exam.

In general, the optimal testing regime from a score evaluation standpoint appears to be an average score regime. This policy appears to provide the most accurate impression of LSES students' scores; while HSES students score closer to their true ability under a first score regime, I am skeptical of this result for the reasons stated above and am more inclined to trust results predicted under an average score regime.[37] Moreover, there appears to be a tighter distribution of scores (conditional on ability) for both socioeconomic groups under this policy than under any other.

VI. Discussion

The advantages and disadvantages associated with different policies make it difficult to unambiguously recommend one particular score evaluation policy. On the one hand, some policies do appear to reduce the advantage that high-SES students hold in college admissions. On the other, these policies do so in part by discouraging students from testing or applying to college, and may raise the costs that low-SES

[37]While the average difference between students' performance and their ability is still fairly low under an average score regime, it is far more intuitive than under a first score regime.

students face. While ensuring a level playing field for applicants is important, doing so at the cost of accessibility may not be worthwhile.

There are several areas where future research may build upon the work presented here. Chief among these is including multiple metrics along which students may form their expectations or be evaluated. While the SAT is one of the primary measures that colleges use in evaluating applicants, it is by no means the only one. Grades, essays, and extracurricular activities are just a small subset of the items that colleges may choose to evaluate during the application process. While quantifying the attractiveness of students' essays or extracurriculars may not be practicable, GPA data is available from ELS. It should therefore be possible to have colleges evaluate students along two metrics simultaneously. Similarly, while the PSAT may provide the single clearest signal of students' test-taking ability, GPA will likely inform their beliefs as well. It may, for instance, influence whether students put much stock in their PSAT score as an accurate indicator of their ability.

This paper also is focused on short-term outcomes rather than long-term outcomes. Students evaluate their likelihood of admission to selective universities without taking into account that these universities have both selectivity and admission targets. That is, if an admissions regime changes the scores submitted to these universities, they may eventually adjust their selectivity to account for this. Drastic decisions may be less likely in the short term, as selectivity and SAT scores play into colleges' rankings, but may occur over a period of time. As a result, students facing an alternative score regime may be able to anticipate lower admissions thresholds, ceteris paribus.

Figure 1.1: 2-Test VC Model

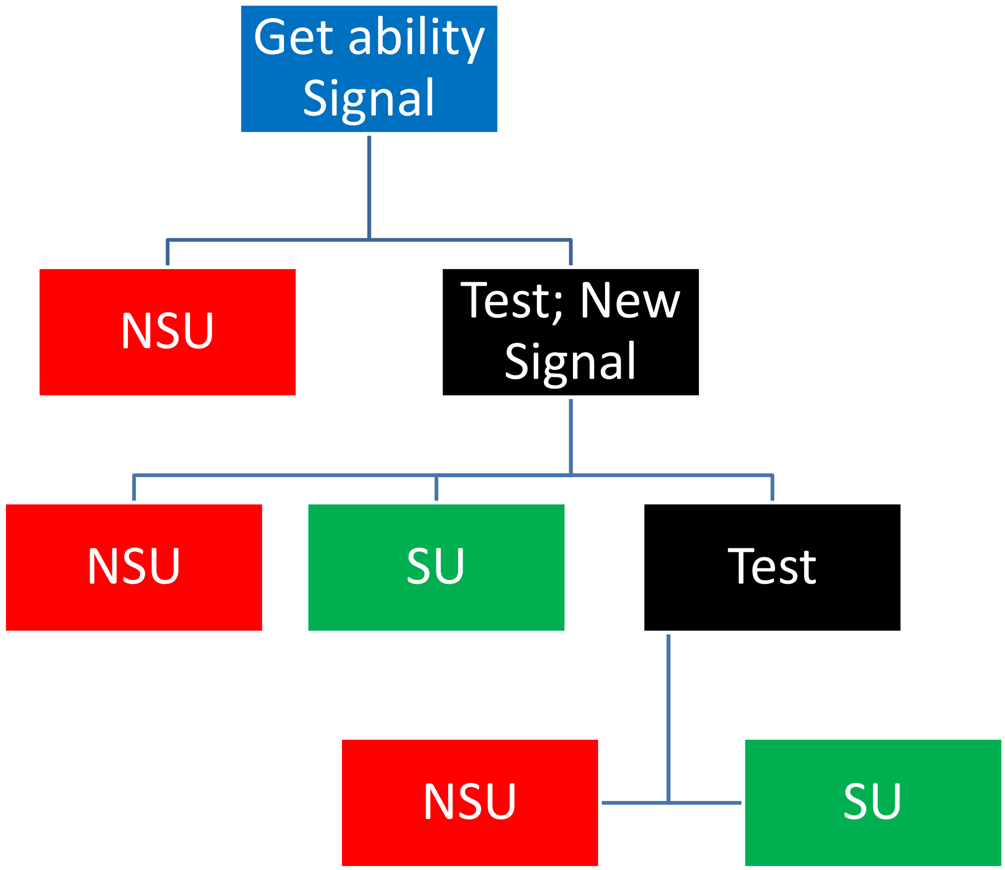


Figure 1.2: 2-Test Leeds Model

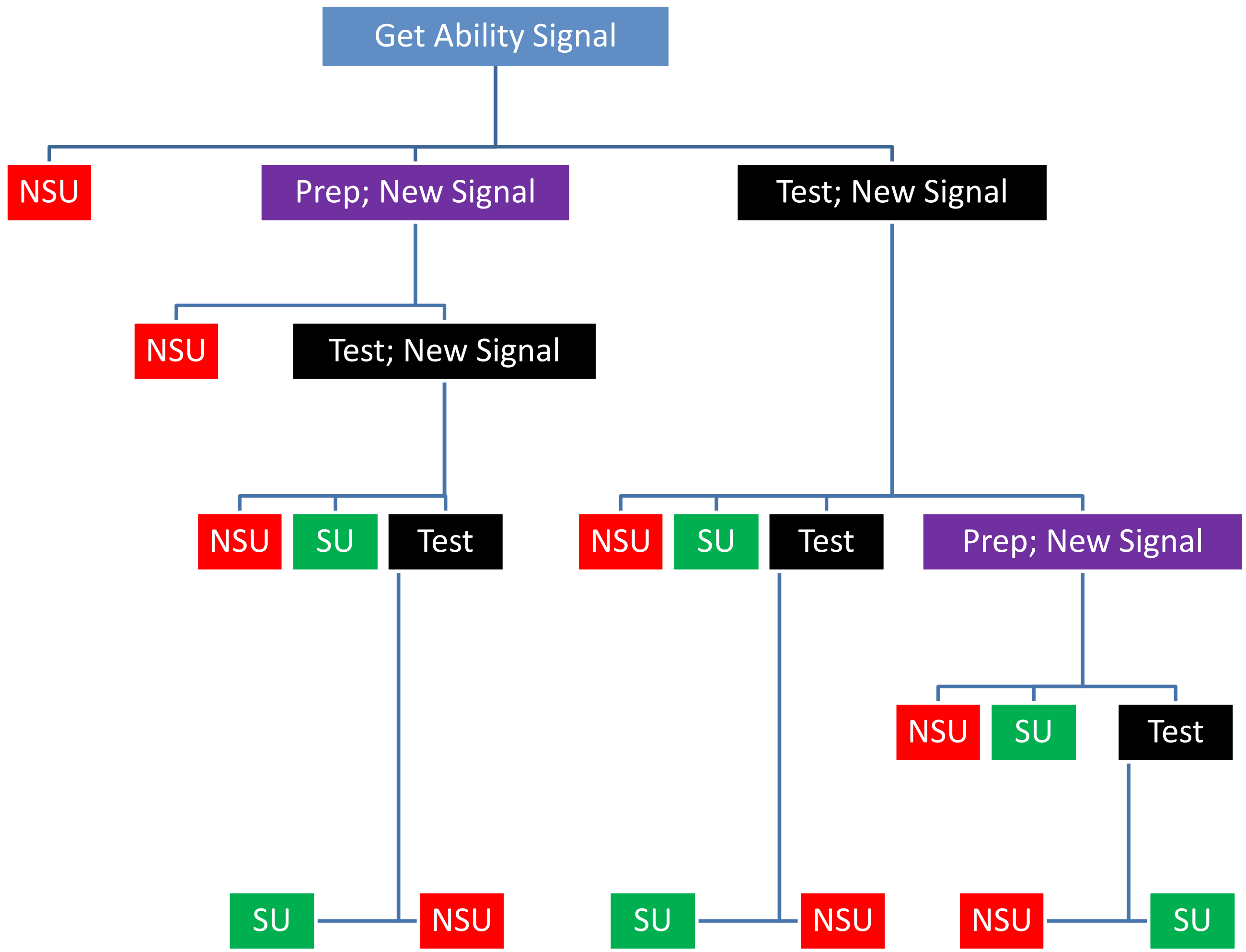
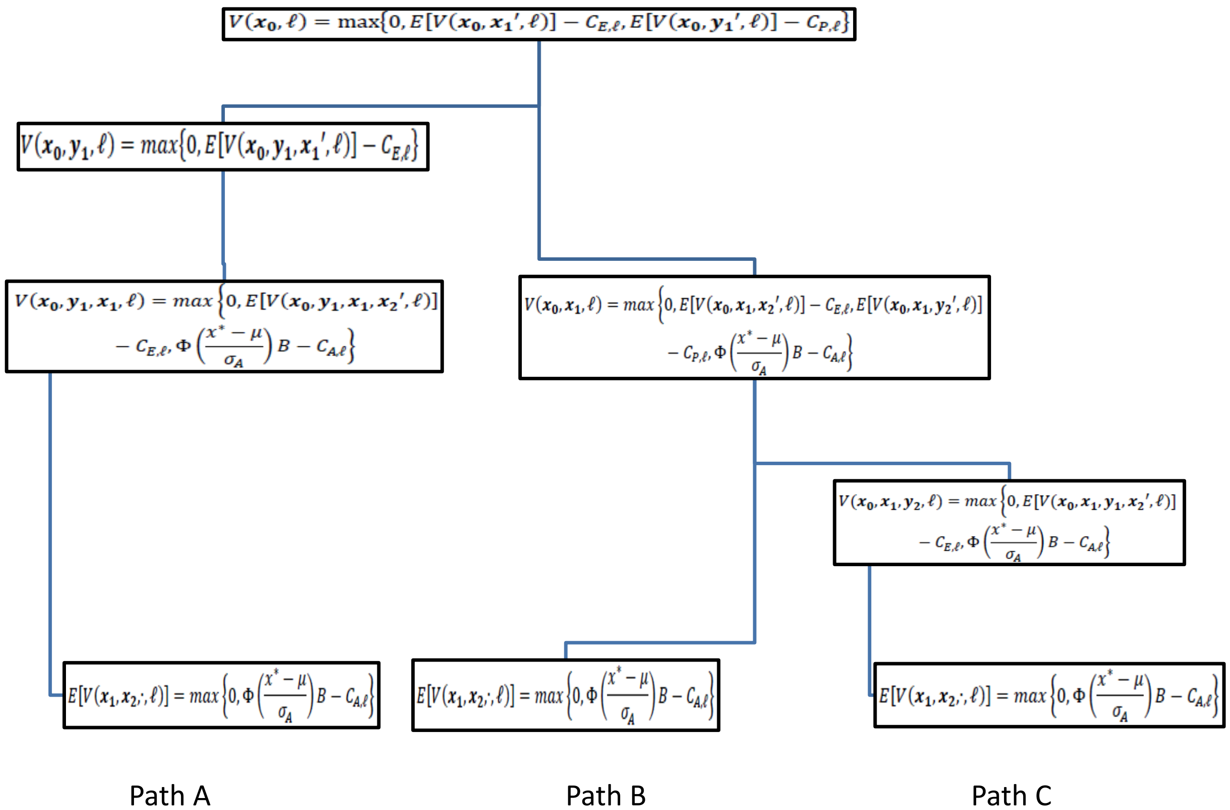


Figure 1.3: Maximization Problems



Paths A, B, and C show the respective maximization problems faced by individuals who test prior to their first exam, second exam, or not at all. μ and σ_A depend on the distribution of SAT scores at students' state flagships and thus vary only at the state level. Further details on these values are presented in the Appendix.

Table 1.1: Summary Statistics by Tests Taken

Table 1.1.A: Demographics

	0 SATs	1 SAT	2+ SATs
% Male	49.7	50.3	46.9
% White	54.2	62.7	64.5
% Black	14.0	7.5	8.2
% Asian	5.6	14.2	14.9
% Hispanic	21.6	10.8	9.2
% Other	4.5	4.8	3.3
% Father has B.A.	21.6	44.9	50.0
% Mother has B.A.	14.9	38.0	44.7
N	360	730	490

Columns are grouped according to whether zero, one, or two SAT exams are visible in the ACT. Students taking any ACT exams are not included here or in the following tables. "Hispanic" ethnicity includes response codes "Hispanic, no race specified" and "Hispanic, Race Specified." "Other" includes response codes "Amer. Indian/Alaska Native, non-Hispanic," "More than one race, non-Hispanic," and "Native Hawaii/Pac. Islander, non-Hispanic."

Table 1.1.B: Expectations

	0 SATs	1 SAT	2+ SATs
% Expect B.A.	78.5	94.3	98.2
% Expect M.A.+	31.5	55.0	62.6
N	300	660	450

This table does not include students whose education ends with less than high school, a high school diploma, some college, or an Associate's degree. "Expect M.A. +" includes students listing a professional degree.

Table 1.1.C: College Outcomes

	0 SATs	1 SAT	2+ SATs
# Schools Applied	1.26	3.08	3.75
	(1.35)	(2.40)	(2.32)
N	290	680	460
# Schools Accepted	1.39	2.35	2.82
	(0.78)	(1.63)	(1.79)
N	200	640	450
% at Public 4-year	11.6	44.9	50.1
% at Private NFP 4-year	6.9	26.5	36.0
% at Private FP 4-year	1.7	1.6	0.7
% at Highly Selective 4-year	1.7	30.6	42.5
% at Selective 4-year	11.6	61.5	78.0
N	170	630	450

Nine schools that liste selectivity levels did not list whether they were public, private not-for-profit, or private for-profit. These schools have been omitted from the table. "Highly Selective" and "Selective" reflect schools' Barron's rankings as of 2006.

Table 1.1.D: Prep Usage

	0 SATs	1 SAT	2+ SATs
% Any Prep	26.1	60.6	69.6
% Comm or Tutor	7.9	21.1	27.4
% HS Course	13.9	23.9	31.7
% Commercial Course	6.5	20.1	24.8
% Tutoring	8.9	11.0	14.2
% Books	36.3	66.3	70.0
% Computer	26.7	31.8	37.8

"Any prep" refers to any of the methods listed above. "Comm or Tutor" refers to usage of a commercial tet prep service or a private tutor. ELS does not specify whether the tutor is associated with a commercial service, high school, or other organization. Video test prep has been omitted due low resposne rates.

Table 1.2: Summary Statistics by PSAT Score Reporting

Table 1.2.A: Demographics

	Non-Takers	Takers, No Score	Score	Total
% Male	55.3	50.1	46.2	52.8
% White	46.6	52.3	61.3	50.8
% Black	14.9	13.7	9.3	13.6
% Asian	8.4	14.7	12.5	10.8
% Hispanic	22.3	13.5	12.7	18.3
% Other	7.7	5.9	4.2	6.6
% Father has B.A.	16.5	41.4	41.3	24.2
% Mother has B.A.	13.7	34.7	34.9	23.1
N	4180	2280	1580	8670

Columns are grouped according to whether students have a PSAT score reported in the data, do not have a PSAT score in the data but reported taking the PSAT, or do not have a PSAT score and do not report taking the PSAT. Students taking any ACT exams are not included here or in the following tables. "Hispanic" ethnicity includes response codes "Hispanic, no race specified" and "Hispanic, Race Specified." "Other" includes response codes "Amer. Indian/Alaska Native, non-Hispanic," "More than one race, non-Hispanic," and "Native Hawaii/Pac. Islander, non-Hispanic."

Table 1.2.B: Expectations

	Non-Takers	Takers, No Score	Score	Total
% Expect B.A.	61.4	91.4	92.2	75.7
% Expect M.A.+	23.4	51.0	52.4	36.6
N	3940	2080	1420	7440

This table does not include students whose education ends with less than high school, a high school diploma, some college, or an Associate's degree. "Expect M.A. +" includes students listing a professional degree.

Table 1.2.C: College Outcomes

	Non-Takers	Takers, No Score	Score	Total
# Schools Applied	1.09	2.85	2.93	1.94
	(1.40)	(2.34)	(2.38)	(2.11)
N	3860	2050	1430	7340
# Schools Accepted	1.32	2.16	2.36	1.86
	(0.82)	(1.50)	(1.65)	(1.39)
N	2220	1860	1280	5370
% at Public 4-year	16.3	44.4	42.1	33.5
% at Private NFP 4-year	5.4	22.9	27.2	17.6
% at Private FP 4-year	3.1	2.3	1.3	2.4
N	1750	1810	1250	4800
% at Highly Selective 4-year	3.2	28.1	30.8	19.7
% at Selective 4-year	13.3	56.7	60.0	41.8
N	1760	1810	1250	4820

Nine schools that liste selectivity levels did not list whether they were public, private not-for-profit, or private for-profit. These schools have been omitted from the table. “Highly Selective” and “Selective” reflect schools’ Barron’s rankings as of 2006.

Table 1.2.D: Prep Usage

	Non-Takers	Takers, No Score	Score	Total
% Any Prep	18.9	60.5	55.6	36.5
% Comm or Tutor	5.3	21.3	20.0	12.2
% HS Course	10.9	25.6	24.8	19.0
% Commercial Course	5.2	17.8	19.5	12.7
% Tutoring	6.9	11.7	11.8	9.6
% Books	30.9	61.1	62.5	48.3
% Computer	17.1	35.2	33.0	26.8

“Any prep” refers to any of the methods listed above. “Comm or Tutor” refers to usage of a commercial tet prep service or a private tutor. ELS does not specify whether the tutor is associated with a commercial service, high school, or other organization. Video test prep has been omitted due low resposne rates.

Table 1.3: SES and SAT Sittings by PSAT Score Reporting

Table 1.3.A: Socioeconomic Quartiles

	Non-Takers	Takers, No Score	Score	Total
Lowest Quartile	38.8	16.3	15.2	28.6
2nd Quartile	29.4	19.0	20.5	25.0
3rd Quartile	20.3	27.7	25.4	23.2
Highest Quartile	11.5	37.1	38.9	23.2
N	4810	2280	1580	

Table 1.3.B: Observed SAT Sittings

	Non-Takers	Takers, No Score	Score	Total
Zero Tests	86.8	35.2	22.6	61.5
One Test	10.9	49.2	46.4	27.4
Two Tests	2.3	15.6	31.1	11.0
N	4810	2280	1580	8670

Columns are grouped according to whether students have a PSAT score reported in the data, do not have a PSAT score in the data but reported taking the PSAT, or do not have a PSAT score and do not report taking the PSAT. Students taking any ACT exams are not included. Socioeconomic quartiles are determined based on respondents' parents' income, employment, and education.

Table 1.4: Prep Usage by SES Quartile

	Quartile 1	Quartile 2	Quartile 3	Quartile 4
% Any Prep	50.0	45.1	55.3	63.6
% Comm or Tutor	13.8	12.7	18.8	27.2
% HS Course	23.8	18.3	23.6	29.0
% Commercial Course	13.2	11.3	18.5	26.1
% Tutoring	9.0	7.4	9.7	16.2
% Books	64.7	53.7	59.1	68.3
% Computer	34.1	31.6	31.8	34.1

"Any prep" refers to any of the methods listed above. "Comm or Tutor" refers to usage of a commercial tet prep service or a private tutor. ELS does not specify whether the tutor is associated with a commercial service, high school, or other organization. Video test prep has been omitted due low resposne rates.

Table 1.5: Data Moments

Moment	All			LSES			HSES		
	μ	σ	N	μ	σ	N	μ	σ	N
High Comp.	1071.57	204.16	1220	1009.99	188.55	670	1145.23	197.76	560
High Math	540.84	111.06	1220	511.01	105.27	670	576.42	107.37	560
High Verbal	530.73	106.28	1220	498.89	97.47	670	568.81	103.90	560
Last Comp.	1052.18	203.82	1220	994.42	189.93	670	1121.26	198.45	560
Last Math	530.72	111.57	1220	503.74	105.90	670	562.99	109.68	560
Last Verbal	521.46	107.38	1220	490.68	99.34	670	558.27	105.11	560
Ever Tested	0.7743	-	1580	0.6898	-	960	0.9070	-	610
Selective App	0.4046	-	1580	0.2956	-	960	0.5759	-	610
Used/Plans to Prep	0.2004	-	1580	0.1546	-	960	0.2724	-	610
Observed Tests	1.0850	0.7277	1580	0.9357	0.7433	960	1.3197	0.6357	610
Observed Math	0.9905	0.6650	1580	0.8600	0.6791	960	1.1958	0.5865	610
Observed Verbal	0.9734	0.6515	1580	0.8413	0.6610	960	1.1811	0.5786	610

"High" scores refer to the highest visible SAT scores as of the 2006 follow-up, while "Last" scores refer to the most recent score available as of the 2006 follow-up. High composite scores reflect students' highest math and verbal scores, while last composite scores reflect the most recent math and verbal scores available. HSES students consist of the top SES quartile, while LSES students consist of all other students.

Table 1.6: Fixed Values

Parameter	Interpretation	Value
B	Utility from attending SU over NSU	1000
σ_0	Variance of noise associated with PSAT	40
σ_E	Variance of noise associated with the SAT and test prep	30

B is listed in utils rather than in dollars. σ_0 and σ_E are in units of SAT points.

Table 1.7: Calibrated Parameter Values

Parameter	Description	LSES	HSES
$C_{E,\ell}$	Cost of testing	27	24
$C_{P,\ell}$	Cost of prep	51	45
$C_{A,\ell}$	Cost of applying to SU	37	51
Y_1^P	Bonus from prep, no retesting	18	10
Y_2	Bonus from retesting, no prep	14	23
Y_2^P	Bonus from retesting and prep	24	33

$C_{E,\ell}$, $C_{P,\ell}$, and $C_{A,\ell}$ are all listed in utils rather than in dollars and should be interpreted relative to the parameter B listed in the preceding table. Y_1^P, Y_2, and Y_2^P are in units of SAT points.

Table 1.8: Moment Matching

Table 1.8.A: Score Moment Matching

SES	Spec.	mMHS	sMHS	mVHS	sVHS	mMLS	sMLS	mVLS	sVLS
LSES	Level	58.7	9.4	56.8	8.9	57.8	9.6	55.7	9.1
	% Diff.	0.1492	-0.1115	0.1388	-0.0911	0.1476	-0.0922	0.1356	-0.0800
HSES	Level	62.7	9.5	60.9	9.6	62.0	9.6	60.3	9.7
	% Diff.	0.0876	-0.1187	0.0703	-0.0789	0.1016	-0.1291	0.0806	-0.0746

Table 1.8.B: Behavioral Moment Matching

SES	Spec.	mSU	mPU	mMT	sMT	mVT	sVT
LSES	Level	0.317	0.157	0.453	0.699	0.452	0.452
	% Diff.	0.0737	0.0134	-0.4729	0.297	0.0506	-0.3157
HSES	Level	0.610	0.271	0.798	0.713	0.768	0.768
	% Diff.	0.0595	-0.0060	-0.3329	0.2153	-0.0289	0.3279

Table 1.8.C: Overall

	Scores	Behavior	All
LSES	0.1173	0.3323	0.4497
HSES	0.0720	0.2691	0.3411

The first row in each table lists the simulated level of each parameter, while the second lists the percentage difference of simulated data from the actual data. Table 1.7.A matches the mean of students' high Math scores, the standard deviations of students' high Math scores, the mean of students' high Verbal scores, the standard deviation of students' high Verbal scores, the mean of students' last Math scores, the standard deviation of students' last Math scores, the mean of students' last Verbal scores, and the standard deviation of students' last Verbal scores. Table 1.7.B matches the percentage of students who apply to selective universities, the percentage of students who use test prep, the mean number of "observed" Math scores, the standard deviation of "observed" Math scores, the mean number of "observed" Verbal scores, and the standard deviation of "observed" Verbal scores. Table 1.7.C lists the sum of the squared entries from columns from Table 1.7.A and Table 1.7.B.

Table 1.9: Counterfactual Outcomes

Table 1.9.A: Counterfactual Testing

Policy	Test_L	Test_H	Retest_L	Retest_H	Prep_L	Prep_H
High Section	0.335	0.626	0.278	0.576	0.157	0.271
High Test	0.328	0.625	0.216	0.564	0.173	0.194
Last Score	0.328	0.625	0.173	0.511	0.171	0.184
Average Score	0.323	0.600	0.077	0.401	0.245	0.191
First Score	0.323	0.600	0	0	0.245	0.160

Columns refer to the percentage of LSES students and of HSES students who ever take the SAT, the percentage of LSES students and of HSES students who retake the SAT, and the percentage of LSES students and HSES students who use test prep services. Rows refer to each of the possible score evaluation policies, in increasing order of restrictiveness.

Table 1.9.B: Counterfactual Applications

Policy	SU_L	SU_H	$\text{E}\left[\text{Admit}\right]_\text{L}$	$\text{E}\left[\text{Admit}\right]_\text{H}$	$\frac{\text{E}[\text{Admit}]_\text{L}}{\text{SU}_\text{L}}$	$\frac{\text{E}[\text{Admit}]_\text{H}}{\text{SU}_\text{H}}$
High Section	0.321	0.604	0.127	0.311	0.395	0.515
High Test	0.317	0.600	0.120	0.303	0.379	0.505
Last Score	0.310	0.604	0.116	0.289	0.374	0.478
Average Score	0.306	0.564	0.111	0.259	0.363	0.459
First Score	0.308	0.564	0.112	0.238	0.364	0.422

Columns refer to the percentage of LSES and of HSES students who apply to the selective university, the percentage of LSES and of HSES students who expect to be admitted, and the expected rate of admissions.

Table 1.9.C: Counterfactual Costs

Policy	TC_L	TC_H	$\frac{TC_L}{SU_L}$	$\frac{TC_H}{SU_H}$	$\frac{TC_L}{E[Admit]_L}$	$\frac{TC_H}{E[Admit]_H}$
High Section	23,658	25,206	76.6	68.1	193.9	132.2
High Test	22,692	22,851	74.2	62.1	195.5	122.9
Last Score	21,456	21,789	71.8	58.9	191.9	123.2
Average Score	22,431	20,001	76.0	57.8	209.2	125.9
First Score	20,433	13,017	68.8	37.6	189.2	89.1

Columns refer to the total costs associated with testing and with test prep paid by LSES and HSES students. The first two columns present these costs in levels, the next two present total costs relative to the number of students who apply to the selective university, and the final two columns present total costs relative to the expected number of students admitted to the selective university.

Table 1.9.D: Score Accuracy

Policy.	$\mu_L(\delta)$	$\mu_H(\delta)$	$\sigma_L(\delta)$	$\sigma_H(\delta)$
High Section	54.0	62.8	44.4	43.6
High Test	46.9	57.8	41.4	43.8
Last Score	39.4	44.6	47.9	44.6
Average Score	32.3	22.4	41.2	36.9
First Score	35.2	3.5	43.9	42.8

Columns refer to the mean amount by which LSES students' scores will overstate their true ability levels, the mean amount by which HSES students' scores will overstate their true ability levels, the standard deviation of the amount by which LSES students' scores will overstate their true abilities, and the standard deviation of the amount by which HSES students' scores will overstate their true abilities.

Chapter One Bibliography

[1] Christopher Avery and Thomas J. Kane. Student Perceptions of College Opportunities: The Boston COACH Program. In Caroline M. Hoxby, editor, *College Choices: The Economics of Where to Go, When to Go, and How to Pay for It*, pages 355–394. National Bureau of Economic Research, Inc., 2004.

[2] Martha Bailey and Susan Dynarski. Inequality in College Entry and Completion. In G.J. Duncan and R.J. Murnane, editors, *Whither Opportunity? Rising Inequality and the Uncertain Life Chances of Low-Income Children*. Russell Sage, 2011.

[3] Dan Black and Jeffrey Smith. Evaluating the Returns to College Quality with Multiple Proxies for Quality. *Journal of Labor Economics*, 24(3):701–728, 2006.

[4] The College Board. College Board SAT and SAT Subject Test FAQs. `http://press.collegeboard.org/sat/faq`, November 2012.

[5] The College Board. SAT®Score-Use Practices by Participating Institution. `http://professionals.collegeboard.com/profdownload/sat-score-choice.pdf`, August 2012.

[6] John Bound, Brad Hershbein, and Bridget Terry Long. Playing the Admissions Game: Student Reactions to Increasing College Competition. *Journal of Economic Perspectives*, 23(4):119–146, 2009.

[7] William G. Bowen, Matthew M. Chingos, and Michael S. McPherson. *Crossing the Finish Line: Completing College at America's Public Universities*. Princeton University Press, 2009.

[8] Derek C. Briggs. The Effect of Admissions Test Preparation: Evidence from NELS:88. *Chance*, 14(1):10–18, 2001.

[9] Donald T. Campbell. Assessing the Impact of Planned Social Change. In *Occasional Paper Series*, number 8. Public Affairs Center, Dartmouth College, December 1976.

[10] Stacy Dale and Alan B. Krueger. Estimating the Return to College Selectivity Over the Career Using Administrative Earnings Data. NBER Working Paper 17159, June 2011.

[11] Stacy Berg Dale and Alan B. Krueger. Estimating the Payoff to Attending a More Selective College: An Application of Selection on Observables and Unobservables. *The Quarterly Journal of Economics*, 117(4):1491–1527, 2002.

[12] Stephen L. DesJardins, Dennis A. Ahlburg, and Brian P. McCall. An Integrated Model of Application, Admission, Enrollment, and Financial Aid. *Journal of Higher Education*, 77(3):381–429, 2006.

[13] Ben Domingue and Derek C. Briggs. Using Linear Regression and Propensity Score Matching to Estimate the Effect of Coaching on the SAT. *Multiple Linear Regression Viewpoints*, 35(1):12–29, 2009.

[14] National Association for College Admission Counseling. Report of the Comission on the Use of Standardized Tests in Undergraduate Admission. Technical report, National Association for College Admission Counseling, September 2008.

[15] Ben B. Hansen. Full Matching in an Observational Study of Coaching for the SAT. *Journal of the American Statistical Association*, 99(467):609–618, September 2004.

[16] Mark Hoekstra. The Effect of Attending the Flagship State University on Earnings: A Discontinuity-Based Approach. *The Review of Economics and Statistics*, 91(4):717–724, 2009.

[17] Bengt Holmstrom and Paul Milgrom. Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *Journal of Law, Economics, and Organization*, 7:24–52, 1991.

[18] C. Kirabo Jackson. A Little Now for a Lot Later: A Look at a Texas Advanced Placement Incentive Program. *Journal of Human Resources*, 47(1):591–639, 2010.

[19] Julie S. Nathan and Wayne J. Camara. Score Change When Retaking the SAT I: Reasoning Test. Research Notes RN-05, The College Board, September 1998.

[20] Derek Neal. The Design of Performance Pay in Education. NBER Working Paper 16710, January 2011.

[21] Donald E. Powers and Donald A. Rock. Effects of Coaching on SAT I: Reasoning Test Scores. *Journal of Educational Measurement*, 36(2):93–118, 1999.

[22] Mitchell L. Stevens. *Creating a Class: College Admissions and the Education of Elites*. Harvard University Press, 2007.

[23] Patricia Sullivan. Test-Prep Pioneer Stanley H. Kaplan Dies at 90. *The Washington Post*, August 25 2009.

[24] Jacob L. Vigdor and Charles T. Clotfelter. Retaking the SAT. *Journal of Human Resources*, 38(1), 2003.

Chapter 2

Does Student Effort Respond to Incentives? Evidence from Automatic College Admissions

Co-authored with

Lindsay Daugherty
RAND Corporation

Paco Martorell
RAND Corporation

Isaac McFarlin, Jr.
Univ. of Michigan

Abstract

It is common to observe high school seniors shirking once they have been admitted to a sufficiently preferred college. Guaranteed college admissions programs, which admit many students at once to similar sets of schools, provide a quasi-experimental setting in which to observe the impact of college admission on student effort. Such programs may either incentivize shirking or motivate students to prepare for demanding academic curricula. Therefore, their effect on student effort is theoretically ambiguous. In 1996, Texas passed the "Top Ten Percent"' law, guaranteeing admission to any public university in Texas for any Texan student in the top ten percent of her class at the end of her junior year. We use regression discontinuity methods and data from a large, urban school district to show that students barely qualifying for automatic admission are more likely to enroll at a flagship institution than those barely missing the cutoff. Qualifying students act as if they are "overinsured," getting slightly lower grades in their senior year, while narrowly ineligible students reduce their course-taking. As lower effort in high school may result in worse college outcomes or lower wages, it is our interest to reduce these incentives. Shirking might be countered by explicitly conditioning acceptances on senior year performance.

I. Introduction

Automatic admissions programs are one of several tools that colleges can use to recruit students from a diverse set of backgrounds. Some of these students may come from disadvantaged backgrounds, some may underestimate their competitiveness or be unaware of their academic options, and others may have their preferences influenced by being proactively selected for admission. Colleges usually extend offers of automatic admission early in the application cycle, possibly to avoid preemption or to ensure that students are able to complete application requirements.

These offers may function as a form of insurance in the college admissions process. Students who might end up in a negative state (rejection from preferred colleges, inability to gain access to college, or ignorance of the process, e.g.) are instead given the option of attending a (potentially) attractive institution. As a result, students who are aware of automatic admissions offers face a guaranteed level of utility provided that they meet set requirements – the analogue of premiums in an insurance model.

Once admitted to college, students face issues of moral hazard. Typically, agents who face moral hazard may be more likely to engage in risky behavior and less likely to protect themselves against negative outcomes (see Pauly (1974) and Akerlof and Katz (1989), e.g.). In our context, college admission implies that students may rationally choose to study less once they have been admitted to a sufficiently attractive institution. In many cases, students do not receive college admissions offers until well into their senior years, making this shirking little more than a nuisance for teachers. Students with automatic admissions offers, though, may be guaranteed admission extremely early in their senior year, which could induce year-long shirking. As a result, early admissions offers may function as overinsurance. Highly qualified students in particular face little incentive to exert extra effort in school, as the odds of having acceptances revoked are fairly low.

The previous literature is vague on the potential impact of admissions offers on student effort levels. Lindo, Sanders, and Oreopoulos (2010) find that the "negative" intervention of warning college freshmen with low GPAs may have multiple effects – it may both discourage students from continuing in school and encourage higher effort levels in those who remain, with ambiguous impacts on eventual graduation. However, they note that the impact of a negative intervention may be qualitatively different from that of a positive intervention. Angrist, Lang, and Oreopoulos (2009) find that positive interventions have little effect, but find that effects vary significantly by student subgroups. Students' responses to incentives matter because additional schooling may have high pecuniary returns and nonpecuniary returns both inside and outside of the labor market (see Oreopoulos (2006) and Oreopoulos and Salvanes (2011), e.g.). If these returns depend on both the quantity and the quality of schooling, shirking may prevent students from capturing them.

We intend to test our hypothesis – that automatic admissions offers may induce suboptimal effort – using a model based on Texas's Top Ten Percent Law and data from an urban Texas school district. As students may invoke automatic admissions offers at the end of their junior year, they may adjust both their course selection and their effort levels. While previous studies have examined aspects of the Top Ten Percent Law ranging from college enrollment to strategic high school enrollment and early high school performance, none have studied the impact of this law on senior year outcomes.

We find that students who barely qualify for automatic admission tend to receive slightly lower grades than comparable students who do not qualify. We also find evidence that some students who do not qualify for automatic admission may be discouraged from further academic effort, taking substantially fewer courses both overall and at the Honors or AP level. There is also some evidence that White and Asian students are more likely to graduate with Recommended-level diplomas (required as part of the guaranteed admission policy), displacing both less-valuable Minimum diplomas and more-valuable Distinguish diplomas.

The literature suggests that offering partial insurance and engaging in monitoring may solve problems relating to moral hazard. As our outcomes are consistent with overinsurance and moral hazard, it may therefore be optimal to condition automatic admissions offers on senior year performance or to extend increasingly selective admissions offers to students with higher GPAs or class rank.

The remainder of the paper proceeds as follows. Section II discusses the Texas Top Ten Percent Law in further detail. Section III examines several related studies, some dealing with academic responses to incentives, others addressing the Top Ten Percent Law itself. Section IV lays out the RD model to be used. Section V discusses the dataset used and how it was set up for this study. Section VI contains the results and interpretation of our analysis. Section VII concludes and presents several extensions for further study.

II. The Texas Top Ten Percent Law

In 1996, the United States Court of Appeals for the Fifth Circuit ruled in the case of *Hopwood v. Texas* that the University of Texas's race-based affirmative action policies were unconstitutional on Equal Protection grounds. Later that year, in an attempt to circumvent the ruling, the Texas Legislature passed Texas House Bill 588, guaranteeing admission to any public university in Texas for any Texan student in the top ten percent of her graduating class.[38] This law (hereinafter the "Top Ten Percent Law") took effect at the beginning of the 1997-1998 school year, and has remained in effect since.[39] While other states – most notably Florida and California – have adopted similar percentage plans, they do not guarantee admission to any single campus. Texas's Top Ten Percent Law, however, allows students to select the campus to which they are admitted (though they may still need to apply to specific programs within each university).

Two campuses in particular – The University of Texas at Austin and Texas A&M College Station, hereinafter "UT" and "A&M" – may be particularly attractive to

[38]While students must still submit a complete application, including SAT or ACT scores, presence in the Top Ten Percent will override any concerns arising from these scores. The law applies to students at both public and private high schools.

[39]This is despite the United States Supreme Court's 2003 affirmation of the constitutionality of certain types of race-based Affirmative Action.

these students. These two campuses, the flagships of their respective university systems, have significantly higher *U.S. News and World Report* rankings than any other public schools in Texas. Given their statuses as highly-respected and selective institutions, a guarantee of admission to either of these universities will dominate many other admissions offers (let alone labor force participation or home production), and for even the most talented and ambitious students will provide a valuable fallback.[40] As many students begin making college application decisions at the beginning of their senior year of high school, the Top Ten Percent offer becomes official at the end of their junior year.

Upon learning their class rank at the end of their junior year, eligible students may react in one of three ways. Some will maintain the study habits that got them into the Top Ten Percent. These students may be unaware of the impact of their class rank or may simply be comfortable with their existing routine. Others will increase their study effort. These students may not have expected to go to college or may be worried about their preparedness. Finally, some students will lower their study effort. These students may be more worried about where they attend than about how they will perform when they are there. The Top Ten Percent Law does not apply to any student ranked outside of the Top Ten Percent, potentially leading to vastly different outcomes between nearly indistinguishable students. This provides an intuitive basis to study student outcomes in a regression discontinuity ("RD") framework.

[40]Credit-constrained students might forgo selectivity in favor of affordability and/or proximity to home. Several studies find that poor and/or minority students are less likely to apply to UT and A&M than economically secure and/or white students are. For our purposes, it is sufficient to assume that students are admitted to their most-preferred university.

While the law guarantees admission, it makes no mention of tuition, fees, or other affordability concerns. Data from the Integrated Postsecondary Education Data System suggest that in-state applicants for the 2011-2012 school year who receive some form of financial aid and come from families earning less than $30,000 could expect to pay an average net price of approximately $5,000 at A&M and $9,000 at UT. Students from more affluent backgrounds could expect to pay more than this. Some students may therefore choose to enroll in less-selective in-state institutions rather than in UT or A&M.[41] While A&M offers a scholarship for Top Ten Percent students, UT does not.

The Top Ten Percent law has implications both for students' access to college and their choice of where to attend. Bailey and Dynarski (2011) find that differences in college enrollment and completion by both race and socioeconomic status have been growing. Much of this can be attributed to stagnant rates of enrollment and completion among low-SES students and underrepresented minorities and growing rates among high-SES and white students. Given the *de facto* racial segregation of many Texas high schools, the Top Ten Percent law is designed to ensure that a large number of low-income and underrepresented minority students have guaranteed college prospects. It also ensures that students of all backgrounds stay in Texas -– as Hoxby (2009) documents, the growing globalization of the college application process makes attending school in other states much easier and complicates colleges' (implicit or explicit) mandates to serve their local and state communities. The Top

[41] As Texan undergraduates are not eligible for tuition reciprocity from other states, credit-constrained students are unlikely to substitute towards out-of-state colleges and universities will be unlikely.

Ten Percent law may preempt this process and ensure that students remain in Texas.

The United States Supreme Court recently ruled in *Fisher v. University of Texas* that race-sighted affirmative action policies are permissible only if universities can demonstrate that race-blind policies would not effectively achieve the same goals. As universities often wish to maintain a racially and socioeconomically diverse student body, many may examine the feasibility of programs such as the Top Ten Percent law, making it crucial to understand all effects of the law.

III. Literature and Background

Several studies have been published on the impact of the Top Ten Percent Law. A large subset focuses on college outcomes. For instance, Long and Tienda (2010) examine how several student characteristics have changed at a number of Texas public schools, finding evidence that the Top Ten Percent Law slowed or stopped steady gains in SAT scores at UT and A&M, crowded out students who fell short of the Top Ten Percent at these institutions, and benefited institutions such as Texas Tech, which were able to attract some of the students crowded out of UT and A&M. Domina (2007) runs fixed-effects regressions on panel data from 1993 to 2002 to examine school-level changes in college applications and in high school performance, finding evidence that the Top Ten Percent Law increased applications to selective Texas universities by students in low-performing high schools and improved school-wide student attendance and advanced course-taking. These results do not separate students by class; so, for instance, a negative effect on seniors' attendance and course-taking might be dominated by a positive effect among other students. Cortes and Zhang (2011) study the incentive effects of the Top Ten Percent Law, finding that students in low-performing high schools, regardless of race, are more likely to attend UT or A&M and perform better on 10th grade achievement exams.

Both Niu and Tienda (2010) and Daugherty, Martorell, and McFarlin (2013) use RD designs to estimate the impact of membership in the Top Ten Percent on flagship university attendance and other outcomes. Niu and Tienda find particularly sharp impacts among Hispanic students and those at predominantly minority high schools; DMM find that students of all racial groups are more likely to attend a flagship university but that overall college-going is largely unaffected. These findings imply that UT and A&M's student bodies will be largely composed of students in the top decile of their high school. Highly qualified students who do not meet this class rank requirement will be crowded into private schools or lower-tier public schools in Texas or into schools in other states.

Several studies examine the incentive effects of other programs on academic outcomes. Jackson (2010a and 2010b) uses randomized rollout to study the impact of another Texas innovation – a program paying students and their teachers for above-average performance on Advanced Placement exams – finding positive impacts in areas ranging from SAT scores to college attendance. These gains were not accompanied by negative unintended consequences, as students still performed comparably in non-AP classes, and were not subject to fadeout in college, suggesting that study habits practiced late in high school will carry over to college. Just as crucially, he states that the monetary rewards from this program are almost certainly lower than the psychic and opportunity costs of the higher effort needed to secure them. This provides evidence that improving students' estimates of the gains from education may be the driving factor in their altered behavior. Two papers by Cornwell, Lee, and Mustard (2005 and 2009) examine how the Georgia HOPE State Merit Scholarship affects student behavior, finding that recipients of the scholarship whose GPAs put them on the margin of eligibility were less likely to enroll in math and science courses, more likely to declare an Education major, less likely to sign up for a full credit load, and more likely to withdraw from courses.[42]

Several other papers study the impact of academic interventions on college students. Lindo, Sanders, and Oreopoulos (2010) find using an RD model that a "negative" intervention – putting underachieving students on academic probation – has mixed positive and negative effects. Students who were put on academic probation after their first semester at college were more likely than comparable students not on probation to drop out prior to their second semester, but those who remained at college saw their GPAs increase by more than the control groups'. The impact of probation on eventual graduation was ambiguous, possibly because only students who had previously been placed on probation could be suspended from the university. Angrist, Lang, and Oreopoulos (2009) find that positive interventions – additional tutoring resources and financial incentives – have only limited effects on student

[42]The latter two findings are attributed to the fact that scholarship eligibility is evaluated at benchmarks corresponding to fixed credit levels. This gives students who are at risk of becoming ineligible incentives to push the checkpoint into later semesters, ensuring additional semesters of eligibility.

performance. In particular, female students receiving both additional tutoring and financial incentives saw their GPAs increase, while male students and female students receiving only one of the two interventions did not. This suggests that analysis by subgroups may yield stronger results than pooled analysis.

There is also an extensive literature on the impact and effectiveness of various types of affirmative action policies. Fryer, Loury, and Yuret (2007) provide evidence that race-blind affirmative action policies such as Texas's may select students for admission based on traits correlated more with race than with future achievement and may reduce the impact of effort on students' probabilities of admission. They argue instead that race-sighted affirmative action is the most effective admission policy conditional on a targeted level of diversity. Fryer and Loury (2005) argues several additional points in an attempt to dispel the "mythology" surrounding affirmative action. Among these are the myths that affirmative action necessarily undercuts investment incentives, is best deployed as early as possible, and always helps its beneficiaries. Taken together, these conclusions have several implications for our work. It is possible that percentage plans such as Texas's may encourage or discourage student effort. It is also possible that the timing of Top Ten eligibility may affect these incentives and the timing of student effort. If the Top Ten Percent law does discourage effort, then it may be worthwhile to examine policy options that preserve the goals of the law while minimizing unintended consequences.

This paper contributes to the literature in several ways. In particular, while most other studies focus on outcomes such as college choice or class composition and a few examine whether students work harder to achieve Top Ten Percent eligibility, we choose to focus on the remainder of students' high school careers. There are several reasons why this time period matters. First, some of these students, despite offers of admission, may not be able to attend college due to budget constraints, academic unpreparedness, or other factors. Students who underachieve in their senior years and do not attend college may find their employment prospects or earnings harmed. Second, students who underachieve in their senior years are more likely to be underprepared for college and may do worse in college as a result, while those who overachieve might outperform prior expectations. It is also possible that underachieving in college may prompt some students to drop out, facing poorer career options (and sizeable loans).

IV. Regression Discontinuity

In our analysis, we use an RD framework to measure the impact of Top Ten Percent status on several outcomes. In particular, following DMM, we estimate

$$Y_i = \theta_i * \mathrm{TTP}_i + f(CR_i) + X_i\beta + \epsilon_i$$

where Y_i is one of several outcomes proxying for effort in a student's senior year, TTP_i is an indicator of Top Ten Percent status, CR_i is student i's class rank in percentage terms (the "running variable"), $f(\cdot)$ is a flexible function mapping class rank to various outcomes, X_i is a vector of observable student characteristics,[43] and ϵ_i is a mean-zero, normally-distributed error term. In this setup, $E[\theta_i]$, the coefficient of interest, measures the average treatment effect associated with Top Ten Percent status.

This effect is composed of two principal factors – the effect on students who barely qualify for guaranteed admission and the effect on students who narrowly miss the cutoff for guaranteed admission. It is possible, for instance, that students who qualify for automatic admission do not lower their effort level at all and that any "negative" effects we find are due to increased performance among non-qualifiers who work harder both in an attempt to qualify for guaranteed admission during their senior year and to strengthen their college admissions profile if they are unable to qualify. To better interpret our results, we examine both the extent to which performance varies among students and how students' performance in their senior year compares to their performance in previous years.

Several assumptions must hold for $E[\theta_i]$ to represent a true ATE. These are that 1) student characteristics, especially the variables in X_i, must trend smoothly through the cutoff for Top Ten Percent status; 2) there are no simultaneous or confounding treatments; and 3) students near the cutoff must be randomly assigned to treatment. Imbens and Lemieux (2008) and Lee (2008) formalize the first two assumptions, while McCrary (2008) does the same for the third.

[43] In the regression results presented below, X_i consists of dummies for race, sex, whether the student was ever eligible for free or reduced-price lunch, ever in special education classes, or ever classified as possessing limited English proficiency. There are also year and school of graduation fixed effects. Coefficients and standard errors on these variables are not listed, but are available upon request.

Evidence on the first assumption is shown in Table 1. These results come from local linear regressions of various student characteristics on class rank, an indicator for Top Ten Percent status, and an interaction term allowing for different slopes on either side of the Top Ten Percent cutoff. None of the coefficients on the Top Ten Percent indicator is statistically significant, suggesting that these traits do not vary discontinuously at the cutoff. We can therefore be fairly confident that our results are not due to dramatically different student characteristics around the cutoff for guaranteed admission.

As for the second assumption, while some students may have their own incentives for maintaining or graduating with a high class rank, it is hard to imagine a more powerful incentive than admission to UT or A&M. Even if there is one, it would be most likely to improve outcomes within the Top Ten Percent; this might weaken results from positive discontinuities, but would strengthen those from negative discontinuities, as students would have to maintain Top Ten Percent status for longer.[44] The issue of nonrandom assignment is dealt with in the next section. For now, suffice it to say that while assignment may not be entirely random, it may be "as good as" random.

Rather than imposing a functional form on the data, we run RD estimation using local linear analysis. We use bandwidths of ten class rank percentiles, five class rank percentiles, and Imbens-Kalyanaraman ("IK") bandwidths (which vary by outcome and by subgroup). Each of these has its own advantages and drawbacks. At one

[44]College scholarships that account for class rank at time of enrollment might cause a similar effect.

extreme, estimates using bandwidths of ten percentiles tend to be less affected by noise within any particular percentile bin but may be more affected by nonlinearities in the data. At the other extreme, IK bandwidths are typically extremely small; therefore, estimates using these bandwidths provide accurate linear approximations but are extremely susceptible to noise in observations near the cutoff for guaranteed admission. We therefore prefer bandwidths of ten class rank percentiles as our primary specifications, using smaller bandwidths to confirm our intuition and determine the impact of nonlinearities in our data.

In running the RD estimation, we examine several outcomes, all of which may proxy for effort. First, we examine whether students in the Top Ten Percent are more likely to graduate with at least a Recommended diploma or with a Distinguished diploma.[45] We also examine whether students take more classes or more courses worth five grade points in the neighborhood of the Top Ten Percent cutoff and whether these values constitute positive or negative changes from the previous year.[46] We also examine whether students who are barely eligible for guaranteed admission have higher senior year GPAs than those who are barely ineligible and how these

[45]The district has three levels of diplomas: Minimum, Recommended, and Distinguished. Students graduating with a Minimum diploma must complete 22 credits; 4 in English Language Arts, 3 in Mathematics, 2.5 in Social Studies, 0.5 in Economics, 2 in Science, 1 in an elective consisting of World History, World Geography, or Approved Science, 1 in Physical Education, 0.5 in Health Education, 0.5 in Speech, 1 in Fine Arts, and 6 in various electives. Students obtaining a Recommended diploma must take one additional credit each in Mathematics and Social Studies and two more each in Science and World Languages, but are exempted entirely from the former elective requirement and from one credit of the latter. Students completing a Distinguished diploma must complete one additional World Language credit above the Recommended standards, but are allowed to take one less elective credit; these students must also complete their choice of four advanced-level courses specified in state guidelines.

[46]Most courses are worth four grades points; those worth five are generally at the honors or Advanced Placement level.

GPAs compare to their cumulative GPA at the end of 11th grade. As a Recommended or Distinguished diploma is necessary to qualify for Top Ten Percent eligibility, the likelihood of obtaining one of these degrees should at least weakly rise. It is unclear, however, how any other outcomes would be affected.[47]

Analysis was run on subsamples according to race, gender, economic background, school college-sending patterns, and likelihood of attending college based on characteristics other than class rank. Intuitively, this can be expected to yield more accurate estimates for individual subgroups; different groups of students may have offsetting reactions to Top Ten Percent status, or a single group's reaction could be drowned by other groups' unresponsiveness. Our prior was that minorities and students from disadvantaged backgrounds or underachieving schools would on average have more to gain from putting forth higher effort and increasing their college readiness, while well-to-do students attending successful school – those more likely to treat UT and A&M as "safety" schools – would be less likely to put forth high effort in their senior years.

The chief drawback of segmenting the data is the reduction in sample size. Even using a large data set, running local estimates on very specific subsamples may result in a sample size too small (and standard errors too large) for valid inference. It is also possible that some students attended lower-quality high schools in hopes of achieving Top Ten Percent eligibility, according to findings by Cullen, Long, and

[47] As it is easy to picture some students working harder and others shirking, a heterogeneous effects framework yielding an average treatment effect rather than a single treatment effect is certainly appropriate here.

Reback (2011). If true, it is unclear what effect this would have on RD estimates, but it would almost certainly increase standard errors.

V. Data

We run our estimation using transcript data from a large, urban, majority-minority school district in Texas. This district (hereinafter "the district") is relatively poor, with a large number of students eligible for free or reduced-price lunch. About half of the district's students are Hispanic, about a quarter each are African-American or White, and the remainder are of other races.

Data were initially listed at the course level, with multiple observations per student. Each school year was a separate data file. Within each year, variables indicated the semester, the school at which each course was taken, the course name, and the course number. Every course taken at every public school in the district between the 1999-2000 and 2010-2011 school years was a separate observation. Sixteen of these schools were standard-curriculum public high schools of various racial makeups and qualities. The remainder fit into several broad categories: "alternative" schools, special education programs, vocational and technical schools, centers for at-risk youth, and middle schools.[48]. Notably, the data did not contain information on class rank,

[48] We assume that courses listed at middle schools were generally of a remedial quality or geared towards students with learning disabilities, but the data do not indicate for certain that this is the case. Fortunately, these account for a relatively small number of observations and therefore can be safely dropped.

though they did contain data on demographics, class, degrees awarded, and course grades on a 0-100 scale.

It took several steps to put the data into a form from which class rank could be recovered. First, several courses were listed as being "Local Credit Only." These included most "athletic" courses, several courses offered only at specific high schools, and several highly advanced language courses. These courses did not contribute to students' GPAs, and were therefore dropped from the data. Courses taken at middle schools or at the middle school level were also dropped; while this was not mentioned in the district guidelines, it is logical that high schools would not count courses taken below the high school level. Together, Local Credit Only and middle school courses comprise between five and ten percent of courses taken in most school years.

Next, grades were converted into a uniform grade point scale. Each course could have several grade entries – one corresponding to a semester grade, which could be split into up to three component "cycle" grades. First, any courses with uninterpretable non-numeric grades were dropped.[49] Where necessary, cycle grades were averaged to obtain a semester grade. These grades were then converted into a five-point grade point scale according to district guidelines.[50] At this point it was possible to compute semester and cumulative GPAs.

[49]While semester grades of "79." and "I00" could be easily be inferred to read "79" and "100" respectively, semester grades such as "W" or "*" would be dropped. These corrections apply to several hundred course-level observations per year, compared with a full sample of over 200,000 courses taken per year.

[50]100-90 was given an A grade, 89-80 a B, 79-70 a C, and everything else an F. A's were worth four points (five if the course was a five-point course); B's were worth three (four); C's were worth two (three); F's were worth zero regardless of the course.

With each student assigned a cumulative GPA for each semester, it was possible to rank students' GPAs by year, semester, high school, and grade level. Ties were broken using randomly assigned student ID numbers. We then merged in data on demographics, degrees, and college outcomes. Some students who could not be merged these individuals were dropped from further analysis. As students had separate observations for demographic data in each semester in which they appeared, we assigned some students values for race and gender.[51] When running separate analyses by racial background, we grouped students into group of "White and Asian" and "Non-Asian Minority." This allowed us to both run valid inferences by racial subgroups and account for the relatively small number of students who could not be classified as White, African-American, or Hispanic. Students who had ever been eligible for free or reduced-price lunch were labeled as "economically disadvantaged."

In performing our analysis we omitted students who did not have valid observations for gender, ethnicity, economic status, or limited English proficiency, who had not graduated, had graduated with Special Education degrees, or who had missing data on Special Education status. We omitted students who did not attend a standard-curriculum high school, as class rank at other institutions might not be as meaningful due to altered curricula, selection on classmates, or smaller class sizes.

[51] Given the slippery definitions of racial groups, some students were listed as belonging to different groups in different semesters. If one ethnicity was listed in a majority of a student's observations, we assigned her to that ethnicity. If no ethnicity appeared in a majority of observations, we declared students Hispanic if any observations stated that they were Hispanic. Students who were listed as White and as one other ethnicity (presumably with some observations missing) were assigned the other ethnicity. Students with multiple non-White and non-Hispanic ethnicities listed were labeled as being two or more races. Students with multiple genders listed were assigned the modal gender; the 70 students with an equal number of male and female observations were dropped from our analysis.

We omitted those students who were never listed as being in 11th grade, as the running variable could not be defined for these students.[52] Finally, we omitted students who were listed as taking no courses or having no GPA in their senior years. The modified data thus consists of 20,000 student-level observations across 16 high schools over nine years; restricting further to observations within 10 class rank percentiles of the cutoff for guaranteed admission gives us a final sample of 4,196 students.

Given the large amount of data and cleaning, some measurement error is unavoidable. Some of this could be due to data entry or keystroke error. It is also possible that some students may be assigned incorrect class ranks. Finally (and most worryingly), teachers may have assigned grades in ways that systematically affect their students' GPAs or fudge the definition of "Top Ten Percent" (labeling a student with a class rank of 10.6%, for instance, as being in the Top Ten Percent). If any of these sources of measurement error is serious enough, it could invalidate the results of the RD setup.

Fortunately, this is unlikely to be the case. The largest threat to an RD framework is if measurement error is systematically present around the cutoff. Measurement error attributable to data entry or to coding should not be systematic in this way. If anything, these sources of measurement error would bias RD estimates towards zero, strengthening most of our results. Measurement error in grading, though, may be designed to put specific students in the Top Ten Percent, which would be a distinct

[52]It is unclear whether these students transferred into the school district, skipped a grade, were victims of a clerical error, or were dropped from the sample, but the end result each case is the same.

threat to RD analysis, according to McCrary (2008). It is unclear, though, whether this threat would materialize. Certainly renegade teachers can affect the cardinal distribution of GPAs, but in equilibrium the rank-ordering around the cutoff may not be significantly altered,[53] so according to Lee (2008) the assignment to treatment may be "as good as" random. Similarly, while fudging the precise definition of "Top Ten Percent" may result in systematically spurious assignment to treatment, this will if anything bias results towards zero rather than towards a positive or negative outcome.

A more likely scenario is that teachers feel pressured or obligated to give the benefit of the doubt to "good" students on the cusp between grades. This may not threaten our results, though. If such pressure occurs uniformly among these students, class ranks (and the thus makeup of the Top Ten Percent) will be largely unaffected. Even if teachers are able to select which marginal "good" students are in the Top Ten Percent, it should be possible to guess the impact of such selection on RD estimates. In particular, given a choice between two students with roughly equal grades, it seems likely that a teacher would choose the more motivated and harder-working one for Top Ten Percent status, as that student would be more likely to succeed in future academic settings. As a result, any selection into Top Ten Percent status based on relevant unobservables would bias RD estimates toward positive (rather

[53] For the rank ordering to be affected, teachers would need to a) know which students are near the cutoff, b) be able to credibly affect those students' course grades, and c) either make certain that a single course grade can affect the rank ordering of GPAs or coordinate with other teachers to affect multiple grades. Of course, this assumes that there is no disagreement among teachers about how to influence the rank ordering of GPAs. As this is unlikely to be the case, and given the vast number of moving parts (to say nothing of ethical issues) involved, it appears unlikely that teachers will be able to have large and systematic effects on class rankings.

than negative or zero) outcomes. If this is the case, positive RD estimates might be threatened, but negative RD estimates would be strengthened. The net effect of all forms of measurement error would therefore be ambiguous for positive estimates, but would bias negative estimates towards zero. As a result, while the barely Top Ten eligible students in our sample may be disproportionately motivated, we are confident that our results will not falsely imply that moral hazard is taking place.

VI. Results

We find some evidence that students who are barely eligible for guaranteed admission exert lower effort in their senior year than their ineligible peers. This difference is attributable more to a drop in their effort level rather than to an increase in effort among ineligible students.

We begin by examining students' college-going patterns. As the primary goal of the Top Ten Percent law is to achieve diversity at public colleges and universities, we should expect students to modify their college enrollment habits accordingly. Conversely, if the Top Ten Percent law does not have a direct effect on college-going patterns it will be difficult to argue that it has an indirect effect on student effort. We show that while the Top Ten Percent law does not significantly affect college access it does affect college choice, primarily by shifting students into UT and A&M.

In studying how students react to their Top Ten Percent status (or lack thereof) over their senior year, we examine both twelfth-grade outcomes and how those twelfth-grade outcomes differ from eleventh-grade outcomes. Running regression discontinuity analysis on twelfth-grade outcomes illustrates how eligible students' outcomes differ from those of ineligible students; it does not say why these outcomes differ. If, for instance, we were to find that eligible twelfth-graders got substantially lower grades than ineligible twelfth-graders did, additional analysis would still be necessary to determine whether this constitutes shirking on the part of eligible students or extra effort on the part of ineligible students. Examining the difference in twelfth-grade outcomes from eleventh-grade outcomes allows us to more conclusively state which of these is likely to be the case.

Students in the Top Ten Percent may react to their status along several margins. First, they may target a different type of degree. Students on track for Minimum diplomas may focus on earning Recommended diplomas in order to take advantage of their Top Ten Percent status, those on track for Recommended diplomas may increase their effort in order to be more competitive college students, and those on track for Distinguished diplomas may reduce their effort due to the lower-stakes nature of their college admissions process.

Next, Top Ten-eligible students may take a different mix of courses. Some may take alter the number of courses that they take. Top Ten-eligible students may take more courses or fewer courses than comparable ineligible students. Eligible students may take additional courses if they wish to become more competitive at the college level or may take fewer courses if they feel that additional effort has been disincentivized. Ineligible students may also react to the TTP cutoff–in particular, some may take fewer courses either in an effort to improve their grades or out of discouragement. We therefore examine both the number of courses taken in 12th grade and the change in this number from the previous year.

Students may also alter the type of courses that they take. In particular, students may choose between five-credit AP or honors courses and four-credit standard courses.[54] Top Ten-eligible students may take more five-credit courses if their taste for risk increases as a result of guaranteed admission or may take fewer if they prefer to consume leisure. Ineligible students may take more five-credit courses if they believe that such courses will improve their GPA or believe that such courses make them more attractive applicants in the standard college admissions procedure. They may take fewer five-credit courses if they are especially risk-averse and prefer a moderate GPA with certainty over risking a low GPA.[55] We examine both the total number of credits that these students take in their final year and the change in this number from the previous year.

Finally, conditional on taking the same set of courses, Top Ten-eligible students may exert different effort levels than their ineligible peers within these courses. On the one hand, they may exert greater effort in order to improve their college readiness; on the other, they may begin to shirk if their grades do not affect their college admissions prospects. Similarly, ineligible students may exert greater effort in an attempt to become eligible prior to graduation or may become discouraged and exert

[54]We treat single-semester honors courses as being worth 2.5 credits and standard courses as two credits.

[55]As any grade a standard course is treated the same as the next grade down an honors or AP course, students opting to take standard courses due to risk aversion must believe that there is a chance that they would do particularly poorly in the corresponding honors course.

less effort than previously. We examine students' first semester, second semester and full-year grades over the course of 12th grade. We do so because the effects of guaranteed admission may vary over time. For instance, the relative impact of automatic admission offers on students' grades may be larger in the second semester as course material builds on itself or may be larger in the first semester while ineligible students have strong incentives to study their way to eligibility. To determine whether any discontinuities are more attributable to eligible or to ineligible students, we also examine the difference between students' cumulative GPAs at the end of their junior year and their senior-year GPAs.

a. College Outcomes

We present estimates of college outcomes in Table 2 and Figure 1. Top Ten eligibility does not appear to affect college access. Students are no more likely to attend any college, to attend a four-year college, to attend a selective college, or to attend a selective college. Surprisingly, Top Ten eligibility does not make students any less likely to attend an out-of-state university.

Top Ten eligibility does affect college choice along two important margins. Students who are eligible for guaranteed admission enroll in private universities as significantly lower rates, and are over five percentage points – over 50 percent – more likely to attend UT or A&M. This finding is corroborated both by previous literature (Niu and Tienda and Long and Tienda, e.g.) and by a 2009 law exempting UT from

Top Ten Percent requirements in filling the final 25 percent of their in-state entering class.[56] Readers interested in learning more about the Top Ten Percent law's effect on college choice are advised to read Daugherty, Martorell, and McFarlin (2011), which provides a fuller analysis of this topic by student subgroups.

b. Recommended and Distinguished Diplomas

Top Ten Percent status has the clearest implications for the odds of acquiring (at least) a Recommended diploma, as students who graduate with a Minimum diploma are ineligible for automatic admission to college, regardless of their class rank.[57] This is balanced, however, by the fact that many eligible students are on track for a Recommended diploma anyway. Among the full sample of students, over 95 percent obtain a Recommended or Distinguished diploma near the cutoff for Top Ten eligibility. Even among economically disadvantaged students, nearly 95 percent of those who are barely ineligible for guaranteed admission obtain one of these two diplomas. As a result, any increases in the probability of obtaining such a degree will be small – there may even be spurious negative estimates of $E\left[\theta_i\right]$ if there is any noise in the relatively small treated group.

[56] UT is required to admit the top one percent of high school graduates, then the top two percent, and so on until 75 percent of its in-state quota is filled. It may then use its discretion in filling the remainder. As this law did not go into effect until 2011, its passage does not affect our results.

[57] Top Ten Percent students who do graduate with a Minimum diploma are still eligible for admission to the university of their choice if they score over 1500 out of 2400 on the SAT or satisfy the ACT's College Readiness Benchmarks. That said, the set of students who simultaneously qualify for Top Ten Percent status, choose to obtain a Minimum diploma anyway, score suitably on either exam, and then apply to college cannot be a large one.

Table 3 contains the results of regression discontinuity analysis using the jump in the probability of obtaining at least a Recommended diploma as the outcome of interest. Columns (1) and (2) have bandwidths of five class rank percentiles, columns (3) and (4) have bandwidths of ten class rank percentiles, and column (5) uses an IK optimal bandwidth. Coefficients are generally positive and nearly always statistically insignificant. Graphs for the full sample and for two subgroups are presented in Figure 2. In the full-sample graph, there is little evidence of a discontinuity. It is worth noting, however, that students in the 11th percentile obtain Recommended or Distinguished diplomas at a rate distinctly below trend. This is even more visible in the graph for economically disadvantaged students, and is a likely contributor to the one statistically significant coefficient in Table 3. It is unclear exactly why this is the case; one possibility is that some fraction of students who are ineligible become discouraged and either drop out or get Minimum-level diplomas. There is some graphical evidence of a positive discontinuity among White and Asian Students, but the point estimate of two percentage points is not statistically significant.

While there is no clear discontinuity in the probability that students graduate with at least a Recommended-level diploma, Table 4 shows a statistically significant drop in the probability that students get a Distinguished-level diploma using a bandwidth of ten class rank percentiles. This drop is stronger when focusing on several particular subgroups, such as female students and students with a high probability of enrolling in college. Graphical evidence, however, suggests that these discontinuities may be due to nonlinearities in students' probabilities of graduating with Distinguished diplomas. The top graph in Figure 3 illustrates this – as students advance into more competitive class ranks, the probability that they graduate with a Distinguished diploma begins to rise more sharply. Accordingly, linear predicted values with a bandwidth of ten class rank percentiles have a much steeper slope than those with a bandwidth of five class rank percentiles. Using the smaller bandwidth shrinks our estimates by over an order of magnitude, making them statistically insignificant in the process. Perhaps the most convincing case of Top Ten eligibility overinsuring students appears among White and Asian students, who are over eight percentage points less likely to obtain a Distinguished diploma when using a ten-point bandwidth. Using a five-point bandwidth shrinks these estimates as well, but there is less evidence of nonlinearities among Top-Ten eligible students – instead, it appears that much of the difference is due to noise among ineligible students.

c. Courses Taken

Course-taking patterns suggest that while students do not significantly alter the number of courses they take, they do take fewer honors or AP classes. Table 5 lists the impact of Top Ten eligibility on the number of courses taken in 12th grade. Point estimates using a 10-percentile bandwidth are uniformly negative but insignificant. Figure 4 shows that there is not much variation in course-taking – even if our estimates were significant, they are so small as to render any analysis almost meaningless. Estimates using an IK optimal bandwidth are somewhat larger and positive, but appear to reflect noise in our sample rather than trends in the data.

It is possible that Top Ten-eligible and ineligible students could take the same number of courses in their senior year because one of the two groups fundamentally altered their course-taking patterns. However, both Table 6 and Figure 5 suggest otherwise. Estimates using bandwidths of ten percentiles and five percentiles are statistically insignificant. While estimates using an IK optimal bandwidth are statistically significant in several cases, this does not appear to reflect actual trends in the data. Notably, students who are just barely ineligible for guaranteed admission take fewer classes in their senior year than in the previous year. There are three possible explanations for this. First, students who barely miss the cutoff for guaranteed admission may take fewer classes due to discouragement. Second, they may take fewer classes in an effort to improve their GPA – if there is a quality-quantity tradeoff in course-taking, then students who are able to focus intensely on a small number of courses may expect to do better in those courses than if they had a heavier course load. Finally, students who barely missed the cutoff for guaranteed admission may be more likely to have taken a suboptimally high number of courses during their junior year. These students may therefore rationally adjust their course-taking downwards.

Students who wish to take a set number of courses have additional margins along which they may adjust their effort levels. In particular, students may take additional five-credit honors or AP courses or may take less challenging four-credit courses. We therefore examine the number of credits that students take to account for differences in course difficulty. This metric does not reveal systematic discontinuities either––though the point estimates in Table 7 are almost frequently negative, they are rarely statistically significant. Figure 6 similarly shows no clear evidence of major discontinuities. Even among the two group with statistically significant indicators for Top Ten Percent status–non-Asian minority students and students whose high schools send few students to college–it appears that the statistical significance in Table 7 may be an artifact of nonlinearities.

Point estimates for the difference in credits taken between 11th grade and 12th grade are contained in Table 8. Point estimates are generally positive but statistically insignificant. Results are statistically significant only when using Imbens-Kalyanaraman bandwidths, which tend to be narrower than the other two bandwidths that we use. Closer inspection of the data shows that much of this effect may be due unexpected drop-offs in course-taking among students who barely miss eligibility for automatic admission. This provides additional evidence that students who narrowly fail to qualify for automatic admission may be particularly discouraged from future academic effort.

d. Course Grades

Table 9 shows that students who are barely eligible for guaranteed admission do only slightly worse in their senior year than those who are barely ineligible. This result holds across nearly all subgroups when using a ten percentile bandwidth – the lone exception being students who are highly likely to attend college anyway. Results become statistically insignificant when the bandwidth is reduced to five percentiles, though point estimates fall farther in some cases than in others. Figure 8 shows graphs for the overall sample and for three of the groups most affected by Top Ten eligibility. In each of the subgroups there is a clear break in the steady upward trend as students rise into higher class rank percentiles. However, even if the differences are statistically significant, they are fairly small. The largest discontinuity in Table 7 – a difference of close to .15 grade points among students attending low-sending schools – does not reflect a huge drop in GPA. For context, a student who takes six courses (somewhat above par for students near the cutoff) and gets five A's and one B would do .17 grade points worse than a student who gets six A's in the same set of courses. It is worth noting, however, that guaranteed admission almost certainly affects individuals heterogeneously, and that this effect may be substantially stronger in the five percent of individuals around the cutoff who are induced to attend UT or A&M.

This drop in GPA is larger in the first semester of 12th grade than in the second, as shown in Tables 10 and 11. Point estimates in Table 10 are approximately .01 grade points larger in magnitude than those in Table 9 and in some cases are more than .02 grade points larger. Point estimates in Table 11 are smaller and less likely to be statistically significant, though they do still reflect lower performance among eligible student than among ineligible students at the cutoff. This image is consistent with students' slightly reducing their effort levels (either by getting lower grades or taking fewer honors or AP courses) as soon as they are admitted to their preferred university. Students who are admitted under the Top Ten Percent law are able to shirk earlier than those admitted under the normal timeline. Students who are admitted during the normal admissions process are often unable to shirk until the second semester of 12th grade, which explains why the GPA gap is smaller in the second semester than in the first.

To verify that the difference in GPAs at the cutoff reflects lower effort among Top Ten-eligible students rather than increased effort among ineligible students, we take the difference in students' 12th grade GPAs and their cumulative GPA at the end of 11th grade and run regression discontinuity estimation on this analysis. If guarantees of admission cause shirking, we should observe that students who are eligible for guaranteed admission have lower GPAs in 12th grade than in previous semesters. Similarly, if Top Ten ineligibility causes students to work harder in an effort to become eligible, we should observe that such students have higher GPAs in 12th grade than in previous semesters.

Evidence on the nature of GPA gaps is mixed. While the point estimates in Table 12 are mostly statistically insignificant, graphical analysis will be more useful for determining whether GPA gaps are due to higher effort among ineligible students or lower effort among eligible students. Several of these graphs are contained in Figure 9. Among the full sample of students there is some evidence for both of these. In several cases, however, there may be clearer evidence that students who qualify for automatic admission see their grades fall rather than the reverse. Ineligible non-Asian minority students and students with low probabilities of attending college perform similarly in 12th grade to previous years, while corresponding eligible students have lower GPAs in 12th grade. In schools that send few students to college, most students near the cutoff for guaranteed admission do worse in 12th grade than in previous years, but this decrease is more pronounced among eligible students.

VII. Conclusion

While students do not react uniformly to the Top Ten Percent Law, there is evidence that they do respond to academic incentives for effort. Top Ten eligibility may induce White and Asian students to graduate with Recommended diplomas at greater rates, reducing both the number who graduate with Minimum diplomas and the number who graduate with Distinguished diplomas. These students act as though they are overinsured – they do enough to meet the guidelines for admission, but since they are protected against poor results in their college search, they are free to shirk with little chance of being punished.

Students respond along other margins as well. While students as a whole do not alter the number of courses that they take, there is evidence that a small group of students who do not qualify for guaranteed admission may do so, possibly out of discouragement or as a way to capitalize on a quantity-quality tradeoff in course-taking. We find similar results when examining the number of credits that students take. Students' grades provide some evidence for both of these possibilities–non-qualifiers get slightly better grades in 12th grade than in previous years, while qualifiers get slightly worse grades.

The challenge that colleges and high schools face is therefore to combat overinsurance and moral hazard by maintaining programs that expand college access while limiting the negative incentives from these programs. One way of doing this might be to offer only partial insurance. States could do this by adopting percent plans similar to California's and Florida's. Presenting students with a series of cutoffs and increasingly valuable prizes rather than providing access to flagship universities at a single cutoff might keep students better engaged for longer. Students would be motivated to work harder in part to gain automatic access to more prestigious universities and in part because their worst-case scenario in the college application process would no longer be quite as rosy. Even this, however, does not address the fact that students, once admitted to the school of their choice, have little incentive to keep working.

A more practical solution might involve additional monitoring, changing automatic admissions programs to conditional admissions programs. For instance, colleges could request to view students' GPAs midway through their senior year. Students who underperform or do not take a sufficiently challenging set of courses would have admissions offers at certain schools revoked. While this would require additional work on the part of colleges, it would affect all admitted students, regardless of their intended destination.

This paper may understate the impact of Top Ten Percent eligibility. In part, this is due to measurement error of various sorts; while unfortunate, it is unlikely that other sources of transcript data would be qualitatively better. Similarly, while some students may be incorrectly assigned high school class ranks, there is no evidence that this is systematic in either direction around the cutoff, and therefore does not drastically affect our results. The results presented above should therefore be taken as lower bounds of the true effects of Top Ten Percent eligibility.

One possible extension of this work involves using data from other school districts. While the district analyzed here is useful for analyzing the responses of minorities, low-income students, and those who might not otherwise have planned on attending college, these groups are also the ones least likely to think strategically about college admissions. These students may therefore react less dramatically than those from wealthier or higher-achieving school districts. Analyzing the responses of other students may thus provide not only a clearer picture of strategic thinking, but also a better idea of the relative impacts of the treatment of college admission versus the treatment of a lower-stakes admissions process.

Figure 2.1: Enrollment

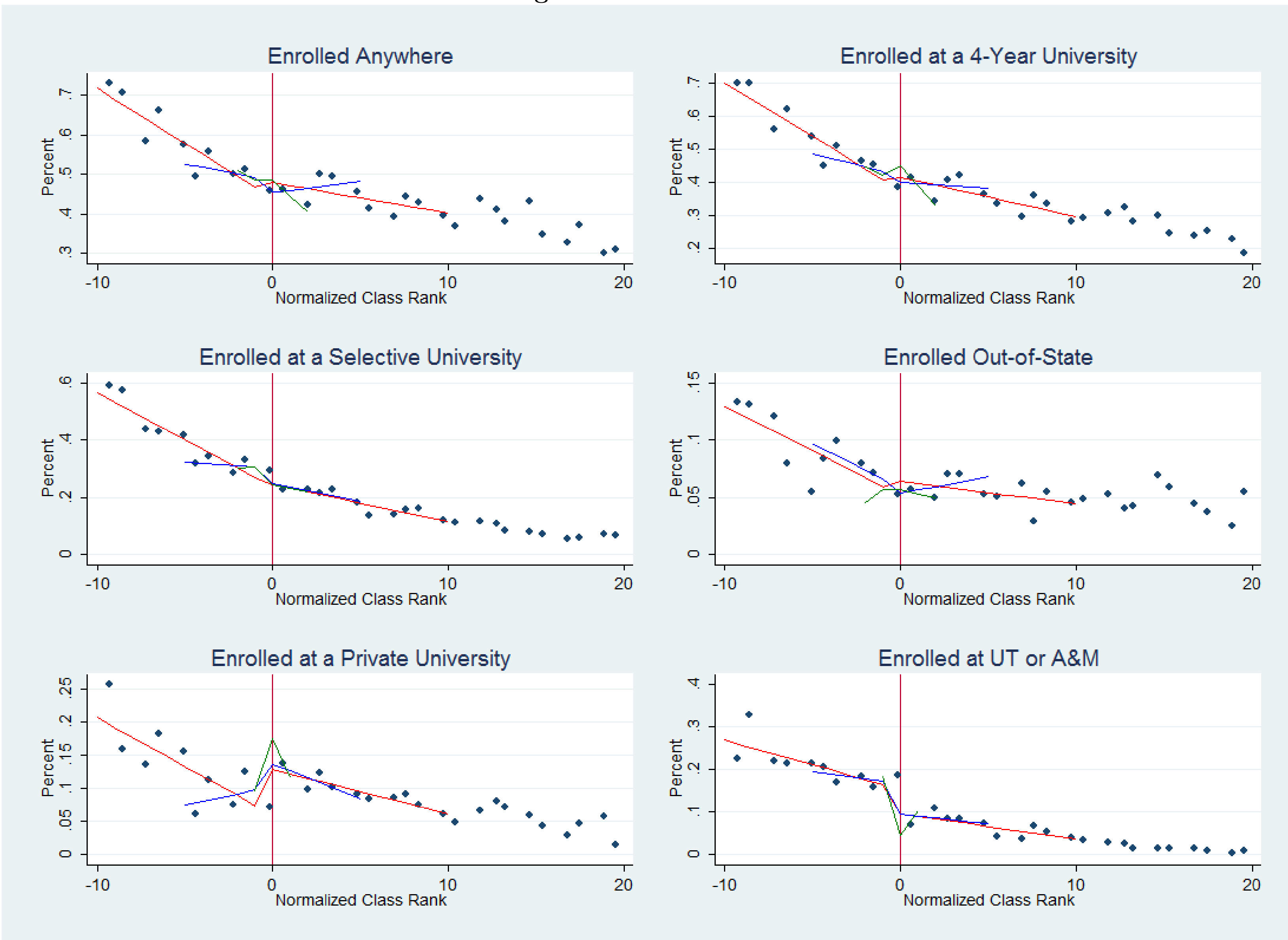


Figure 2.2: Recommended or Distinguished Diplomas

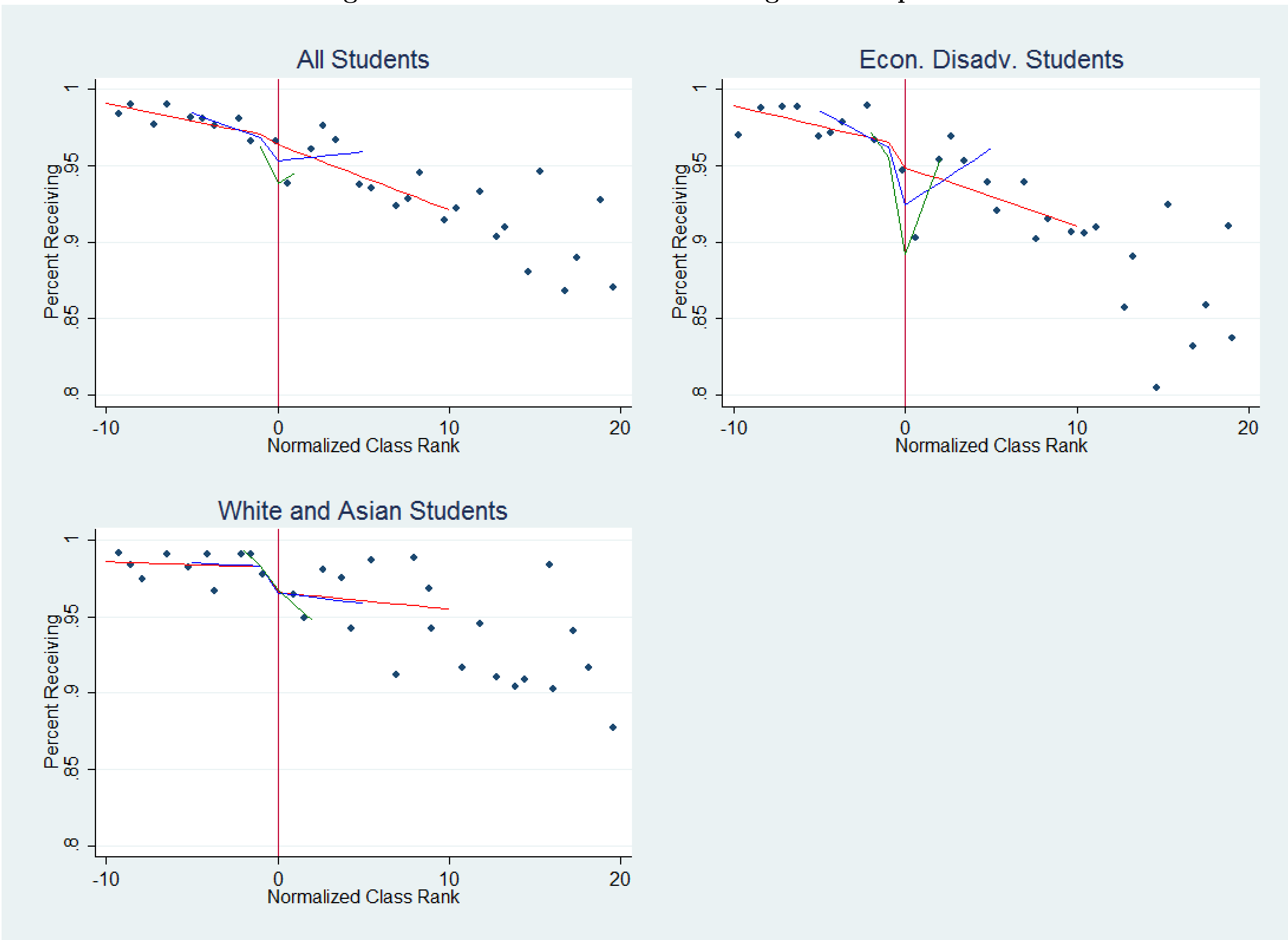


Figure 2.3: Pr(Distinguished Diplomas)

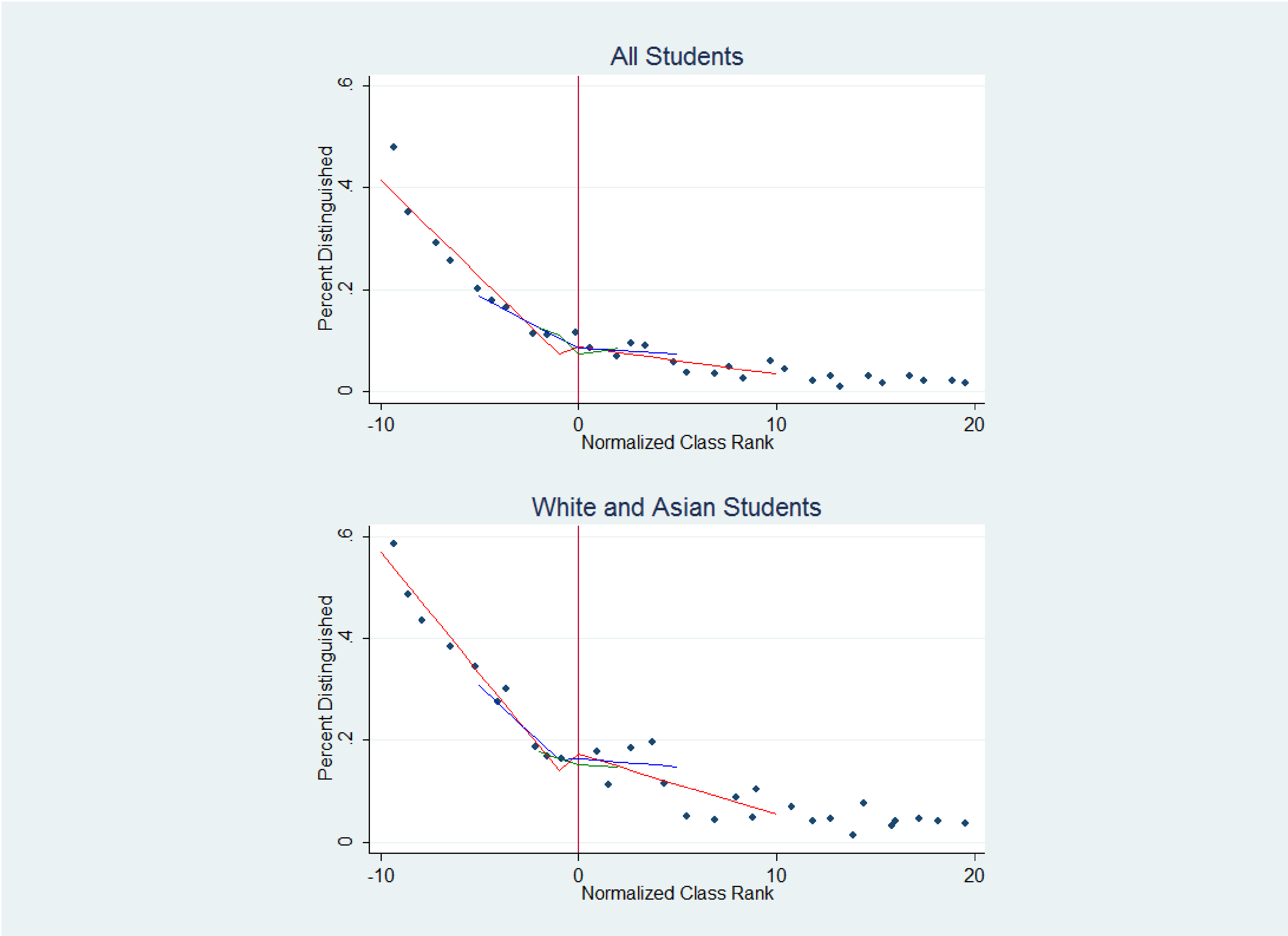


Figure 2.4: Classes Taken

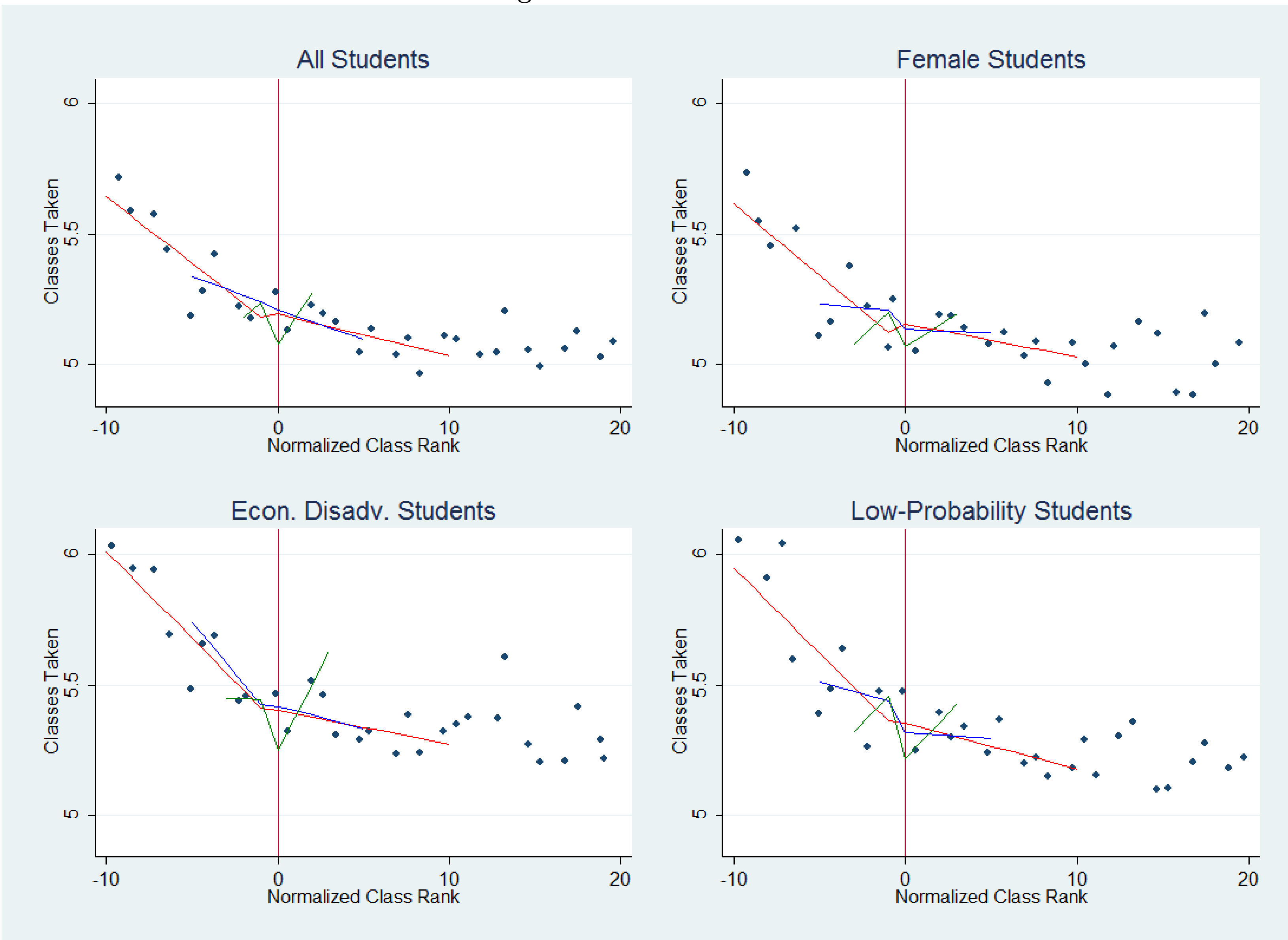


Figure 2.5: Difference in Classes Taken

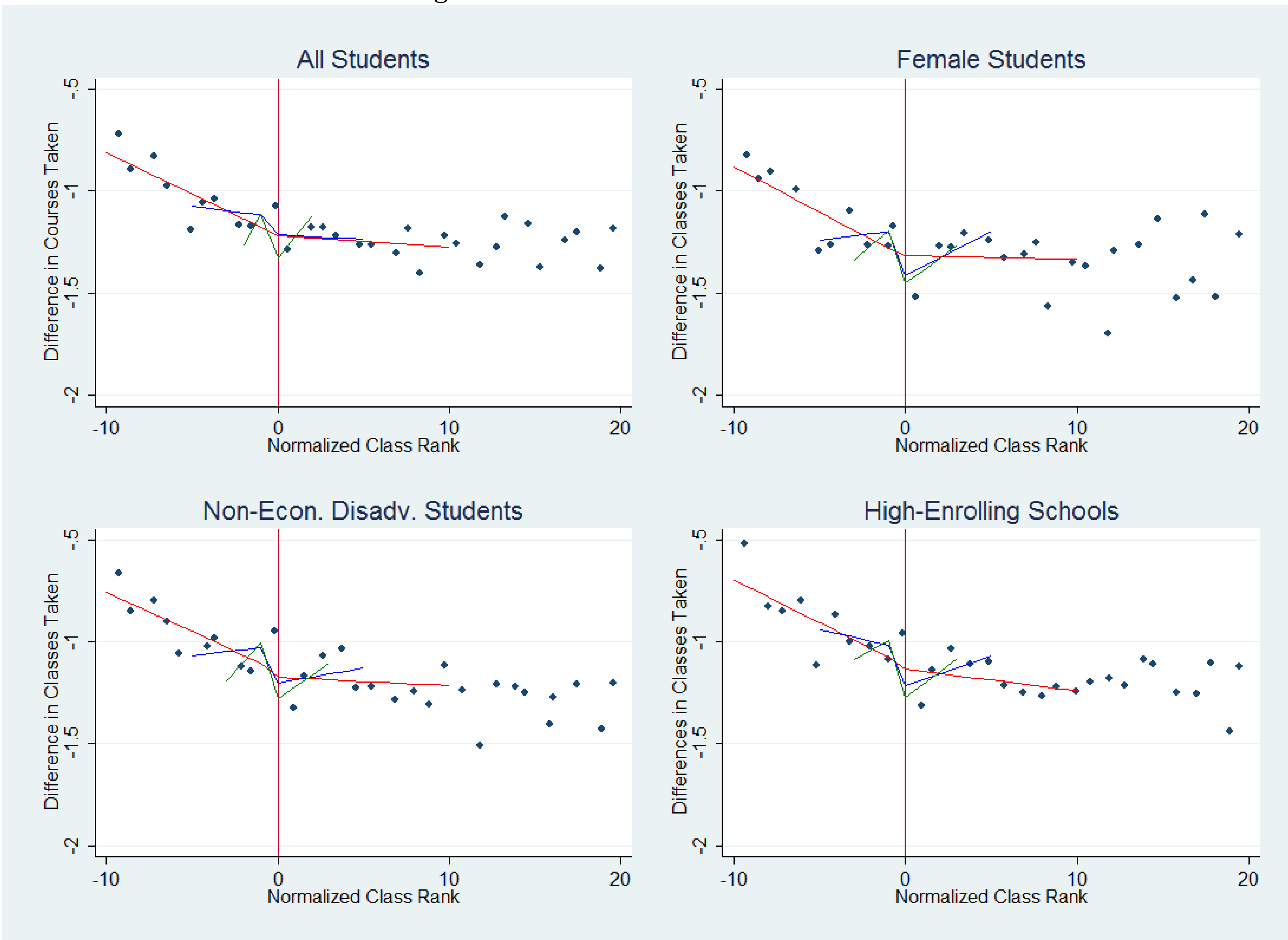


Figure 2.6: Honors/AP Credits Taken

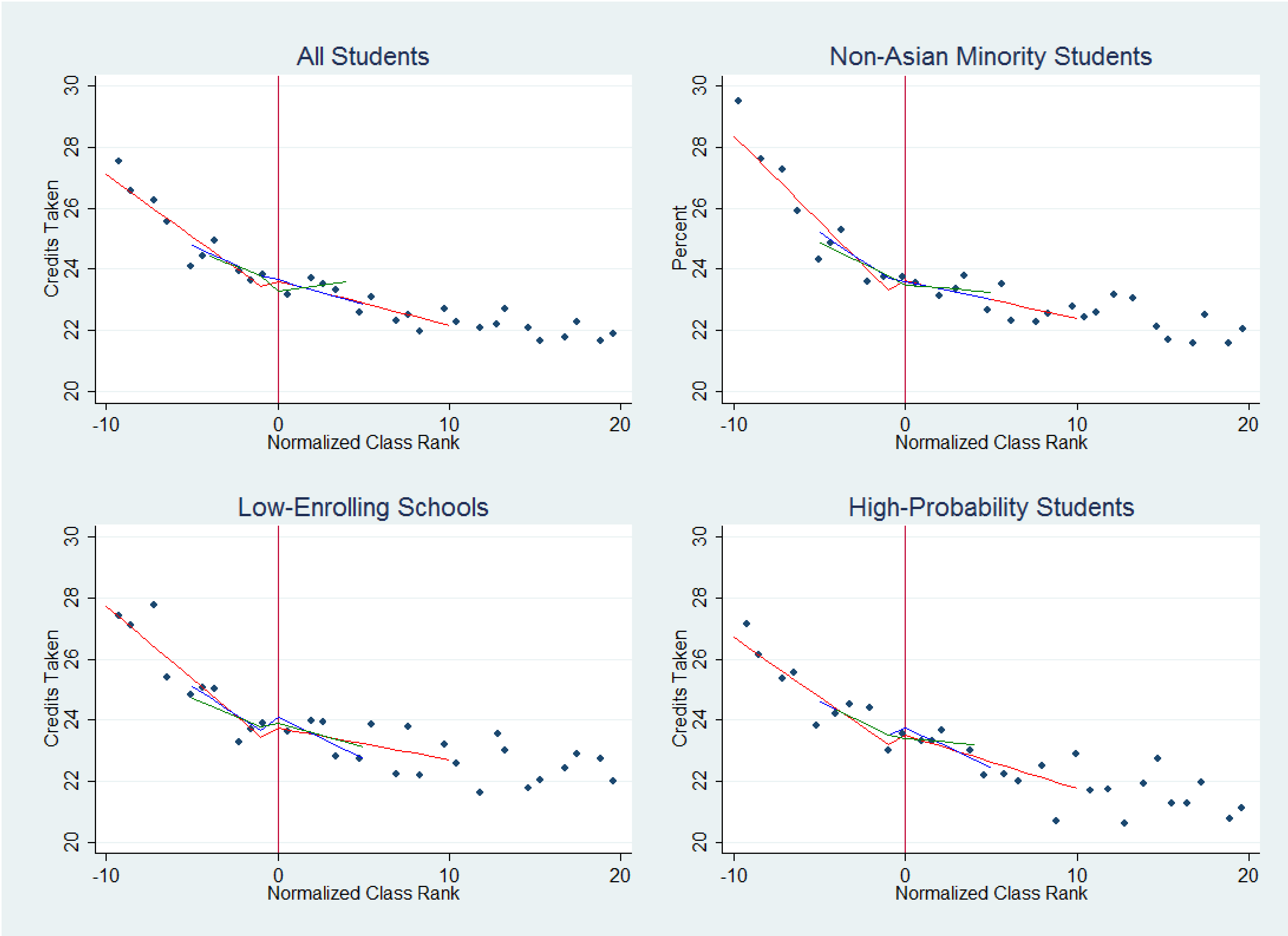


Figure 2.7: Difference in Honors/AP Credits Taken

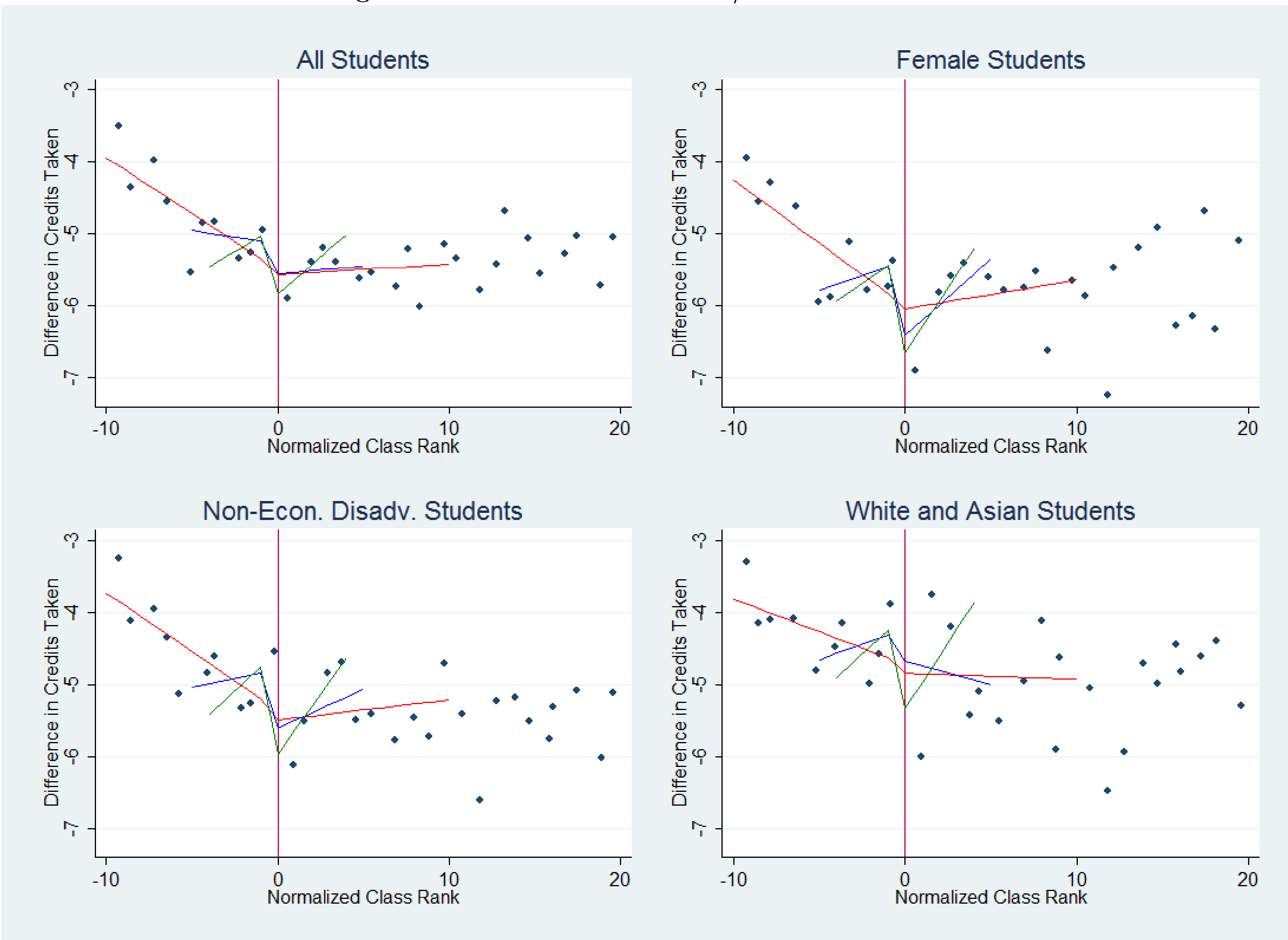


Figure 2.8: Senior Year GPA

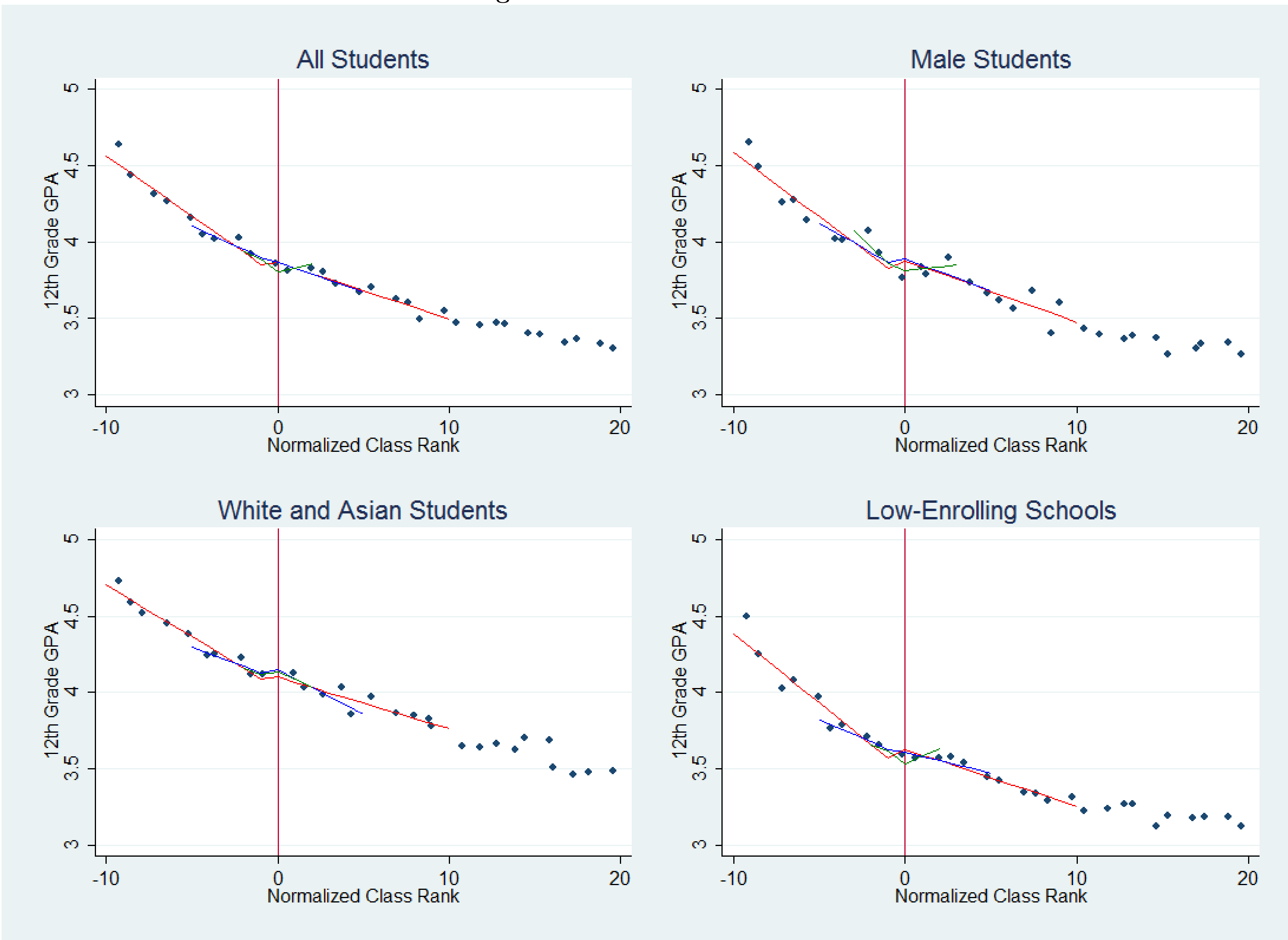


Figure 2.9: Difference in GPA

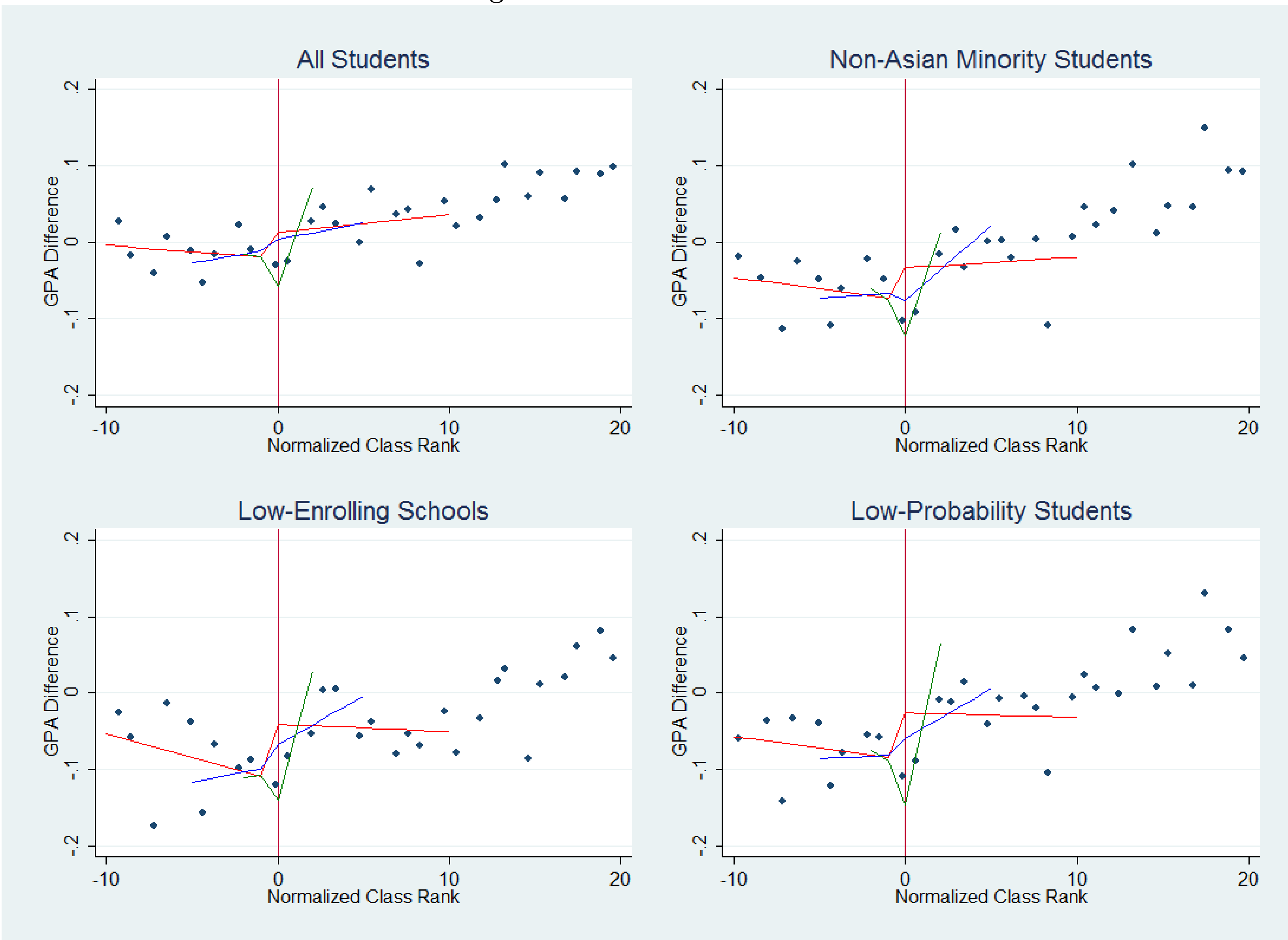


Table 2.1: Continuity of Student Characteristics

	(1)	(2)	(3)
Male	-0.0197	0.0247	0.0283
	(0.0624)	(0.0414)	(0.0295)
White	-0.0671	0.0025	-0.0029
	(0.0632)	(0.0425)	(0.0303)
African-American	0.0486	0.0257	-0.0125
	(0.0666)	(0.0357)	(0.0250)
Other Ethnicity	0.0612	0.0230	0.0067
	(0.0724)	(0.0390)	(0.0275)
Economically Disadvantaged	0.0588	0.0199	0.0110
	(0.0647)	(0.0434)	(0.0308)
Limited English Proficiency	0.0032	-0.0072	-0.0207
	(0.0405)	(0.0254)	(0.0175)
Special Ed	0.0006	0.0052	-0.0023
	(0.0153)	(0.0111)	(0.0080)
Bandwidth	5	10	IK

Table 2.2: First-Stage Outcomes

	(1)	(2)	(3)	(4)	(5)	(6)
Enrolled, Any University	0.00287	0.0270	-0.0403	-0.0337	-0.0228	-0.0283
	(0.0436)	(0.0394)	(0.0307)	(0.0279)	(0.0703)	(0.0593)
	$N = 2107$		$N = 4196$		$N = 963$	
Enrolled, 4-Year University	0.0174	0.0149	-0.0406	-0.0335	-0.0504	-0.0517
	(0.0430)	(0.0377)	(0.0303)	(0.0268)	(0.0672)	(0.0544)
	$N = 2107$		$N = 4196$		$N = 1029$	
Enrolled, Selective University	0.0549	0.0541	-0.0088	-0.0050	0.0654	0.0541
	(0.0384)	(0.0331)	(0.0273)	(0.0238)	(0.0603)	(0.0492)
	$N = 2107$		$N = 4196$		$N = 1037$	
Enrolled, Out-of-State	0.0049	0.0064	-0.0121	-0.0115	0.0132	0.0142
	(0.0209)	(0.0208)	(0.0154)	(0.0151)	(0.0363)	(0.0344)
	$N = 2107$		$N = 4196$		$N = 931$	
Enrolled, Private University	-0.0329	-0.0329	-0.0689 ***	-0.0657 ***	-0.0895 *	-0.0847 *
	(0.0269)	(0.0266)	(0.0194)	(0.0191)	(0.0521)	(0.0481)
	$N = 2107$		$N = 4196$		$N = 827$	
Enrolled, UT or A&M	0.0717 **	0.0701 **	0.0582 ***	0.0575 ***	0.1314 ***	0.1197 ***
	(0.0291)	(0.0271)	(0.0208)	(0.0197)	(0.0476)	(0.0418)
	$N = 2107$		$N = 4196$		$N = 807$	
Bandwidth	5		10		IK	
Controls?	No	Yes	No	Yes	No	Yes

Table 2.3: Probability of Obtaining at Least a Recommended-Level Diploma

	(1)	(2)	(3)	(4)	(5)	(6)
All	0.0109	0.0134	0.0046	0.0088	0.0240	0.0273
	(0.0177)	(0.0166)	(0.0119)	(0.0113)	(0.0316)	(0.0278)
	$N = 2107$		$N = 4196$		$N = 788$	
Male	-0.0083	-0.0082	0.0010	0.0032	0.0120	0.0049
	(0.0290)	(0.0266)	(0.0195)	(0.0183)	(0.0418)	(0.0344)
	$N = 743$		$N = 1475$		$N = 336$	
Female	0.0223	0.0229	0.0068	0.0116	0.0285	0.0295
	(0.0223)	(0.0213)	(0.0151)	(0.0144)	(0.0399)	(0.0337)
	$N = 1364$		$N = 2721$		$N = 572$	
Not Economically Disadvantaged	-0.0079	-0.0111	-0.0041	-0.0045	-0.0009	0.0061
	(0.0193)	(0.0179)	(0.0138)	(0.0131)	(0.0331)	(0.0236)
	$N = 1173$		$N = 2328$		$N = 484$	
Economically Disadvantaged	0.0318	0.0481 *	0.0142	0.0245	0.0478	0.0734 *
	(0.0314)	(0.0289)	(0.0203)	(0.0186)	(0.0515)	(0.0236)
	$N = 934$		$N = 1868$		$N = 427$	
White or Asian	0.0170	0.0205	0.0165	0.0206	0.0048	0.0212
	(0.0221)	(0.0186)	(0.0150)	(0.0136)	(0.0349)	(0.0260)
	$N = 957$		$N = 1954$		$N = 424$	
Non-Asian Minority	0.0043	0.0082	-0.0058	-0.0019	0.0337	0.0368
	(0.0267)	(0.0255)	(0.0180)	(0.0172)	(0.0436)	(0.0391)
	$N = 1150$		$N = 2242$		$N = 516$	
High-Enrolling School	-0.0008	0.0079	0.0064	0.0099	-0.0003	0.0135
	(0.0191)	(0.0168)	(0.0134)	(0.0124)	(0.0288)	(0.0240)
	$N = 1304$		$N = 2585$		$N = 563$	
Low-Enrolling School	0.0300	0.0286	0.0021	0.0081	0.0651	0.0410
	(0.0344)	(0.0337)	(0.0224)	(0.0212)	(0.0602)	(0.0505)
	$N = 803$		$N = 1611$		$N = 368$	
High Enrollment Probability	0.0022	0.0220	0.0043	0.0025	0.0051	0.0101
	(0.0172)	(0.0168)	(0.0119)	(0.0115)	(0.0280)	(0.0240)
	$N = 1031$		$N = 2096$		$N = 413$	
Low Enrollment Probability	0.0168	0.0328	0.0069	0.0187	0.0317	0.0623
	(0.0305)	(0.0285)	(0.0202)	(0.0189)	(0.0512)	(0.0449)
	$N = 1076$		$N = 2100$		$N = 501$	
Bandwidth	5		10		IK	
Controls?	No	Yes	No	Yes	No	Yes

Table 2.4: Probability of Obtaining a Distinguished-Level Diploma

	(1)	(2)	(3)	(4)	(5)	(6)
All	-0.0028 (0.0257)	-0.0019 (0.0222)	-0.0525 *** (0.0193)	-0.0496 *** (0.0166)	0.0267 (0.0392)	0.0221 (0.0340)
	$N = 2107$		$N = 4196$		$N = 983$	
Male	-0.0743 (0.0537)	-0.0160 (0.0480)	-0.0347 (0.0380)	-0.0233 (0.0329)	-0.0237 (0.0894)	0.0741 (0.0730)
	$N = 743$		$N = 1475$		$N = 360$	
Female	0.0307 (0.0269)	0.0084 (0.0233)	-0.0585 *** (0.0213)	-0.0625 *** (0.0183)	0.0522 (0.0348)	0.0258 (0.0306)
	$N = 1364$		$N = 2721$		$N = 641$	
Not Economically Disadvantaged	-0.0058 (0.0414)	-0.0166 (0.0360)	-0.0708 ** (0.0304)	-0.0667 *** (0.0329)	0.0291 (0.0614)	0.0082 (0.0522)
	$N = 1173$		$N = 2328$		$N = 612$	
Economically Disadvantaged	0.0092 (0.0241)	0.0082 (0.0234)	-0.0275 (0.0194)	-0.0214 (0.0188)	0.0105 (0.0319)	0.0378 (0.0254)
	$N = 934$		$N = 1868$		$N = 445$	
White or Asian	-0.0392 (0.0480)	-0.0362 (0.0430)	-0.0803 ** (0.0348)	-0.0802 ** (0.0301)	-0.0026 (0.0719)	0.0144 (0.0616)
	$N = 957$		$N = 1957$		$N = 530$	
Non-Asian Minority	0.0290 (0.0243)	0.0260 (0.0220)	-0.0093 (0.0184)	-0.0081 (0.0173)	0.0391 (0.0362)	0.0249 (0.0280)
	$N = 1150$		$N = 2242$		$N = 511$	
High-Enrolling School	-0.0069 (0.0393)	-0.0091 (0.0343)	-0.0574 ** (0.0287)	-0.0510 ** (0.0247)	0.0448 (0.0578)	0.0387 (0.0505)
	$N = 1304$		$N = 2585$		$N = 694$	
Low-Enrolling School	0.0028 (0.0168)	0.0008 (0.0169)	-0.0454 *** (0.0158)	-0.0440 *** (0.0159)	0.0100 (0.0100)	0.0021 (0.0105)
	$N = 803$		$N = 1611$		$N = 315$	
High Enrollment Probability	-0.0203 (0.0459)	-0.0200 (0.0401)	-0.0609 * (0.0388)	-0.0703 ** (0.0290)	0.0203 (0.0668)	0.0120 (0.0567)
	$N = 1031$		$N = 2096$		$N = 559$	
Low Enrollment Probability	0.0130 (0.0208)	0.0154 (0.0202)	-0.0115 (0.0166)	-0.0081 (0.0162)	0.0170 (0.0144)	0.0618 ** (0.0256)
	$N = 1076$		$N = 2100$		$N = 457$	
Bandwidth	5		10		IK	
Controls?	No	Yes	No	Yes	No	Yes

Table 2.5: Number of Courses Taken in 12th Grade

	(1)	(2)	(3)	(4)	(5)	(6)
All	0.0660	0.0215	-0.0628	-0.0541	0.2140	0.2135
	(0.1060)	(0.0980)	(0.0746)	(0.0688)	(0.1476)	(0.1331)
	$N = 2100$		$N = 4182$		$N = 1128$	
Male	-0.1150	-0.0662	-0.0049	-0.0300	0.0642	0.1573
	(0.1817)	(0.1733)	(0.1270)	(0.1192)	(0.2154)	0.(1897)
	$N = 741$		$N = 1468$		$N = 519$	
Female	0.0657	0.0418	-0.0877	-0.0732	0.1919	0.1169
	(0.1306)	(0.1208)	(0.0921)	(0.0848)	(0.1675)	(0.1460)
	$N = 1359$		$N = 2714$		$N = 938$	
Not Economically Disadvantaged	0.0397	0.0016	-0.0668	-0.0936	0.1370	0.0733
	(0.1388)	(0.1302)	(0.0985)	(0.0931)	(0.1695)	(0.1538)
	$N = 1171$		$N = 2321$		$N = 838$	
Economically Disadvantaged	-0.0698	0.0175	-0.0562	-0.0108	0.1865	0.2215
	(0.1594)	(0.1501)	(0.1107)	(0.1040)	(0.2076)	(0.1806)
	$N = 929$		$N = 1861$		$N = 643$	
White or Asian	0.0504	0.0284	-0.0328	-0.0495	0.3427	0.3284 *
	(0.1598)	(0.1475)	(0.1088)	(0.0991)	(0.2165)	(0.1805)
	$N = 954$		$N = 1948$		$N = 577$	
Non-Asian Minority	-0.0403	-0.0417	-0.1303	-0.1123	0.0887	-0.0026
	(0.1397)	(0.1300)	(0.1010)	(0.0946)	(0.1701)	(0.1506)
	$N = 1146$		$N = 2234$		$N = 789$	
High-Enrolling School	0.0992	0.1011	-0.0444	-0.0298	0.1608	0.1860
	(0.1412)	(0.1297)	(0.0984)	(0.0905)	(0.1824)	(0.1671)
	$N = 1301$		$N = 2577$		$N = 880$	
Low-Enrolling School	-0.1417	-0.1513	-0.0937	-0.1199	0.1979	0.1419
	(0.1532)	(0.1470)	(0.1104)	(0.1048)	(0.2019)	(0.1764)
	$N = 799$		$N = 1605$		$N = 444$	
High Enrollment Probability	-0.0821	-0.1280	-0.1076	-0.1145	0.0112	-0.0274
	(0.1480)	(0.1392)	(0.1046)	(0.0975)	(0.1798)	(0.1623)
	$N = 1029$		$N = 2090$		$N = 763$	
Low Enrollment Probability	0.1024	0.1464	-0.0586	-0.0260	0.3066	0.3379 **
	(0.1489)	(0.1384)	(0.1047)	(0.0973)	(0.1873)	(0.1599)
	$N = 1071$		$N = 2092$		$N = 758$	
Bandwidth	5		10		IK	
Controls?	No	Yes	No	Yes	No	Yes

Table 2.6: Difference in Courses Taken

	(1)	(2)	(3)	(4)	(5)	(6)
All	0.0895 (0.1176)	0.1213 (0.1114)	0.0023 (0.0810)	0.0179 (0.0777)	0.3505 ** (0.1699)	0.3330 ** (0.1567)
	$N = 2100$		$N = 4182$		$N = 1119$	
Male	-0.1892 (0.2066)	-0.0386 (0.1942)	0.0490 (0.1390)	0.0681 (0.1354)	0.0607 (0.2564)	0.2511 (0.2062)
	$N = 741$		$N = 1468$		$N = 551$	
Female	0.2282 (0.1413)	0.2081 (0.1361)	-0.0100 (0.0986)	0.0003 (0.0948)	0.3316 * (0.1803)	0.2201 (0.1625)
	$N = 1359$		$N = 2714$		$N = 999$	
Not Economically Disadvantaged	0.1843 (0.1460)	0.1463 (0.1411)	0.0285 (0.1031)	0.0256 (0.1001)	0.3627 ** (0.1813)	0.3334 * (0.1721)
	$N = 1171$		$N = 2321$		$N = 845$	
Economically Disadvantaged	-0.0252 (0.1910)	0.0476 (0.1787)	-0.0322 (0.1283)	0.0218 (0.1222)	0.0840 (0.2376)	0.1125 (0.2059)
	$N = 929$		$N = 1861$		$N = 737$	
White or Asian	0.1040 (0.1646)	0.0844 (0.1580)	0.0117 (0.1107)	0.0123 (0.1067)	0.5136 ** (0.2318)	0.4976 ** (0.2020)
	$N = 954$		$N = 1948$		$N = 558$	
Non-Asian Minority	0.0673 (0.1640)	0.0814 (0.1534)	-0.0143 (0.1151)	-0.0123 (0.1102)	0.1427 (0.2133)	0.0770 (0.1858)
	$N = 1146$		$N = 2234$		$N = 752$	
High-Enrolling School	0.1789 (0.1481)	0.1918 (0.1415)	0.0179 (0.1009)	0.0411 (0.0979)	0.3383 * (0.1960)	0.3172 * (0.1777)
	$N = 1301$		$N = 2577$		$N = 974$	
Low-Enrolling School	-0.0565 (0.1921)	-0.0674 (0.1808)	-0.0249 (0.1342)	-0.0496 (0.1269)	0.1159 (0.2332)	0.1272 (0.2012)
	$N = 799$		$N = 1605$		$N = 597$	
High Enrollment Probability	0.0971 (0.1519)	0.0266 (0.1442)	0.0263 (0.1062)	-0.0048 (0.1027)	0.3383 * (0.1960)	0.3070 * (0.1790)
	$N = 1029$		$N = 2090$		$N = 682$	
Low Enrollment Probability	0.0811 (0.1784)	0.1347 (0.1664)	-0.0028 (0.1213)	0.0190 (0.1154)	0.1877 (0.2214)	0.2287 (0.1855)
	$N = 1071$		$N = 2092$		$N = 868$	
Bandwidth	5		10		IK	
Controls?	No	Yes	No	Yes	No	Yes

Table 2.7: Number of Credits Taken in 12th Grade

	(1)	(2)	(3)	(4)	(5)	(6)
All	-0.1344	-0.0515	-0.5731	-0.4652	0.2843	0.3106
	(0.5036)	(0.4702)	(0.3542)	(0.3290)	(0.5629)	(0.5148)
	$N = 2100$		$N = 4182$		$N = 1930$	
Male	-1.0201	-0.6501	-0.3391	-0.3409	-0.5675	-0.0898
	(0.8742)	(0.8449)	(0.6078)	(0.5748)	(0.8560)	(0.7698)
	$N = 741$		$N = 1468$		$N = 895$	
Female	0.2875	0.1601	-0.6495	-0.5326	0.6394	0.3178
	(0.6133)	(0.5776)	(0.4345)	(0.4040)	(0.6954)	(0.6237)
	$N = 1359$		$N = 2714$		$N = 1247$	
Not Economically Disadvantaged	0.1223	-0.1438	-0.5067	-0.6012	0.4324	0.0804
	(0.6691)	(0.6272)	(0.4764)	(0.4460)	(0.7002)	(0.6411)
	$N = 1171$		$N = 2321$		$N = 1217$	
Economically Disadvantaged	-0.5702	-0.0470	-0.6855	-0.3388	-0.1527	0.3830
	(0.7606)	(0.7180)	(0.5261)	(0.4939)	(0.8185)	(0.7235)
	$N = 929$		$N = 1861$		$N = 974$	
White or Asian	-0.0414	-0.0796	-0.4278	-0.4250	0.5466	0.5190
	(0.7816)	(0.7115)	(0.5325)	(0.4804)	(0.9078)	(0.7663)
	$N = 954$		$N = 1948$		$N = 835$	
Non-Asian Minority	-0.2508	-0.2775	-0.8689 *	-0.7567 *	0.0406	-0.1054
	(0.6566)	(0.6192)	(0.4733)	(0.4464)	(0.6760)	(0.6132)
	$N = 1146$		$N = 2234$		$N = 1248$	
High-Enrolling School	0.2667	0.3173	-0.4496	-0.3275	0.4932	0.5658
	(0.6842)	(0.6287)	(0.4768)	(0.4371)	(0.7634)	(0.6951)
	$N = 1301$		$N = 2577$		$N = 1221$	
Low-Enrolling School	-0.7812	-0.8592	-0.7809	-0.8115 *	-0.3392	-0.4869
	(0.7167)	(0.6919)	(0.5138)	(0.4899)	(0.7075)	(0.6485)
	$N = 799$		$N = 1605$		$N = 936$	
High Enrollment Probability	-0.5135	-0.7035	-0.6834	-0.7137	-0.1590	-0.3612
	(0.7248)	(0.6775)	(0.5119)	(0.4736)	(0.7801)	(0.7033)
	$N = 1029$		$N = 2090$		$N = 1020$	
Low Enrollment Probability	0.2270	0.5902	-0.6105	-0.3428	0.6376	1.0258
	(0.7014)	(0.6553)	(0.4914)	(0.4578)	(0.7687)	(0.6732)
	$N = 1071$		$N = 2092$		$N = 1047$	
Bandwidth	5		10		IK	
Controls?	No	Yes	No	Yes	No	Yes

Table 2.8: Difference in Credits Taken

	(1)	(2)	(3)	(4)	(5)	(6)
All	0.4146	0.5769	0.0679	0.1380	0.9514	1.0219 *
	(0.5398)	(0.5107)	(0.3721)	(0.3557)	(0.6378)	(0.5881)
	$N = 2100$		$N = 4182$		$N = 1788$	
Male	-0.8839	-0.1246	0.2429	0.3784	-0.2827	0.4659
	(0.9386)	(0.8924)	(0.6349)	(0.6163)	(0.9227)	(0.8077)
	$N = 741$		$N = 1468$		$N = 941$	
Female	1.0521	0.9754	0.0328	0.0518	1.3767 *	0.9518
	(0.6529)	(0.6273)	(0.4550)	(0.4360)	(0.7772)	(0.7049)
	$N = 1359$		$N = 2714$		$N = 1155$	
Not Economically Disadvantaged	0.8215	0.6349	0.1384	0.1234	1.4331 *	1.2779 *
	(0.6712)	(0.6505)	(0.4762)	(0.4622)	(0.7601)	(0.7287)
	$N = 1171$		$N = 2321$		$N = 1032$	
Economically Disadvantaged	-0.0912	0.3187	-0.0126	0.2241	0.2042	0.5187
	(0.8767)	(0.8122)	(0.5872)	(0.5543)	(0.8830)	(0.7697)
	$N = 929$		$N = 1861$		$N = 1142$	
White or Asian	0.4560	0.3624	0.1229	0.1041	1.3095	1.3257
	(0.7658)	(0.7370)	(0.5149)	(0.4974)	(0.9143)	(0.8356)
	$N = 954$		$N = 1948$		$N = 797$	
Non-Asian Minority	0.3405	0.4405	-0.0057	0.0097	0.5296	0.5474
	(0.7479)	(0.6949)	(0.5253)	(0.4997)	(0.7662)	(0.6791)
	$N = 1146$		$N = 2234$		$N = 1131$	
High-Enrolling School	0.7444	0.7910	0.0715	0.1755	1.2349	1.2179
	(0.6790)	(0.6516)	(0.4654)	(0.4509)	(0.7853)	(0.7417)
	$N = 1301$		$N = 2577$		$N = 1191$	
Low-Enrolling School	-0.1254	-0.1314	0.0545	-0.0648	0.0951	0.0146
	(0.8863)	(0.8275)	(0.6145)	(0.5722)	(0.8633)	(0.7711)
	$N = 799$		$N = 1605$		$N = 990$	
High Enrollment Probability	0.4216	0.0934	0.1709	0.0243	1.1091	0.9249
	(0.7046)	(0.6724)	(0.4924)	(0.4770)	(0.8254)	(0.7615)
	$N = 1029$		$N = 2090$		$N = 860$	
Low Enrollment Probability	0.4022	0.6565	0.0658	0.1373	0.5887	0.7287
	(0.8141)	(0.7516)	(0.5547)	(0.5221)	(0.8382)	(0.7167)
	$N = 1071$		$N = 2092$		$N = 1405$	
Bandwidth	5		10		IK	
Controls?	No	Yes	No	Yes	No	Yes

Table 2.9: 12th Grade GPA

	(1)	(2)	(3)	(4)	(5)	(6)
All	-0.0193	-0.0107	-0.0914 ***	-0.0879 ***	0.0417	0.0321
	(0.0463)	(0.0340)	(0.0322)	(0.0241)	(0.0743)	(0.0511)
	$N = 2095$		$N = 4176$		$N = 955$	
Male	-0.0850	-0.0216	-0.1334 **	-0.0894 **	-0.0650	-0.0283
	(0.0833)	(0.0631)	(0.0572)	(0.0437)	(0.1153)	(0.0767)
	$N = 740$		$N = 1467$		$N = 430$	
Female	0.0137	-0.0114	-0.0695 *	-0.0865 ***	0.0877	0.0609
	(0.0554)	(0.0405)	(0.0387)	(0.0290)	(0.0973)	(0.0626)
	$N = 1355$		$N = 2709$		$N = 558$	
Not Economically Disadvantaged	-0.0117	-0.0238	-0.0846 **	-0.0799 ***	0.0775	0.0214
	(0.0591)	(0.0430)	(0.0416)	(0.0309)	(0.0914)	(0.0575)
	$N = 1167$		$N = 2316$		$N = 561$	
Economically Disadvantaged	-0.0160	0.0069	-0.1120 **	-0.1012 ***	-0.0519	0.0169
	(0.0656)	(0.0551)	(0.0452)	(0.0385)	(0.0936)	(0.0713)
	$N = 928$		$N = 1860$		$N = 561$	
White or Asian	-0.0635	-0.0417	-0.0809 **	-0.0712 **	-0.0494	0.0063
	(0.0591)	(0.0474)	(0.0411)	(0.0331)	(0.0829)	(0.0600)
	$N = 951$		$N = 1945$		$N = 522$	
Non-Asian Minority	0.0119	0.0116	-0.0879 *	-0.0712 **	0.0510	0.0467
	(0.0589)	(0.0483)	(0.0416)	(0.0348)	(0.0881)	(0.0699)
	$N = 1144$		$N = 2231$		$N = 593$	
High-Enrolling School	-0.0154	0.0028	-0.0573	-0.0497 *	0.0365	0.0142
	(0.0571)	(0.0398)	(0.0402)	(0.0290)	(0.0848)	(0.0534)
	$N = 1297$		$N = 2573$		$N = 668$	
Low-Enrolling School	-0.0299	-0.0311	-0.1481 ***	-0.1478 ***	0.0506	0.0263
	(0.0627)	(0.0612)	(0.0430)	(0.0419)	(0.1034)	(0.0882)
	$N = 798$		$N = 1603$		$N = 369$	
High Enrollment Probability	-0.0105	-0.0197	-0.0441	-0.0675 **	0.0041	-0.0138
	(0.0593)	(0.0432)	(0.0417)	(0.0314)	(0.0824)	(0.0529)
	$N = 1026$		$N = 2087$		$N = 575$	
Low Enrollment Probability	-0.0477	0.0070	-0.1145 ***	-0.0986 ***	0.0106	0.1082
	(0.0586)	(0.0525)	(0.0413)	(0.0371)	(0.0960)	(0.0825)
	$N = 1069$		$N = 2089$		$N = 488$	
Bandwidth	5		10		IK	
Controls?	No	Yes	No	Yes	No	Yes

Table 2.10: 12th Grade GPA (1st Semester Only)

	(1)	(2)	(3)	(4)	(5)	(6)
All	-0.0395 (0.0478)	-0.0323 (0.0352)	-0.1042 *** (0.0334)	-0.0994 *** (0.0252)	-0.0155 (0.0716)	-0.0311 (0.0492)
	$N = 2093$		$N = 4171$		$N = 1046$	
Male	-0.1235 (0.0824)	-0.0703 (0.0619)	-0.1588 *** (0.0572)	-0.1132 *** (0.0433)	-0.1654 (0.1128)	-0.1380 * (0.0750)
	$N = 739$		$N = 1466$		$N = 379$	
Female	0.0013 (0.0587)	-0.0188 (0.0435)	-0.0735 * (0.0412)	-0.0901 *** (0.0312)	0.0558 (0.0945)	0.0191 (0.0627)
	$N = 1354$		$N = 2705$		$N = 629$	
Not Economically Disadvantaged	-0.0232 (0.0599)	-0.0273 (0.0445)	-0.1022 ** (0.0426)	-0.0939 *** (0.0322)	0.0220 (0.0856)	-0.0229 (0.0563)
	$N = 1167$		$N = 2315$		$N = 628$	
Economically Disadvantaged	-0.0456 (0.0684)	-0.0324 (0.0577)	-0.1214 *** (0.0472)	-0.1111 *** (0.0402)	-0.1266 (0.0959)	-0.0714 (0.0730)
	$N = 926$		$N = 1856$		$N = 564$	
White or Asian	-0.0630 (0.0598)	-0.0414 (0.0493)	-0.0993 ** (0.0421)	-0.0876 *** (0.0343)	-0.0808 (0.0844)	-0.0195 (0.0639)
	$N = 951$		$N = 1945$		$N = 529$	
Non-Asian Minority	-0.0224 (0.0614)	-0.0239 (0.0501)	-0.0970 ** (0.0436)	-0.1079 *** (0.0365)	-0.0266 (0.0843)	-0.0303 (0.0669)
	$N = 1142$		$N = 2226$		$N = 643$	
High-Enrolling School	-0.0241 (0.0594)	0.0000 (0.0420)	-0.0736 * (0.0420)	-0.0623 ** (0.0308)	-0.0016 (0.0825)	-0.0229 (0.0540)
	$N = 1296$		$N = 2571$		$N = 744$	
Low-Enrolling School	-0.0696 (0.0647)	-0.0764 (0.0624)	-0.1566 *** (0.0444)	-0.1556 *** (0.0429)	-0.0437 (0.1002)	-0.0690 (0.0876)
	$N = 797$		$N = 1600$		$N = 399$	
High Enrollment Probability	-0.0206 (0.0604)	-0.0242 (0.0454)	-0.0591 (0.0428)	-0.0768 ** (0.0330)	-0.0086 (0.0836)	-0.0231 (0.0552)
	$N = 1026$		$N = 2086$		$N = 557$	
Low Enrollment Probability	-0.0776 (0.0616)	-0.0305 (0.0549)	-0.1255 *** (0.0435)	-0.1118 *** (0.0388)	-0.0843 (0.0935)	0.0036 (0.0791)
	$N = 1067$		$N = 2085$		$N = 543$	
Bandwidth	5		10		IK	
Controls?	No	Yes	No	Yes	No	Yes

Table 2.11: 12th Grade GPA (2nd Semester Only)

	(1)	(2)	(3)	(4)	(5)	(6)
All	0.0014	0.0040	-0.0798 **	-0.0813 ***	0.0914	0.0842
	(0.0527)	(0.0421)	(0.0365)	(0.0298)	(0.0862)	(0.0652)
	N = 2044		N = 4087		N = 925	
Male	-0.0587	0.0115	-0.1157 *	-0.0717	0.0232	0.0586
	(0.0967)	(0.0814)	(0.0664)	(0.0562)	(0.1309)	(0.0951)
	N = 733		N = 1450		N = 478	
Female	0.0341	-0.0083	-0.0632	-0.0865 **	0.1193	0.0844
	(0.0620)	(0.0490)	(0.0432)	(0.0350)	(0.1084)	(0.0763)
	N = 1311		N = 2637		N = 567	
Not Economically Disadvantaged	-0.0031	-0.0330	-0.0719	-0.0758 **	0.0989	0.0193
	(0.0668)	(0.0520)	(0.0469)	(0.0374)	(0.1052)	(0.0706)
	N = 1141		N = 2268		N = 565	
Economically Disadvantaged	0.0140	0.0412	-0.1001 *	-0.0899 *	0.0307	0.1081
	(0.0781)	(0.0692)	(0.0539)	(0.0487)	(0.1105)	(0.0915)
	N = 903		N = 1819		N = 561	
White or Asian	-0.0731	-0.0500	-0.0700	-0.0642	-0.0306	0.0122
	(0.0679)	(0.0562)	(0.0470)	(0.0399)	(0.0939)	(0.0694)
	N = 930		N = 1904		N = 520	
Non-Asian Minority	0.0500	0.0425	-0.0783	-0.0957 **	0.1349	0.1231
	(0.0695)	(0.0608)	(0.0491)	(0.0435)	(0.1089)	(0.0908)
	N = 1114		N = 2183		N = 583	
High-Enrolling School	-0.0256	-0.0111	-0.0543	-0.0486	0.0403	0.0063
	(0.0623)	(0.0467)	(0.0437)	(0.0341)	(0.0904)	(0.0625)
	N = 1269		N = 2516		N = 706	
Low-Enrolling School	0.0213	0.0225	-0.1319 **	-0.1335 **	0.1488	0.1321
	(0.0809)	(0.0793)	(0.0557)	(0.0548)	(0.1341)	(0.1172)
	N = 775		N = 1571		N = 358	
High Enrollment Probability	-0.0030	-0.0175	-0.0292	-0.0603	0.0222	-0.0059
	(0.0675)	(0.0520)	(0.0474)	(0.0379)	(0.0983)	(0.0676)
	N = 1007		N = 2052		N = 528	
Low Enrollment Probability	-0.0162	0.0385	-0.1090 **	-0.0917 **	0.0954	0.1946 **
	(0.0700)	(0.0648)	(0.0494)	(0.0461)	(0.1082)	(0.0971)
	N = 1037		N = 2035		N = 560	
Bandwidth	5		10		IK	
Controls?	No	Yes	No	Yes	No	Yes

Table 2.12: Difference in GPAs

	(1)	(2)	(3)	(4)	(5)	(6)
All	-0.0110	-0.0081	-0.0335	-0.0332	0.0340	0.0284
	(0.0361)	(0.0339)	(0.0249)	(0.0236)	(0.0570)	(0.0510)
	$N = 2095$		$N = 4176$		$N = 1025$	
Male	-0.0013	-0.0070	-0.0541	-0.0423	0.0405	0.0982
	(0.0662)	(0.0640)	(0.0448)	(0.0430)	(0.0976)	(0.0874)
	$N = 740$		$N = 1467$		$N = 398$	
Female	-0.0157	-0.0144	-0.0252	-0.0288	0.0325	0.0417
	(0.0427)	(0.0403)	(0.0297)	(0.0283)	(0.0723)	(0.0618)
	$N = 1355$		$N = 2709$		$N = 620$	
Not Economically Disadvantaged	-0.0070	-0.0227	-0.0210	-0.0254	0.0880	0.0254
	(0.0454)	(0.0435)	(0.0320)	(0.0307)	(0.0744)	(0.0625)
	$N = 1167$		$N = 2316$		$N = 484$	
Economically Disadvantaged	-0.0162	0.0089	-0.0517	-0.0418	-0.0166	0.0297
	(0.0579)	(0.0544)	(0.0391)	(0.0374)	(0.0870)	(0.0726)
	$N = 928$		$N = 1860$		$N = 525$	
White or Asian	-0.0451	-0.0399	-0.0207	-0.0184	0.0126	0.0266
	(0.0490)	(0.0474)	(0.0338)	(0.0327)	(0.0698)	(0.0639)
	$N = 951$		$N = 1945$		$N = 480$	
Non-Asian Minority	0.0106	0.0122	-0.0422	-0.0488	0.0307	0.0373
	(0.0510)	(0.0482)	(0.0356)	(0.0341)	(0.0784)	(0.0708)
	$N = 1144$		$N = 2231$		$N = 606$	
High-Enrolling School	-0.0018	-0.0002	-0.0088	-0.0091	0.0311	0.0018
	(0.0424)	(0.0398)	(0.0298)	(0.0285)	(0.0626)	(0.0551)
	$N = 1297$		$N = 2573$		$N = 673$	
Low-Enrolling School	-0.0277	-0.0157	-0.0738 *	-0.0699 *	0.0355	0.0356
	(0.0636)	(0.0615)	(0.0428)	(0.0413)	(0.1062)	(0.0906)
	$N = 798$		$N = 1603$		$N = 377$	
High Enrollment Probability	-0.0079	-0.0151	0.0035	-0.0053	0.0143	-0.0034
	(0.0459)	(0.0439)	(0.0324)	(0.0312)	(0.0660)	(0.0590)
	$N = 1026$		$N = 2087$		$N = 507$	
Low Enrollment Probability	-0.0211	0.0076	-0.0607	-0.0548	0.0448	0.0982
	(0.0544)	(0.0520)	(0.0374)	(0.0632)	(0.0990)	(0.0874)
	$N = 1069$		$N = 2089$		$N = 452$	
Bandwidth	5		10		IK	
Controls?	No	Yes	No	Yes	No	Yes

Chapter Two Bibliography

[25] George A. Akerlof and Lawrence F. Katz. Workers' Trust Funds and the Logic of Wage Profiles. *The Quarterly Journal of Economics*, 104(3):525–536, August 1989.

[26] Joshua Angrist, Daniel Lang, and Philip Oreopoulos. Incentives and Services for College Achievement: Evidence from a Randomized Trial. *American Economic Journal: Applied Economics*, 1(1):136–163, January 2009.

[27] Martha Bailey and Susan Dynarski. Inequality in College Entry and Completion. In G.J. Duncan and R.J. Murnane, editors, *Whither Opportunity? Rising Inequality and the Uncertain Life Chances of Low-Income Children*. Russell Sage, 2011.

[28] Christopher M. Cornwell, Kyung Hee Lee, and David B. Mustard. Student Responses to Merit Scholarship Retention Rules. *Journal of Human Resources*, 40(4):895–917, 2005.

[29] Christopher M. Cornwell, Kyung Hee Lee, and David B. Mustard. The Effects of State-Sponsored Merit Scholarships on Course Selection and Major Choice in College. Institute for the Study of Labor (IZA) Discussion Paper Series No. 1953, January 2009.

[30] Kalena Cortes and Lei Zhang. The Incentive Effects of the Top 10% Plan. Under Review, 2011.

[31] Julie Berry Cullen, Randall Reback, and Mark Long. Jockeying for Position: High School Student Mobility and Texas' Top-Ten Percent Rule. NBER Working Paper 16663, January 2011.

[32] Lindsay Daugherty, Paco Martorell, and Isaac McFarlin. Percent Plans, Automatic Admissions, and College Entry. January 2011.

[33] Thurston Domina. Higher Education Policy as Secondary School Reform: Texas Public High Schools After Hopwood. *Educational Evaluation and Policy Analysis*, 29(3):200–217, September 2007.

[34] Roland G. Fryer, Jr. and Glenn C. Loury. Affirmative Action and Its Mythology. *Journal of Economic Perspectives*, 19(3):147–162, 2005.

[35] Roland G. Fryer, Jr., Glenn C. Loury, and Tolga Yuret. An Economic Analysis of Color-Blind Affirmative Action. *Journal of Law, Economics, and Organization*, 24(2):319–355, December 2007.

[36] Caroline M. Hoxby. The Changing Selectivity of American Colleges. *Journal of Economic Perspectives*, 23(4):95–118, 2009.

[37] Guido Imbens and Karthik Kalyanaraman. Optimal Bandwidth Choice for the Regression Discontinuity Estimator. *Review of Economic Studies*, 79(3):933–959, June 2012.

[38] Guido W. Imbens and Thomas Lemieux. Regression Discontinuity Designs: A Guide to Practice. *Journal of Econometrics*, 142(2):615–635, 2008.

[39] C. Kirabo Jackson. A Little Now for a Lot Later: An Evaluation of a Texas Advanced Placement Incentive Program. *Journal of Human Resources*, 45(3):591–639, 2010.

[40] C. Kirabo Jackson. A Stitch in Time: The Effects of a Novel Incentive-Based High-School Intervention on College Outcomes. NBER Working Paper 15722, February 2010.

[41] David Lee. Randomized Experiments from Non-Random Selection in U.S. House Elections. *Journal of Econometrics*, 142(2):675–697, 2008.

[42] Jason M. Lindo, Nicholas J. Sanders, and Philip Oreopoulos. Ability, Gender, and Performance Standards: Evidence from Academic Probation. *American Economic Journal: Applied Economics*, 2(2):95–117, April 2010.

[43] Mark Long and Marta Tienda. Changes in Texas Universities' Applicant Pools After the Hopwood Decision. *Social Science Research*, 39(1):48–66, 2010.

[44] Justin McCrary. Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test. *Journal of Econometrics*, 142(2):698–714, 2008.

[45] Sunny Xinchun Niu and Marta Tienda. The Impact of the Texas Top 10 Percent Law on College Enrollment: A Regression Discontinuity Approach. *Journal of Policy Analysis and Management*, 29(1):84–110, 2012.

[46] Philip Oreopoulos. The Compelling Effects of Compulsory Schooling: Evidence from Canada. *The Canadian Journal of Economics*, 39(1):22–52, February 2006.

[47] Philip Oreopoulos and Kjell G. Salvanes. Priceless: The Nonpecuniary Benefits of Schooling. *The Journal of Economic Perspectives*, 25(1):159–184, 2011.

[48] Mark V. Pauly. Overinsurance and Public Provision of Insurance: The Roles of Moral Hazard and Adverse Selection. *The Quarterly Journal of Economics*, 88(1):44–62, February 1974.

Chapter 3

Iowa's National Scholars Award and the Efficiency of Merit Aid: A Regression Discontinuity Analysis

Co-authored with

Stephen L. DesJardins
University of Michigan

Abstract

The costs of attending college have been rising steadily over the past thirty years, making financial aid both an important determinant of college choice for many students and a subject of concern for financial aid offices and state governments. In this paper, we estimate the effect of rule-based merit aid assignment on students' enrollment decisions at the University of Iowa. Iowa evaluates many students using an admissions score that is a linear combination of their high school GPA, class rank, core high school courses, and ACT test score. Students from out-of-state who meet a specific threshold on the admissions score qualify for the National Scholars Award (NSA), presently worth nearly $20,000 (approximately one-fifth of tuition) over a four-year period. We employ a regression discontinuity model to take advantage of award assignment criteria, finding that the award does increase the odds that students enroll at Iowa. This result is robust for several applicant subsamples and passes falsification checks using Iowa residents, who are not eligible for the award. Preliminary analysis of an earlier, tiered version of the current single-valued award suggests that the NSA could be effectively targeted towards very high-achieving candidates.

I. Introduction

Over the last three deca-des, the cost of attending college has soared – outpacing inflation and the oft-cited costs of medical care.[58] Recent state budget cuts have exacerbated this trend. Meanwhile, college attendance has been rising. These two patterns have put many financial aid officers in a quandary: more students require financial aid, and each requires more aid to cover the same share of college expenses.

The primary role of financial aid officers is to provide money to students. Some of this money is portable for students – National Merit Scholarships and Pell grants, for instance, may be used at any four-year college. Some aid, however, is institution-specific. Individual colleges may use this aid, which can be awarded based on financial need or based on merit, to attract particularly desirable students. Because colleges have limited financial aid budgets, it is worth asking how aid can be distributed to achieve institutional objectives such as attracting a high-quality class and distributing aid in an efficient and equitable manner. Perhaps providing aid to a large number of students is effective, or perhaps it makes more sense to target a small group of students who meet specific criteria (e.g., merit or need).

To examine some of these issues we analyze data from the University of Iowa. Iowa awards its National Scholars Award (henceforth, NSA) to out-of-state applicants with admissions test scores above a specific threshold that is equal to a linear combination of high school GPA, class rank, core courses in high school, and ACT score. These admissions scores are mechanically generated and allow

[58]Taken from `http://www.freakonomics.com/2011/10/27/cost-of-college-on-the-rise-again/`

little room for student manipulation. As a result, students with similar admissions scores may have very different probabilities of being offered the NSA.[59] Because the NSA is awarded to a large number of students and is (as of 2012) worth over $4,800 per year for up to four years, a regression discontinuity (henceforth, RD) analysis of NSA receipt is potentially very informative about the impact of this merit-based aid on student enrollment decisions. For comparison, tuition for out-of-state freshmen in 2012-2013 was set at just over $26,000, meaning that the NSA represents a substantial decrease in costs for these students but does not lower tuition to the level used for in-state students.

Using data on students' application profiles and scholarship receipt, we find that receiving the NSA leads to a statistically significant increase in the enrollment rates of admitted students. This finding is robust to multiple bandwidth specifications and multiple subsamples of out-of-state students. Falsification tests also demonstrate that for Iowa residents, who are not eligible for the NSA, we cannot reject the null hypothesis of no effect. A final specification compares an earlier, tiered version of the NSA where students received different award amounts, to the flat award regime currently in place. We find statistically significant positive discontinuities at the threshold for the award's highest tier, suggesting that recalibrating the award structure may allow the University of Iowa to better target this funding to the highest-achieving students.

The remainder of the paper is structured as follows: Section II outlines the use of admissions scores in awarding the NSA and provides an overview of the relevant

[59]Students' admissions scores may be updated after they submit their applications, but we view their scores at the time that they apply. As a result, some students who do not appear eligible are offered the NSA. Despite this, the probability of a student being offered the NSA rises by over 70 percentage points at the cutoff for NSA receipt.

literature. Section III explains the regression discontinuity model to be estimated and analyzes potential threats to the validity of our RD estimates. Section IV describes the data used in our analysis. Section V presents our results, and Section VI provides a discussion of the policy implications and the conclusion.

II. Background/Literature

Since 1999, the University of Iowa has made admissions and merit aid decisions based on a fixed metric. The initial measure used was known as the Admissions Index Score (or "AIS") and defined as:

$$AIS_i = HSR\%_i + 2 * ACT_i \tag{1}$$

where $HSR\%_i$ refers to the percentile of student i's high school rank (from zero to 99) and ACT_i is her Composite ACT score.[60] In theory, this value could range from a minimum of one to a maximum of 171, but among the 52,968 applicants with valid AIS scores in our sample the maximum score was a 171 and the minimum score was a 27, with a median of 126.

Beginning with applicants for the fall of 2009, the university began using a new index, the Regent Admission Index (or "RAI"), to help make admissions decisions. This index contains four components and is defined as:

$$RAI_i = HSR\%_i + 2 * ACT_i + 20 * HSGPA_i + 5 * CORE_i \tag{2}$$

[60]Because the majority of University of Iowa test-takers take the ACT rather than the SAT, applicants with SAT scores have these scores converted to ACT scores using equi-percentile method concordance tables.

where $HSR\%_i$ and ACT_i are defined as in (1), $HSGPA_i$ is student i's high school grade point average on a 4.0 scale, and $CORE_i$ is the number of high school courses completed in core subject areas.[61] Among the 19,629 applicants in our sample with RAI values, the maximum score on the RAI was 410, the minimum was 93, and the median was 299.[62]

The Iowa Board of Regents currently states that students graduating from Iowa high schools will be automatically admitted to the University of Northern Iowa, Iowa State University, or the University of Iowa's College of Liberal Arts and Sciences if they have an RAI of at least 245 and have completed a pre-specified number of core courses in high school.[63] Out-of-state students are required to have an RAI score of 255 for automatic admission to the University of Iowa's College of Liberal Arts, whereas the College of Engineering requires a score of 265.[64] Students with an RAI score below 245 or who do not have all of the information necessary to calculate their RAI are evaluated on an individual basis,

[61] The website of the Board of Regents of Iowa refers to these as "English, mathematics, natural science, social science, and foreign language" courses, with single-semester courses counting as half a course for RAI purposes. A full list of the NCES codes of qualifying courses may be found at `http://www.regents.iowa.gov/RAI/NCES.pdf`.

[62] Students within 5 RAI points of the cutoff for NSA eligibility have a mean ACT score of 24.4 with a standard deviation of 3.1 below the NSA cutoff and a mean ACT value of 24.8 with a standard deviation of 2.9 above the cutoff. Corresponding GPA values take on a mean of 3.49 below the cutoff with and a mean of 3.55 above the cutoff; both groups have a standard deviation of approximately 0.19.

[63] To provide some perspective, a student with an RAI value of 410 would be exceptional – a student ranked first in her high school class, with a perfect GPA and ACT score, would have had to complete 32 core courses to obtain this score. On the other hand, a score of 93 is very poor, corresponding (among other possibilities) to an ACT score of 12, a last-place class rank, a GPA of 1.0, and only 10 core courses completed. 245, on the other hand, corresponds almost exactly to a straight-B student with an ACT score of 18, a class rank in the 50th percentile, and 20 completed core courses.

[64] The Business School has only ACT and GPA requirements, while the Nursing school requires minimum ACT scores of 28 overall and 25 in science, a GPA of 3.8 out of 4.0, and "no deficiencies in the minimum High School Course Requirements."

and we omit such students from our analysis.[65] The AIS has been, and the RAI is currently used to award institutional financial aid, including the NSA.

The NSA simultaneously addresses the University of Iowa's goals of student quality and diversity by targeting qualified students from other states. Initially, the NSA was awarded in tiers, with the lowest award offered to students with an AIS score of at least 129, a higher amount offered to students with a score of at least 140, and the highest amount offered to students scoring at least 154. Beginning in 2005, the award was changed to a flat rate, with all enrollees scoring at least a 129 receiving the same amount. In 2009, when the RAI was implemented, the cutoff for NSA eligibility was set at 290 on the RAI scale. The rule-based nature of the award therefore makes it possible to evaluate the causal impact of financial aid at the cutoff for NSA receipt and to examine whether, and if so how, the effect of the award has changed over time.

Several previous studies have also examined the impact of monetary awards on college attendance. Dynarski (2000) and Cornwell, Mustard, and Sridhar (2006; henceforth, CMS) both studied the impact of the Georgia HOPE scholarship on college enrollment and found different results – Dynarski found a large impact, whereas CMS found relatively small effects, driven primarily by college choice. Cornwell, Lee, and Mustard (2005 and 2009) also studied the impact of the Georgia HOPE on academic outcomes while students were enrolled in college, finding that recipients were more likely to take easier courses and lighter course loads with the goal of delaying GPA checkpoints and thereby extending the receipt of the scholarship. Dynarski (2003) and Garibaldi, Giavazzi, Ichino, and Rettore (2011) examined the effect of tuition on time to college completion; Dynarski used

[65] Many high schools choose not to disclose class ranks, for instance.

a difference-in-differences approach and Garibaldi, Giavazzi, Ichino, and Rettore employed a RD design. Both papers found that students inside higher tuition brackets were significantly more likely to graduate early.

The paper most similar to ours is Van Der Klaauw (2002). He employs a RD design to capitalize on discontinuities in the amount of financial aid offered to otherwise similar applicants to a highly selective East Coast university, with the goal being to determine the causal impact of financial aid on student enrollment. He finds an enrollment elasticity with respect to financial aid of .86 among applicants who filed for financial aid. He also finds a larger impact on higher ability students.

Our paper adds to the literature in several ways. First, the time period analyzed is qualitatively different than that analyzed by Van Der Klaauw. While an excellent and oft-cited paper, he examines twenty-year old data, and the cost of college attendance has risen considerably since then. According to the National Center for Education Statistics, the cost of tuition, room, and board at four-year public institutions has risen by over 30 percent between the year prior to Van Der Klaauw's sample and the first year of our data. During the years of our sample, 2004-2011, costs have risen by more than 20 percent. Additionally, over a similar period the number of recent high school graduates (defined as individuals 16-24 receiving a high school degree in the past 12 months) has increased by almost seven percent. It is ex ante unclear how the impact of financial aid on college enrollment has changed over this time. If more low-income or credit-constrained students are applying to college, financial aid may have a larger effect on 1) whether students enroll in college at all and 2) where they choose to enroll. However, if employers are demanding higher levels of education from prospective employees, demand for college degrees may be relatively inelastic. If so, financial aid would still affect where applicants choose to enroll but would not alter their decision to attend college. Similarly, if employers place a large premium on their employees' having a degree from a highly selective school, financial aid packages may affect neither college attendance nor college choice.[66]

We also expect to find different results from those cited by Van Der Klaauw because the University of Iowa is a different type of institution than the "University X" he studied. Where University X is a private, East Coast university, Iowa is a public, Midwestern institution; University X is also somewhat more selective and significantly more urban than Iowa. For these reasons, Iowa's students may be qualitatively different from those attending University X, and may therefore respond differently to financial aid offers.

III. Model and Theory

We implement a fuzzy regression discontinuity model (henceforth, FRD) in our analysis of students' enrollment choices. FRD models are defined by two stages. The first estimates the probability of receiving a given "treatment" (denoted by W_i below):

$$Pr\left(W_i = 1\right) = f\left(SC_i\right) + \gamma * \mathbf{1}\left(SC_i \geq c\right) + \nu_i \tag{3}$$

Here W_i is an indicator of treatment status, in this case, whether an admitted applicant received the NSA. SC_i is a student's admission score (either RAI or AIS, depending on the year), c is the cutoff value for NSA receipt, and ν_i is a mean-

[66] For an example of how this might work, see Hershbein (2011).

zero error term. The probability of receiving the NSA is a continuous function of a student's admissions score (the "running variable"), with a discontinuity at c. In sharp regression discontinuity (i.e., SRD) models, $f(SC_i) = 0$ and $\gamma = 1$, implying that treatment is completely determined by whether the running variable has crossed the threshold value. FRD models, however, are generalized so that the jump in the probability of treatment at the cutoff may be less than one. As a result, while crossing the cutoff value has a substantial and statistically significant effect on the probability of treatment, it is not the sole determinant of treatment status.

The second stage in the estimation process is presented formally as:

$$Y_i = g\left(SC_i\right) + X_i\beta + \delta_i W_i + \epsilon_i \tag{4}$$

where Y_i indicates whether student i has, conditional on admission, enrolled at the University of Iowa, X_i is a vector of student characteristics, and ϵ_i is a mean-zero error term. We expect the probability of enrollment to vary continuously with students' admissions scores and discontinuously as students cross the threshold for NSA receipt. In SRD models it is possible to estimate the effect of treatment using the second stage equation and the fact that $W_i = \mathbf{1}\left(SC_i \geq c\right)$, but FRD models require both stages to account for the non-deterministic nature of the running variable.In particular, whereas SRD point estimates and standard errors are equivalent to the treatment effect (δ_i), FRD second-stage point estimates must be weighted by the probability of receiving treatment and are therefore equal to $\frac{E[\delta_i]}{E[\gamma]}$. FRD estimates and standard errors are also equivalent to those from Two-Stage Least Squares ("2SLS"), where the treatment equation is the first stage in

2SLS and the outcome equation is the second stage.

According to Imbens and Lemieux (2008), McCrary (2008), and Lee (2008), for $\frac{E[\delta_i]}{E[\gamma]}$ to represent a true causal estimate several assumptions must hold. First, there cannot be any confounding treatments at the cutoff. Second, student characteristics must not vary systematically around the cutoff. Third, students may not alter their treatment status (i.e., whether they received the NSA) by manipulating the running variable. All of these are potential concerns, addressed below.

First, in addition to the NSA, Iowa offers many scholarships to their prospective students. Fortunately (from an analytic perspective), the NSA appears to be the only scholarship explicitly awarded based on the AIS or RAI. As demonstrated in Figure 3.1, the probability of an out-of-state student getting any other award trends continuously through the cutoffs for the NSA. The bottom two graphs in Figure 3.1 are for students applying for Fall 2004 entry and thus have three cutoffs – one for each level of the NSA. The first cutoff is for students receiving at least a low-value NSA, the second is for students receiving at least a medium-value NSA, and the third is exclusively for students receiving a high-value NSA. As the probability of receiving any other award trends continuously through the NSA cutoff, and since many more students receive the NSA than receive any other award, it is unlikely that other awards (i.e., aid "treatments") are driving our results. Furthermore, because being offered another award does not preclude receiving the NSA, students qualified for the NSA should all receive the same treatment in terms of their NSA award.

It is unlikely that scholarships or admissions offers from other schools would affect the validity of our RD estimates. The only other schools using the AIS or

RAI are Iowa State University and the University of Northern Iowa, neither of which offers a scholarship analogous to the NSA. Since other schools do not use these score indices, it is unlikely that policies at other schools will discontinuously impact students' enrollment decisions. For example, while public universities in Texas will admit any Texan high school student in the top ten percent of her high school class, class rank is only one element of AIS or RAI. Depending on their other characteristics, students in the top ten percent of their high school class may have a wide range of AIS or RAI values. Since Texan high school students do not become eligible for automatic admission at a uniform AIS or RAI value, this admissions guarantee should not discontinuously affect student enrollment decisions and should therefore not affect the validity of our RD estimates. By similar logic, our estimates should remain unaffected by any other awards based on GPA, class rank, or standardized test scores.

It is also possible that shifts in student characteristics around the NSA cutoff could affect the validity of our results. For example, consider a hypothetical case in which students admitted to the Business School enroll with probability one and all other admitted applicants enroll with probability 0.20, and NSA receipt causes more students to apply to the Business School but does not change the probability that any given applicant enrolls. In this case, failing to control for these between-group differences will falsely imply that the NSA affects the probability of enrollment rather than affecting the types of applications that the university receives.

We therefore run density tests of the running variable by subgroup to determine whether student characteristics trend continuously through the cutoff for eligibility. For each of four subgroups, we pool across three sets of years (depending on whether the NSA was awarded at a flat rate based on RAI, at a flat rate based on AIS, or in tiers based on AIS). If students who are narrowly eligible for the NSA are significantly different from those who are narrowly ineligible, it would be difficult to argue that our results represent the causal impact of NSA receipt on enrollment decisions. They could just as easily represent different tastes for attending the University of Iowa, as in the example above.

To run these density checks, we first determined the number of applicants belonging to particular subgroups at each RAI or AIS value. We focused on four subgroups – applicants from Illinois, nonresident applicants from states other than Illinois, applicants to the College of Liberal Arts, and White applicants. Using these large subgroups allows us to focus on meaningful variation in the number of applicants and provides a clear picture of this variation to the reader. For each of these subgroups, we ran regressions of the number of applicants at each RAI or AIS value on RAI or AIS values, an indicator for RAI or AIS values above the threshold for NSA eligibility, and an interaction of RAI or AIS values with this indicator. These regressions were done for a large number of potential bandwidths – regressions using RAI used every possible bandwidth between five RAI points and 30 RAI points, regressions using a single AIS cutoff used every possible bandwidth between five AIS points and 20 AIS points, and regression using a tiered NSA structure used every possible bandwidth between five AIS points and ten AIS points.

Graphical results of these density tests are contained in Figures 3.2.A, 3.2.B, and 3.2.C. Figure 3.2.A uses the years 2009-2011, Figure 3.2.B uses the years 2005-2008, and Figure 3.2.C uses the year 2004. These graphs largely support a causal interpretation of our RD results. In Figure 3.2.A, three of the four graphs show either no statistically significant discontinuity or one that is only significant at very small bandwidths (11 RAI points or fewer). As smaller bandwidths are much more susceptible to noise, and since our estimates tend to rise sharply in magnitude as bandwidths become very small, this is not a great threat to our RD estimates. To the extent that it does affect our estimates, the fact that the statistically significant discontinuities are positive may imply downward bias, strengthening our RD results, though the precise effect depends on the type of students attracted as well as their number. A more significant concern is the top-right graph, which shows a noticeable discontinuity in the number of applicants from states other than Illinois at the cutoff for NSA eligibility. It is not clear why this discontinuity is negative – some students who learn that they are eligible for the NSA may be tempted to apply to more selective schools, but this should affect the number of students who enroll rather than the number who apply. Regardless, a decrease in the number of applicants at the cutoff may bias our results for this subgroup upwards. While this will not automatically be the case, it is worth noting when we present our RD results.

Figure 3.2.B contains no statistically significant discontinuities at any of the bandwidths we use. This bodes particularly well for our RD estimates in 2005-2008. Figure 3.2.C, however, contains a large number of discontinuities at a large number of bandwidths. The clearest pattern is that applicants from Illinois, White applicants, and applicants to the College of Liberal Arts are all discontinuously more likely to apply at the middle cutoff for NSA eligibility. Students may also be less likely to apply from states other than Illinois or to the College of Liberal Arts at the lowest cutoff for NSA eligibility. That said, the magnitude of our results makes it difficult to attach any particular interpretation to them, especially since

Figure 3.2.C appears to contain more noise than Figures 3.2.A and 3.2.B. However, it is worth noting that these discontinuities exist and may affect our RD estimates – though only at the low and middle tiers of NSA eligibility – in 2004.

The largest threat to the validity of our estimates is that students may manipulate their AIS or RAI in order to qualify for the NSA. Fortunately, the margins along which high school seniors may affect these scores are fairly small and involve some degree of error. For example, students may retake the ACT in hopes of obtaining a higher score. However, students' ACT scores are subject to uncertainty – it is unclear whether they will improve their score and, if they do, by what margin they will improve. Other inputs for AIS and RAI are even harder to manipulate. Substantially improving one's GPA or class rank is a challenging proposition. By the time students apply to college, they have already received three years' worth of grades and have only one more year to improve their (cumulative) GPA. Even if they do improve their GPA, it is unclear whether this will result in an improvement in their class rank.[67] While particularly motivated students may take additional courses in hopes of improving their RAI, it is unclear that many students are likely to be this proactive and strategic. Even among this subset of students, additional courses may prove overwhelming and may lower some students' GPAs and class rankings. Given these difficulties in manipulating score indices, we are relatively confident that students' score indices are "as good as" random.

Nonetheless, students who improve their ACT score or class rank will see their AIS and RAI rise, and those who increase their GPA or class rank or who

[67] Altering one's class rank is made all the more difficult by the fact that class rank is an ordinal measure, not a cardinal one. If one's peers are similarly motivated and all improve their GPAs by the same amount, their rank ordering will remain unchanged.

take more courses will see their RAI rise. Similarly, those who do poorly in their senior year will see those metrics fall. In these cases, NSA receipt is based on students' highest RAI or AIS scores – students may become eligible for the NSA by performing well, but will not lose their award if they perform poorly. As we observe students' scores at the time of application, some students who appear ineligible for the NSA will still receive the award, although the reverse does not hold. As a result, the probability of NSA receipt increases continuously as admissions scores increase and then jumps to one at the cutoff. This necessitates using a FRD design, but does not invalidate our results. The fact that some students who do not appear eligible will receive the NSA and may therefore be induced to enroll at Iowa may bias our results towards zero. If so, our results will represent a lower bound of the impact of NSA eligibility on enrollment decisions.

IV. Data

The data for our analysis consists of admissions and financial aid records from the University of Iowa. Eight cohorts of students are included, beginning with applicants for the fall 2004 semester and continuing through applicants for fall 2011. The admissions data includes information on students' state or country of residence, admissions score, enrollment decision, and race/ethnicity, as well as the college to which they applied and the semester for which they applied.[68] The financial aid data indicate which scholarships (if any) a student received, but they do not provide scholarship amounts. As a result, we can determine the average

[68] Additional data, such as students' ACT scores or transfer credits, are available and may be used in future extensions or robustness checks.

effect of receiving the "National Scholars Award" for each of the eight cohorts, but are not currently able to provide an estimate of the NSA (or any other aid) elasticity of enrollment.

We restrict our data on a number of dimensions. First, we exclude students who do not have a recorded value for the relevant admissions score. Next, our analysis is restricted to new entering freshmen applicants because the NSA is awarded only to freshmen enrolling directly after high school. As a result, transfer applicants and those with applications in multiple years are not relevant to our analysis and are therefore dropped from the data. We also dropped students who applied to multiple colleges within the same year. For example, students who apply to both the College of Liberal Arts and Sciences and the College of Engineering have two applications on record but can enroll in only one college. If one such student were to receive the NSA and enroll in the College of Engineering, both application records would reflect NSA eligibility but only the application record for the College of Engineering would indicate that the student enrolled. As a result, this student would make the NSA appear less attractive to applicants than it actually is (at least within the College of Liberal Arts). If this student instead chose not to enroll at either school, it is unclear whether we should count their decision twice in the data (since both colleges failed to attract the applicant) or once (since one applicant chose not to enroll).[69] Furthermore, such applicants may be disproportionately likely to enroll, as their application profile implies a willingness to attend their second-choice college within the University of Iowa. Because these applicants make up 1.2 percent of the applicant pool (and 2.3

[69]If we count this applicant only once, we also have to determine which application to consider her primary application, or whether to somehow combine multiple application profiles.

percent of applications), dropping them from our analysis should not undermine the validity of our results.

Table 3.1 provides summary statistics for the data used in our analysis. We divide our data into three time periods based on how the NSA is awarded and provide statistics on the pooled sample and for each time period.

Our overall sample consists of 119,381 applicants, with 13,487 applying in 2004, 56,385 applying in 2005-2008, and 49,509 applying in 2009-2011. 35,470 students received the NSA over the eight years of our sample – 3,126 in 2004, approximately 3,750 annually from 2005-2008, and approximately 5,750 annually from 2009-2011. While RAI scores are available only in the final period, AIS scores remained relatively stable – the mean AIS value was 121.9 in 2004, 122.8 in 2005-2008, and 123.8 in 2009-2011.

Approximately 70 percent of the students in our sample are from states other than Iowa. Of these, approximately two-thirds are from Illinois, ten percent are international, seven percent are from Minnesota, and four percent are from Wisconsin. Missouri, California, and Nebraska are the only other states with over one percent of the out-of-state applicants in our data. Approximately 90 percent of all applicants applied to the College of Liberal Arts and Sciences, partly because the Tippie College of Business did not admit freshmen until fall 2008. Approximately two percent applied to the business school, eight percent to the College of Engineering, and under one-half of one percent applied to the College of Nursing.[70] Just over seventy-five percent of all applicants are listed as White, four and a half percent as Asian-American, and approximately four percent each

[70]If we restrict our analysis to the 2008 cohort and later, just over three percent apply to Business, eight percent to Engineering, and one-half of one percent to Nursing.

as African-American or Hispanic. International students, who comprise nearly eight percent of our sample, are listed as a separate ethnicity, making it difficult to assign true ethnicities to these students.

Strong regression discontinuity estimates require large amounts of data near the cutoff for receiving treatment. Having too little data could undermine valid results in a variety of ways. First, if there was (hypothetically) exactly one student with an RAI value of 290 (the cutscore), the percentage of students enrolling with an RAI of 290 would appear in our data as either zero (if she did not enroll) or one (if she did). Neither represents the actual probability that a random student with an RAI of 290 will enroll. This problem is largely solved by pooling multiple admission cohorts, ensuring that enough students are at each RAI position to provide an accurate probability of enrollment.

The second potential problem arising from small sample sizes has to do with optimal bandwidth selection. If our estimates are based on using a very small number of data points, our results will be very susceptible to statistical "noise." If we were to use only RAI values of 288, 289, 290, and 291 in our analysis, noise at any of these four values will have an extremely large effect on our estimates. An unexpectedly large enrollment rate among students with an RAI of 289 could then make the impact of the NSA appear extremely small (or even negative). It is therefore important to have enough students near the cutoff for NSA eligibility to reduce the chances of obtaining spurious results.

While we would optimally prefer to estimate regression discontinuity models on many subsets of the data, the above sample size issues prevent us from doing so. For example, we are unable to estimate the effect of NSA receipt on subgroups such as applicants to the Tippie School of Business or Asian-American applicants because there are too few such students, especially near the NSA threshold, to make accurate inferences. We are, however, able to separately examine the effect of NSA receipt on nonresident applicants, applicants from Illinois, applicants pooled from all states other than Illinois or Iowa, White applicants, and applicants to the College of Liberal Arts.

Our estimates of the effect of the tiered NSA regime on enrollment may not reflect the true impact of the NSA in 2004. This is partly due to sample size issues; we have only one year of data in which the NSA was awarded at a tiered rate compared to four years of data when the NSA was awarded at a flat rate based on the AIS and three years in which it was awarded on a flat rate based on RAI. Our estimates in 2004 may therefore be less stable than those in later years. Our difficulty in estimating the effects of the NSA for 2004 is further exacerbated by the tiered nature of the NSA in this year. If we use too large a bandwidth, the data points we use will capture not only individuals near the cutoff for eligibility for a particular NSA tier but also individuals around the cutoffs for higher or lower NSA tiers. For instance, if we examine individuals far from the cutoff for the lowest NSA tier, some of these individuals will have qualified not only for the low-level NSA, but also for the medium-level NSA. The inclusion of students in both the lowest and next highest tier would likely produce spurious results.

V. Results

We find evidence that eligibility for the NSA does increase the probability that an admitted applicant will enroll at the University of Iowa. This finding holds when conducting subgroup analysis on students from Illinois, from other states, on White students, and on applicants to the College of Liberal Arts. These results are stronger when using larger bandwidths, but hold for some narrower bandwidths as well. Falsification tests using in-state students, who are not eligible for the NSA, reveal no statistically significant effect, implying that our estimates are due to NSA receipt rather than to some unobserved factor.

Graphical evidence of the effect of NSA receipt is displayed in Figures 3.3, 3.4, and 3.5. Figure 3.3 indicates the discontinuity in enrollment at the NSA cutoff using a bandwidth of 30 RAI points and pooling over three admissions cycles (2009, 2010, and 2011) in which the NSA was awarded at a flat rate based on RAI. Reading across rows, the graphs provide a description of the enrollment yield among 1) all nonresident applicants, 2) applicants from Illinois, 3) nonresident applicants from states other than Illinois, 4) in-state applicants, 5) nonresident White applicants, and 6) nonresident applicants to the College of Liberal Arts. Figure 3.4 presents an analogous set of graphs, using a bandwidth of 20 AIS points and pooling over four admissions cycles (2005, 2006, 2007, and 2008) in which the NSA was awarded at a flat rate based on AIS. Figure 3.5 estimates are for the one available year in which NSA receipt was tiered (2004) and uses a bandwidth of only 10 AIS points to avoid interference from discontinuities at neighboring award tiers. Slopes are allowed to differ on opposite sides of the NSA cutoffs, and the graphs present unconditional estimates (given the particular subgroup in consideration). RAI and AIS scales in Figure 3.3 and Figure 3.4 are normalized so that the cutoff is assigned a value of zero. In Figure 3.5, since there are three NSA tiers, we opted not to normalize AIS values.

A number of patterns emerge. First, there is a visible jump in the predicted probability of enrollment in five of six graphs in both Figures 3.3 and 3.4. As these graphs represent various samples of nonresident students, they provide evidence that NSA eligibility does induce enrollment. The bottom-left graphs in these figures, representing enrollment rates among Iowa residents, do not display a noticeable jump at the cutoff for NSA eligibility. As the NSA is awarded only to nonresidents, these bottom-left graphs serve as an important falsification test. If these two graphs had indicated a jump in enrollment rates at the NSA cutoff, it would be impossible to attribute this shift to the NSA and would raise the question of whether some other factor is affecting enrollment rates. This would pose a serious threat to the interpretation of the other five graphs. However, this does not appear to be the case. Since nonresidents display a noticeable jump in enrollment at the cutoff for NSA eligibility while Iowa residents do not, we are able to interpret this jump in probability as the effect of the NSA.

Figure 3.5, which displays the predicted probability of enrollment in 2004, is harder to interpret. There appear to be different effects at the cutoffs for each NSA tier, and these effects vary substantially from graph to graph within Figure 3.5. There is a lack of a consistent effect on enrollment at the cutoff for the lowest level of NSA eligibility. This could be due to heterogeneous responses to small amounts of aid. For example, if White applicants tend to be wealthier than non-White applicants, a low amount of aid would intuitively have a smaller impact on White applicants than on non-White applicants. Alternatively, inconsistencies in the impact of NSA eligibility could be due to small sample sizes and noise. The largest sample used in creating the local linear estimates in these graphs is 2,131, an average of fewer than 200 students per AIS point. If ten students at a particular AIS value unexpectedly choose to enroll, this would mean a shift of over five percentage points in the enrollment yield at that AIS value.

There are two main reasons why small sample sizes would pose problems in Figure 3.5 but not in Figure 3.4. First, in Figure 3.4, we pool across four years of data, meaning that a fixed amount of noise in our data will have approximately one-fourth the effect in Figure 3.4 than it would in Figure 3.5. Assuming that we have four times as many applicants at each AIS value, having ten students with the same AIS value unexpectedly enroll would increase the enrollment yield at that value by only 1.25 percentage points rather than five percentage points. Second, since Figure 3.4 uses only a single cutoff, we are able to use larger bandwidths. A local linear regression using 20 AIS values will be better able to withstand noise at any one AIS value than a similar regression using only 10 AIS values.

There is a drop in the predicted probability of enrollment at the cutoff for the medium level of NSA eligibility. This drop appears in all six graphs within Figure 3.5. Without data on scholarship amounts, it is hard to interpret this effect. While students with higher admissions scores will be more likely to receive awards or admission at other institutions, these increased likelihoods should not cause discontinuities in the probability of enrollment at Iowa. Even if many scholarships and admissions offers become available at a single RAI level, it seems curious that this would coincidentally happen at the cutoff for the medium level of NSA eligibility. As the NSA does not crowd out other awards, students at this cutoff should receive more financial aid than students below the cutoff – and it seems implausible that the causal effect of receiving additional scholarship money would be to discourage students from enrolling. Even if receiving this award makes students more likely to apply to preferred institutions (perhaps due to a boost in confidence), it seems highly unlikely that the net effect of a scholarship would be to drive students away. A more likely interpretation is that the additional award money at this cutoff induces more students who are marginally interested in the University of Iowa to apply – an interpretation backed by many of our density checks. If these students enroll at lower rates than the general nonresident population, the result may be a drop in enrollment yield. Alternatively, the drop in the predicted probability of enrollment may be due to small sample sizes and noise. Additional data would help determine which of these two factors is causing the drop in enrollment.

At the cutoff for the highest level of NSA eligibility, however, there appears to be a nearly uniform positive jump in the predicted probability of enrollment. As students with higher RAI values will generally have larger choice sets of colleges and are more likely to be admitted to a school that they prefer to the University of Iowa (ceteris paribus), financial aid may play a disproportionately large role in determining which school these students attend. Curiously, students from states other than Illinois (and, of course, Iowa) do not appear to react very strongly to the highest level of eligibility. It is unclear exactly why these students would be less responsive to this tier of NSA eligibility – perhaps they were more likely in 2004 to receive scholarships from schools in their respective states, or perhaps Illinois students were better informed of the value of the NSA and applied specifically with the award in mind. Obtaining data from additional years in which the NSA was awarded at a tiered rate would also reduce the possibility that these results are due to small sample sizes.

Tables 3.2.A, 3.2.B, and 3.2.C contain the point estimates and standard deviations associated with the discontinuities in Figures 3.2, 3.3, and 3.4. These estimates confirm the intuition presented above – that the NSA does have a statistically significant impact on enrollment decisions and has stronger effects on minority applicants and those from states other than Illinios. Figure 3.3 is drawn from column (4) in Table 3.2.A, Figure 3.4 drawn from column (4) in Table 3.2.B, and Figure 3.5 is drawn from columns (2), (4), and (6) in Table 3.2.C. As these tables reflect information contained in the graphs, they do not contain any additional regressors. Although some columns have bandwidths too small for reliable estimates, larger bandwidths produce statistically significant discontinuities. Whereas large bandwidths may reduce variance at the expense of introducing bias, the discontinuities in Figures 3.3 and 3.4 do not appear to contain substantial bias relative to those from using narrow bandwidths. The one case in which student behavior appears to change substantially at larger bandwidths is the top-right graph in Figure 3.3, which represents non-resident students from states other than Illinois in the admissions cycles from 2009-2011.

The top rows of Tables 3.2.A and 3.2.B indicate that NSA receipt increases nonresident students' probability of enrollment by five to seven percentage points – a notable increase from a baseline of approximately 25 percent. Students from states other than Illinois are affected between 1.5 and 3.5 percentage points more than those from Illinois. This may be because students from Illinois enroll at a higher baseline rate prior to receiving the NSA, making marginal enrollees from Illinois quite different from those from other states. The magnitude of this effect in Figure 3.2.A may be due to the drop in applications from applicants from states other than Illinois at the cutoff for NSA eligibility, but the fact that a (smaller) gap appears in Figure 3.2.B suggests that these students are indeed more receptive to NSA receipt. The impact of NSA receipt on White students is approximately 0.75 percentage points lower than for nonresidents overall, which makes sense if White students are better able to afford college to begin with. The impact on applicants to the College of Liberal Arts relative to all nonresident applicants is unclear – Liberal Arts applicants in 2005-2008 are more affected by NSA receipt than the full nonresident sample, while those in 2009-2011 are less so.

The fourth row of Tables 3.2.A, 3.2.B, and 3.2.C provide another version of the falsification check discussed above. Since Iowa residents are not eligible for the NSA, we estimate the first stage of 2SLS on out-of-state students (as in the first row) and use the predicted probability of treatment obtained from stage one in our second stage estimation. The resulting second stage estimates are both close to zero and statistically insignificant, meaning that nonresident students show no effects from crossing the cutoff for NSA eligibility. This lends credibility to the causal interpretation of our estimates – that it is the NSA causing an increase in the probability of enrollment for nonresidents and not some unobserved factors.

Estimates in Table 3.2.C, representing the 2004 admissions cycle, are largely statistically insignificant, most likely because sample sizes are too small for accurate inference. Columns (3) and (4) provide further evidence of a negative discontinuity at the cutoff for the medium level of NSA eligibility. This may be partly due to the fact that most density checks show an increase in applicants at this threshold, but the actual magnitude is unclear. If these additional applicants are only marginally interested in attending the University of Iowa, this could explain the apparent negative impact of the NSA at this point. The second entries in columns (5) and (6) imply that Illinois residents react particularly strongly to receiving the highest level of NSA. Given the small sample sizes used in Table 3.2.C, however, these results are more susceptible to bias and/or imprecision than those in Tables 3.2.A and 3.2.B. The implications of these estimates are therefore less clear than those presented earlier.

We also conducted regression discontinuity analysis using additional covariates. By doing so, we hope to both control for sampling variation and to lower the error associated with our treatment estimates. In doing so we allow the level and slope of the probability of enrollment to vary according to student characteristics. We include three types of variables among these additional covariates: indicator variables, single interactions, and double interactions. Indicators such as $ILLINOIS_i$ take on a value of one if applicant i is from Illinois and zero otherwise. Single interactions such as $ILLINOIS_i * RAI_i$ are equal to applicant i's RAI if she is from Illinois and zero otherwise. Double interactions such as $ILLINOIS_i * \mathbf{1}(RAI_i \geq 290) * RAI_i$ are equal to applicant i's RAI if she is from Illinois and has an RAI of at least 290 and zero otherwise.

The covariates used here fall into three broad categories. The first contains indicators, single interactions, and double interactions for students applying to the Tippie College of Business, the College of Engineering, and the College of Nursing (with the College of Liberal Arts as the omitted category). The second category contains indicators, single interactions, and double interactions for respondents who reported their ethnicity as Native American, African-American, Hispanic, Asian-American, International, or Pacific Islander, as well as those who reported no ethnicity (with White as the omitted category). The third category consists of indicators, single interactions, and double interactions for students whose home state is Illinois, Minnesota, Wisconsin, Nebraska, California, or Missouri (with any other state as the omitted category).[71]

Results using these additional covariates are presented in Tables 3.3.A, 3.3.B,

[71]Only international students did not have a state of residence listed. As these students are already captured in the ethnicity variables, we do not include variables for students who did not list a state of residence.

and 3.3.C. Whereas point estimates in these tables are slightly lower than in the results not including covariates, there is very little change in statistical significance. In Table 3.3.A, including additional covariates reduces the impact of NSA eligibility by over half a percentage point, from 6.95 to 6.27. Including covariates for Illinois residents reduces the size of the impact (in percentage terms) slightly more; lowering the effect of NSA eligibility from 5.99 to 5.29. The impact on the other subgroups is smaller – the impact of NSA eligibility on non-Illinois residents falls to 9.24, that on White applicants falls to 5.84, and that on Liberal Arts applicants falls to 5.46. The impact on non-Illinois residents and on White applicants falls by approximately a third of a percentage point, whereas that on Liberal Arts applicants falls by just over one-tenth. Curiously, in Table 3.3.B, adding covariates has a noticeable impact only on non-Illinois residents and Liberal Arts applicants – all other point estimates remain very close to their values in Table 3.2.B.

VI. Conclusion

Employing a fuzzy regression discontinuity design on data from the University of Iowa, we find evidence that students who receive the National Scholars Award are significantly more likely to enroll than their counterparts who do not receive this award. Evidence is mixed on whether students from Illinois are affected more than nonresident students from other states. The predicted probability that White students enroll at the University of Iowa rises by less than that of non-White students. Most importantly, falsification tests do not reveal any statistically significant effects among (ineligible) Iowa residents, lending credibility to the interpretation of our results as causal average treatment effects. These results are fairly consistent whether the NSA is awarded based on RAI or on AIS, though its precise impact on students from Illinois versus those from other states is unclear. Data from 2004 is too limited for strong inference, but it does appear that students at the highest tier of eligibility were significantly more likely to enroll at the University of Iowa than students with lower AIS scores.

This information can be of use to admissions officers and other policy makers at the university. First, our findings suggest that the award does achieve its stated goal of increasing nonresident student enrollment. However, since many states have had to reduce funding for their universities, it may make sense to target awards more narrowly in order to save money. Given the NSA's large impact on enrollment, and given preliminary evidence from 2004 that this effect applies to the highest-achieving students, it may be possible to reweight award amounts in such a way that the University of Iowa both attracts extremely high-achieving students and reduces its NSA expenditures. For example, rather than having a $4,000 scholarship that is offered to over 40 percent of nonresident applicants,[72] it might make sense for Iowa to target two specific groups of students. The first is extremely high-achieving students. By reducing the number of students eligible for the NSA and/or returning to a tier system, Iowa could potentially reduce its scholarship outlays while simultaneously offering more attractive aid packages to elite applicants. Remaining funding could be better targeted toward financially needy students or to minority students, who are more responsive to NSA receipt than more affluent or White students.[73] This would allow Iowa to

[72]In recent years, more of these scholarships have been awarded – in 2010, over 47 percent of applicants were offered the NSA.

[73]Alternatively, it could be used as need-based aid, as ethnicity is almost certainly correlated with

simultaneously focus on recruiting a talented and diverse student body, while reducing the amount of money spent in doing so. Future research could include policy simulations to determine an optimal award structure, and experiments could actually test whether this optimal structure works in practice.

There are several other possible extensions of this work. The first involves computing financial aid elasticities of enrollment. As data on scholarship values become available, we will incorporate it into extensions of this line of inquiry. Another extension involves using these elasticities to simulate counterfactual aid policies. While our current results suggest qualitative adjustments to Iowa's financial aid policies may be in order, we cannot yet recommend exactly how much aid to allocate to each group of students. With some additional assumptions and data it would be possible to suggest a range of possible actions and to predict how they might impact Iowa's enrollment yield and financial aid outlays.

Finally, one flaw in using RAI to evaluate students is that many students do not have valid RAI values. In particular, many high schools do not provide their students' class rank. While admissions scores like the RAI are designed to sacrifice some intensity of evaluation in favor of administrative efficiency, metrics that do not apply to large swaths of students fail to either sacrifice intensity or achieve efficiency. As a result, there should be some mechanism by which students who are missing a component of RAI may be quickly evaluated. Design of such a mechanism would preserve the intent of RAI while greatly reducing the amount of time spent on student evaluation.

---

financial need and it may be that need, rather than ethnicity *per se* is causing differential responses to NSA receipt across ethnic groups. Some research on this topic is possible using self-reported family income values from SAT and ACT questionnaires, but FAFSA or similar data would provide more reliable estimates.

While students may have idiosyncratic tastes for college selectivity, location, size, and any additional number of factors, they ought to have a limited willingness to pay for each of these things. By systematically awarding merit aid to high-achieving students, the University of Iowa has been able to attract a large number of students who might not have otherwise enrolled. Due to its structured nature, the NSA has allowed us a window into student decision-making, and may be of great use in setting future education policy and in attracting an optimal mix of students to the University of Iowa.

Figure 3.1: Award Rate Comparison

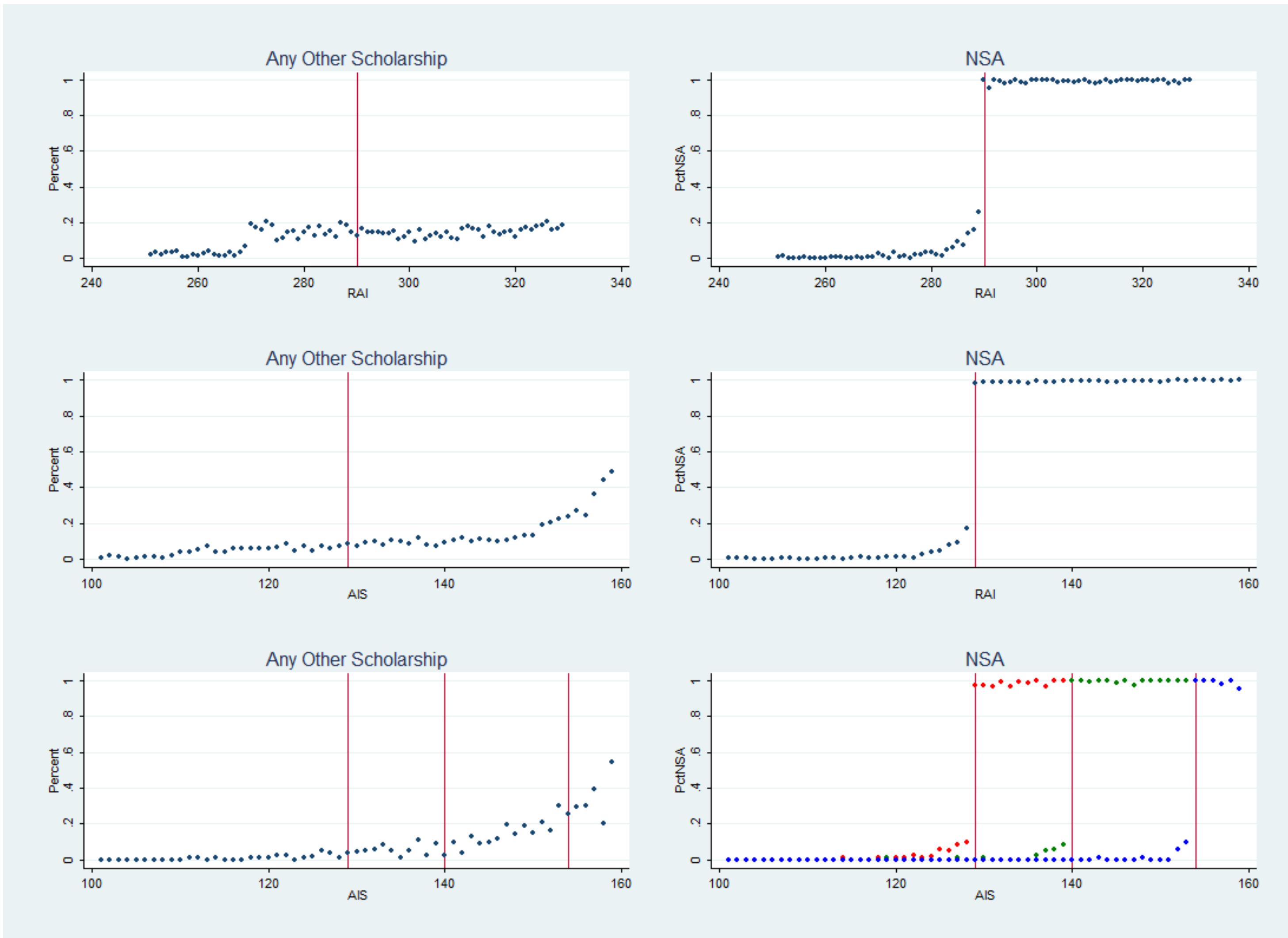


Figure 3.2.A: McCrary Density Tests, 2009-2011

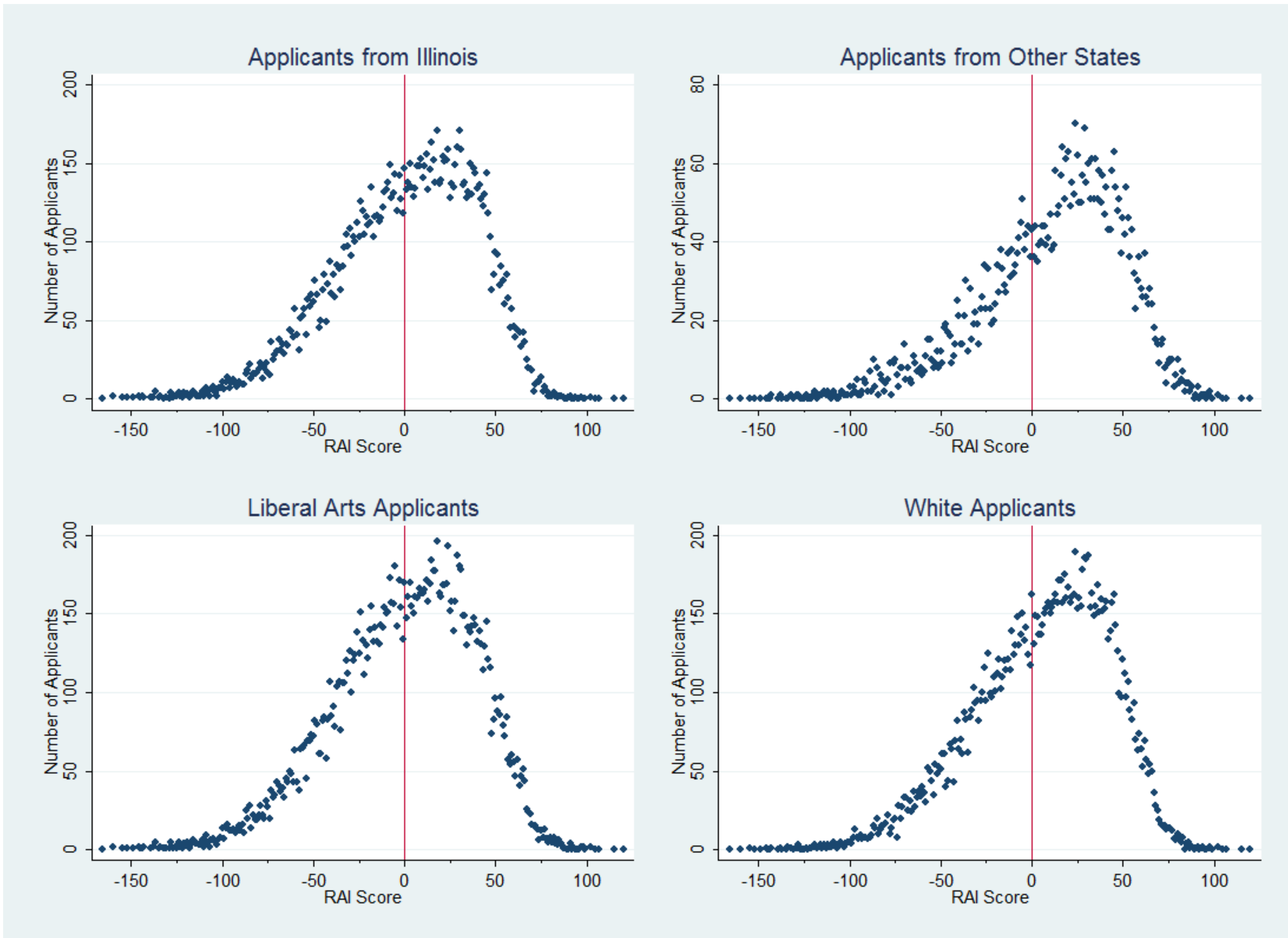


Figure 3.2.B: McCrary Density Tests, 2005-2008

Figure 3.2.C: McCrary Density Tests, 2004

Figure 3.3: RAI Graphs, Bandwidth 30

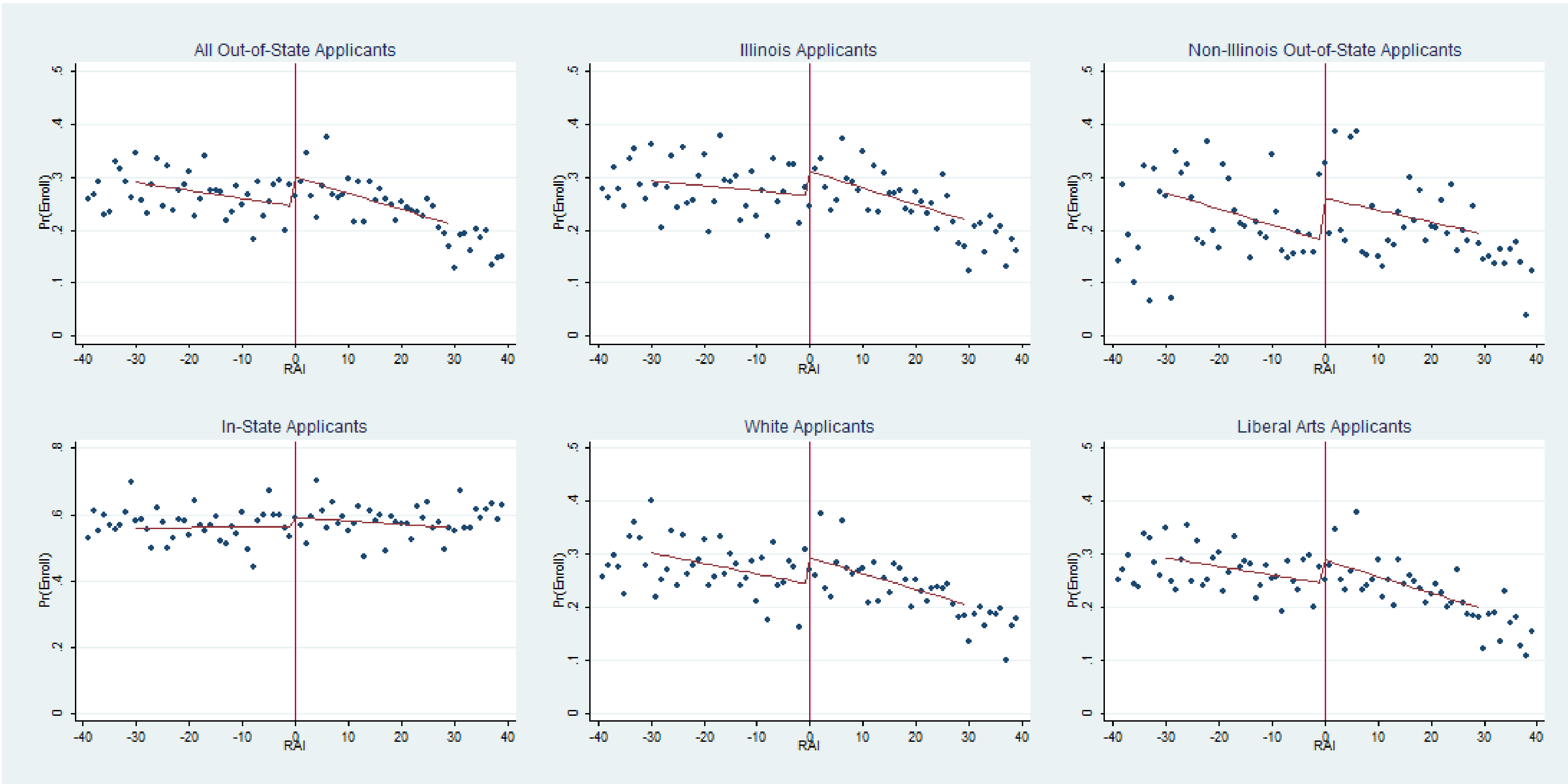


Figure 3.4: AIS Flat Rate Graphs, Bandwidth 20

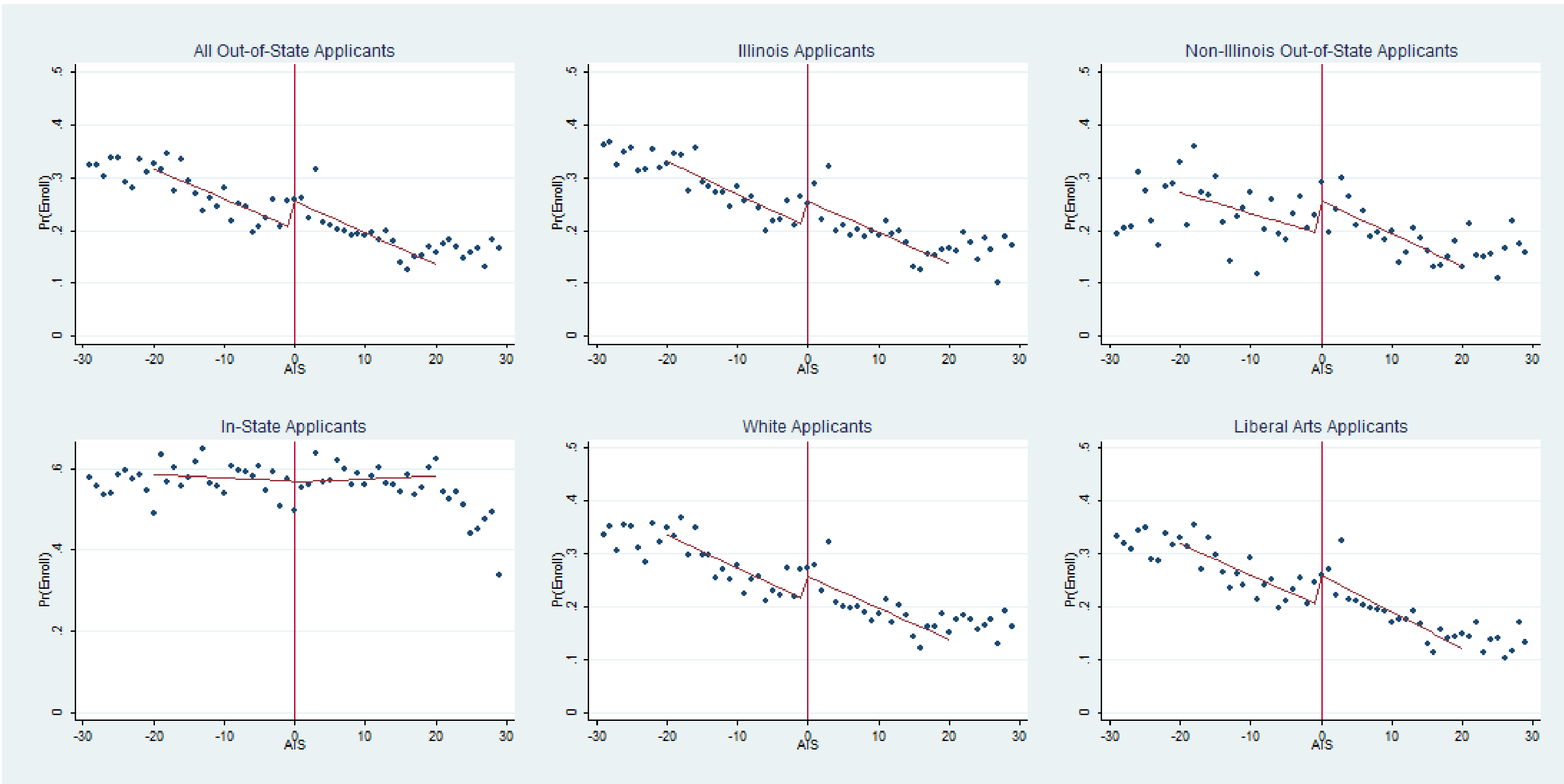


Figure 3.5: AIS Tiered Graphs, Bandwidth 10

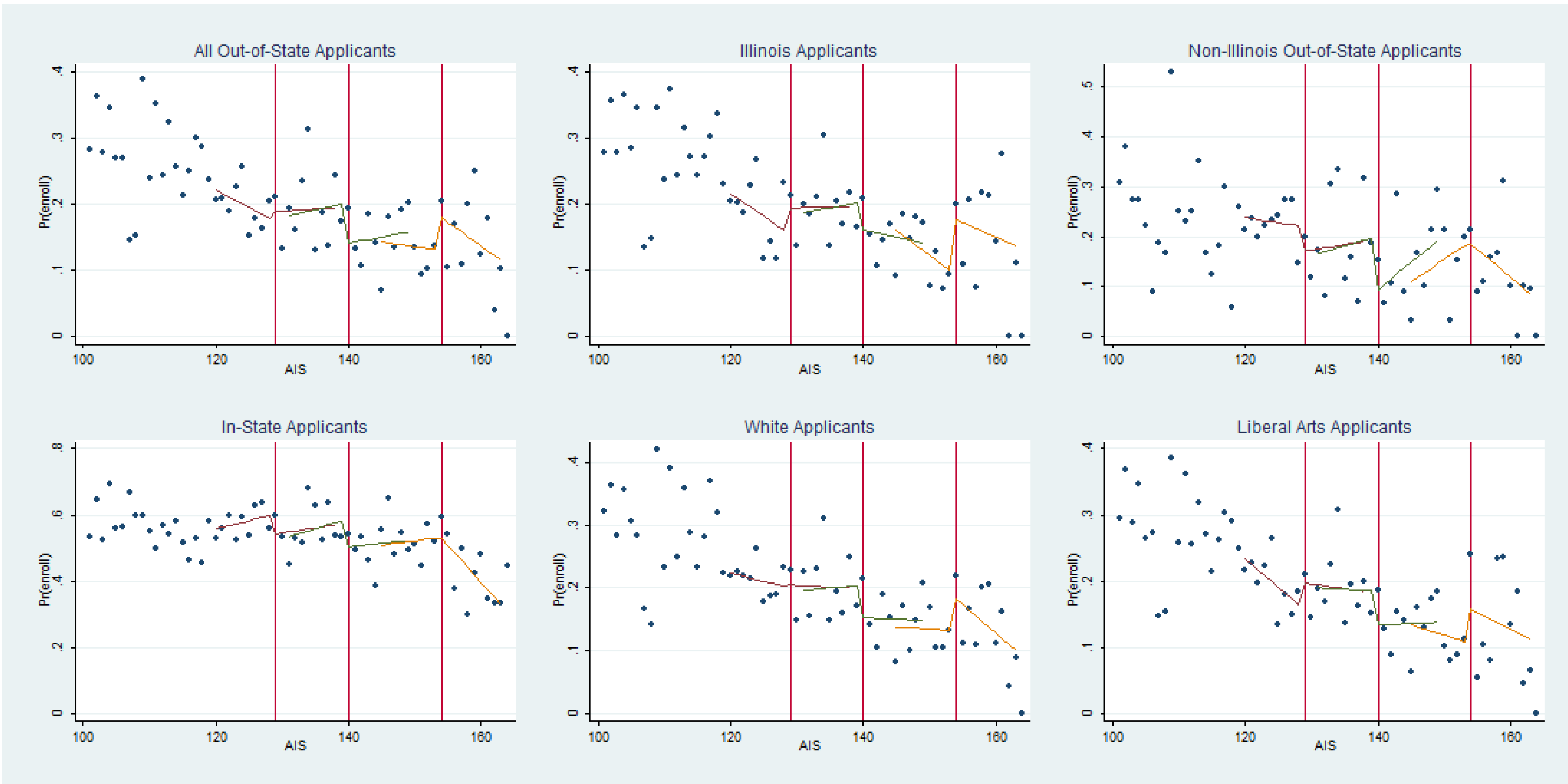


Table 3.1: Summary Statistics

	2004	2005-2008	2009-2011	Overall
N(Applicants)	13,487	56,385	49,509	119,381
N(Nonresident)	8,288	38,091	37,197	83,576
N(AnyNSA)	3,126	15,066	17,278	35,470
AIS Score	121.9	122.8	123.8	123.1
	(2.39)	(23.6)	(24.3)	(23.9)
RAI Score	-	-	294.6	294.6
			(38.1)	(38.1)
Liberal Arts	0.923	0.913	0.877	0.899
Business	0.004	0.014	0.031	0.020
Engineering	0.070	0.072	0.087	0.078
White	0.837	0.808	0.693	0.763
Black	0.040	0.034	0.041	0.037
Hispanic	0.027	0.033	0.044	0.037
Asian	0.048	0.045	0.043	0.045
International	0.022	0.040	0.136	0.078
Iowa	0.386	0.324	0.249	0.300
Illinois	0.424	0.462	0.448	0.452
Minnesota	0.047	0.053	0.052	0.052
Missouri	0.010	0.009	0.009	0.009
Nebraska	0.012	0.009	0.007	0.008
Wisconsin	0.029	0.031	0.029	0.030

Table 3.2: RD Estimates Without Additional Regressors

Table 3.2.A: 2009-2011 Pooled

Sample	BW = 5		BW = 10		BW = 20		BW = 30	
	1st Stage	2SLS	1st Stage	2SLS	1st Stage	2SLS	1st Stage	2SLS
Non-IA	0.7158	0.0478	0.7641	0.0253	0.8398	0.0566*	0.8704	0.0695***
	(0.0348)	(0.0666)	(0.0305)	(0.0435)	(0.0278)	(0.0298)	(0.0238)	(0.0242)
	$N = 1930$		$N = 3737$		$N = 7241$		$N = 10624$	
IL Res.	0.7234	0.0519	0.7639	-0.0231	0.8381	0.0370	0.8681	0.0599**
	(0.0315)	(0.0700)	(0.0270)	(0.0491)	(0.0263)	(0.0342)	(0.0224)	(0.0284)
	$N = 1482$		$N = 2885$		$N = 5544$		$N = 8124$	
Non-IL, IA	0.6921	0.0409	0.7617	0.1756*	0.8424	0.1090*	0.8765	0.0958**
	(0.0482)	(0.1208)	(0.0441)	(0.0904)	(0.0353)	(0.0547)	(0.0308)	(0.0424)
	$N = 448$		$N = 852$		$N = 1697$		$N = 2500$	
IA Res.	0.7158	0.0737	0.7641	-0.0099	0.8398	0.0293	0.8704	0.0303
	(0.0348)	(0.0430)	(0.0305)	(0.0552)	(0.0278)	(0.0378)	(0.0238)	(0.0303)
	$N_1 = 1930$, $N_2 = 907$		$N_1 = 3737$, $N_2 = 1766$		$N_1 = 7241$, $N_2 = 3488$		$N_1 = 10624$, $N = 5092$	
White	0.7328	0.0482	0.7824	0.0268	0.8570	0.0539	0.8852	0.0617***
	(0.0361)	(0.0985)	(0.0308)	(0.0610)	(0.0272)	(0.0372)	(0.0229)	(0.0292)
	$N = 1529$		$N = 2951$		$N = 5737$		$N = 8437$	
Liberal Arts	0.7212	0.0316	0.7741	0.0211	0.8475	0.0468	0.8766	0.0559***
	(0.0310)	(0.0640)	(0.0296)	(0.0437)	(0.0271)	(0.0301)	(0.0233)	(0.0253)
	$N = 1734$		$N = 3334$		$N = 6446$		$N = 9376$	

Table 3.2.B: 2005-2008 Pooled

Sample	BW = 5		BW = 10		BW = 15		BW = 20	
	1st Stage	2SLS	1st Stage	2SLS	1st Stage	2SLS	1st Stage	2SLS
Non-IA	0.8068	0.0210	0.8540	0.0524*	0.8817	0.0487**	0.9001	0.0587***
	(0.0241)	(0.0289)	(0.0233)	(0.0279)	(0.0220)	(0.0222)	(0.0205)	(0.0202)
	$N = 4660$		$N = 8794$		$N = 12646$		$N = 16070$	
IL Res.	0.8048	0.0217	0.8597	0.0554	0.8894	0.0476*	0.9065	0.0543**
	(0.0244)	(0.0382)	(0.0252)	(0.0325)	(0.0238)	(0.0257)	(0.0218)	(0.0226)
	$N = 3438$		$N = 6554$		$N = 9416$		$N = 11952$	
Non-IL, IA	0.8098	0.0184	0.8378	0.0397	0.8597	0.0506*	0.8813	0.0704**
	(0.0254)	(0.0464)	(0.0208)	(0.0395)	(0.0188)	(0.0298)	(0.0181)	(0.0275)
	$N = 1222$		$N = 2240$		$N = 3230$		$N = 4118$	
IA Res.	0.8068	-0.0054	0.8540	-0.0016	0.8817	0.0131	0.9001	-.0029
	(0.0241)	(0.0543)	(0.0233)	(0.0411)	(0.0220)	(0.0323)	(0.0205)	(0.0282)
	$N_1 = 4660$, $N_2 = 2589$		$N_1 = 8794$, $N_2 = 4776$		$N_1 = 12646$, $N_2 = 6883$		$N_1 = 16070$, $N_2 = 8777$	
White	0.8112	0.0276	0.8584	0.0460	0.8880	0.0420*	0.9061	0.0513**
	(0.0262)	(0.0318)	(0.0244)	(0.0291)	(0.0228)	(0.0244)	(0.0212)	(0.0229)
	$N = 3874$		$N = 7286$		$N = 10510$		$N = 13379$	
Liberal Arts	0.8194	0.0378	0.8620	0.0642**	0.8877	0.0577**	0.9050	0.0652***
	(0.0243)	(0.0281)	(0.0219)	(0.0283)	(0.0205)	(0.0220)	(0.0191)	(0.0201)
	$N = 4287$		$N = 8095$		$N = 11524$		$N = 14549$	

Table 3.2.C: 2004

Sample	BW = 5, Low		BW = 10, Low		BW = 5, Med		BW = 10, Med		BW = 5, High		BW = 10, High	
	1st Stage	2SLS	1st Stage	2SLS	1st Stage	2SLS	1st Stage	2SLS	1st Stage	2SLS	1st Stage	2SLS
	0.8590	-0.0133	0.8829	0.0170	0.8975	-0.0476	0.9286	-0.0647*	0.9020	0.0556	0.9392	0.0527
Non-IA	(0.0053)	(0.0613)	(0.0098)	(0.0332)	(0.0044)	(0.0475)	(0.0118)	(0.0389)	(0.0185)	(0.0621)	(0.0240)	(0.0382)
	$N = 1169$		$N = 2131$		$N = 1064$		$N = 1963$		$N = 622$		$N = 1198$	
	0.8747	0.0113	0.8970	0.0442	0.9103	-0.0221	0.9312	-0.0455	0.9590	0.0894*	0.9768	0.0849**
IL Res.	(0.0286)	(0.0948)	(0.0200)	(0.0487)	(0.0138)	(0.0403)	(0.0095)	(0.0350)	(0.0151)	(0.0462)	(0.0111)	(0.0404)
	$N = 862$		$N = 1552$		$N = 742$		$N = 1401$		$N = 399$		$N = 778$	
	0.8244	-0.0817	0.8485	-0.0561	0.8703	-0.1140	0.9211	-0.1183*	0.8123	-0.0005	0.8748	-0.0074
Non-IL, IA	(0.0722)	(0.0923)	(0.0399)	(0.0632)	(0.0242)	(0.0817)	(0.0301)	(0.0658)	(0.0499)	(0.1168)	(0.0535)	(0.0650)
	$N = 307$		$N = 579$		$N = 322$		$N = 562$		$N = 223$		$N = 420$	
	0.8590	-0.1007	0.8829	-0.0649	0.8975	0.0060	0.9286	-0.0902*	0.9020	0.0202	0.9392	-0.0034
IA Res.	(0.0053)	(0.0795)	(0.0098)	(0.0487)	(0.0044)	(0.0340)	(0.0118)	(0.0460)	(0.0185)	(0.0520)	(0.0240)	(0.0580)
	$N_1 = 1169,\ N_2 = 734$		$N_1 = 2131,\ N_2 = 1409$		$N_1 = 1064,\ N_2 = 750$		$N_1 = 1963,\ N_2 = 1378$		$N_1 = 622,\ N_2 = 512$		$N_1 = 1198,\ N_2 = 950$	
	0.8349	-0.0277	0.8711	0.0051	0.8905	-0.0319	0.9259	-0.0569	0.8879	0.0811	0.9397	0.0538
White	(0.0071)	(0.0610)	(0.0140)	(0.0343)	(0.0081)	(0.0525)	(0.0125)	(0.0410)	(0.0175)	(0.0579)	(0.0245)	(0.0366)
	$N = 972$		$N = 1791$		$N = 925$		$N = 1693$		$N = 525$		$N = 1025$	
	0.8636	0.0218	0.8906	0.0479	0.9085	-0.0197	0.9334	-0.0552	0.9143	0.0433	0.9459	0.0547
Liberal Arts	(0.0064)	(0.0572)	(0.0087)	(0.0306)	(0.0103)	(0.0367)	(0.0125)	(0.0337)	(0.0258)	(0.0895)	(0.0251)	(0.0545)
	$N = 1079$		$N = 1971$		$N = 962$		$N = 1763$		$N = 525$		$N = 1022$	

Table 3.3: RD Estimates With Additional Regressors

Table 3.3.A: 2009-2011 Pooled

Sample	BW = 5		BW = 10		BW = 20		BW = 30	
	1st Stage	2SLS	1st Stage	2SLS	1st Stage	2SLS	1st Stage	2SLS
Non-IA	0.7158	0.0493	0.7641	0.0171	0.8398	0.0523*	0.8704	0.0627**
	(0.0348)	(0.0679)	(0.0305)	(0.0437)	(0.0278)	(0.0290)	(0.0238)	(0.0237)
	$N = 1930$		$N = 3737$		$N = 7241$		$N = 10624$	
IL Res.	0.7234	0.0621	0.7639	-0.0280	0.8381	0.0339	0.8681	0.0529*
	(0.0315)	(0.0710)	(0.0270)	(0.0486)	(0.0263)	(0.0327)	(0.0224)	(0.0274)
	$N = 1482$		$N = 2885$		$N = 5544$		$N = 8124$	
Non-IL, IA	0.6921	0.0263	0.7617	0.1585*	0.8424	0.1032*	0.8765	0.0924**
	(0.0482)	(0.1148)	(0.0441)	(0.0854)	(0.0353)	(0.0533)	(0.0308)	(0.0416)
	$N = 448$		$N = 852$		$N = 1697$		$N = 2500$	
IA Res.	0.7158	0.0620	0.7641	0.0082	0.8398	0.0325	0.8704	0.0330
	(0.0348)	(0.0422)	(0.0305)	(0.0519)	(0.0278)	(0.0367)	(0.0238)	(0.0298)
	$N_1 = 1930$, $N_2 = 907$		$N_1 = 3737$, $N_2 = 1766$		$N_1 = 7241$, $N_2 = 3488$		$N_1 = 10624$, $N = 5092$	
White	0.7328	0.0523	0.7824	0.0282	0.8570	0.0512	0.8852	0.0584**
	(0.0361)	(0.0931)	(0.0308)	(0.0591)	(0.0272)	(0.0361)	(0.0229)	(0.0284)
	$N = 1529$		$N = 2951$		$N = 5737$		$N = 8437$	
Liberal Arts	0.7212	0.0325	0.7741	0.0078	0.8475	0.0425	0.8766	0.0546**
	(0.0310)	(0.0681)	(0.0296)	(0.0443)	(0.0271)	(0.0295)	(0.0233)	(0.0244)
	$N = 1734$		$N = 3334$		$N = 6446$		$N = 9376$	

Table 3.3.B: 2005-2008 Pooled

Sample	BW = 5		BW = 10		BW = 15		BW = 20	
	1st Stage	2SLS	1st Stage	2SLS	1st Stage	2SLS	1st Stage	2SLS
	0.8068	0.0228	0.8540	0.0496*	0.8817	0.0476**	0.9001	0.0586***
Non-IA	(0.0241)	(0.0287)	(0.0233)	(0.0275)	(0.0220)	(0.0221)	(0.0205)	(0.0200)
	$N = 4660$		$N = 8794$		$N = 12646$		$N = 16070$	
	0.8048	0.0218	0.8597	0.0554	0.8894	0.0502*	0.9065	0.0573**
IL Res.	(0.0244)	(0.0395)	(0.0252)	(0.0330)	(0.0238)	(0.0260)	(0.0218)	(0.0226)
	$N = 3438$		$N = 6554$		$N = 9416$		$N = 11952$	
	0.8098	0.0147	0.8378	0.0336	0.8597	0.0440	0.8813	0.0650**
Non-IL, IA	(0.0254)	(0.0404)	(0.0208)	(0.0367)	(0.0188)	(0.0290)	(0.0181)	(0.0266)
	$N = 1222$		$N = 2240$		$N = 3230$		$N = 4118$	
	0.8068	-0.0056	0.8540	-0.0027	0.8817	0.0145	0.9001	0.0019
IA Res.	(0.0241)	(0.0546)	(0.0233)	(0.0394)	(0.0220)	(0.0317)	(0.0205)	(0.0273)
	$N_1 = 4660,\ N_2 = 2589$		$N_1 = 8795,\ N_2 = 4776$		$N_1 = 12647,\ N_2 = 6883$		$N_1 = 16071,\ N_2 = 8777$	
	0.8112	0.0260	0.8584	0.0434	0.8880	0.0228*	0.9061	0.0515**
White	(0.0262)	(0.0333)	(0.0244)	(0.0294)	(0.0228)	(0.0247)	(0.0212)	(0.0227)
	$N = 3874$		$N = 7286$		$N = 10510$		$N = 13379$	
	0.8194	0.0407	0.8620	0.0605**	0.8877	0.0536**	0.9050	0.0625***
Liberal Arts	(0.0243)	(0.0264)	(0.0219)	(0.0273)	(0.0205)	(0.0216)	(0.0191)	(0.0198)
	$N = 4287$		$N = 8095$		$N = 11524$		$N = 14549$	

Table 3.3.C: 2004

Sample	BW = 5, Low		BW = 10, Low		BW = 5, Med		BW = 10, Med		BW = 5, High		BW = 10, High	
	1st Stage	2SLS	1st Stage	2SLS	1st Stage	2SLS	1st Stage	2SLS	1st Stage	2SLS	1st Stage	2SLS
	0.8590	-0.0194	0.8829	0.0138	0.8975	-0.0330	0.9286	-0.0528	0.9020	0.0490	0.9392	0.0566
Non-IA	(0.0053)	(0.0659)	(0.0098)	(0.0355)	(0.0044)	(0.0528)	(0.0118)	(0.0399)	(0.0185)	(0.0724)	(0.0240)	(0.0389)
	$N = 1169$		$N = 2131$		$N = 1064$		$N = 1963$		$N = 622$		$N = 1198$	
	0.8747	0.0178	0.8970	0.0429	0.9103	-0.0120	0.9312	-0.0315	0.9590	0.0640	0.9768	0.0893**
IL Res.	(0.0286)	(0.0981)	(0.0200)	(0.0510)	(0.0138)	(0.0406)	(0.0095)	(0.0340)	(0.0151)	(0.0519)	(0.0111)	(0.0408)
	$N = 862$		$N = 1552$		$N = 742$		$N = 1401$		$N = 399$		$N = 778$	
	0.8244	-0.1625	0.8485	-0.0754	0.8703	-0.0937	0.9211	-0.0990	0.8123	-0.0795	0.8748	-0.0297
Non-IL, IA	(0.0722)	(0.0912)	(0.0399)	(0.0551)	(0.0242)	(0.1118)	(0.0301)	(0.0748)	(0.0490)	(0.1499)	(0.0535)	(0.0577)
	$N = 307$		$N = 579$		$N = 322$		$N = 562$		$N = 223$		$N = 420$	
	0.8590	-0.0943	0.8829	-0.0674	0.8975	-0.0069	0.9286	-0.0919*	0.9020	0.0226	0.9392	-0.0203
IA Res.	(0.0053)	(0.0668)	(0.0098)	(0.0463)	(0.0044)	(0.0341)	(0.0118)	(0.0476)	(0.0185)	(0.0531)	(0.0240)	(0.0661)
	$N_1 = 1169,\ N_2 = 734$		$N_1 = 2131,\ N_2 = 1409$		$N_1 = 1064,\ N_2 = 750$		$N_1 = 1963,\ N_2 = 1378$		$N_1 = 622,\ N_2 = 512$		$N_1 = 1198,\ N_2 = 950$	
	0.8349	-0.0375	0.8711	0.0003	0.8905	-0.0276	0.9259	-0.0513	0.8879	0.0533	0.9397	0.0551
White	(0.0071)	(0.0690)	(0.0140)	(0.0359)	(0.0081)	(0.0559)	(0.0125)	(0.0424)	(0.0175)	(0.0675)	(0.0245)	(0.0422)
	$N = 972$		$N = 1791$		$N = 925$		$N = 1693$		$N = 525$		$N = 1025$	
	0.8636	0.0200	0.8906	0.0489	0.9085	-0.0049	0.9334	-0.0474	0.9143	0.0423	0.9459	0.0376
Liberal Arts	(0.0064)	(0.0585)	(0.0087)	(0.0331)	(0.0103)	(0.0404)	(0.0125)	(0.0361)	(0.0258)	(0.0760)	(0.0251)	(0.0495)
	$N = 1079$		$N = 1971$		$N = 962$		$N = 1763$		$N = 525$		$N = 1022$	

Chapter Three Bibliography

[49] Christopher Cornwell, David B. Mustard, and Deepa J. Sridhar. The Enrollment Effects of Merit-Based Financial Aid: Evidence from Georgia's HOPE Program. *Journal of Labor Economics*, 24(4):761–786, 2006.

[50] Christopher M. Cornwell, Kyung Hee Lee, and David B. Mustard. Student Responses to Merit Scholarship Retention Rules. *Journal of Human Resources*, 40(4):895–917, 2005.

[51] Christopher M. Cornwell, Kyung Hee Lee, and David B. Mustard. The Effects of State-Sponsored Merit Scholarships on Course Selection and Major Choice in College. Institute for the Study of Labor (IZA) Discussion Paper Series No. 1953, January 2009.

[52] Susan Dynarski. Hope for Whom? Financial Aid for the Middle Class and Its Impact on College Attendance. *National Tax Journal*, 53(3):629–662, September 2000.

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

[54] Pietro Garibaldi, Andrea Ichino, Francesco Giavazzi, and Enrico Rettore. College Cost and Time to Complete a Degree: Evidence from Tuition Discontinuities. Posted Online by Review of Economics and Statistics, 2011.

[55] Brad Hershbein. *Worker Signals Among New College Graduates: The Role of Selectivity and GPA*. PhD thesis, University of Michigan, October 2011.

[56] Guido W. Imbens and Thomas Lemieux. Regression Discontinuity Designs: A Guide to Practice. *Journal of Econometrics*, 142(2):615–635, 2008.

[57] David Lee. Randomized Experiments from Non-Random Selection in U.S. House Elections. *Journal of Econometrics*, 142(2):675–697, 2008.

[58] Justin McCrary. Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test. *Journal of Econometrics*, 142(2):698–714, 2008.

[59] Wilbert Van Der Klaauw. Estimating the Effect of Financial Aid Offers on College Enrollment: A Regression Discontinuity Approach. *International Economic Review*, 43(4):1249–1287, November 2002.

Appendix: Flagship Characteristics

School Name	UGPop	ANP	SAT	ACT	SAT25M	SAT75M	SAT25V	SAT75V	ACT25	ACT75	μ_{SAT}	σ_{SAT}
U. of Alabama	24,882	$16,255	1,200	4,447	500	650	500	620	22	29	1135	200.15
U. of Arizona	30,592	$12,185	4,816	2,999	490	620	480	610	21	27	1100	192.74
U. of Arkansas	17,247	$9,987	1,247	4,101	520	630	500	610	23	28	1130	163.09
U. of California-Berkeley	25,540	$15,589	3,894	1,567	630	760	600	730	27	32	1360	192.74
U. of Colorado-Boulder	26,648	$18,377	2,869	4,436	540	650	520	630	24	28	1170	163.09
U. of Connecticut	17,345	$14,877	3,080	865	580	670	550	640	25	29	1220	133.43
U. of Delaware	17,507	$13,376	4,144	1,265	560	660	540	650	24	28	1205	155.67
U. of Florida	32,660	$11,579	4,404	1,953	590	690	570	670	24	30	1260	148.26
U. of Georgia	25,947	$9,693	4,281	2,624	560	670	560	660	25	29	1225	155.67
U. of Hawaii-Manoa	13,912	$10,484	1,524	470	510	610	480	580	21	25	1090	148.26
U. of Idaho	9,573	$13,253	772	1,143	490	610	480	600	20	26	1090	177.91
U. of Illinois-UC	31,540	$15,610	1,761	5,967	690	780	540	660	26	31	1335	155.67
Indiana U.-Bloomington	32,367	$10,342	5,541	4,067	540	650	510	630	24	29	1165	170.50
U. of Iowa	21,176	$14,245	395	4,078	540	685	450	630	23	28	1152.5	240.92
U. of Kansas	20,343	$14,768	-	3,522	-	-	-	-	22	28	1145	170.50
U. of Kentucky	19,927	$12,916	834	4,048	500	630	490	620	22	28	1120	192.74
Louisiana State U.	23,685	$10,629	767	4,700	530	650	510	630	23	28	1160	177.91
U. of Maine	9,183	$15,075	1,713	222	480	600	480	590	21	27	1075	170.5
U. of Maryland-College Park	26,922	$13,625	3,685	-	610	710	580	680	-	-	1290	148.26
U. of Massachusetts-Amherst	21,373	$16,145	4,433	898	560	650	530	630	24	28	1185	140.85
U. of Michigan-Ann Arbor	27,027	$14,074	2,147	4,977	670	770	630	730	29	33	1400	148.26
U. of Minnesota-Twin Cities	33,607	$16,019	817	4,882	610	740	540	690	25	30	1290	207.56
U. of Mississippi	14,159	$12,516	1,100	3,038	470	590	460	590	20	27	1055	185.33
U. of Missouri-Columbia	24,834	$15,759	575	5,880	530	650	530	650	23	28	1180	222.39
U. of Montana	13,335	$13,937	854	1,271	490	600	490	600	21	26	1090	163.09
U. of Nebraska-Lincoln	19,383	$13,108	346	3,877	520	670	510	660	22	28	1180	222.39
U. of Nevada-Reno	14,185	$14,127	2,211	1,370	470	590	470	580	20	26	1055	170.50
U. of New Hampshire	12,458	$18,439	2,805	13	510	610	490	590	-	-	1100	148.26
Rutgers U.-New Brunswick	30,351	$15,905	5,839	-	560	680	520	630	-	-	1195	170.50
U. of New Mexico	22,476	$10,272	725	3,198	480	620	480	620	19	25	1100	207.56
Binghamton U.	14,746	$14,031	2,186	680	620	700	580	670	26	30	1285	126.02
UNC-Chapel Hill	18,579	$11,952	3,762	1,362	610	710	590	700	27	31	1305	155.67
U. of North Dakota	11,139	$11,952	-	1,895	-	-	-	-	21	26	1090	148.26

Appendix (cont.)

School Name	UGPop	ANP	SAT	ACT	SAT25M	SAT75M	SAT25V	SAT75V	ACT25	ACT75	μ_{SAT}	σ_{SAT}
Ohio State U.	42,082	$19,082	2,996	6,057	600	700	540	660	26	30	1250	163.09
U. of Okahoma-Norman	19,779	$15,289	1,436	3,444	540	655	510	650	23	29	1177.5	189.03
U. of Oregon	19,528	$14,669	3,158	608	499	611	492	607	22	27	1104.5	168.28
Pennsylvania State U.	38,594	$21,342	5,851	966	560	670	530	630	25	29	1195	155.67
U. of Rhode Island	13,093	$14,667	2,825	478	490	590	470	570	21	26	1060	148.26
USC-Columbia	21,385	$15,487	2,926	2,506	560	650	530	630	24	29	1185	140.85
U. of South Dakota	7,220	$12,961	51	1,159	460	580	460	550	20	26	1025	155.67
U. of Tennessee	21,392	$13,758	848	3,977	530	640	520	640	24	29	1165	170.50
U. of Texas-Austin	38,420	$14,629	6,118	3,776	580	710	540	670	25	31	1250	192.74
U. of Utah	23,371	$8,773	444	2,829	513	650	490	630	21	27	1141.5	205.34
U. of Vermont	11,593	$13,864	2,061	789	550	640	540	640	24	29	1185	140.85
U. of Virginia	15,595	$11,590	3,197	1,129	630	740	610	720	28	32	1350	163.09
U. of Washington-Seattle	29,307	$8,739	4,927	1,752	580	700	520	650	24	30	1225	185.33
West Virginia U.	22,303	$8,344	2,297	2,638	480	580	460	560	21	26	1040	148.26
U. of Wisconsin	30,170	$14,940	1,226	5,185	620	740	530	670	26	30	1280	192.74
U. of Wyoming	10,079	$10,776	236	1,340	490	610	490	600	21	27	1095	170.50

All data computed for the 2010-2011 entering class. UGPop = Undergraduate Population; ANP = Average Net Price for entering in-state students awarded Title IV financial aid; SAT = number of enrollees submitting an SAT score; ACT = number of enrollees submitting an ACT score; SAT(##)j = (##)th percentile of enrollee scores on SAT section j; ACT(##) = (##)th percentile of enrollees' composite ACT scores; $\mu_{\text{SAT}} = \frac{\text{SAT25M+SAT75M+SAT25V+SAT75V}}{2}$; $\sigma_{\text{SAT}} = \frac{mathrmSAT75M+SAT75V-\mu_{\text{SAT}}}{\Phi^{-1}(0.75)}$. Alaska is not listed because its state flagship is open admission; Washington, DC is not listed because it does not have a flagship university. While New York technically does not have a flagship, it has an extensive system of public universities; as a result, use Binghamton University, the most university with the highest SAT interquartile range among enrollees, as the de facto flagship.