Top Percent Plans and Student Beliefs about STEM Ability

Ryan Mather*

October 8, 2024

Abstract

Top percent plans give automatic admission to students who excelled relative to their peers in high school, but do not take into account how they perform relative to students in their state as a whole. If students use their relative performance to form beliefs about their own abilities, this suggests that students from relatively low-performing high schools who receive automatic admission may tend to overestimate their abilities. This paper suggests a new method of evaluating the effects of the top ten percent rule in Texas by estimating a student's class rank. I find that all students who receive top ten admission become more likely to initially major in STEM and to persist in their STEM degrees. However, if a high school student does better compared to their peers than they would at a different high school, this does decrease the benefit of receiving top ten admission for persisting in STEM. One possible explanation for this is that students can get biased views from their time in high school about how their abilities compare to other students, and then become discouraged when surrounded by potentially higher-performing peers in college.

1 Introduction

In the case *Students for Fair Admissions v. Harvard*, the Supreme Court ruled that race-based affirmative action programs are unconstitutional. This raises the question of how else a college could encourage diversity on its campus if it were not allowed to directly consider race in this way. One prominent race-neutral alternative is a top percent plan, which gives guaranteed admission at certain universities to students whose high school class ranks are within some highest percent.

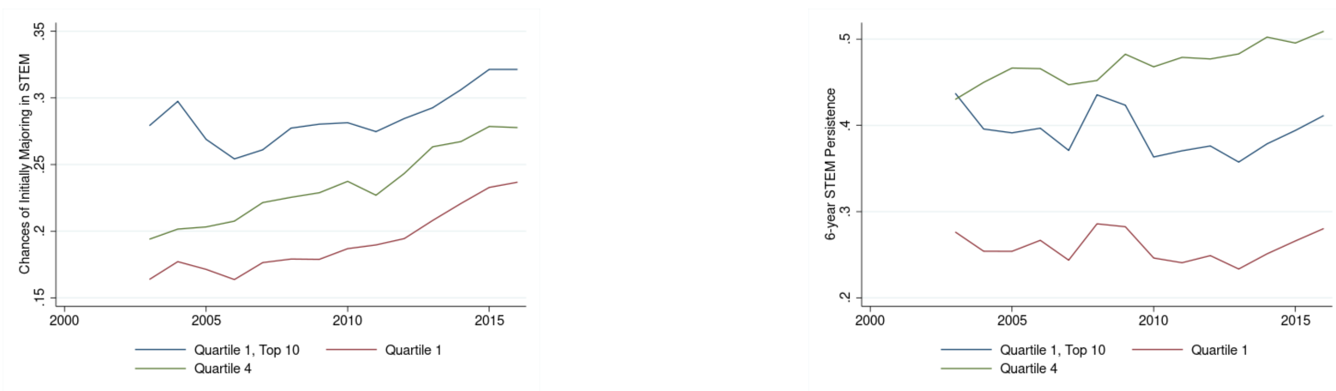
This is a unique form of affirmative action, because all of the disadvantaged students that these plans help come from high schools where they outperformed their peers, as demonstrated by being near the top of their graduating classes. Compared to an affirmative action policy that helps the highest performing students in certain demographic groups no matter what high schools they came from, top percent plans will admit more students who performed well compared to their peers and less of the high-performing students who did worse compared to their peers. If students use their relative performance in high school as a measurement of their

*I am grateful to have received feedback on this draft of the paper from Matt Wiswall, Chris Taber, Chao Fu, and participants at the labor reading group. I received helpful feedback on earlier versions of this idea from Paulo Martellini, Rasmus Lentz, Corina Mommaerts, and Naoki Aizawa. Conversations with Andrew Smith and Kensuke Sakamoto were also very helpful. Funding for purchasing the data was provided by the University of Wisconsin Madison. The conclusions of this research do not necessarily reflect the opinions or official position of the Texas Education Agency, the Texas Higher Education Coordinating Board, the Texas Workforce Commission or the State of Texas.

own ability, this could make it so that top percent plans will give admission to students with more upward bias in their beliefs about their own abilities.

This is especially important for STEM because persistence rates in STEM are low, and survey evidence suggests that students generally overestimate their chances of graduating with a STEM degree. Specifically, 19.8% of students enter college believing that a degree in science or math will be their most likely outcome, and yet only 7.4% of students actually graduate with one (Stinebrickner and Stinebrickner, 2014). In some contexts, students which do not persist in STEM at an upper-tier university may have persisted at lower-ranked universities (Bleemer (2020), Arcidiacono et al. (2016),Black et al. (2022)).[1] This motivates the research question for this paper, which is how top percent plans and a student's beliefs affect their chances of success in STEM programs.

Figure 1 presents suggestive evidence that some students who receive admission from the top ten percent rule in Texas may overestimate their chances of graduating with a STEM degree. I divided Texas public high schools into four quartiles based on their math standardized test scores in 2011, where quartile four has the highest test scores and quartile one has the lowest. The first graph shows that students from the lowest quartile who qualify for top ten admission are more likely to initially major in STEM than students from the highest quartile. However, conditional on initially majoring in STEM, they are less likely than students from the highest quartile to persist in a STEM degree and graduate within 6 years. Section 2.2 gives more details on how these figures are constructed.



(a) Chances of Initially Majoring in STEM

(b) 6-year STEM Persistence

Figure 1: The Chances of Majoring in STEM and Persisting in STEM until Graduating
Note: The second panel is plotted just over students who initially major in STEM

Given that students with class ranks above the ten percent cutoff receive the benefit of the top ten rule, a regression discontinuity approach seems like a promising way to find out how this rule might work together with student beliefs. Unfortunately, the data does not include a measure of class rank. Instead, for students who applied to public universities within the state of Texas, it records whether they were admitted based on the top ten percent rule. Black et al. 2022 get around this issue by using a differences-in-differences approach around the beginning of the top ten rule, and predict who among the students admitted under the top ten rule would not have been admitted otherwise using machine learning methods. Some limitations of this

[1] Bleemer (2020) looks at students who received automatic admission from a top percent plan in California between 2001 and 2011, and shows that they are somewhat less likely to earn degrees in STEM fields. Arcidiacono et al. (2016) look at an earlier affirmative action policy in California, and finds that some disadvantaged students with lower academic preparation who were admitted into top ranked California universities under the program would have been more likely to persist in STEM if they had attended lower ranked universities. Black et al. (2022) finds a similar effect for the top ten percent rule in Texas, but it is not statistically significant.

approach are that it requires a major shift in policy like the beginning of the top ten rule, and also that it relies upon a parallel trends assumption. Daugherty et al. (2014) solved this issue by looking at information from a single school district in Texas for which they had class rank information, and use a regression discontinuity approach. The main limitation here is that there are over 1000 school districts in Texas, and so it would be good to have a broader measurement of the effects.

In order to evaluate the effects of the top ten rule over the entire state of Texas at a time other than the beginning of the rule, I use a different method and attempt to predict a student's class rank directly. The two main challenges with this approach are that I only had access to student grades for a single school year when preparing these results,[2] and that different school districts have different methods of aggregating student grades into class ranks. To solve the first problem, I assume that the process by which grades are determined does not change from year to year, and use a regression to predict grades for other years based on the single year in which I observed grades. Then, to predict class rank, I search over a set of possible rules that the school district could be using to aggregate grades, and pick the rule which maximizes the likelihood of students falling into the correct "bin" of class rank based upon the admissions decisions of public universities in Texas.

I additionally prepare estimates of grade inflation at high schools by using a method that compares a student's class grade with their score on a standardized test meant to evaluate that class's material. These standardized tests are uniform across the state of Texas, and in this way provide a measure of student ability that is unaffected by high-school specific grade inflation. If students at a high school consistently receive higher class grades than their standardized test scores would suggest, for example, then that school will be found to inflate their student's grades. I find that grade inflation is inversely related with course ability. In other words, schools where students have lower ability tend to give higher grades for the same level ability. This is consistent with what would be expected if high schools graded on a curve,[3] and is interesting in light of research which shows that students can be influenced to continue in STEM or Economics by higher letter grades, even if this higher letter grade does not actually reflect higher ability (e.g., see Ahn et al. (2024), McEwan et al. (2021), and Owen (2010)). If this is true, then the high school grades that students from lower performing schools receive may further contribute to STEM over-confidence through grade inflation.

With estimates of student's grades and class ranks prepared, I use a matching method to estimate the effects of the top ten percent rule.The method tries to compare students that have the same course performance and demographic characteristics, but only one of them received top ten admission because of the different peer groups surrounding each student and the methods that their high schools use to find GPAs.[4]

I find that students who receive automatic admission from the top ten percent rule are more likely to initially major in STEM and to persist in STEM. This appears to hold even for students with the lowest levels of academic preparation in high school. One thing that helps to explain these results is that students who receive automatic admission are more likely to attend one of the flagship universities in Texas, and students at these universities are more likely to major in STEM and persist in STEM than students at other public Texas universities.

[2] I have since gained access to some grade data in an additional school year, which I will intend to use in future revisions of this paper.

[3] This curve could be a formal grade policy by a teacher, or simply the result of less explicit responses by teachers like offering more extra credit or grading less stringently when their students are performing at lower levels.

[4] Assuming that I were able to get the class rank exactly correct for some portion of students around the top ten percent cutoff, another viable strategy would be to use a fuzzy regression discontinuity approach. The first stage of this method was not strong enough for this method to work properly, but section 5 shows the results of attempting it.

I find evidence that if a student attended Texas A&M university as a result of the top ten percent rule, their chances of initially majoring in STEM go up by 25.49 percentage points.

For student's beliefs, I find evidence that a student's performance relative to their peers in high school influences their later decisions in college. No matter how well a high school student does compared to their peers, receiving top ten admission seems to provide similar benefits for how likely that student is to initially major in STEM. However, if a student does better compared to their peers than they would have at a different high school, this does decrease the benefit of receiving top ten admission for persisting in STEM. One possible explanation for this is that if students do better compared to their high school peers than they would compared to students from Texas more broadly, they can become overconfident about how their STEM abilities will compare to other students in college. When they do get to college, then, they may become discouraged and not persist in their STEM degree.

1.1 Related Literature

This paper contributes to literature evaluating the effects of the top ten percent rule in Texas (e.g. Black et al. (2022) and Daugherty et al. (2014)) and top percent plans more generally (e.g. Bleemer (2020)). As described above, the method that I use allows me to estimate the effects of the top ten percent rule at a time after its start and to use data from Texas as a whole, which was not done in these previous papers on Texas. My paper may be closest in focus to Kapor (2020), who also looks at the top ten percent rule in Texas, and estimates how much of the effect of this program comes from how it directly informs students that they *will* be admitted. This is a unique feature of Top Percent plans compared to more standard affirmative action policies, which may raise a student's chances of being admitted but do not provide information about guaranteed admission to any student. Where Kapor (2020) studies the effects of Top Percent plans on a student's beliefs about college admission, I will study the interaction between Top Percent plans and student's beliefs about their own ability, which may be influenced by their ability relative to their peers. If a top percent rule is found to lower a student's chances of graduating in STEM, it is not always clear whether this is because the student anticipates that STEM will be difficult and does not attempt it, or attempts it but does not persist. In this paper I will look at whether a student initially majors in STEM and persists in STEM as separate outcomes to help distinguish between these two stories.

There is also literature which tries to understand how student's form beliefs about their own abilities and how those beliefs affect their choices between majors. Some papers like Stinebrickner and Stinebrickner (2014) and Wiswall and Zafar (2015) use direct information on student's beliefs from surveys. While I do not have survey information, the ERC data includes detailed information on the majors that students at Texas schools have declared in every semester, and section 2.2 shows that a student's chances of persisting in STEM using this data are similar to the chances found by surveys. One advantage of ERC data over the surveys from those papers is that it allows more students to be tracked over a longer period of time. For example, since information on a high school experiences is included, it is possible to study how those experiences contributed to student beliefs.

This paper is more related to research that tries to estimate how student beliefs are formed using information on their grades and peer groups (e.g. Arcidiacono (2004) and Arcidiacono et al. (2021)). I do not know of any papers in this area that focus on the specific interaction between the beliefs that students form about their own abilities and top percent rules. Kapor (2020) is closest, but focuses on how the top ten percent rule directly affects students beliefs about college admissions, whereas this paper will look how receiving top ten admission interacts with the beliefs that students had already based on comparing themselves with their peers. Kinsler and Pavan (2021) show that parents and teachers tend to overestimate the performance of students who are surrounded by lower performing peers, and underestimate the performance of students who are surrounded by high performing peers. I am hoping to compliment these results by showing how a student's peer group can bias the student's own beliefs about their ability.

Finally, this paper fits with other papers that have looked at what determines persistence in STEM degrees and other high-earning fields. Bleemer and Mehta (2021) and Ahn et al. (2019) look at different ways that the grades that students earn in college can lower persistence in these fields. This paper will focus on how grades earned in high school can effect student's decisions in college by influencing their beliefs about their own ability. Arcidiacono et al. (2016) shows that in California, there was a time where students who were admitted to top schools because of affirmative action would have been more likely to persist in STEM if they had attended lower ranked schools. I do not find that result in Texas.

In section 2, I describe STEM interest and enrollment for Texas. The levels shown are closely aligned with survey evidence from the early 2000s measuring STEM interest and persistance in other areas of the country, and I track these trends forward to 2016. Then, section 3 describes the method used to predict a student's grades and class rank. Section 4 describes how these grades can be used to estimate the effects of the top ten percent rule, and section 5 shares the results of those estimates. Finally, section 6 discusses the results and describes next steps for future work.

2 Data and Setting

I am using data from the Texas Schools Project which combines high school, college, and employment records for students that attended attend public schools in Texas. Between 2009 and 2017, information is also available from the National Student Clearinghouse (NSC). An administrative data set like this has the advantage of providing a large sample size, including information for all students who went to public school within the state of Texas. Since Texas educates around 10% of all public school students in the country, this is a substantial sample at the national level as well.

The main disadvantage of administrative data is that it often does not provide any information on students who move outside of the state. For example, if someone graduates college with Texas and then does not show up in the employment records, I cannot see for sure whether they are unemployed or just took a job in a different state. Foote and Stange (2022) describe how this kind of bias can lead to understating the wage benefits of graduating from a flagship university by 10%. For the case of Texas, however, there are several things that can be done to address this problem. First, the NSC data is valuable because it describes students' college experiences if they attended colleges outside the state of Texas. This can be used to tell the difference between students who went to college elsewhere and students who never went to college. In addition, the NSC data set contains a variable that records student majors in college, and can be used to tell whether they majored in STEM.

Unfortunately, I cannot know for sure whether students who went to universities outside Texas persisted with their STEM degrees. The NSC data is only available up until 2017, which is not late enough to see whether students in the high school classes I am looking at graduated within 6 years with a STEM degree. It is possible, however, to provide some possible limits around how bad this bias can be. The results section below will look at two possible cases, one in which all students who initially major in STEM at colleges outside of Texas go on to complete STEM degrees within 6 years, and another case in which none of them do. These two cases will be used as lower and upper bounds for the possible effect of the top ten percent rule.

The way that missing wages for people outside Texas could be a problem is if people who leave to work outside Texas are different in unobservable ways from the people who stayed, and that this creates a bias if the people who left to work outside of Texas are simply omitted from the sample. Andrews et al. (2016) performed a check for this kind of bias using a sample of people who likely went to flagship universities in the state of Texas. Specifically, they used Census data for recent college graduates who had lived in the MSAs surrounding the Texas flagship universities five years earlier as a proxy for the group of people who graduated from those flagship universities. Then, the Census data allowed them to track the earnings of these graduates even if they moved outside of Texas, and they found that the distributions of in-state and out-of-state earnings were similar. This provides encouraging evidence that people who move outside of Texas are not receiving disproportionately higher or lower paying jobs. Foote and Stange (2022) directly confirmed their result that these two groups have similar earnings using national data from the U.S. Census Longitudinal Employer-Household Dynamics (LEHD).

This suggests that in the future, I should be able to address the problem of missing wage data by ignoring observations for people with missing wage data. For the results presented here, however, I go with the smaller goal of estimating effects on being employed in Texas and wages in Texas. In other words, people with missing wage data are not excluded from the sample, but are recorded as being not employed in Texas and earning no wages in Texas. These are not the main estimates that I will ultimately want in the long term, but they are relevant estimates for Texas policymakers seeking the economic well-being of Texas through the top ten percent rule.

2.1 The Top Ten Percent Rule

The top ten percent rule was created as an alternative to affirmative action. In 1996, the Fifth Circuit court ruled in *Hopwood v. Texas* that a students' race could not be considered for admissions decisions. Shortly afterword, in 1998, the top ten percent rule was passed. It guaranteed that any student within the top ten percent of their high school classes who applied for admission at a public Texas university would be accepted. Since some high schools have large populations of students from demographic groups that affirmative action often tries to help, the idea was that providing admission to the top ten percent of every high school class would provide admission to disadvantaged students in a way that is similar to affirmative action.

This rule largely remains to present day, even though in 2003 the Supreme Court overturned the *Hopwood v. Texas* decision and allowed race to be considered again for admission into college. One change to the rule is that the University of Texas at Austin has been allowed to slightly change the specific class rank threshold that guarantees admission. For the other flagship university in Texas, Texas A&M, the top ten percent rule remains unchanged.

A final change in Texas which I think may be relevant is that in 2013, Texas A&M Unversity began a program called "25 by 25" with the goal of increasing their total engineering student enrollment to 25,000 by 2025. This represents a massive change in a relatively short period of time. A full 50 percent of the student growth required to meet Texas A&M's goal is intended to come purely through increasing the retention of student in the College of Engineering, which illustrates how low STEM persistence rates might be if no action were taken.[5] Already, there are 50% more Texas A&M engineering students on campus than would have been in the absence of this policy.[6] To address this unique situation at Texas A&M, in the results section

[5] By contrast, only 14 percent of the growth is intended to occur by actually bringing new graduate and undergraduate students onto campus.

[6] My source for the information in this paragraph is a website for Texas A&M: https://engineering.tamu.edu/25by25/common-questions.html.

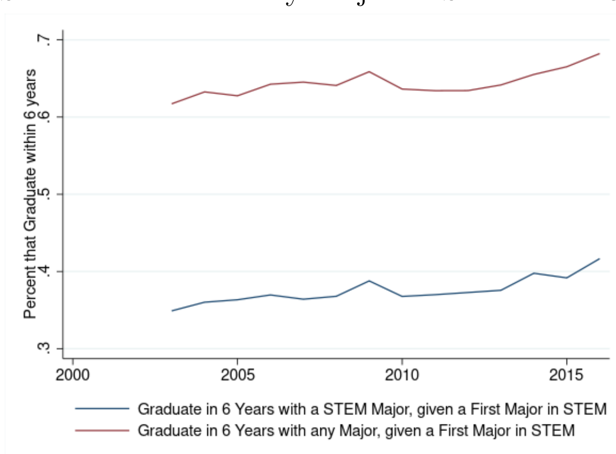
I will include a control for attending Texas A&M when I compare students that did and did not receive admission under the top ten percent rule.

2.2 Documenting STEM Interest and Persistance

In this section I describe the trends in STEM interest, measured by a student's first recorded major being in STEM, and STEM persistence, which here means initially majoring in STEM and then graduating with a STEM degree. My sample for these graphs is students who graduated from a public high school in Texas and by the next year had begun attending a 4-year Texas university. For most graphs I limit the sample to students whose first attended university was public, but figure 19 in the appendix shows these results in certain years for students at private universities. Student majors are recorded for each semester that the student is enrolled, and I record a student's "first majors" as their majors listed the first semester that their major no longer "undeclared" or missing.

A student "persists" in STEM if they initially major in STEM and then earn a STEM degree within 6 years.[7] In this section, the graphs related to persisting in STEM only include information from students who initially majored in STEM. Figure 2 shows that among students who initially major in STEM, only around 40% go on to graduate with STEM within 6 years. Most of of these students do still graduate within 6 years, though, showing that many students who do not persist in STEM switch their major to something else.

Figure 2: Chances that Students who Initially Major in STEM will Graduate Within 6 Years



One disadvantage of this data compared to a survey is that it does not record a student's originally intended degree path for certain. Students might enter college and then changed their mind before indicating their first major. For this reason, the fraction of people who persist in STEM degrees presented here is likely an upper bound, because people that did not persist to the point of recording a STEM major are not included. However, there are two reasons why the major used here still provides a good estimate of the originally intended major. First, majors are generally recorded for students very early on in college. For the data used in the estimation later on, table 5 will show that students are enrolled in college for just .35 semesters on average before having a recorded major. For students that have a STEM major recorded as their first major, this average is even lower at .02.[8]

[7] I measure this as earning a STEM degree within six years after graduating high school. These graphs only use information for students who had begun attending a 4-year university within a year of graduating high school, and that will also be a requirement to count as "persisting in STEM" for the results in section 5.

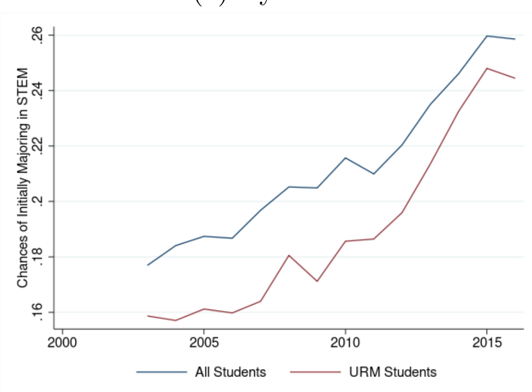
[8] This measurement only includes the Fall and Spring semesters. For example, if a student was enrolled for the

Second, the results shown here are in line with previous survey estimates. Stinebrickner and Stinebrickner (2014) surveyed students beginning college in 2000 and in 2001, and found that 31.1% of students with a stated intention to major in science actually graduated with a major in science. The National Academies of Sciences, Engineering, and Medicine (2016) report similarly found that in 2004, 40% of students who initally intended to major in STEM were able to graduate with a STEM degree within six years.[9] The persistence with STEM degrees shown in figure 2 is between these two estimates around this time. This provides encouraging evidence that the estimates presented here for Texas are externally valid for other parts of the country.

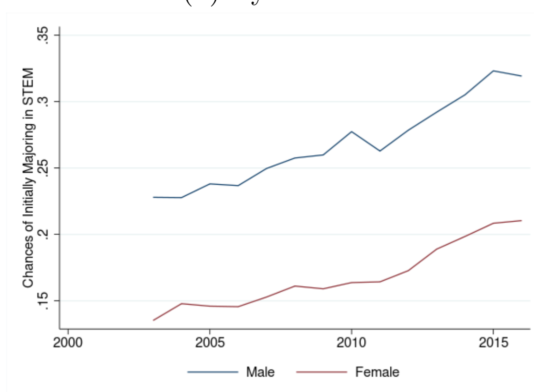
Figures 3 and 5 show that there are significant disparities between demographic groups in their chances of initially declaring a major in STEM and then persisting to graduate with a STEM degree. Under-represented minority (URM) students[10] declare majors in STEM similarly often to average students, but are much less likely to persist. Female students, by contrast, a significantly less likely to declare a major in STEM than males, but their chances of persisting in a STEM degree has been rising over time and is now close to the chances of males. These disparities underline the importance of considering not just access to college for disadvantaged groups, but also their outcomes once getting there.

Figure 3: Chances of Initially Declaring a Major in STEM

(a) By Race



(b) By Gender



The data includes application information for students who applied to public schools, and so it is possible to see who qualified for admission based on the top ten percent rule. Figure 7 shows, as would be expected, that students in the top ten percent of their high school classes who attend public universities are much more likely to declare majors in STEM and persist until graduating. This holds true across different levels of high school quality. To show this, figure 21 in the appendix divides schools into four quartiles based on their standardized test scores in math for the year 2011. In all cases, students in the top 10% are more likely to major in STEM and persist in STEM degrees.

However, there is suggestive evidence that qualifying for admission under the top ten percent rule does not provide the same benefit to students from schools at each quintile of math test scores. Figure 9 takes the difference between the red and blue lines in figure 21, and plots this difference for each quintile as a measure of the expected benefit from being in the top ten Percent. Top ten students from schools with better standardized math test scores

Fall, the Spring, and the Summer, then had a recorded major in the next Fall, this would count as being enrolled for two semesters before having a recorded major.

[9]Specifically, this was 40% of first-time full-time students who enrolled at four-year universities in 2004.

[10]Here, this includes black, hispanic, and native american students.

Figure 5: Chances of Persisting in a STEM Degree until Graduating within 6 years

(a) By Race (b) By Gender

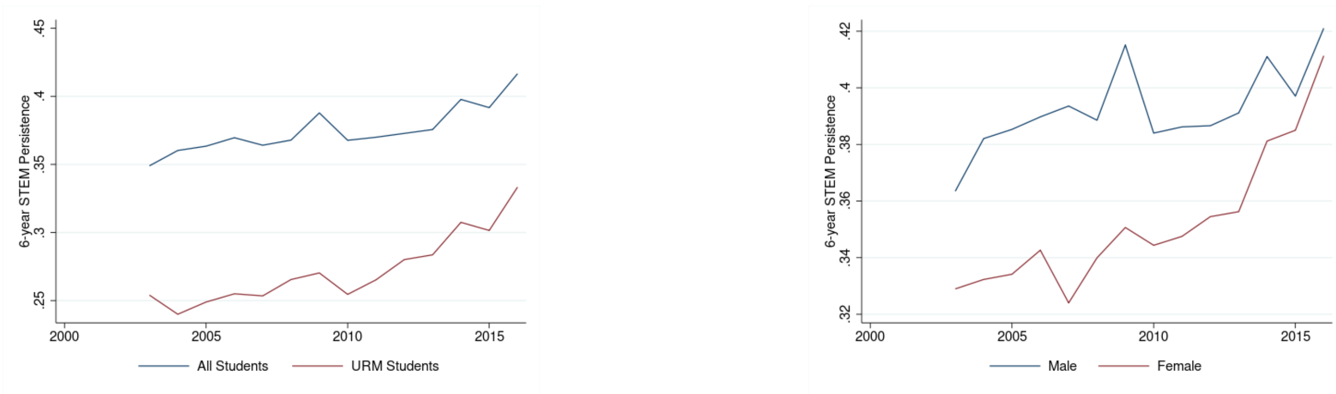
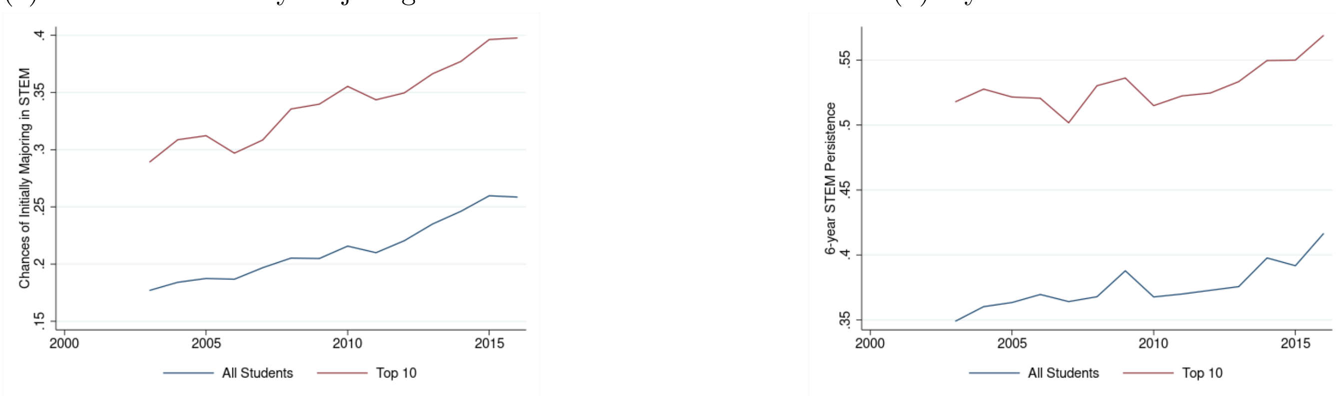


Figure 7: Comparing Students who Received Admission from the Top Ten Percent Rule to the Average Student

(a) Chances of Initially Majoring in STEM (b) 6-year STEM Persistance



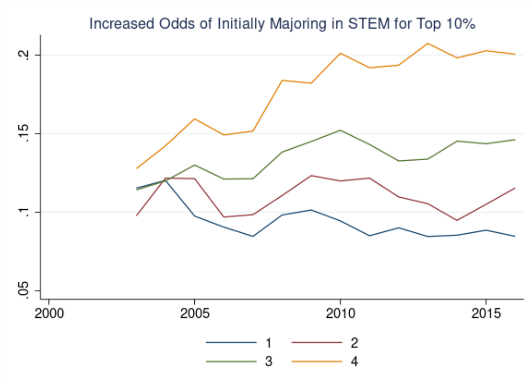
experience greater gains relative to their peers in terms of their chances of initially majoring in STEM and persisting in STEM degrees.

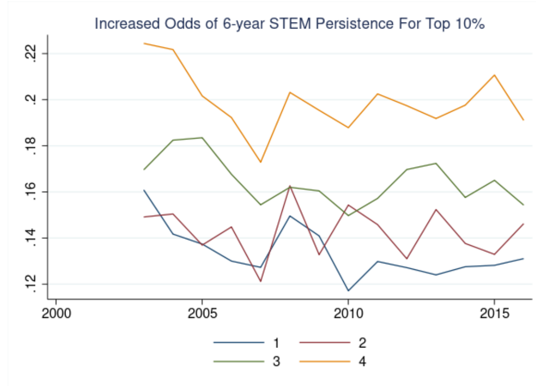
3 Predicting Student Performance in High School Classes

In this paper, I am trying to understand the effects of providing automatic admission to students who have out-performed their peers in their high school classes. The ideal data for answering this question would be two rankings for each student: one that measured how their high school course performance ranked relative to their high school peers, and the other that ranked the student's course performance relative to students across the state of Texas as a whole using a common standard. These rankings are not available in the data, but there are extremely rich measurements in other areas that can be used to estimate what they would be.

This section describes how I will use available data can be used to predict a student's performance in their courses and ranking compared to other students. There are three main steps. At the time of preparing these results, I had access to student grade data only in the 2011-2012 school year, and so the first step toward predicting class rank will be predicting

Figure 9: Comparing the Benefits of being in the Top Ten Percent for Schools with Different Performance Levels on Standardized Tests





student grades when they are not recorded.[11] Then, since each school district in Texas is free to establish its own methods of compiling student grades into a class rank, it will also be necessary to predict what this method is for each school district. Finally, it is important to be able to compare student grades from different school districts, and so the end of this section describes how I will do that.

3.1 Predicting Grades

This section describes how student grades are predicted when they are not recorded in the data. While I only had access to student grades in the 2011-2012 school year, the information provided for that school year is very detailed. Rather than merely providing a letter grade, the data displays the student's percentage score in each portion of the class. For example, it would be possible to see that a certain student earned a grade of 85% for their first semester of a freshman algebra course.

Additionally, besides the grades almost all of the information about students and classes that can be seen in the 2011-2012 school year is also available in other school years. In the previous example, it would still be possible to see that a particular student took a freshman algebra course in some semester, and even information on whether they passed or failed the class, but not the actual grade of 85%. This allows linear regression models for student grades to be fit using the data that I have in all school years, and then used to predict grades for the school years where those grades are not known.

To do this, I run a few different regression models. The first includes fixed effects for individuals, classes, and sections of classes,[12] as well as fixed effects related to whether a student passed or failed and received credit or not. Unfortunately, there are some grades that this regression cannot be used to predict. For example, if a new student enters the school district or if a new class is added, then it will not be possible to predict grades for that student or class.

For cases that the first regression model can't predict, I use a second regression model that replaces the class and individual fixed effects with more general information. These include variables relating to characteristics of the class (e.g. subject area, whether it counts for college credit, whether the student passed or failed, the school district, etc.), variables with demographic information on the student (e.g. gender, ethnicity, whether or not they

[11]Since then I have gained access to some grade data for another year, and intend to use that data in a future revision of this paper.

[12]For example, there would be a fixed effect for being in the first half of a two semester course.

are economically disadvantaged, etc.), measures about other classes that the student is taking (e.g. the number of courses which count for college credit that a student passed this year), and TAKS standardized test scores from grades seven and eight (these include separate scores for math, reading, science, social studies, and writing). Finally, if this model also fails to predict a grade, a third model is used that removes the TAKS standardized test scores.[13]

The key assumption used here is that the process for determining grades remains consistent over time. Since I am mainly concerned about class rank, it should not be too big of a problem if there is some kind of "fixed effect" for grades from a certain school year, because if the same constant boost is added to everyone's grades then their ranking will remain largely unchanged. For measures of within-high school class rank, these fixed effects can also change by high school without greatly affecting the results. The sort of thing that would be a problem is if in 2011-2012, students' eighth grade TAKS math scores were positively associated with grades in 10th grade math, but in 2012-2013 eighth grade TAKS scores were negatively associated with 10th grade math performance. In other words, it is important that the factors which affect the grades of students taking junior-year classes in 2011-2012 will have similar affects for students who are sophomores in 2011-2012, and will be taking those junior-year classes in 2012-2013.

I want to make sure that I avoid "over-fitting" the data by including too many variables and finding relationships that might not hold elsewhere.[14] To check whether this is an issue, I fit the regression using up to three quarters of the available data for each student, and then check the prediction error for the remaining data. Table 1 shows R^2 for those two samples in each of the models used. For the first model, the R^2 is much better in the estimation sample then the hold out sample, showing that there is some over-fitting. However, the R^2 in the hold out sample is still better than the R^2 that is achievable by the second model, and so it makes sense to use the first model whenever possible.

Table 1: Error from Predicting Grades with the First Regression

Regression Model 1:

	count	R^2
1	1.37e+07	.7581245
2	5279233	.6952929

Regression Model 2:

	count	R^2
Estimation Sample:	8091680	.607
"Hold Out" Sample:	3020745	.607

Regression Model 3:

	count	R^2
1	1.13e+07	.5919065
2	4238954	.5912912

Each student's class rank is based on their GPA, which is generally calculated from the

[13] More details on how these models are prepared are included in the appendix B.1.

[14] One way of addressing the "over-fitting" issue would be to use a machine learning algorithm like lasso, and select variables to minimize prediction error outside of the sample used to estimate the data.In the future, I hope to use some kind of machine learning algorithm like this instead of a linear regression.

grades a student receives before their senior year of high school.[15] Since I had access to true grades recorded in the 2011-2012 school year at the time of preparing these results, I am fully relying on estimated grades for only two out of the three years of high school that class rankings will be based on.

3.2 Predicting Methods for Finding GPAs

With predictions for student grades in place, there still remains the challenge that different high schools use different methods for aggregating grades into the overall GPA that will be used to calculate class rank. Each high school h falls into a school district d, and that school district sets rules R^d for how the grades will be converted into a GPA.[16]

As described in Klopfenstein (2010), these methods vary widely across the over 1000 school districts in Texas. Klopfenstein (2010) dealt with this issue by focusing on a single component of these methods—the additional grade points given to AP classes relative to non-honors and non-remedial courses—and directly surveying over 900 individual high schools. The problem of predicting a student's class rank, though, demands knowing the full method that school districts use to aggregate grades into a GPA. To do this, I take a different approach and estimate the rules R^d for each school district so as to match the data as closely as possible.[17]

From what I have seen, the main areas that these methods tend to differ is in the set of classes S^d that they will allow to count toward a person's class rank, and in how many extra grade points P^d to give to different kinds of classes. For example, the class rank policy for the Sanger school district says that "honors" classes receive an additional 5 point boost and "advanced" classes receive an additional 10 point boost on a 100 point grading scale. Additionally, the class rank will only be calculated using four courses from each of english, math, science, and social studies, as well as courses in languages other than English.

Let $R^d(G_i)$ represent the GPA found by applying the rules R^d to a set of grades G_i. G_i includes the individual grades g_i^c for every class c that a student took by the end of their junior year. Since the true grades are not always observed, the previous section described how estimates $\hat{g}_i^c$ are created to fill in the missing values. Let ϵ_i^c be the errors in these estimates so that $g_i^c = \hat{g}_i^c + \epsilon_i^c$, and let $\hat{G}_i$ describe the set of grades filled in with these estimates where necessary.

The estimated GPA $R^d\left(\hat{G}_i\right)$ includes many separate ϵ_i^c errors. Specifically, as shown in appendix B.2, $R^d(G_i) = R^d\left(\hat{G}_i\right) + \epsilon_i^G$ where

$$\epsilon_i^G = \frac{1}{\sum_{c \in S_i^d} w^c} \sum_{c \in S_i^d} \frac{w^c \epsilon_i^c}{10} \tag{1}$$

Here, S_i^d is the classes that an individual has taken out of the set of classes S^d that their school district uses to calculate GPAs. Each w^c term is the number of high school credits associated

[15] Assuming that new grades are released each semester and that Fall semester grades are not available until late January, Daugherty et al. (2014) argue that the class rank most likely to be reported on applications for admission to the Texas flagship universities is based on grades from the end of a student's junior year. For this reason, I will use the grades that students receive up until the end of their junior year to calculate their GPAs for class rank.

[16] Technically, I think that individual high schools are also allowed by law to establish their own practices for calculating GPAs. However, from what I have seen these rules tend to be set at the district level.

[17] Each school district has their current rules for calculating their GPAs are posted online, and so it would be possible to copy them down for each of the over 1000 school districts. However, even if all the nuances of these different school district rules were captured in a systematic way, there would still remain the challenge that these rules can change over time. This means that the current posted rules may not reflect the rules used during the time periods considered here.

with class c. Notice that the extra grade points P^d that a district decides to give for certain classes do not affect this error.

In general, it is not clear whether the fact that $R^d\left(\hat{G}_i\right)$ includes many separate ϵ_i^c terms will make it more accurate or less. On one hand, if all of these terms are nearly independent, then you might expect the variance of the overall ϵ_i^G to be smaller than the variance of any given ϵ_i^c term because the positive and negative errors would begin to cancel each other out. On the other hand, if the ϵ_i^c terms for the same student in different classes are closely related, then the errors could add onto one another and increase the variance.

To try and resolve this problem, I will assume that each $\epsilon_i^c = \eta_i + \nu_i^c$, where η_i is specific to an individual and ν_i^c changes from class to class. The ν_i^c are assumed to be independent of each other and η_i, and both η_i and ν_i^c are assumed to be independent of the set of classes S_i^d and the credits associated with those classes w^c. I will also assume that η_i is pulled from a normal distribution with mean zero and standard deviation σ^η, and that ν_i^c is pulled from a normal distribution with mean zero and standard deviation σ^ν.

Since ϵ_i^G is a linear combination of these terms, it is also normally distributed.[18] The variance of ϵ_i^G is given by

$$\left(\sigma^G\right)^2 = \sum_{c\in S_i^d}\left(\frac{w^c}{10\sum\limits_{c\in S_i^d} w^c}\right)^2 \mathrm{Var}\left(\epsilon_i^c\right) + 2\sum_{\substack{c,c'\in S_i^d\\ c\neq c'}}\left(\frac{w^c w^{c'}}{10\sum\limits_{c\in S_i^d} w^c}\right)^2 \mathrm{Cov}\left(\epsilon_i^c,\epsilon_i^{c'}\right) \tag{2}$$

Fortunately, since I was able to observe true grades for students in the 2011-2012 school year, I also know what their ϵ_i^c terms would have been if their grades had been estimated. The ϵ_i^c terms for student grades that were not used to estimate any of the regression models are especially valuable, because they provide a more reasonable estimate of what the errors would be for grades that had to be predicted. Using these terms, I can calculate $\mathrm{Var}\left(\epsilon_i^c\right)$ and $\mathrm{Cov}\left(\epsilon_i^c,\epsilon_i^{c'}\right)$ for different classes c and c'. From there, the expected variance of ϵ_i^G can also be calculated, with the details included in appendix section B.2.

Figure 12a shows how the variance of ϵ_i^G changes over time as students earn more high school credits. Since there is no error for classes where the true grades are available, these graphs hold the total credits with true grades fixed at the average number of these credits with true grades. Then, the x-axis plots out what happens as additional credits are taken where the grades must be estimated, assuming that each additional class is worth one half credit. The variance of ϵ_i^G depends on the group of classes S^d that are used to calculate GPAs, and so different lines are included to show each of the possible class groups that will be considered later on. Then, figure 12b shows what would happen if the ϵ_i^c terms were assumed to be independent, and all of the covariances between them were set to zero. The line which assumes independence is much lower, and so assuming independence would lead to an underestimate of the variance of ϵ_i^G.

While I do not have access to each student's class rank, the data does provide information on what ranges a student's class rank must have fallen into, and these can be used to predict the methods that schools use to find GPAs. Specifically, for every student who applied to a public university in Texas, the data records whether they qualified for the top ten percent rule.[19] Certain universities also have other admission rules based on class rank, and if a student

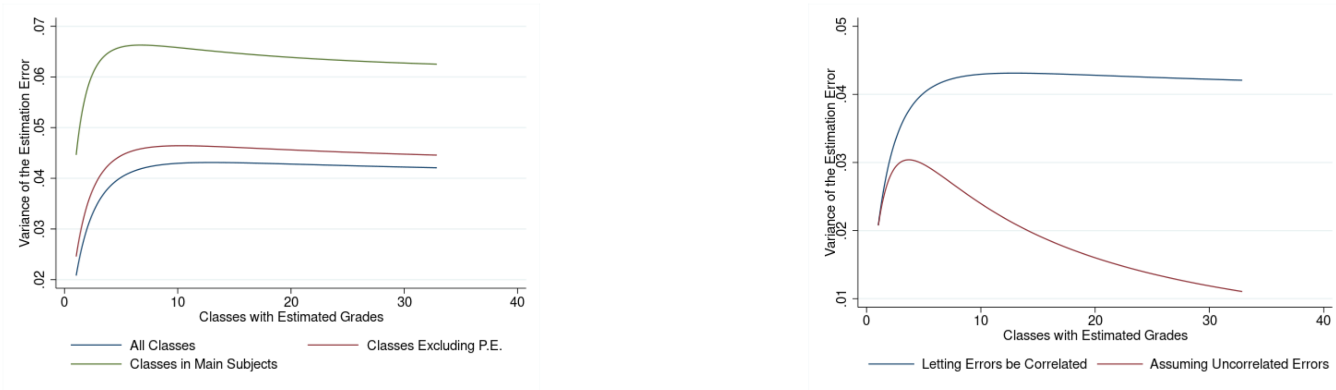
[18] The one exception to this rule is that failing grades below 60% earn 0 grade points. However, it seems unlikely that a student who receives more than one or two failing grades is in contention for special admission rules like the top ten percent.

[19] Technically, the class rank that is used for each student is their class rank at the time that they apply for college. This means that colleges with different application deadlines may have viewed different class ranks if new grades were released in between the two deadlines or if there were changes to the group of students that the class rank was being

Figure 11: Variance of the Estimation Error for GPAs

(a) Using Different Groups of Classes

(b) Assuming that the Errors are Independent



was admitted under these rules then that information is included as well.[20] Therefore, the data allows me to assign a student's class rank into "bins" based upon which thresholds their class rank must have reached or not reached to explain their admissions decisions. For example, if a student was accepted to a certain university because they are in the top 10% of their class, then that student's class rank must have fallen into the bin $B_i = [.9, 1]$.

To predict the grade aggregation method used by a given school, I will attempt to find the rules R^d for each district which maximize the probability that each student's unknown class rank fell within their bin B_i. This probability can be represented as $\mathbb{P}\left(B_i \mid R^d\left(\hat{G}_i\right)\right)$. To do this, I first prepare a grid of possible methods which include different sets of classes S^d to aggregate grades over, and different additional points P^d that can be given to more difficult coursework. Specifically, the sets of classes that I consider are "all classes taken for high school credit," "all classes taken for high school credit excluding physical education," and "all classes taken for high school credit within the standard subject areas of Math, Science, Social Studies, and English and Language Arts." Then, I let two groups of classes earn separate extra weights. These groups are "AP/IB/College Credit courses" and "advanced courses."[21]

Next, my estimate of a district's rules $\hat{R}^d$ is the combination of grid points for S^d and P^d that maximizes

$$\hat{R}^d = \arg\max_{R^d} \sum_{h \in d} \sum_{i=1}^{n^h} \log\left(\mathbb{P}\left(B_i \mid R^d\left(\hat{G}_i\right)\right)\right) \tag{3}$$

where n^h is the number of students with bins in high school h. Additional details on the estimation method are included in appendix section B.2.

calculated over. I am able to correct for this issue somewhat by only using information from applications where the student was applying to be admitted in the upcoming Fall term. This hopefully requires that all of the class ranks used for these applications were found during a similar time.

[20]For example, certain public colleges in Texas also accept students according to a "25% rule" that offers automatic admission to certain students in the top 25% of their high school graduating classes. Additionally, since 2009 UT Austin has been allowed to raise their threshold for automatic admission so that the percentage of students admitted under the top 10 percent rule is around 75% of incoming freshman students. It was 8% in 2011, 9% in 2012, 8% in 2013, 7% in 2014, and since 2019 it as been at 6%. This allows a large degree of precision in finding the class ranks of certain students. If a student was accepted to one school under the top ten percent rule in 2014, but denied automatic admission to UT Austin, then it is likely that their class rank fell in the narrow range of 90% to 93%.

[21]Specifically, the "AP/IB/Collegel Credit courses" can earn 0, .5, 1, or 2 extra grade points, and the "advanced courses" can earn 0, .5, or 1 extra grade points.

Since I am only able to assign students to a "bin" of class rank if they applied to a public university in Texas, the additional weights and sets of classes for each district were found using information from only those students. This should not create a bias, because in order to be fair school districts would need to use the same method to calculate the GPAs of all students regardless of whether they intend on attending a public university or not. Estimates from school districts where very few people apply to public universities are likely to be noisy, and so I ignore districts where too few students did this.[22]

Table 2 shows the estimated rules for the five school districts in Texas that had the most students with bins. Then, in Table 3, I looked at the current true rules posted for these districts and did my best to record the aspects that matched up with what I was estimating. In some cases, previous changes to these rules were recorded in the district's policy, and I included some of these changes that seemed relevant.

Table 2: Estimated Rules for how School Districts find GPAs for Class Ranks

	Additional Grade Points for		Class	Number of Students
ISD Name	AP/IB/College Credit	Advanced	Group	with Bins
Houston	1	1	Standard	3284
Fort Bend	1	1	Standard	2686
Dallas	2	1	Standard	2615
Cypress-Fairbanks	1	1	Standard	2492
Northside (Bexar)	.5	1	Standard	2090

The estimated rules were prepared using data from students who were juniors in 2014, which was well before the current rules recorded in table 3. Additionally, the set of possible rules that I considered was relatively small compared to all the possible ways that a school district could calculate GPAs. That being said, the estimates from table 2 look like a reasonable approximation of the true rules. The estimated "additional grade points" given to certain classes were generally equal to the true values, with the biggest exception being the Northside (Bexar) school district. For the class groups, the estimate for all of the school districts was that they used classes from the "standard" subject areas of Math, Science, Social Studies, and English and Language Arts. This does not exactly match the class groups used by districts in table 3, but most school districts do put restrictions on either the types of classes considered or the number of classes considered. Only three possible class groups were considered in the estimation, and so it would be good to include more possible groups in the future.

While these differences in the methods used to calculate class rank add complexity the problem, they also provide an opportunity to create even more accurate comparison groups for students who received college admission under the top ten percent rule. Because grades can be aggregated in different ways in different school districts, it is possible that two students could have identical high school transcripts and be surrounded by identical peers, but for one of them to be in the top ten percent and the other not because of differences in how their school districts find GPAs.

3.3 Comparing Grades across Texas

In order to find students that are comparable even when they go to school in different districts, I need to be able to find what their GPAs would be if they were all calculated the same way.

[22]Specifically, I only include high schools where at least 10 students who were juniors in 2014 and had a bin. After removing the high schools where this is not true, I also require that the school districts have at least 50 students who were juniors in 2014 and had a bin.

Table 3: Current Rules for how School Districts find GPAs for Class Ranks

ISD Name	Additional Grade Points: AP/IB/ College Credit	Additional Grade Points: Advanced	Class Group	Recorded Changes
Houston	1	1	All high school classes besides some special education classes.	Dual credit courses started receiving an extra point in the Fall of 2015.
Fort Bend	1	.5	Some third and fourth year elective classes that encourage special skills and interests are not included in the GPA.	Students who entered grade 9 before the 2019-2020 school year only received .5 extra points for dual credit courses.
Dallas	2	Pre-AP and certain other courses count for 1.	A set number of classes are used from mathematics, English Language Arts, Science, and Social Studies.	Students entering high school before the 2017-2018 school year used a set amount of classes from any subject. Instead of receiving additional grade points, grades from classes like AP/IB classes were multiplied by different amounts when finding GPAs.
Cypress-Fairbanks	1[1]	Gifted and Talented (Horizons) courses receive 1.[1]	All classes besides classes taken as a student assistant, local credit classes, and classes taken pass/fail. Also, four math and four science courses are used for the GPA.	
Northside (Bexar)	AP receive .8 plus their converted AP test score.[2] College enrollment classes get 1.	Pre-AP classes receive .5 extra grade points.	Their student handbook lists the classes that will be included in class rank.	Students who entered before the fall of 2017 receive 1 additional grade point for AP, pre-AP, honors, and college enrollment classes.

[1] These extra points are added relative to “on-level” coursework.
[2] AP Scores range from 1 to 5, which I think here would be .1 to .5 additional grade points.

In 2008, the Texas legislature gave the Texas Higher Education Board (THECB) the authority to set a uniform method for GPA calculation which all high schools would be required to calculate, and which institutions of higher education would need to use for decisions relating to admissions and financial aid. The method that they came up with was never enforced,[23] but regardless it provides a useful uniform standard to use for comparing student GPAs across Texas. Using this standard, I also calculate a student's rank within the state of Texas as a whole.

One additional concern is that grades received within different high schools may not reflect the same course performance. In the estimation section, I will use matching methods to assess the effects of the top ten percent rule. In order for these methods to work well, each of the matched student pairs ought to have had the same performance in their classes, and only be different because one received top ten percent admission and the other did not. However, different levels of grade inflation in different high schools may mean that two students with identical grades had very different actual performance levels in their courses. For example, if all teachers graded on a curve with a "B" average, then students who have average performance within their school districts would receive B's even if they were did not have performance that was average for Texas as a whole.

Fortunately, certain standardized tests in Texas can be used to estimate the degree of grade inflation in each high school or district. Beginning with students who were in ninth grade during the 2011-2012 school year, Texas began administering standardized tests called STAAR end-of-course (EOC) exams for specific classes. The unique advantage of these standardized tests is that they are designed to cover the same material that is covered by a class, and should measure the same underlying ability as the class grade. 2011-2012 is also the only year in where I had access to true high school grades. By observing the grades students received in their classes and the scores they received on the associated EOC exam, it is possible to get an understanding of how much each school district inflates their grades.

Let g_i^c be student i's grade on class c, and E_i^c be their standardized test score for that class. Then, assume that

$$g_i^c = \beta_0^{gc} + \beta_1^{gc} a_i^c + \pi^{ch} + \eta_i^{gc} \tag{4}$$

$$E_i^c = \beta_0^{Ec} + \beta_1^{Ec} a_i^c + \eta_i^{Ec} \tag{5}$$

where π^{ch} is grade inflation within high school h for class c. Notice that both grades and exam scores can be imperfect measures of a student's ability a_i^c due to the η terms. In writing these equations, I assume that the EOC exam functions as it is supposed to and truly measures the same latent ability a_i^c as the course.

Identifying π^{ch} requires two classes (c and c') where students receive both a class grade and an EOC exam score. In my case, those classes will be Algebra 1 and Biology. Additionally, I will make two assumptions:

1. All students in a school district receive the same grade inflation π^{ch} for each class c.
2. The η terms are independent of the school district, and there is zero covariance between each of the η terms and (i) the other η terms, (ii) the a_i^c terms, (iii) and the π^{ch} terms. The one exception to this rule is that $\text{Cov}\left(\eta_i^{gc}, \eta_i^{gc'}\right)$ can be nonzero.

The exception that $\text{Cov}\left(\eta_i^{gc}, \eta_i^{gc'}\right)$ can be nonzero is useful because students will be taking both of these classes during the same school year, and a "shock" like getting sick could affect their grades in both courses without reflecting anything about their abilities in either. While these assumptions may be strict, they also allow many things to remain unrestricted. For

[23] The Texas Senate voted to repeal this rule in 2009. Even if it had been enforced, individual high schools would have been free to use a different method of calculating GPA for the purposes of assigning class rank.

example, a student's abilities in the two classes a_i^c and $a_i^{c'}$ are separate, and may be either correlated or not. In addition, the level of inflation in a school district π^{ch} is free to depend on the grading inflation $\pi^{c'h}$ that it uses for the other class, as well as the abilities of its students in both classes.

I assume that $\mathbb{E}\left(a_i^c\right) = 0$, $\mathrm{Var}\left(a_i^c\right) = 1$, and $\mathbb{E}\left(\pi_i^{ch}\right) = 0$ for both classes. Since all of the terms are mean zero, the β_0 terms can be found by taking the means of each g_i^c and E_i^c term:

$$\mathbb{E}\left(g_i^c\right) = \beta_0^{gc}$$
$$\mathbb{E}\left(E_i^c\right) = \beta_0^{Ec}$$

The variances and covariances of the g_i^c and E_i^c terms, both conditioning on school districts and not conditioning on school districts, together make up 20 equations. There are also 20 unknowns, so that the system as a whole is just identified. Appendix section B.2 shares details on these equations, and specific solutions for the β_1 terms. Once these terms are estimated, I can use them to identify the π^{ch} terms. Rearranging equations 4 and 5 yields

$$\pi^{ch} = g_i^c - \beta_0^{gc} - \beta_1^{gc}\left(\frac{E_i^c - \beta_0^{Ec}}{\beta_1^{Ec}}\right) - \frac{\beta_1^{gc}}{\beta_1^{Ec}}\eta_i^{Ec} - \eta_i^{gc}$$

Then, since the η terms have expected values of zero and are independent of the school district, taking the expected value within a school district d identifies each π^{ch} term:

$$\pi^{ch} = \hat{\mathbb{E}}\left(g_i^c \mid d\right) - \beta_0^{gc} - \beta_1^{gc}\left(\frac{\hat{\mathbb{E}}\left(E_i^c \mid d\right) - \beta_0^{Ec}}{\beta_1^{Ec}}\right)$$

This can be done separately for the classes c and c' to find the grade inflation in each for all school districts.

Table 4 shows estimates for the β terms, as well as estimates related to the grade inflation. There is a negative covariance between student ability in a school district and the grade inflation for both classes. This is what would be expected if schools graded at least partially on a curve, so that schools with lower-performing students need to inflate their student's grades more to maintain a "B" average.

Table 4: Grade Inflation Results

	Algebra 1	Biology
β_0^{gc}	78.9867	79.3867
β_0^{Ec}	3.8e+03	3.9e+03
β_1^{gc}	7.3977	7.1504
β_1^{Ec}	496.9664	436.4491
$\mathrm{Cov}\left(a_i^c, \pi^{ch}\right)$	-0.2688	-0.5122
$\mathrm{Var}\left(\pi^{ch}\right)$	5.9204	8.2385

The correlation between the grade inflation π^{ch} for algebra 1 and biology was 0.5419. The results in this table were found using all students that took either algebra 1 or biology and had a corresponding EOC exam. This was 431136 students.

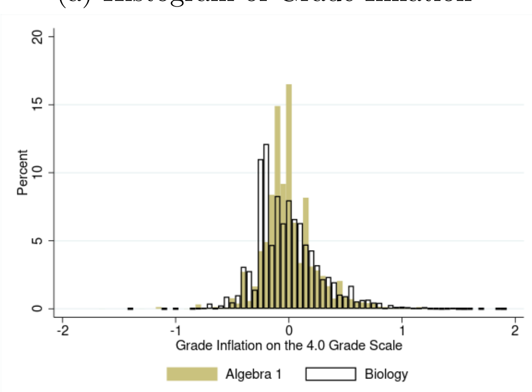
If all classes in Texas had an EOC exam, then it would be possible to use a method like this to identify the grade inflation for every class in every school district. Then, I could subtract the class-specific grade inflation from the grades of each student in all of their classes. Unfortunately, that is not the case, and in the 2011-2012 school year I only have access to two EOC exams. In order to subtract the grade inflation from student's grades, then, I need to strengthen the first assumption above to be that grade inflation π^{ch} is constant within a school district across all classes in that district, and over time.

This is a much stronger assumption, and will not be exactly true in the data. However, it is possible to partially test how good of an approximation this is by comparing the results for grade inflation in the two classes above. First, if grade inflation is approximately constant for all classes in a school district, then π^{ch} and $\pi^{c'h}$ should be highly correlated. This seems true because the correlation between them is large and positive at .54.

Second, if grade inflation is approximately constant for all classes in a school district, then the distribution of grade inflation in Algebra 1 should be similar to the distribution of grade inflation in Biology. Table 4 and figure 14a show that the variance of grade inflation is comparable between Algebra 1 and Biology, with the variance of Biology being slightly larger. There is a stronger covariance between inflation in Biology and unobserved ability than there is between Algebra 1 and unobserved ability, but this is not a problem because I am not assuming that the unobserved abilities are the same in these two classes. Even if grade inflation were exactly equal in these two classes, the unobserved abilities could be different and result in different covariances.

Figure 13: Comparing the Grade Inflation in Algebra 1 and Biology

(a) Histogram of Grade Inflation



(b) Scatter Plot of Grade Inflation

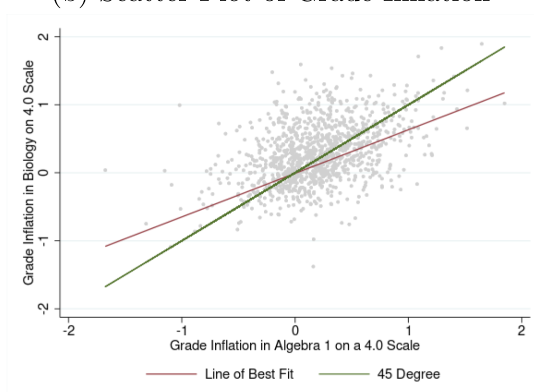


Figure 14b plots the grade inflation in Algebra 1 and Biology. If grade inflation were constant across school districts, then these points should roughly follow the 45 degree line, with errors from the η terms pushing them slightly above or below. The real line of best fit is close to the 45 degree line, especially around the small levels of grade inflation where many of the points are clustered.

Under the assumption that grade inflation is constant across classes, I can combine the results for these two classes to find a single estimate of grade inflation π^h for each district. Specifically, I set $\pi^h = (\pi^{ch}+\pi^{c'h})/2$. Figure 15 shows the distribution of this inflation measurements on the 4.0 GPA scale, and shows that they are generally small. Whenever I need to compare student grades across different school districts in the results section below, I correct for grade inflation by subtracting this district-specific inflation measurement from the grades of all the students in that district. As a robustness check, I also run these results without making the correction for grade inflation in appendix D.

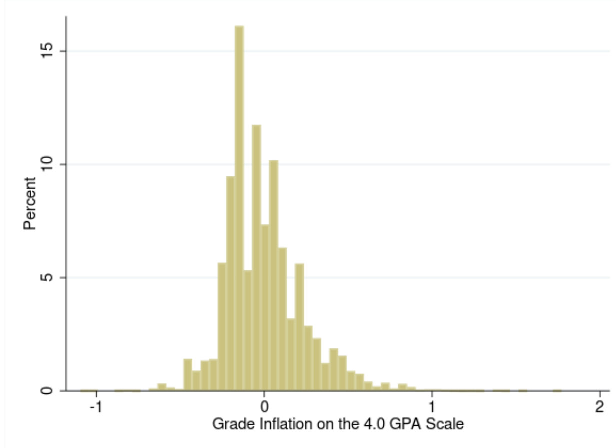


Figure 15: Grade Inflation By School District
Note: The bars in this figure are weighted by the number of students with STAAR EOC tests in each district.

4 Estimation Framework

In this section I will work on a framework for evaluating the effects of the receiving automatic admission under the Top 10% rule.

Let G_i be the set of courses that a student took in high school and their grades in those courses. Their ability A_i is measured by their performance in courses under the uniform grade aggregation method proposed for Texas, $R^U(G_i)$. This ability can be expressed as

$$A_i = h_i + a_i$$

where $h_i = \overline{R^U(G_h)}$ is the average ability for student i's high school and $a_i = R^U(G_i) - h_i$ is the deviation of a student's ability from that average.

Each school district can have their own method $R^d(\cdot)$ of aggregating grades into an overall GPA. Within a high school, students receive a signal on their ability s_i based on their performance compared to the average performance in their high school, where $s_i = R^d(G_i) - \overline{R^d(G_h)}$. Let F_h^s be the distribution of these signals for a high school. The student receives automatic admission under the top ten percent rule if $F_h^s(s_i) \geq .9$.

Because each school district can have their own method of combining grades into a GPA, the signal that a student receives s_i is can be different than the signal a_i that they would receive from the uniform method for finding GPAs. Let ϵ_i^s represent this difference so that

$$a_i = s_i + \epsilon_i^s$$

Let D_i represent receiving automatic admission under the top ten percent rule. Y_i^j represents outcome j for student i. Assume that

$$Y_i^j = \gamma_0^j + X_i\gamma_1^j + D_i\gamma_{2i}^j + f^j(A_i) + \eta_i^j \tag{6}$$

4.1 Additional Structure for Matching Methods

In order to use matching methods, I will make some additional assumptions. First, assume that the distribution of signals s_i follows a normal distribution $N(0, \sigma_h^s)$ for each high school h. This means that the cutoff rule $F_h^s(s_i) \geq .9$ can be rewritten as $\frac{s_i}{\sigma_h^s} \geq 1.28$. $a_i = s_i + \epsilon^s$, so this can be rewritten again as

$$a_i - \epsilon^s \geq 1.28\sigma_h^s$$

Second, assume that a student's high school placement h_i is random conditional on their observed characteristics X_i so that $h_i = X_i\beta + \epsilon_i^h$. These characteristics are meant to capture things that would be true about a student before attending high school, and could influence where they attend high school.[24] Since $a_i = A_i - h_i$, this lets the cutoff be rewritten again as

$$A_i - X_i\beta - \epsilon_i^h - \epsilon_i^s \geq 1.28\sigma_h^s$$

Finally, A_i and σ_h^s are both measured with some error ϵ_i^m because student grades are predicted, and so the cutoff is

$$\hat{A}_i - X_i\beta - 1.28\hat{\sigma}_h^s \geq \epsilon_i^m + \epsilon_i^h + \epsilon_i^s \tag{7}$$

where $\hat{A}_i$ and $\hat{\sigma}_h^s$ are the estimated values for A_i and σ_h^s.

These ϵ terms create variation in who receives the top ten percent rule between similar students. Figure 16 displays the distribution of people who received automatic admission under the top ten percent rule across each of three different class rank measures. The blue bars use the method of aggregating high school grades that were estimated for each school district, and find a student's class rank within their high school. If there were no measurement error ϵ_i^m, these bars would be entirely contained in the top 10% of the ranking. The yellow bars also find a student's class rank within their high school, but use the uniform standard for aggregating grades into GPAs. Therefore, the additional dispersion in the yellow bars compared to the blue bars is coming from the ϵ_i^s term. Finally, the black bars use this uniform standard and find a student's rank within the entire state of Texas. The additional variation in the black bars compared to the yellow bars points to the variation in peer ability level ϵ_i^h. If the distribution of student ability were the same in every high school, there would be no difference between the yellow and black groups of bars.

As would be expected, the students who receive admission under the top ten percent rule are concentrated in the higher rankings for all of these measures. However, the differences between these three groups of bars suggest that certain students who received admission under the top ten percent rule may not have qualified if they had gone to different high schools where the high school GPAs were found differently or their peer group had been different. In order to estimate the effects of the top ten percent rule, I will assume that these ϵ terms are independent of treated and untreated student outcomes Y^j conditional on a student's ability level and demographic characteristics. Specifically, letting Y_0^j represent untreated outcomes and Y_1^j represent untreated outcomes, I am assuming that

$$E\left(Y_0^j \mid X, \hat{A}, \hat{\sigma}_h^s, D = 1\right) = E\left(Y_0^j \mid X, \hat{A}, \hat{\sigma}_h^s, D = 0\right) \tag{8}$$

Given an ability level A_i and set of demographic characteristics X_i, the average effect of receiving automatic admission from the top ten percent rule is

$$\bar{\gamma}_{2i}^j = E\left(Y_i^j \mid X_i, \hat{A}_i, \hat{\sigma}_h^s, D = 1\right) - E\left(Y_i^j \mid X_i, \hat{A}_i, \hat{\sigma}_h^s, D = 0\right) \tag{9}$$

Since students who qualify for the program are not forced to take advantage of it, this term is capturing the results of the choices that students make when two things happen: First, when they have done well enough compared to their peers to be in the top ten percent, and second, when are given the option of automatic admission to public schools. Both of these things could

[24]In the estimation section, X_i includes eighth grade TAKS standardized test scores in math and reading, and also certain demographic characteristics: ethnicity, gender, economic disadvantage status, and whether a student's home language is english, spanish, or something else. These demographic characteristics are recorded while a student was in high school, so it is possible that do not exactly represent a person's characteristics while in middle school. For example, it could be that someone was recorded as economically disadvantaged while in middle school but not in high school.

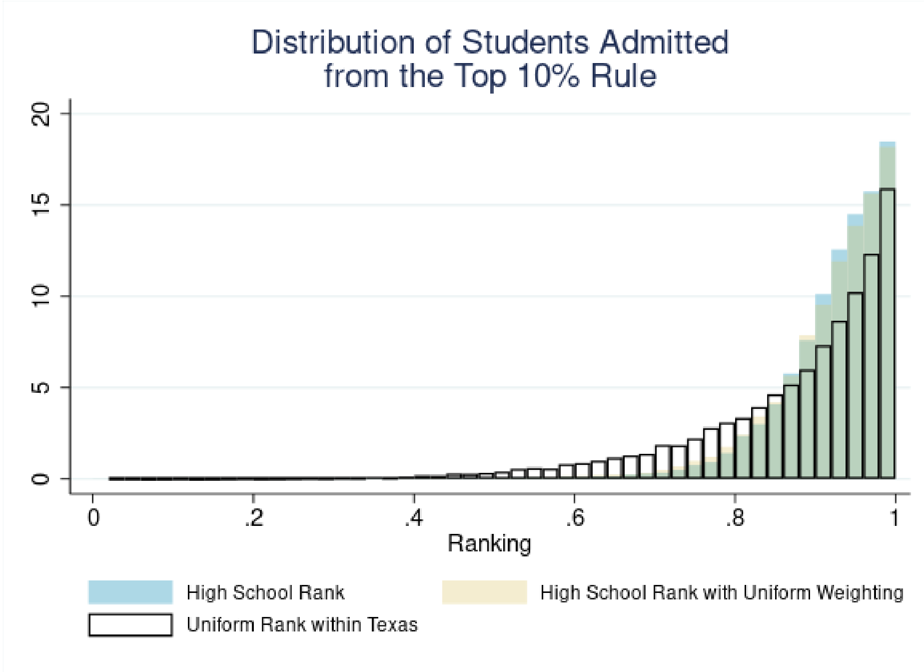


Figure 16: Variation in who Receives Automatic Admission

be significant in determining how students respond. When the matching method identifies two students with the same course performance and demographics, but only one of them received automatic admission, this suggests that one of them has a class rank above 90% and the other has a class rank below 90%.[25]

Receiving automatic admission from the top ten percent rule involves several smaller "treatments" that all might contribute to the overall effect shown above. To understand how these different factors influence the treatment effect, I look at the differences between each student i who received automatic admission under the top ten percent rule and their matched student $\hat{i}$ who did not. Specifically,

$$\gamma^j_{2i\hat{i}} = Y^j_i - Y^j_{\hat{i}}$$

is a measure of the effect of receiving automatic admission for student i. The most obvious possible difference is that students who would not otherwise attend a 4-year university might do so if they received guaranteed admission. Let U_i represent whether a student attended a four year university. Then, the difference in whether these students attended a 4-year university is given by

$$\Delta_{i\hat{i}} U_i = U_i - U_{\hat{i}}$$

A second possible difference is that students who receive automatic admission go to higher-ranked universities. $\bar{A}^u_i$ represents the average ability $\hat{A}_i$ at a student's university,[26] and U^{AM}_i represents whether a student attended Texas A&M university. It is important to include a dummy variable for this university because of the "25 by 25" that it was running during this time which encouraged students to major in engineering. Like before, the differences are given by

$$\Delta_{i\hat{i}} U^{AM}_i = U^{AM}_i - U^{AM}_{\hat{i}}$$

[25] Unfortunately, if a student did not apply for admission to a public university in Texas, then I cannot say for certain that they did not qualify for the top ten percent rule. This is discussed more in appendix C.

[26] $\hat{A}_i$ is still a student's inflation-adjusted GPA from high school using the uniform GPA standard. If a student did not go to a university, then I use the average ability of students who did not go to a university.

and

$$\Delta_{i\hat{i}}\bar{A}^u = \bar{A}^u_i - \bar{A}^u_{\hat{i}}$$

Finally, I will also look at the difference in how well the matched students did in their high school peer groups. Remember that since only one of the matched students received automatic admission under the top ten percent rule, it is likely that this student performed better compared to their high school peers than the other one. If students evaluate their own abilities in light of their peer groups, this could lead these two students to have different beliefs about their own abilities even though they were matched as having the same performance in high school classes. s_i was defined earlier and gives the difference between a student's class performance and the average performance at their high school. Then

$$\Delta_{i\hat{i}}s = s_i - s_{\hat{i}}$$

The effect of automatic admission $\gamma^j_{2i\hat{i}}$ depends on the difference in these variables between student i and student $\hat{i}$. Specifically, let

$$\gamma^j_{2i\hat{i}} = \alpha_0 + \alpha_1\Delta_{i\hat{i}}U_i + \alpha_2\Delta_{i\hat{i}}U^{AM}_i + \alpha_3\Delta_{i\hat{i}}\bar{A}^u + \alpha_4\Delta_{i\hat{i}}s + \omega^j_i \tag{10}$$

I will estimate the α terms using ordinary least squares.

4.2 Fuzzy Discontinuity Method

The challenge with using a regression discontinuity approach is that each student's class rank is measured with some error. As discussed in Davezies and Barbanchon (2017), a regression discontinuity approach will not work if this measurement error is independent of D_i and the treated and untreated outcomes. The errors would remove any discontinuity in whether or not a student receives automatic admission right around the cutoff. However, if some fraction of observations are measured correctly, Battistin et al. (2009) show that this can create a discontinuity in a student's chances of receiving automatic admission right around the cutoff and justify a fuzzy regression discontinuity approach.[27]

In this case, that would mean that a certain portion of students right around the 90% cutoff in class rank had their class rank measured correctly. This could be true even if their GPA was predicted with some amount of error. For example, imagine that right around the 90% cutoff, each person had a GPA that was .1 points higher than the GPA of the person below them. Then, as long as the estimates of their GPAs were all correct to within .05, the class rank based on those estimates would be correct.

If the number of students at each high school were to grow toward infinity, the estimation error would eventually be bigger than the difference between the GPAs of students with similar class ranks. However, there is not a huge number of students at each high school, and so if some class ranks around the 90% cutoff are being predicted correctly, I would argue that the measurement error is functioning like a more ideal type of random assignment in this context. At the very least, if the effects shown here are similar to the results found with matching methods above, that should be taken as evidence that the estimates from the matching methods are credible.

In section 5, I will check to see if there is a discontinuity in the probability of being accepted from the top ten percent rule at a class rank of 90%, and also if students seem to be very similar around the 90% cutoff. Given that these two things hold, a regression discontinuity approach can be used. Remembering that $A_i = h_i + \hat{s}_i + \epsilon^s + \epsilon^m$, I set the

[27] This still requires that the measurement errors which do exist are independent of the treated and untreated outcomes.

equation $f^j(A_i) = \alpha_h + f_D^j\left(F_h^{\hat{s}}(\hat{s}_i)\right) + \epsilon$, where α_h is a fixed effect for each high school. Then, I use the estimation equations

$$D_i = \alpha_h + X_i\alpha_1 + f_D^j\left(F_h^{\hat{s}}(\hat{s}_i)\right) + \eta_D$$

and

$$Y_i = \gamma_h + X_i\gamma_1 + \hat{D}_i\gamma_2 + f_Y^j\left(F_h^{\hat{s}}(\hat{s}_i)\right) + \eta_Y$$

using the same student characteristics X_i as in the matching section above.

5 Results

Tables 5 provide some descriptive statistics on the data used to make these results. Students who do not attend a university are recorded as not majoring in STEM. Similarly, in order to count as "persisting in STEM" a student must initially major in STEM and then graduate with a STEM degree within 6 years.[28] Currently, the code also requires that the student begin attending a university within a year after graduating high school. This makes it so that all the students who are recorded as "persisting in STEM" completed their degrees in the same time frame after high school, but does mean that some students who take a gap year after high school before attending college will be counted as not persisting in STEM even if they earned a STEM degree in 6 years. I plan to change this in the future. All other students who do not meet the conditions described here are counted as not persisting in STEM.

For students who do not attend a university in Texas, I am not able to see whether they graduated with a STEM degree or not. To deal with this, there are two variables for persisting in STEM. The one used for a "lower bound" assumes that all students who initially majored in STEM at universities outside Texas went on to persist in STEM. This will set a lower bound for the effect of the top ten percent rule on STEM persistence. The version used for an "upper bound" assumes the opposite, that no one who initially majored in STEM at universities outside of Texas went on to persist in a STEM degree.

At the time of preparing these results, I only had information on the courses that students took beginning in the 2011-2012 school year, and so I only calculated the class rank for people who were juniors two years later in the 2013-2014 school year. The results in this section are only based on this class of students.

The number of enrolled semesters before declaring a major[29] includes only the Fall and Spring semesters, and not any Summer enrollment. In this section, I will use a student's first recorded major as a guess for their originally intended major when coming into college. One possible concern with this is that a student's first recorded major might also be influenced by things the student learns about their STEM prospects from their initial experiences in college. Fortunately, table 5 provides encouraging evidence that students generally have a recorded major very early on in college. Students that will declare a STEM major spend an average of only .021 semesters without a recorded major.

The variables for "working in Texas 6 years later" and "wages in Texas 6 years later" refer to six years after their junior year of high school. For these students, that would be the year 2020. This is not ideal, because the COVID-19 pandemic was occurring at this time and affected wages and employment. Also, it would be nice to see these things later on, because 2020 would have been just 5 years after these students graduated high school, and may not

[28]Since some students take courses for college credit while in high school, I measure this as earning a STEM degree within six years after graduating high school. A student is counted as persisting in STEM even if they graduate with their STEM degree from a different Texas university than the one that they initially attended.

[29]Here, by "declaring a major" I mean having a major recorded in the data besides "undecided."

Table 5: Summary Statistics

Demographic Characteristics

	count	mean	sd
Race: Asian	223517	0.047	0.212
Race: Black	223517	0.129	0.336
Race: Hispanic	223517	0.499	0.500
Race: White	223517	0.303	0.460
Race: Other	223517	0.020	0.142
Home Language: English	223517	0.714	0.452
Home Language: Spanish	223517	0.252	0.434
Home Language: Other	223517	0.034	0.182
Gender: Male	223517	0.497	0.500
Gender: Female	223517	0.503	0.500
Eligible for Free or Reduced Meals	223517	0.402	0.490

Grades

	count	mean	sd
Weighted GPA: District-Specific GPA Method	223517	2.315	0.730
Weighted GPA: Uniform Method	223517	2.466	0.679
Weighted GPA: Uniform Method and Corrected for Inflation	223517	2.459	0.742
The Standard Deviation of the Grade Signal $\hat{\sigma}_h^s$	223517	0.669	0.086

Outcomes

	count	mean	sd
Went to a Uni	223517	0.436	0.496
First Uni Outside TX	223517	0.063	0.243
First Uni TX Pub Flagship	223517	0.033	0.178
First Major in STEM	223517	0.092	0.289
Persisted in STEM Lower Bound	223517	0.040	0.196
Persisted in STEM Upper Bound	223517	0.029	0.169
Working in Texas 6 yrs later	223517	0.689	0.463
Log Wages in Texas 6 yrs later	223517	6.636	4.569
Enrolled Semesters before Declaring Major	83207	0.350	1.355
Enrolled Semesters before Declaring STEM Major	18177	0.021	0.247

Note: The enrolled semesters before declaring a major shown here includes only the Fall and Spring Semesters-the summer term is not included.

reflect what their earnings will be like in the long term. However, 2020 is the last year that has a full year of data available for wages.

5.1 Matching

In this section I will try to identify the effects of the automatic admission under the top ten percent rule using matching methods. Assuming that the ϵ terms all follow normal distributions with means of zero, the equation 7 for receiving automatic admission from the top ten percent rule can be estimated with a probit. Specifically, let the standard deviation of $\epsilon^m + \epsilon^h + \epsilon^s$ be given by σ_ϵ. Then the probability that a person will receive automatic admission under the top ten percent rule is

$$Pr\left(D = 1\right) = \Phi\left(\frac{1}{\sigma_\epsilon}\hat{A}_i - X_i\frac{\beta}{\sigma_\epsilon} - 1.28\frac{\hat{\sigma}^s_h}{\sigma_\epsilon}\right) \tag{11}$$

After estimating this function with a probit,[30] I can use the fitted values as a person's propensity score for receiving automatic admission. Figure 23 in the appendix shows that there is common support for most observations. In all of the matching results that follow, I require a student to be on the common support in order to be included when finding the the effects of the top ten percent rule.

Table 6 shows the results of estimating equation 9, where each cell shows the average treatment effect $\overline{\gamma}^j_2$ from a separate matching estimation. When I ran these results, I unfortunately forgot to include the standardized test scores in the X_i term for the Mahalanobis matching.[31] However, the propensity score matching estimates did include standardized test scores, and those results are very close to the estimates with Mahalanobis matching. Receiving automatic admission from the top ten percent rule consistently raises a student's chances of attending a university, lowers their chances of attending a university outside of Texas, and raises their chances of attending a Texas flagship university. 6 years after their junior year of high school, students who apply for automatic admission are more likely to be working in Texas and tend to be earning more money in Texas.

Students who qualify for automatic admission from the top ten percent rule are also more likely to major in STEM and to persist in STEM. The estimated lower and upper bounds for the effect on persistence in STEM are close together in every case. This suggests that even though the data does not show whether people who attend colleges outside Texas persist in STEM or not, that does not have a large effect on the results.

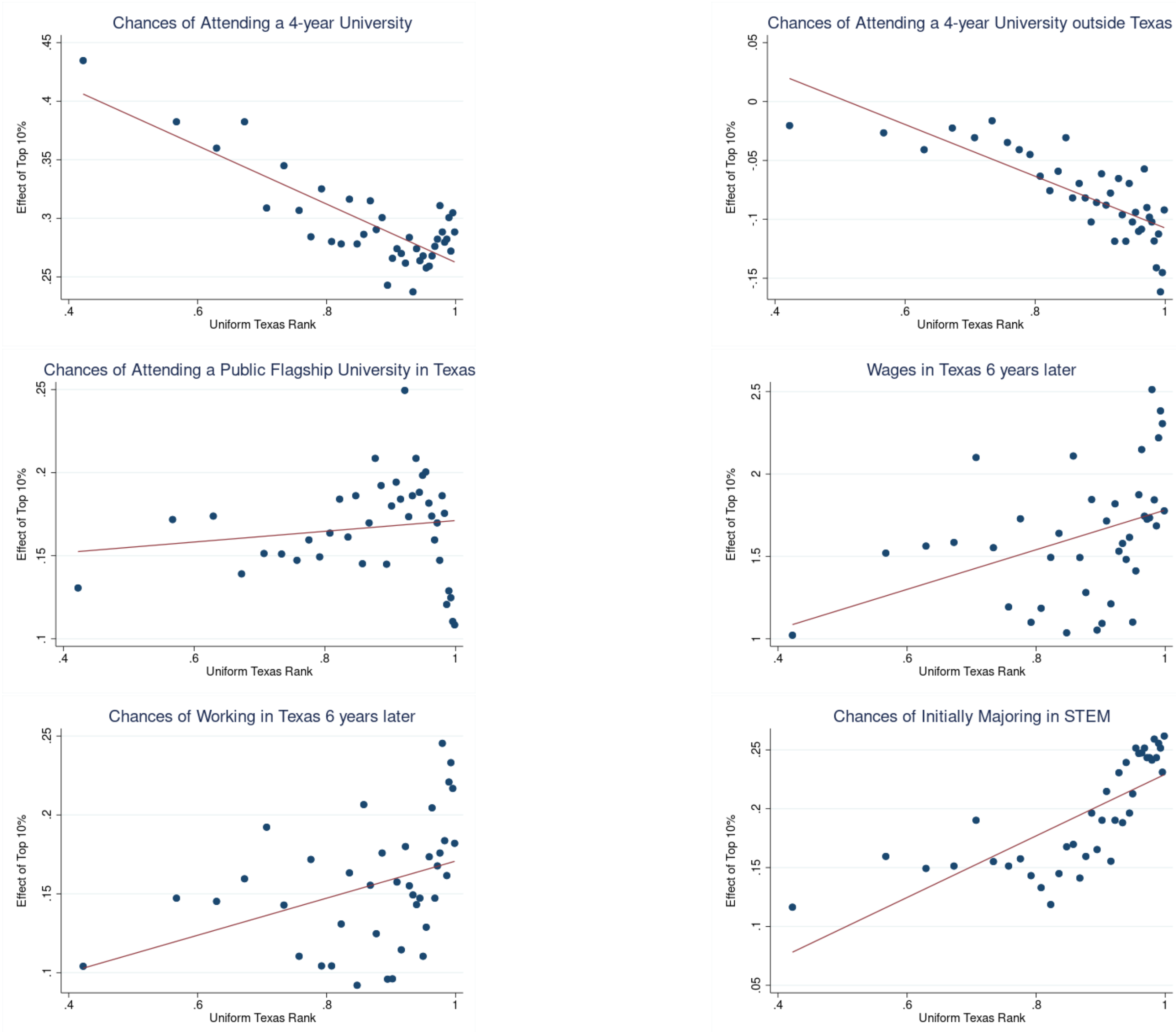
Figure 17 breaks the data into 40 bins of ability $\hat{A}_i$ and shows the average treatment effect $\bar{\gamma}^j_{2i}$ within each bin.[32] The results show that there are often large differences in the effects of receiving automatic admission for students at differing levels of ability. In some cases, like the chances of attending a flagship university in Texas, there does not seem to be a large trend. For the others, though, there appears to be a strong connection between the effects of receiving automatic admission and a student's ability.

In the first graph, the positive effects of the top ten percent rule on attending a 4-year university are largest for students at the lowest ability levels. This makes sense, because students that are on the lower end of academic preparation in Texas could have a harder

[30]Equation 11 is written so that $\hat{A}_i$ and $1.28\hat{\sigma}^s_h$ have the same coefficient $\frac{1}{\sigma_\epsilon}$, but when I actually estimate the probit I allow them to have different coefficients.

[31]The Mahalanobis matching was done with $\hat{A}_i$, $\hat{\sigma}^s_h$, and the other variables in X_i. These include a student's ethnicity, gender, economic disadvantage status, and whether the student's home language is english, spanish, or something else. All of the matched Mahalanobis terms were still required to be on the "common support" of the propensity scores, which did include the standardized test scores.

[32]Figure 22 in the appendix shows the same figures for nearest-neighbor matching.



(a) Chances of Persisting in STEM

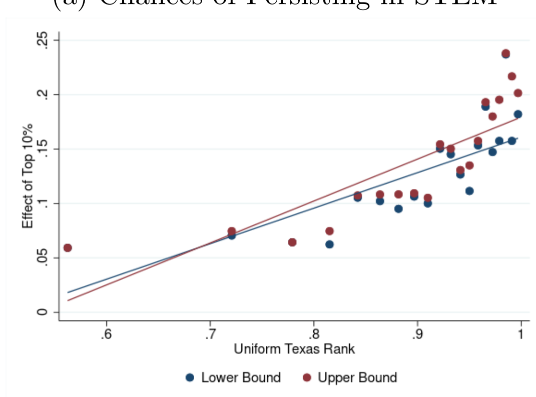


Figure 17: Variation in Treatment Effects Across Ability with Mahalanobis Matching

Table 6: Effects of the Top Ten Percent Rule using Matching

	(1)	(2)	(3)
Went to a Uni	0.2946***	0.2782***	0.2754***
	(0.0057)	(0.0050)	(0.0036)
First Uni Outside TX	-0.0789***	-0.0831***	-0.0827***
	(0.0043)	(0.0040)	(0.0029)
First Uni TX Pub Flagship	0.1670***	0.1516***	0.1524***
	(0.0037)	(0.0036)	(0.0032)
First Major in STEM	0.1954***	0.1977***	0.1943***
	(0.0054)	(0.0049)	(0.0041)
Persisted in STEM Lower Bound	0.1261***	0.1226***	0.1223***
	(0.0042)	(0.0039)	(0.0033)
Persisted in STEM Upper Bound	0.1382***	0.1355***	0.1356***
	(0.0037)	(0.0035)	(0.0031)
Working in Texas 6 yrs later	0.1555***	0.1295***	0.1334***
	(0.0065)	(0.0058)	(0.0046)
Log Wages in Texas 6 yrs later	1.6252***	1.3607***	1.4033***
	(0.0644)	(0.0581)	(0.0457)
Method	Mahalanobis	Nearest Neighbor	Nearest 5 Neighbors
Observations	223517	223517	223517

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

time competing for admission if they are not provided admission automatically. It also makes sense that students at the higher end of academic preparation experience the largest negative effect on their chances of attending schools outside Texas. If those students are not admitted automatically to one of the flagship schools in Texas, schools outside Texas are more likely to be an alternative option for them than they would be for students at the lower end.

On the graphs about initially majoring in STEM and persisting in STEM, receiving automatic admission from the top ten percent rule seems to be more beneficial for students with higher ability levels. This may be because students with good course performance have the necessary abilities to major in STEM, but get discouraged from majoring in STEM if they were not within the top ten percent of their high school peers. The effect of the top ten percent rule is positive for students at all levels of ability, including students at lower ability levels. In the persistence graph, the difference between the dots for the lower and upper bounds is generally larger for students at the higher end of ability. This is what would be expected, because students with high ability may be more likely to attend universities outside of Texas.

Equation 10 can be used to show how different parts of student's experiences contribute to the effect of receiving automatic admission from the top ten percent rule. Table 7 shows the results of estimating this equation using the student pairs from the Mahalanobis matching. A large portion of each effect happens because people who receive automatic admission from the top ten percent rule are more likely to attend four-year universities. Since students need to attend a university in order to major in STEM, this is not surprising. What is more surprising is that attending Texas A&M gives a similarly sized boost. The odds that a student will initially major in STEM go up by the same amount when they move from a different university to Texas A&M as when they begins attending a university in the first place. This is likely because Texas A&M has a program called "25 by 25" that tries to increase the number of engineering majors. Figure 20 in the appendix shows that students at the flagship public schools are more likely to major in STEM and persist in STEM than students at other public universities in Texas.

Table 7: Breaking down the Effect of the Top Ten Percent Rule

	(1) First Major in STEM	(2) Persisted in STEM (Lower Bound)	(3) Persisted in STEM (Upper Bound)
Diff in attending a Uni $(\Delta_{i\hat{i}}U_i)$	0.2573***	0.1042***	0.0853***
	(0.0117)	(0.0097)	(0.0090)
Diff in attending A&M $\left(\Delta_{i\hat{i}}U_i^{AM}\right)$	0.2549***	0.0689***	0.1102***
	(0.0093)	(0.0077)	(0.0071)
Diff in Uni Peers $\left(\Delta_{i\hat{i}}\bar{A}^u\right)$	-0.0309**	0.0176	-0.0097
	(0.0116)	(0.0097)	(0.0090)
Diff in ability compared to HS Peers $(\Delta_{i\hat{i}}s)$	0.0140	-0.0156	-0.0281***
	(0.0096)	(0.0080)	(0.0074)
Constant	0.0952***	0.0891***	0.1111***
	(0.0057)	(0.0048)	(0.0044)
Observations	19567	19567	19567

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The third row of table 7 shows that if the top ten percent rule leads a student to attend a university with higher ability peers, then that will lower their chances of majoring in STEM. Specifically, if the student that received automatic admission is surrounded in college by peers who had high school GPAs that were 1 point higher on average, then that student's chances of initially majoring in STEM decrease by three percentage points.

The fourth row of table 7 is the most interesting. In order to qualify for the top ten percent rule, a student's course performance needs to be better than 90% of their high school peers. As a result, students who receive admission from the top ten rule likely performed better compared to their high school peers than the students they were matched to who did not receive admission from the top ten rule. This may make the student who received the top ten rule relatively more confident in their STEM abilities, even though them and the student they were matched to had the same performance on high school courses.

The fourth row shows some evidence that this is occurring. For the STEM persistence upper bound, a 1 point difference in how a student's GPA compared to their high school peers lowers the effect of receiving automatic admission from the top ten rule by 2.81 percentage points. The effect at the lower bound is also negative, and very close to being statistically significant at the 5% level. The true effect should lie somewhere between these two extremes and be statistically significant. If it were halfway in between, the coefficient would be about -2.19 percentage points.[33] In that case, a 1 point difference in $\Delta_{i\hat{i}}s$ would reduce the effect shown in table 6 from about .13 to .11.

The effect of $\Delta_{i\hat{i}}s$ on initially majoring in STEM is positive but statistically insignificant. This suggests that for students who received top ten admission, having better performance compared to their high school peers than they would have had at other schools does not reduce the positive effects of the top ten rule on initially majoring in STEM. It does, however, reduce the positive effects of the top ten rule on persisting in STEM.

5.2 Fuzzy Regression Discontinuity

Figure 18 shows evidence that there is a small discontinuity in the probability that a student will receive automatic admission from the top ten percent rule. However, this discontinuity disappears when standard errors are clustered at the district level. This is shown in the first row of table 8. As a result, the rest of the regression discontinuity estimation will not work. For completeness, though, I included an updated version of some the results from the first draft below. The rest of the rows in table 8 shows evidence that there are not any discontinuities in other observed variables around the 90% class rank cutoff. Then, table 9 shows the estimated effects of receiving top ten admission, where each cell shows the results of a separate fuzzy regression discontinuity.

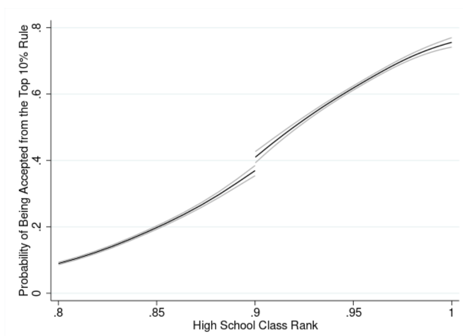


Figure 18: Discontinuity in the Probability of Receiving Automatic Top 10 Admission
Note: The bandwidth used in this figure is .03. Information on a lower bandwidth is shown in table 8.

6 Discussion

Arcidiacono et al. (2016) found that students with lower academic preparation who attended the top ranked campuses in California through an affirmative action program would have had greater chances of graduating with a science degree if they had attended a lower ranked campus. I do not find that result here. Even among the students with the lowest classroom performance in high school, receiving automatic admission from the top ten percent rule was associated with a higher likelihood of initially majoring in STEM and of persisting in a STEM

[33] Another way to guess the true value would be to assume that students who attend universities outside Texas have the same chances of persisting in STEM as the average student in Texas. Remember that the lower bound assumes that all of the students who initially majored in STEM at universities outside of Texas persisted, and the upper bound assumes that none of them persisted. Section 2.2 showed that about 40% of students in Texas who major in STEM persist. Using a weighted average of the effects in 7 with a 40% on the lower bound where all students persist and a 60% weight on the upper bound where all students persist would result in an estimate of -2.31 percentage points.

Table 8: Regression Discontinuity Estimates

	(1)	(2)	(3)
Accepted from Top 10% Rule	0.0286	0.0177	0.0244
	(0.0199)	(0.0280)	(0.0239)
App to Pub Uni in TX	0.0193	0.0139	0.0154
	(0.0170)	(0.0274)	(0.0204)
High School Class Rank	0.0000	0.0000	-0.0000
	(0.0000)	(0.0000)	(0.0000)
Race			
Asian	-0.0094	-0.0250	-0.0178
	(0.0169)	(0.0211)	(0.0183)
Black	0.0119	-0.0102	0.0090
	(0.0180)	(0.0211)	(0.0192)
Hispanic	-0.0025	-0.0101	0.0002
	(0.0354)	(0.0413)	(0.0370)
White	-0.0061	0.0301	-0.0021
	(0.0345)	(0.0402)	(0.0360)
Other	0.0062	0.0152	0.0107
	(0.0053)	(0.0087)	(0.0066)
Elig for Meals	-0.0010	-0.0149	-0.0047
	(0.0304)	(0.0334)	(0.0316)
Home Language			
Spanish	-0.0042	-0.0207	-0.0049
	(0.0285)	(0.0336)	(0.0302)
Other	-0.0016	-0.0121	-0.0085
	(0.0130)	(0.0177)	(0.0145)
Gender			
Male	-0.0297	-0.0329	-0.0356
	(0.0155)	(0.0251)	(0.0195)
Female	0.0297	0.0329	0.0356
	(0.0155)	(0.0251)	(0.0195)
Bandwidth	0.0448	0.0150	0.0578
Degree	1	1	2
Observations	223517	223517	223517

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Standard errors are clustered at the District level, and each cell shows the results of a separate regression discontinuity estimate. Columns (1) and (3) show optimally chosen bandwidths. The bandwidth of .015 was picked in the last draft of this paper as a smaller bandwidth where the effect of attending a public flagship university in table 9 was still statistically significant. “App to Pub Uni in TX” indicates whether a student applied to a public university in Texas.

Table 9: Effects of the Top Ten Percent Rule using Fuzzy Regression Discontinuity

	(1)	(2)	(3)	(4)	(5)	(6)
Went to a Uni	0.1236	0.0573	0.0525	-0.0837	0.1500	0.0597
	(0.4403)	(0.4266)	(1.1948)	(1.1078)	(0.6345)	(0.5876)
First Uni TX Pub Flagship	0.0922	0.1060	1.6476	1.4323	0.4252	0.4101
	(0.2433)	(0.2387)	(2.5288)	(1.9902)	(0.5520)	(0.5345)
First Uni Outside TX	-0.1204	-0.1311	-0.2179	-0.2969	-0.0918	-0.1220
	(0.2018)	(0.2025)	(0.9352)	(0.8588)	(0.4205)	(0.4104)
First Major in STEM	0.2223	0.2490	1.6502	1.6349	0.6377	0.6983
	(0.2734)	(0.2688)	(2.7874)	(2.4665)	(0.7047)	(0.7031)
Persisted in STEM Lower Bound	0.4550	0.4745	0.8940	0.8909	0.8646	0.9237
	(0.2491)	(0.2517)	(1.5971)	(1.4170)	(0.7174)	(0.7318)
Persisted in STEM Upper Bound	0.3805	0.3982	0.9755	0.9755	0.6381	0.6981
	(0.2105)	(0.2113)	(1.6385)	(1.4581)	(0.5360)	(0.5520)
Working in Texas 6 yrs later	-0.2504	-0.3100	-1.5604	-1.6537	-1.0243	-1.0779
	(0.3551)	(0.3614)	(3.1938)	(2.9423)	(1.1597)	(1.1510)
Log Wages in Texas 6 yrs later	-3.5512	-3.9522	-16.3000	-16.9005	-11.1404	-11.3750
	(3.7042)	(3.7706)	(32.9280)	(29.9323)	(12.1206)	(11.8553)
Bandwidth	0.0640	0.0631	0.0150	0.0150	0.0715	0.0751
Demographic Controls	No	Yes	No	Yes	No	Yes
High School Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
Degree	1	1	1	1	2	2
Observations	223517	223517	223517	223517	223517	223517

Standard errors in parentheses

Standard Errors are clustered at the District level.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

All of the columns besides (3) and (4) show optimally chosen bandwidths. In the last draft of the paper, the bandwidth of .015 for columns (3) and (4) was picked as a smaller bandwidth at which the effect associated with attending a public flagship university would still be statistically significant.

degree. One reason for this difference could be that the top state schools in California are generally higher in national rankings than the top state schools in Texas. According to the US News National University rankings, the Berkeley, Los Angeles, Santa Barbara, Irvine, San Diego, and Davis campuses of the University of California, as well as the University of Southern California, are all ranked more highly than the University of Texas at Austin. This may mean that it is more difficult to graduate with a STEM degree from a top school in California, and so students with lower academic preparation that do not persist at those schools might have persisted if they had gone to a top school in Texas.

Second, Arcidiacono et al. (2016) used data between 1995 and 1997. Black et al. (2022) look at the effect of the top ten percent rule in Texas in 1998, and do find that the students who received admission from the top ten percent rule had slightly lower rates of graduating in STEM.[34] It may be that the effect of the top ten percent rule on STEM persistence has changed in Texas since then, and is now positive. The figures in section 2.2 show that the probability that a student will major in STEM and persist in STEM has been increasing in Texas over time. Also, figure 20 shows that the probability that a student will major in STEM has been growing quickly for flagship universities.

A next step for future work will be to take these estimates and use them in a model which incorporates student beliefs more directly. Appendix E includes more details on what a model like this could look like.

7 Conclusion

This paper looked at the connection between student beliefs and the effects of the top ten percent rule on student success in STEM programs. To deal with the issue that data on class ranks is not available for students in Texas, I predicted class ranks using estimates for student grades and for the ways that high schools gathered grades together into a GPA. The results showed that students who receive automatic admission from the top ten percent rule are more likely to initially major in STEM and persist in STEM.

I also found evidence that a student's performance relative to their peers in high school affects their decisions in college. If a student does better compared to their high school peers than they would have at another high school, this does not seem to change the effect of receiving top ten admission on whether they initially major in STEM. However, it does decrease the benefits of receiving top ten admission for persisting in STEM. One possible explanation for this is that students who do better compared to their high school peers than they would have at another school can enter college overconfident about how their STEM abilities will stack up against other students, and become discouraged when they perform worse compared to their new college peers.

[34]This effect was not statistically significant.

References

Thomas Ahn, Peter Arcidiacono, Amy Hopson, and James R. Thomas. Equilibrium grade inflation with implications for female interest in stem majors. Working Paper 26556, NBER, dec 2019.

Tom Ahn, Peter Arcidicacono, Amy Hopson, and James Thomas. Equilibrium grading policies with implications for female interest in stem courses. Econometrica, 92(3):849–880, may 2024.

Rodney J. Andrews, Jing Li, and Michael F. Lovenheim. Quantile treatment effects of college quality on earnings. The Journal of Human Resources, 51(1):200–238, 2016.

Peter Arcidiacono. Ability sorting and the returns to college major. Journal of Econometrics, 121(1-2):343–375, jul 2004.

Peter Arcidiacono, Esteban M Aucejo, and V. Joseph Hotz. University differences in the graduation of minorities in stem fields: Evidence from california. American Economic Review, 106(3):525–562, 2016.

Peter Arcidiacono, Esteban Aucejo, Arnaud Maurel, and Tyler Ransom. College attrition and the dynamics of information revelation. Working Paper, 2021.

Erich Battistin, Agar Brugiavini, Enrico Rettore, and Guglielmo Weber. The retirement consumption puzzle: Evidence from a regression discontinuity approach. The American Economic Review, 99(5):2209–2226, dec 2009.

Sandra Black, Jeff Denning, and Jesse Rothstein. Winners and losers? the effect of gaining and losing access to selective colleges on education and labor market outcomes. American Economic Journal: Applied Economics, 2022.

Zachary Bleemer. Top percent policies and the return to postsecondary selectivity. Working Paper, sep 2020.

Zachary Bleemer and Aashish Mehta. College major restrictions and student stratification. dec 2021.

Lindsay Daugherty, Paco Martorell, and Isaac McFarlin Jr. Percent plans, automatic admissions, and college outcomes. IZA Journal of Labor Economics, 3(10):1–29, 2014.

Laurent Davezies and Thomas Le Barbanchon. Regression discontinuity design with continuous measurement error in the running variable. Journal of Econometrics, 200(2):260–281, oct 2017.

Andrew Foote and Kevin M. Stange. Attrition from administrative data: Problems and solutions with an application to postsecondary education. Working Paper 30232, NBER, jul 2022.

Florian Heiss and Viktor Winschel. Likelihood approximation by numerical integration on sparse grids. Journal of Econometrics, 144(1):62–80, may 2008.

Adam Kapor. Distributional effects of race-blind affirmative action. Working Paper, apr 2020.

Josh Kinsler and Ronni Pavan. Local distortions in parental beliefs over child skill. Journal of Political Economy, 129(1):81–100, 2021.

Kristin Klopfenstein. The ap arms race: Is grade-weighting to blame?, apr 2010. URL https://cpb-us-e2.wpmucdn.com/sites.utdallas.edu/dist/0/221/files/seminar_klopfenstein_advanced_placement_v2.pdf.

Patrick J. McEwan, Sheridan Rogers, and Akila Weerapana. Grade sensitivity and the economics major at a women's college. AEA Papers and Proceedings, 111:102–106, 2021.

National Academies of Sciences, Engineering, and Medicine. Barriers and Opportunities for 2-Year and 4-Year STEM Degrees: Systemic Change to Support Students' Diverse Pathways. National Academic Press, 2016.

Ann L. Owen. Grades, gender, and encouragement: A regression discontinuity analysis. The Journal of Economic Education, 41(3):217–234, jul 2010.

Ralph Stinebrickner and Todd R. Stinebrickner. A major in science? initial beliefs and final outcomes for college major and dropout. The Review of Economic Studies, 81(1 (286)): 426–472, 2014.

Matthew Wiswall and Basit Zafar. Determinants of college major choice: Identification using an information experiment. The Review of Economic Studies, 82(2):791–824, apr 2015.

Appendix

A Additional Figures

Figure 19 shows that students whose first university in Texas is a private school tend to declare majors in STEM less often, and are also less likely to persist in those majors. While information on a student's major at the time of graduating from a private schools is available before 2010, their major while attending is only visible starting in 2010.

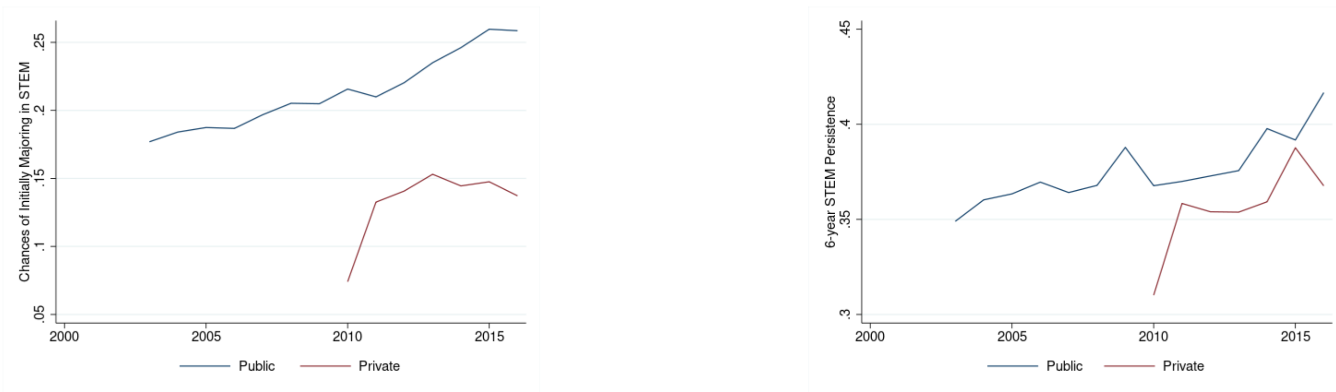


Figure 19: Public and Private Universities

Figure 20 shows that students at flagship public universities tend to major in STEM and persist in STEM more often than students at other public universities in Texas. In 2013, Texas A&M began its "25 by 25" program which was intended to greatly increase the number of engineering majors. This may help to explain the large increase in students' chances of initially majoring in STEM at flagship universities that began around this time.

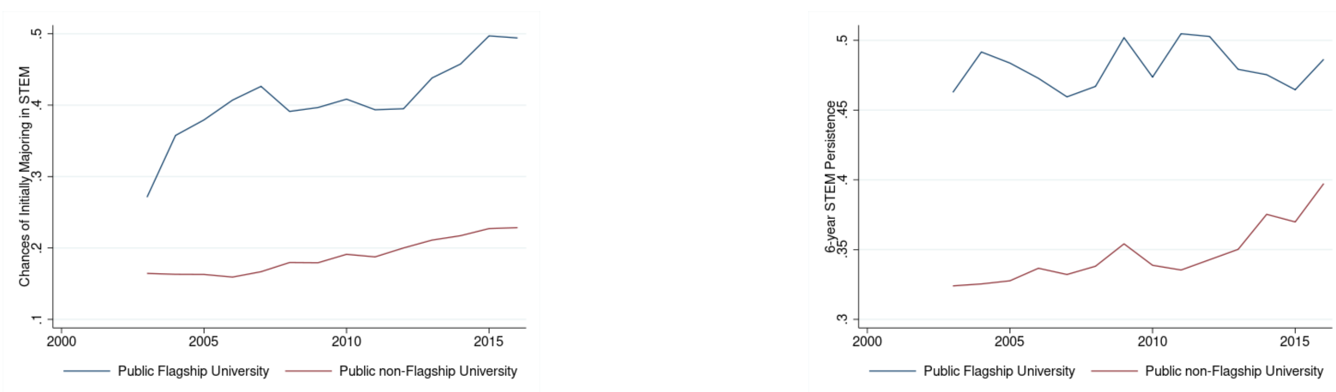


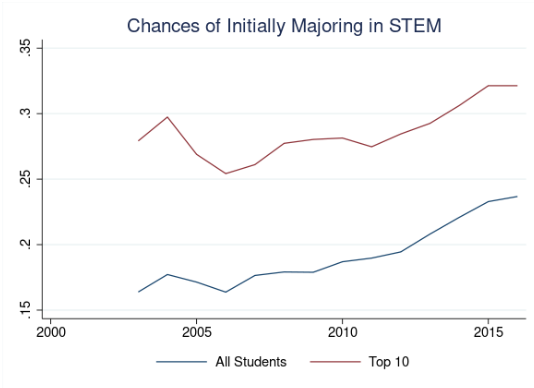
Figure 20: Comparing Flagship Public Universities to non-Flagship Public Universities

Figure 21 divides schools into four quartiles based on their TAKS standardized test scores in math for the year 2011. In all cases, students in the top 10% are more likely to major in STEM and persist in STEM degrees.

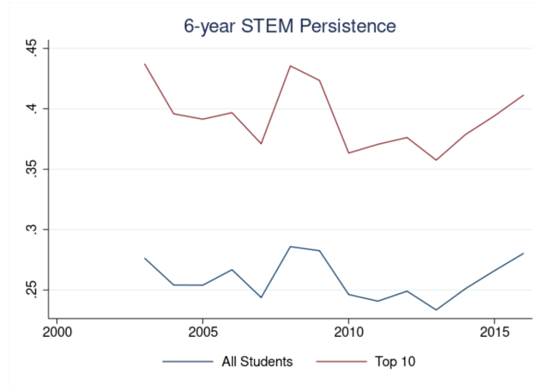
B Additional Information on Data Preparation

B.1 Predicting Course Grades

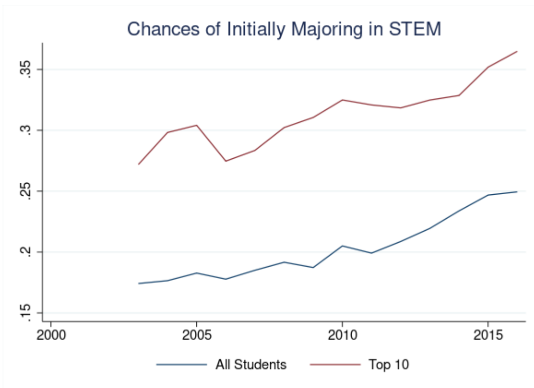
This section provides some additional details about the linear regression models used to predict grades in section 3.1. The second regression model that was used includes seventh and eighth



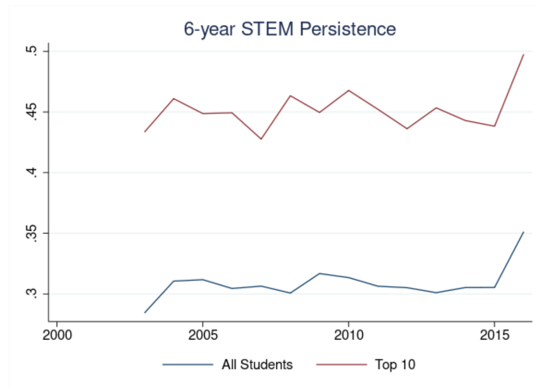
(a) Quartile 1



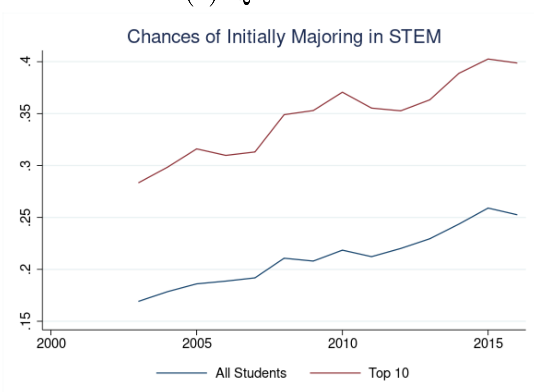
(b) Quartile 1



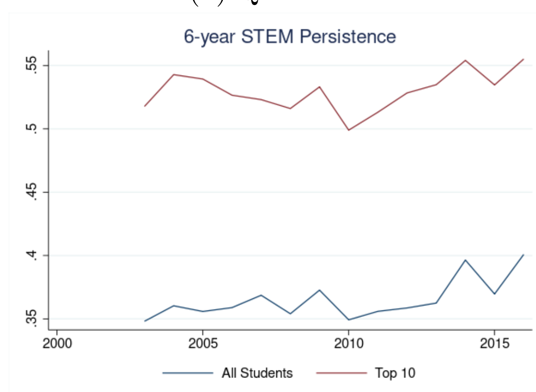
(c) Quartile 2



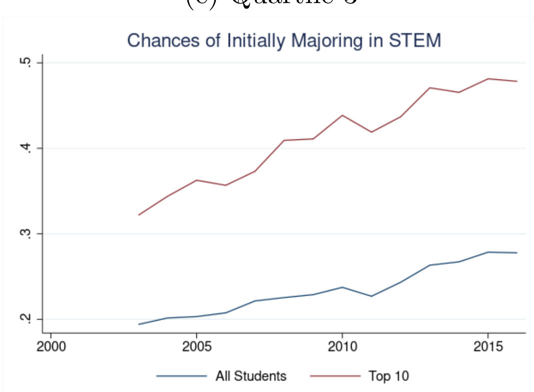
(d) Quartile 2



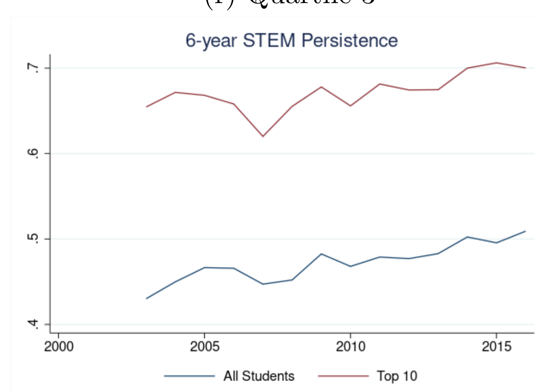
(e) Quartile 3



(f) Quartile 3

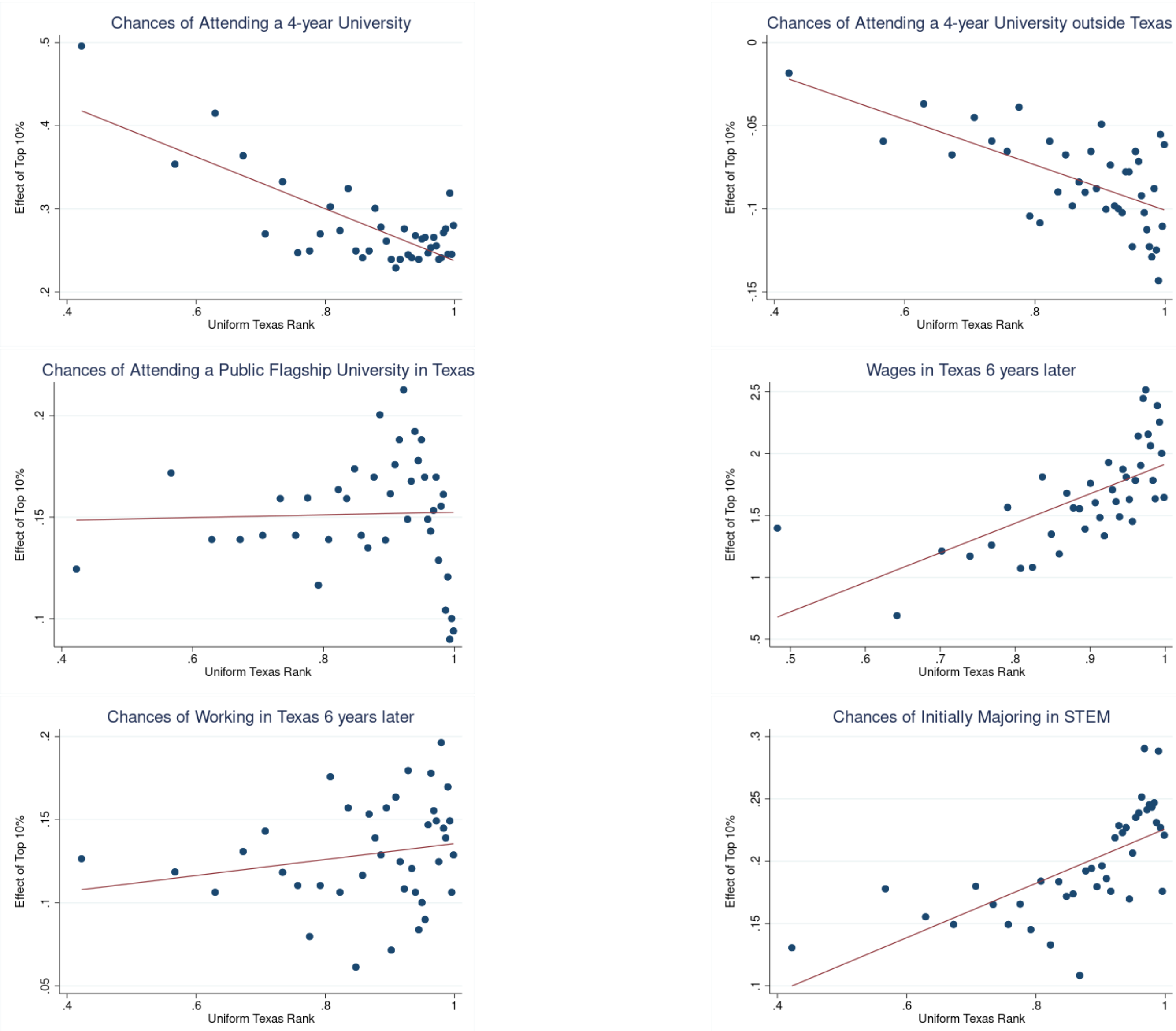


(g) Quartile 4



(h) Quartile 4

Figure 21: Students in the Top 10% At all Quartiles of High School



(a) Chances of Persisting in STEM

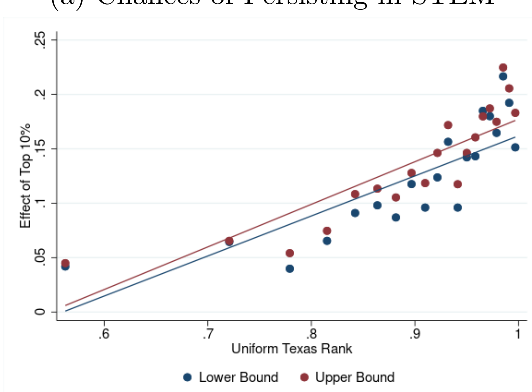


Figure 22: Variation in Treatment Effects Across Ability with Nearest Neighbor Matching
Note: I accidentally transferred the wrong file from the ERC computer, so the "Wages in Texas 6 years later" figure does not have the correction made for grade inflation. All of the other figures do.

Table 10: How Often the Three Regression Models were Used

Regression	Total Predicted Grades (millions)	% of All Predicted Grades
1	26.356	98.051
2	0.164	0.609
3	0.360	1.340

If a grade had the necessary information to be predicted with regression 1, then that was used. If not, then regression 2 was used, and after that regression 3.

grade TAKS standardized test scores, but not every student has recorded scores for these tests. For the students that have other TAKS standardized test scores recorded for prior grade levels, I fix this issue by using a lasso algorithm to predict their seventh and eighth grades test scores with the TAKS scores that they do have. These prior TAKS test scores could go back as early as the third grade. Unfortunately, it is not possible to do this for students that never took a TAKS test before high school. This could happen, for example, if the students moved to Texas for high school from another state. The third regression deals with cases like these by not including TAKS standardized test scores.

Using these regression models, I predict the grades data for all of the 2011-2012 school year. Also, for all students that were juniors in the 2011-2012 through 2013-2014 school years, I predict their grades data in the 2012-2013 through the 2014-2015 school years. Table 10 shows how often each model ended up being used to estimate grades when the true grades were not available. In most cases, it was possible to use the first model, which included fixed effects for individuals, classes, and sections of classes,[35] as well as fixed effects related to whether a student passed or failed and received credit or not. The second and third model both used more general information about students and classes instead of fixed effects, and the only difference between them was that the second model included TAKS standardized test scores.

Not all of the grades represented in table 10 were used later on in the paper. The only ones that were used to estimate things later on were grades for students who were juniors in the 2013-2014 school year. However, the other predictions will hopefully be valuable in the future, because I now have access to some additional data that will allow me to estimate things later on using grades for students who were juniors in other years as well. This additional data would also include some information on grades for the 2010-2011 school year, which would allow me to test the prediction error when the models are used to predict a different year of data.

B.2 Predicting GPA Rules

In this section, I will provide some additional details about how the variance of the error term for GPAs is estimated, and also about how a district's rules for finding GPAs are estimated in equation 3. First, GPAs are a credit-weighted average of grades in individual classes:[36]

$$R^d\left(G_i\right) = \frac{1}{\sum\limits_{c \in S_i^d} w^c} \sum_{c \in S_i^d} \left(\left(4 - \left(100 - g_i^c\right) \left(\frac{1}{10} \right) \right) + p^c \right) w^c \tag{12}$$

[35] For example, there would be a fixed effect for being in the first half of a two semester course.

[36] Here and throughout this section, each portion of a course which receives its own grade counts as a "class." For example, a typical year-long Algebra 1 course would include two classes: the first semester of Algebra 1, and the second semester of Algebra 1.

Here, p^c is the additional grade points given to class c according the school district's rules on extra grade points P^d. Remember that S_i^d is the set of classes that an individual has taken out of the set of classes S^d that their school district uses to calculate GPAs, and each w^c term is the number of high school credits associated with class c.

Since not all of a student's true grades are recorded, I have to use estimated grades $\hat{G}_i$ instead of the true grades G_i. For the estimated grades, each unknown grade g_i^c is replaced with the estimate $\hat{g}_i^c$, where $g_i^c = \hat{g}_i^c + \epsilon_i^c$. Substituting in this expression and rearranging yields

$$R^d\left(G_i\right) = R^d\left(\hat{G}_i\right) + \frac{1}{\sum\limits_{c \in S_i^d} w^c} \sum_{c \in S_i^d} \frac{w^c \epsilon_i^c}{10} \tag{13}$$

For simplicity, call the second term involving the errors ϵ_i^G, so that $R^d\left(G_i\right) = R^d\left(\hat{G}_i\right) + \epsilon_i^G$.

In the paper, I assumed that $\epsilon_i^c = \eta_i + \nu_i^c$, where η_i and ν_i^c are pulled from independent normal distributions with means of zero and standard deviations of σ^η and σ^ν. The variance of ϵ_i^G is

$$\left(\sigma^G\right)^2 = \sum_{c \in S_i^d} \left(\frac{w^c}{10 \sum\limits_{c \in S_i^d} w^c} \right)^2 \operatorname{Var}\left(\epsilon_i^c\right) + 2 \sum_{\substack{c,c' \in S_i^d \\ c \neq c'}} \frac{w^c w^{c'}}{\left(10 \sum\limits_{c \in S_i^d} w^c\right)^2} \operatorname{Cov}\left(\epsilon_i^c, \epsilon_i^{c'}\right) \tag{14}$$

Since student's real grades are available for some of the classes that they take, many of these ϵ_i^c terms are zero. Let $\hat{S}_i^d$ represent the classes that need to use estimated grades out of the full list S_i^d of relevant classes that they that students took. Then, $\operatorname{Var}\left(\epsilon_i^c\right) = \left(\sigma^\eta\right)^2 + \left(\sigma^\nu\right)^2$ and $\operatorname{Cov}\left(\epsilon_i^c, \epsilon_i^{c'}\right) = \left(\sigma^\eta\right)^2$, so this can be rewritten as

$$\left(\sigma^G\right)^2 = \left(\left(\sigma^\eta\right)^2 + \left(\sigma^\nu\right)^2\right) \sum_{c \in \hat{S}_i^d} \left(\frac{w^c}{10 \sum\limits_{c \in S_i^d} w^c} \right)^2 + 2\left(\sigma^\eta\right)^2 \sum_{\substack{c,c' \in \hat{S}_i^d \\ c \neq c'}} \frac{w^c w^{c'}}{\left(10 \sum\limits_{s \in S_i^d} w^c\right)^2} \tag{15}$$

To find the expected variance of ϵ_i^G for students in general, I need to take the expectation of this over individuals i. To make this simpler, in the paper I assumed that η_i and ν_i^c were independent of the set of classes S_i^d and the credits associated with those classes w^c. Then, the expected variance is

$$\left(\sigma^G\right)^2 = \left(\left(\sigma^\eta\right)^2 + \left(\sigma^\nu\right)^2\right) \mathbb{E}_i \left(\sum_{c \in \hat{S}_i^d} \left(\frac{w^c}{10 \sum\limits_{c \in S_i^d} w^c} \right)^2 \right) + 2\left(\sigma^\eta\right)^2 \mathbb{E}_i \left(\sum_{\substack{c,c' \in \hat{S}_i^d \\ c \neq c'}} \frac{w^c w^{c'}}{\left(10 \sum\limits_{s \in S_i^d} w^c\right)^2} \right) \tag{16}$$

where the expectations are being taken over individuals.

I calculate $\left(\sigma^\eta\right)^2$ as

$$\left(\sigma^\eta\right)^2 = \mathbb{E}_i \left(\mathbb{E}_{\substack{c,c' \in \hat{S}_i^d \\ c \neq c'}} \left(\operatorname{Cov}\left(\epsilon_i^c, \epsilon_i^{c'}\right)\right)\right) = \sum_{\substack{c,c' \\ c \neq c'}} \operatorname{Cov}\left(\epsilon_i^c, \epsilon_i^{c'}\right) \mathbb{E}_i \left[\frac{\mathbb{1}\left(c, c' \in \hat{S}_i^d\right)}{|\hat{S}_i^d| \left(1 - |\hat{S}_i^d|\right)} \right] \tag{17}$$

Calculating this using all possible pairs of classes taken by all students would take a very long time. To speed things up, I use one school year of data for a random sample of students, and only include their 750 most-taken classes.[37] Then, $(\sigma^{\nu})^2$ is calculated as

$$(\sigma^{\nu})^2 = \mathbb{E}_i \left[\mathbb{E}_{c\in \hat{S}_i^d} \left[\mathrm{Var}\left(\epsilon_i^c\right)\right]\right] - (\sigma^{\eta})^2 = \sum_c \mathrm{Var}\left(\epsilon_i^c\right) \mathbb{E}_i \left[\frac{\mathbb{1}\left(c \in \hat{S}_i^d\right)}{|\hat{S}_i^d|}\right] - (\sigma^{\eta})^2 \tag{18}$$

The $\mathrm{Var}\left(\epsilon_i^c\right)$ and $\mathrm{Cov}\left(\epsilon_i^c, \epsilon_i^{c'}\right)$ above are found using information from the 2011-2012 school year, because that is the only year that the ϵ_i^c terms could be observed when these results were being prepared. Instead using all of the ϵ_i^c values of that year, I used only the ones from the sample of data that was not used to estimate the model. Hopefully these ϵ_i^c terms will be a better representation of the errors that I get when predicting the unknown grades.

Even though the error terms were only available in the 2011-2012 school year, I could calculate the expectations in equations 17 and 18 using data from the school years where grades needed to be predicted. In other words, I had to assume that the variances and covariances of epsilon terms from different classes were the same over time, but could try to weight these terms to better match the classes that students take in the years where their grades needed to be estimated.

Table 11 shows the resulting estimates for σ^G. Since these estimates depend on the set of classes S^d that a school district uses to calculate GPAs, they are calculated separately for different possible sets of classes.

Table 11: Estimates for the Standard Deviation of ϵ_i^G

Class Group	σ^G
All Classes	0.2116
No P.E. Classes	0.2187
Standard	0.2614

These terms were estimated using data from at least 8000 students.

Let F^h represent the cdf of GPAs within high school h. The class rank of a student in district d with grades G_i is given by $F^h\left(R^d\left(G_i\right)\right)$. Since I do not observe the true set of grades G_i for each student, I also do not know the exact GPA values needed to qualify for the top ten percent rule. If I could see the true data for grades, this would just be the grade of the student at the 90th percentile. However, the errors in my prediction of grades make it so that I do not know which student is at the 90th percentile, and even if I did then I still would not know their grade for certain.

One way of dealing with the measurement error is to estimate the $F^h\left(R^d\left(\cdot\right)\right)$ cdf using convolution methods. Remember that $R^d\left(G_i\right) = R^d\left(\hat{G}_i\right) + \epsilon_i^G$, and that ϵ_i^G is normally distributed. Since $R^d\left(\hat{G}_i\right)$ and ϵ_i^G are independent, the cdf of $R^d\left(G_i\right)$ is given by

$$F^h\left(t\right) = \int_{-\infty}^{\infty} H\left(u\right)\left(\frac{1}{\sigma^G}\right)\phi\left(\frac{t-u}{\sigma^G}\right)du \tag{19}$$

[37] Before taking this random sample, I limit the data for each student to school years where they took at least two classes. Then, I find the 750 most-taken classes using 1 school year of information for 10,000 students. Not all of these students necessarily took at least two classes in the final list of 750, and so before taking the expectation I remove students that did not take at least two of these classes.

where H^h is the cdf of the estimated GPAs $R^d\left(\hat{G}_i\right)$. The empirical version of this cdf $\hat{H}^h$ can be used as an approximation. One of the useful things about the empirical cdf is that it is zero for all values below the minimum GPA in that high school, and one for all values above the maximum GPA. This allows 19 to be simplified as

$$\hat{F}^h(t) = \left(1 - \Phi\left(\frac{\max\left(R^d\left(\hat{G}_i\right)\right) - t}{\hat{\sigma}^G}\right)\right) + \int_{\min\left(R^d\left(\hat{G}_i\right)\right)}^{\max\left(R^d\left(\hat{G}_i\right)\right)} \hat{H}^h(u)\left(\frac{1}{\hat{\sigma}^G}\right)\phi\left(\frac{t-u}{\hat{\sigma}^G}\right) du \quad (20)$$

When I try later on to estimate the rules R^d that each high school uses to calculate GPAs, this expression will need to be evaluated separately for every high school, and for every set of rules that is considered. To do this quickly, I calculate the integral here using sparse grid quadrature with weights from Heiss and Winschel (2008).

In terms of the class rank "bins" that I can observe in the data, the fact that a student's class rank falls into the bin $B_i = \left[\tau_i^L, \tau_i^U\right)$ means that $\tau_i^L < F\left(R^d(G_i)\right) \leq \tau_i^U$. A student qualifies for the top ten percent rule if $F^h\left(R^d(G_i)\right) > .9$. For each possible class rank boundary τ other than 0 or 1, I use estimated cdf $\hat{F}^h$ to find the GPA associated with that class rank. Specifically, this is the GPA $R^d(G)_{[\tau]}$ which satisfies

$$\hat{F}^h\left(R^d(G)_{[\tau]}\right) = \tau \quad (21)$$

For example, $R^d(G)_{[.9]}$ is the minimum GPA that would allow a student to qualify for the top ten percent rule.

Then, the probability that a student's class rank was in the bin $B_i = \left[\tau_i^L, \tau_i^U\right)$ given their estimated grades $\hat{G}_i$ and district rules R^d is

$$\begin{aligned}
\mathbb{P}\left(B_i \mid R^d\left(\hat{G}_i\right)\right) &= \mathbb{P}\left(R^d(G)_{\left[\tau_i^L\right]} \leq R^d\left(\hat{G}_i\right) + \epsilon_i^G < R^d(G)_{\left[\tau_i^U\right]}\right) \\
&= \Phi\left(\frac{R^d(G)_{\left[\tau_i^U\right]} - R^d\left(\hat{G}_i\right)}{\hat{\sigma}^G}\right) - \Phi\left(\frac{R^d(G)_{\left[\tau_i^L\right]} - R^d\left(\hat{G}_i\right)}{\hat{\sigma}^G}\right)
\end{aligned}$$

With this probability prepared, I have everything that I need to estimate equation 3 by trying many possible district rules R^d and choosing the one which maximizes the likelihood.

B.3 Removing Grade Inflation

This section provides some additional details on how equations 4 and 5 can be used to identify the grade inflation π^{ch} for class c in high school h. Taking the covariances of the g_i^c and E_i^c terms, both conditioning on school districts d and not conditioning on school districts, gives 20 equations. These are $\mathrm{Var}(g_i^c)$, $\mathrm{Var}(E_i^c)$, $\mathrm{Cov}(g_c^c, E_i^c)$, $\mathbb{E}(\mathrm{Var}(g_i^c \mid d))$, $\mathbb{E}(\mathrm{Var}(E_i^c \mid d))$, and $\mathbb{E}(\mathrm{Cov}(g_i^c, E_i^c \mid d))$ for both c and c', as well as covariance terms between classes: $\mathrm{Cov}\left(g_i^c, g_i^{c'}\right)$, $\mathrm{Cov}\left(g_i^c, E_i^{c'}\right)$, $\mathrm{Cov}\left(g_i^{c'}, E_i^c\right)$, $\mathrm{Cov}\left(E_i^c, E_i^{c'}\right)$, $\mathbb{E}\left(\mathrm{Cov}\left(g_i^c, g_i^{c'} \mid d\right)\right)$, $\mathbb{E}\left(\mathrm{Cov}\left(g_i^c, E_i^{c'} \mid d\right)\right)$, $\mathbb{E}\left(\mathrm{Cov}\left(g_i^{c'}, E_i^c \mid d\right)\right)$, and $\mathbb{E}\left(\mathrm{Cov}\left(E_i^c, E_i^{c'} \mid d\right)\right)$.

Estimates for the variances and covariances conditional on a school district could be noisy for small districts. To get around this issue, I will calculate the conditional variance and covariance terms using the equivalent expressions from the law of total variance and covariance. For example, I will calculate $\mathbb{E}(\mathrm{Var}(g_i^c \mid d))$ as

$$\mathbb{E}(\mathrm{Var}(g_i^c \mid d)) = \mathrm{Var}(g_i^c) - \mathrm{Var}(\mathbb{E}(g_i^c \mid d))$$

and $\mathbb{E}\left(\operatorname{Cov}\left(g_i^c, g_i^{c'} \mid d\right)\right)$ as

$$\mathbb{E}\left(\operatorname{Cov}\left(g_i^c, g_i^{c'} \mid d\right)\right) = \operatorname{Cov}\left(g_i^c, g_i^{c'}\right) - \operatorname{Cov}\left(\mathbb{E}\left(g_i^c \mid d\right), \mathbb{E}\left(g_i^{c'} \mid d\right)\right)$$

The equations that condition on d are especially useful. For example, in

$$\mathbb{E}\left(\operatorname{Var}\left(g_i^c \mid d\right)\right) = \mathbb{E}\left(\left(\beta_1^{cg}\right)^2 \operatorname{Var}\left(a_i^c \mid d\right) + \operatorname{Var}\left(\pi_i^{ch} \mid d\right) + 2\beta_1^{gc}\operatorname{Cov}\left(\pi^{ch}, a_i^c \mid d\right) + \operatorname{Var}\left(\eta_i^{gc} \mid d\right)\right)$$

the $\operatorname{Var}\left(\pi_i^{ch} \mid d\right)$ and $\operatorname{Cov}\left(\pi^{ch}, a_i^c \mid d\right)$ are both zero because π^{ch} is constant within each school district d. Additionally, since η_i^{gc} is independent of d, $\mathbb{E}\left(\operatorname{Var}\left(\eta_i^{gc} \mid d\right)\right) = \operatorname{Var}\left(\eta_i^{gc}\right)$. This allows the above expression to be rewritten more simply as

$$\mathbb{E}\left(\operatorname{Var}\left(g_i^c \mid d\right)\right) = \left(\beta_1^{cg}\right)^2 \mathbb{E}\left(\operatorname{Var}\left(a_i^c \mid d\right)\right) + \operatorname{Var}\left(\eta_i^{gc}\right)$$

These 20 equations involve 20 unknowns, and so the full system is just identified. There are β_1^{gc}, β_1^{Ec}, $\operatorname{Cov}\left(\pi^{ch}, a_i^c\right)$, $\mathbb{E}\left(\operatorname{Var}\left(a_i^c \mid h\right)\right)$, $\operatorname{Var}\left(\eta_i^{gc}\right)$, $\operatorname{Var}\left(\eta_i^{Ec}\right)$, and $\operatorname{Var}\left(\pi^{ch}\right)$ terms for both c and c', as well as the terms involving both tests: $\operatorname{Cov}\left(a_i^c, a_i^{c'}\right)$, $\mathbb{E}\left(\operatorname{Cov}\left(a_i^c, a_i^{c'} \mid d\right)\right)$, $\operatorname{Cov}\left(a_i^c, \pi^{c'h}\right)$, $\operatorname{Cov}\left(a_i^{c'}, \pi^{ch}\right)$, $\operatorname{Cov}\left(\pi^{ch}, \pi^{c'h}\right)$, and $\operatorname{Cov}\left(\eta_i^{gc}, \eta_i^{gc'}\right)$.

Solving this system of equations for the β_1 terms gives the following results:

$$\beta_1^{Ec} = \sqrt{\hat{\operatorname{Var}}\left(\mathbb{E}\left(E_i^c \mid d\right)\right) + \frac{\hat{\mathbb{E}}\left(\operatorname{Cov}\left(g_i^c, E_i^c \mid d\right)\right)}{\hat{\mathbb{E}}\left(\operatorname{Cov}\left(g_i^c, E_i^{c'} \mid d\right)\right)}\hat{\mathbb{E}}\left(\operatorname{Cov}\left(E_i^c, E_i^{c'} \mid d\right)\right)}$$

$$\beta_1^{Ec'} = \sqrt{\hat{\operatorname{Var}}\left(\mathbb{E}\left(E_i^{c'} \mid d\right)\right) + \frac{\hat{\mathbb{E}}\left(\operatorname{Cov}\left(g_i^{c'}, E_i^{c'} \mid d\right)\right)}{\hat{\mathbb{E}}\left(\operatorname{Cov}\left(g_i^{c'}, E_i^c \mid d\right)\right)}\hat{\mathbb{E}}\left(\operatorname{Cov}\left(E_i^c, E_i^{c'} \mid d\right)\right)}$$

$$\beta_1^{gc} = \frac{\hat{\mathbb{E}}\left(\operatorname{Cov}\left(g_i^c, E_i^{c'} \mid d\right)\right)}{\hat{\mathbb{E}}\left(\operatorname{Cov}\left(E_i^c, E_i^{c'} \mid d\right)\right)}\beta_1^{Ec}$$

$$\beta_1^{gc'} = \frac{\hat{\mathbb{E}}\left(\operatorname{Cov}\left(g_i^{c'}, E_i^c \mid d\right)\right)}{\hat{\mathbb{E}}\left(\operatorname{Cov}\left(E_i^c, E_i^{c'} \mid d\right)\right)}\beta_1^{Ec'}$$

where the hats indicate that these quantities can be calculated directly from the data.

The purpose of attempting to correct the grades data for grade inflation is to make the grade more accurately reflect a student's true performance in the course. If I could make this correction perfectly, then in the matching exercise two matched students would have the same high school course performance, even if one student's school decided to give them all A's and the other student's school decided to give them all B's. In the propensity score matching estimation, I also control for eighth grade standardized test scores in math and reading.[38] Once I have matched students on these test scores and their course performance, the idea is that the only remaining difference between the student that received the top ten percent rule and the one that did not will be how their shared performance stacked up against their high school peer group.

There are at least two ways why people might disagree with what I am doing here. First, people might say that I should not be correcting for grade inflation at all, and that the actual grades that students receive are what is relevant. For example, Ahn et al. (2019) found that providing students better grades in STEM classes for the same performance would help to

[38] Unfortunately, I did not include these standardized test scores in the Mahalanobis matching. I plan on doing this in the future.

make female students persist in STEM at rates closer to their male peers, and also increase the total number of students that enroll in STEM. The idea here is that even if a student did better than anyone else in their school in a STEM class, they would still find it discouraging if their letter grade were a B rather than an A. This seems reasonable, and in appendix D I also show a version of the results that does not correct for grade inflation. The fact that these results are close to the results that did correct for grade inflation provide encouraging evidence that grade inflation may not be having a big effect.

Second, people could think that I am not doing enough to correct for grade inflation. They would probably say that the actual course performance of students (independently of their grades) is a better indicator of how students will perform in college classes, and so students should be matched based on their standardized test scores and inflation-corrected grades. Some of the assumptions that I have to make to correct for grade inflation are very strict, like the assumption that grade inflation is the same for all classes in a school district. There could be a concern about how much of the true grade inflation this is capturing. To those people I would say that what I am doing is at least a step in the right direction. Given that it is a step in the right direction, the fact that it produces very similar results compared to the uncorrected grade data should also be encouraging, because this suggests that grade inflation is not having a very large effect.

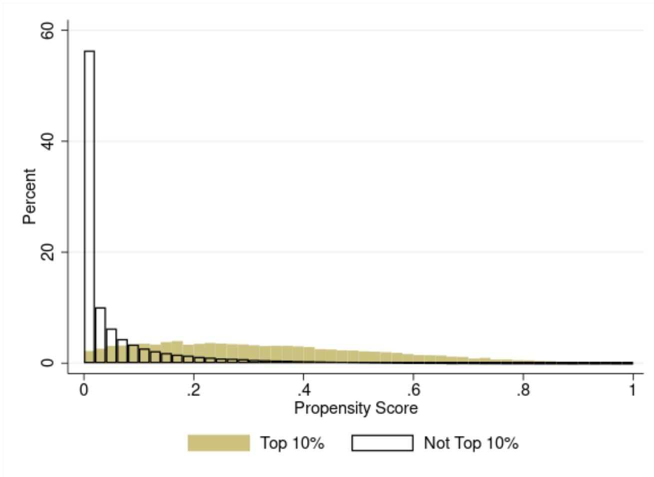
C Additional Information on the Results

Figure 23 shows that there is common support for propensity score matching. It does not look like there are very many observations that are not in the top ten percent for the higher propensity scores, but it is important to remember that these account for about 90% of the students in the data. The last graph in the figure shows what happens when the propensity scores are estimated just for students that went to college, and there it is more clear that there is common support.

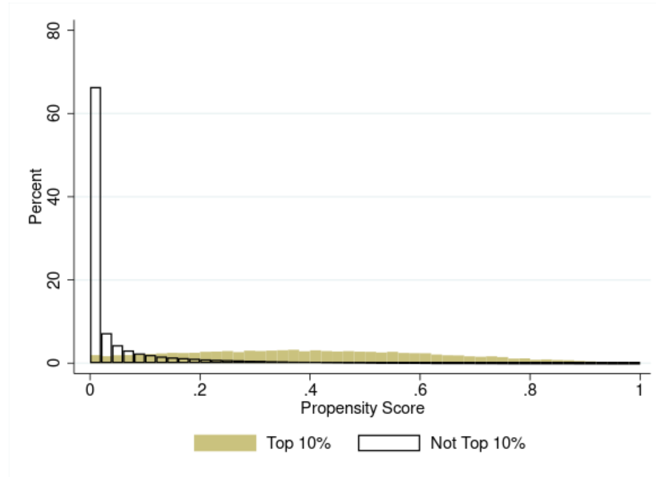
For all students who applied to public universities in Texas, the data shows whether they qualified for the top ten percent rule even if they did not actually attend a public university. Unfortunately, it does not include this information for students who did not apply to public universities. To deal with this, I mark those students as not receiving automatic admission under the top ten percent rule. This is true, because even if any of those students were in the top ten percent of their class rank, they would still not receive automatic admission unless they actually apply to a public university in Texas. This means that the matching results in section 5 are specifically measuring the effect of "receiving automatic admission under the top ten percent rule" rather than "qualifying for automatic admission under the top ten percent rule."

One concern that people may have about this is that if a student qualified for automatic admission under the top ten percent rule but did not apply to any public universities in Texas, that student might be different in unobservable ways from other students with the same course performance and demographic ability. For example, if students think that public schools are an acceptable place to learn STEM but provide a lower quality education in non-STEM fields, then high-performing students who plan on majoring in something other than STEM might not apply to any public universities in Texas. This could put an upward bias on the effect of the top ten percent rule on majoring in STEM found by matching, because students who received admission under the top ten percent rule could be matched to students that also qualified for the top ten percent rule, but did not apply to any public colleges because they did not want to major in STEM.

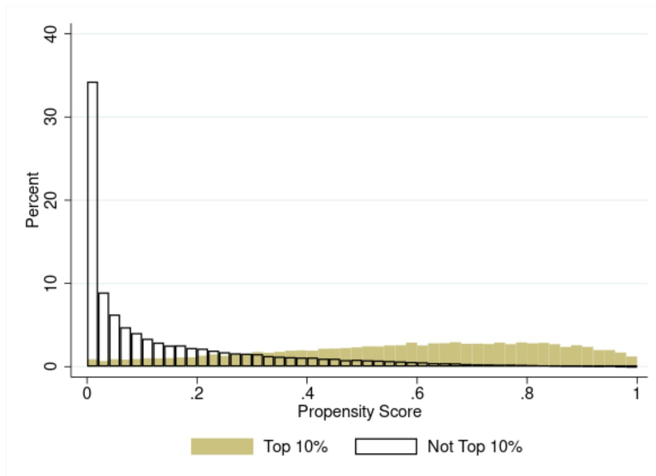
If this bias exists, it should only affect the matching results for people who are matched with students that qualified for the top ten percent rule but did not apply to any public university



(a) Not Conditional on Attending College and Corrected for Grade Inflation



(b) Not Conditional on Attending College and Not Corrected for Grade Inflation



(c) Conditional on Attending College and Not Corrected for Grade Inflation

Figure 23: The Common Support Condition for Propensity Score Matching

in Texas because they did not want to major in STEM. Fortunately, a large portion of students predicted to be in the top ten percent by their class rank applied to some public university in Texas. Figure 24 shows the probability that a student will apply for the fall term at a public university in Texas after their junior year in high school. As the figure shows, about 75% of the people predicted to be in the top ten percent did apply to a public university in Texas.

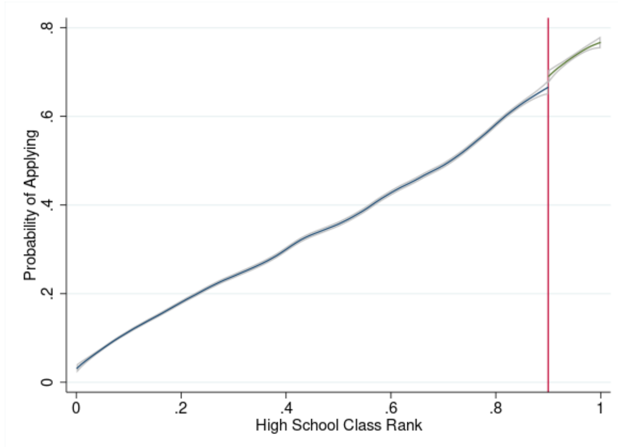


Figure 24: Discontinuity in the Probability of Applying to a Public University in Texas
Note: The bandwidth used in this figure is .03. Information on the discontinuity around the 90% cutoff at a lower bandwidth is shown in table 8.

In the future, there are several things that I would like to do to either limit the possibility for this kind of bias in the matching estimation or show that it is unlikely to cause a problem. First, since I know the majors for all students who went to college, I could compare whether students who who apply to public universities in Texas have similar chances of initially majoring in STEM. If they were similar, that would lower concerns that students avoid applying to public universities in Texas because they do not want to major in STEM. Second, I could redo all of the results using only the students that did apply to a public university in Texas, since the data directly shows whether these students qualified for the top ten percent rule.

There are also some things that I could do to try and understand the effects of qualifying for the top ten percent rule rather than receiving automatic admission under the top ten percent rule. First, I could see what percent of all students have applications to public universities in Texas that show that they are in the top ten percent. I know that ten percent of students qualified for the top ten percent rule, and so the difference between this number and 10% should tell me how many students qualified for the top ten percent rule but did not apply to a public university. This number could maybe be used to place some kind of a bound on how different the estimates shown here could be from estimates that directly measured the effects of qualifying for the top ten percent rule. Second, I could redo all of the results assuming that all students with a predicted class rank above the 90% cutoff qualified for the top ten percent rule.

D Not Correcting for Grade Inflation

This section presents some of the results from section 5 without the correction that was made for grade inflation. Since the correction for grade inflation was done at the school district level, students' class ranks within their high school were not affected. This means that all of the fuzzy regression discontinuity results are not changed.

Figure 25 shows the distribution of students admitted from the top 10% rule using three different "ranking" options without correcting for inflation. Compared to the figure 16 which corrects for inflation, the distribution of students in the top 10% of their classes shifts slightly to the right for the uniform rank within Texas. This suggests that even if two students qualified under the top ten percent rule with identical GPAs that were calculated in the same way, their grades might still reflect different ability levels if their school districts used different amounts of grade inflation.

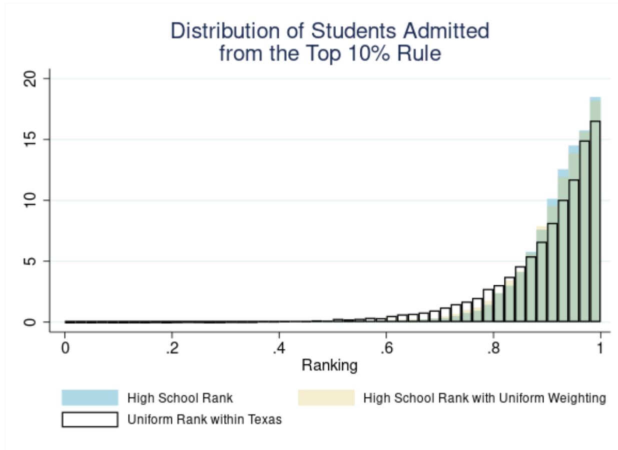


Figure 25: Variation in who Receives Automatic Admission

However, the difference between these two graphs is small, and removing the correction for grade inflation effects the results of the propensity score matching estimation very little. Table 12 shows the results of the matching methods used in table 6, but without the correction for grade inflation. For mahalanobis matching, many of the effects and standard errors are the same to the fourth decimal point. The propensity score matching methods have slightly different results, but the changes are not large.

Table 13 shows a version of table 7 without correcting for grade inflation. The results are similar. In both tables, if a high school student does better compared to their peers than they would at a different high school, this decreases the benefits of receiving top ten admission for persisting in STEM.

E Possible Model

This section suggests a model with two universities. One of them represents the public flagship universities in Texas and the other represents public universities in Texas that are not the flagship schools. All students apply to both universities, and they are accepted or rejected based on standards chosen by the university. Then, students choose between the university options that they have been accepted to and decide whether or not to attempt majoring in STEM. Initially majoring in STEM and then persisting in a STEM degree both require effort, and this effort depends upon a student's true ability which they do not know for sure. After making their initial decision between majors, students see their true ability and decide whether or not to persist in STEM. If they do not persist in STEM, they will graduate with a non-STEM degree.

The possible model suggested here builds on the framework shown in section 4. Remember that $A_i = h_i + a_i$, which can be rewritten as

$$A_i = X_i\beta + \epsilon_i^h + s_i + \epsilon_i^s$$

$X_i\beta$ is the expected average ability of a student's high school peers given their demographics X_i, and s_i is the difference between a student's ability and the average ability at their high

Table 12: Effects of the Top Ten Percent Rule using Matching and not Correcting for Inflation

	(1)	(2)	(3)
Went to a Uni	0.2946***	0.2863***	0.2821***
	(0.0057)	(0.0055)	(0.0042)
First Uni Outside TX	-0.0789***	-0.0826***	-0.0814***
	(0.0043)	(0.0044)	(0.0033)
First Uni TX Pub Flagship	0.1669***	0.1617***	0.1627***
	(0.0037)	(0.0037)	(0.0033)
First Major in STEM	0.1954***	0.1936***	0.1925***
	(0.0054)	(0.0053)	(0.0044)
Persisted in STEM Lower Bound	0.1261***	0.1231***	0.1206***
	(0.0042)	(0.0042)	(0.0035)
Persisted in STEM Upper Bound	0.1382***	0.1382***	0.1352***
	(0.0037)	(0.0037)	(0.0032)
Working in Texas 6 yrs later	0.1555***	0.1584***	0.1547***
	(0.0065)	(0.0063)	(0.0050)
Log Wages in Texas 6 yrs later	1.6250***	1.6595***	1.6191***
	(0.0644)	(0.0633)	(0.0503)
Method	Mahalanobis	Nearest Neighbor	Nearest 5 Neighbors
Observations	223517	223517	223517

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

school using the school's method for finding GPAs. I will assume that students know $X_i\beta$ and s_i, but that they do not know ϵ_i^h or ϵ_i^s. For simplicity, let $\epsilon_i^{hs} = \epsilon_i^h + \epsilon_i^s$.

The signals that students receive are normally distributed within each school, and I will assume that they all have the same standard deviation σ_h^s. Students receive automatic admission from the top percent rule if

$$\frac{s_i}{\sigma_h^s} \geq \tau^T$$

In Texas, the τ^T threshold is set so that the top ten percent qualify, which is $\tau^T = 1.28$ in this model since signals are normally distributed.

Before the students apply, each universities u decides a threshold for accepting applications τ_u^a and how much effort c_u^S their STEM program will require in order to graduate with a STEM degree. A student who is not automatically accepted from the top ten percent rule is admitted to the school if

$$A_i > \tau_u^a$$

This means that the students who receive this kind of regular admission learn something about their ϵ_i^{hs} term. For example, if a student is accepted into the non-flagship school nFS but not into the flagship school FS, then they know that

$$\tau_{nFS}^a \geq X_i\beta + s_i + \epsilon_i^{hs} < \tau_{FS}^a$$

Table 13: Breaking down the Effect of the Top Ten Percent Rule without Correcting for Inflation

	(1) First Major in STEM	(2) Persisted in STEM (Lower Bound)	(3) Persisted in STEM (Upper Bound)
Diff in attending a Uni $(\Delta_{i\hat{i}}U_i)$	0.2720***	0.1209***	0.1011***
	(0.0117)	(0.0098)	(0.0090)
Diff in attending A&M $\left(\Delta_{i\hat{i}}U_i^{AM}\right)$	0.2635***	0.0829***	0.1260***
	(0.0095)	(0.0080)	(0.0074)
Diff in Uni Peers $\left(\Delta_{i\hat{i}}\bar{A}^u\right)$	-0.0192	0.0204*	-0.0107
	(0.0115)	(0.0097)	(0.0089)
Diff in ability compared to HS Peers $(\Delta_{i\hat{i}}s)$	0.0139	-0.0225*	-0.0343***
	(0.0115)	(0.0097)	(0.0090)
Constant	0.0753***	0.0739***	0.0977***
	(0.0054)	(0.0046)	(0.0042)
Observations	19564	19564	19564

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

and ϵ_i^{hs} must be between $\tau_{nFS} - X_i\beta - s_i$ and $\tau_{FS} - X_i\beta - s_i$. Let E_i^{hs} represent the set of possible ϵ_i^{hs} values after a student observes their admissions decisions.

All students place the same value v_u on attending each university u. If students graduate with a STEM degree they will earn w^S in wages afterword, and if students graduate with a non-STEM degree they will earn wages w^N afterword.

Students who initially major in STEM see their actual ability A_i, and see whether it is worth it to major in STEM based upon the effort cost c_u^S. Specifically, a student will persist in STEM if

$$w^S - c_u^S A_i \geq w^N$$

Students know this when selecting between schools and majors. The value of attending university u and majoring in STEM is

$$V_u^S = v_u - c^e + w^N + \mathbb{E}\left(\left(w^S - c_u^S A_i - w^N\right)\mathbb{1}\left\{w^S - c_u^S A_i - w^N \geq 0\right\}\right)$$

where c^e is an effort cost associated with initially majoring in STEM. Since $A_i = X_i\beta + s_i + \epsilon_i^{hs}$ and the students know $X_i\beta$ and s_i, this expectation is being taken over ϵ_i^{hs}, given that ϵ_i^{hs} must be in E_i^{hs}. Then, the value of attending university u and majoring in something other than STEM is

$$V_u^N = v^u + w^N$$

The non-college option has a value of zero. Let the set of universities that a student is accepted into be U_i. Students choose a university u and a major field f so as to maximize their expected value V_u^f.

Universities pick τ_u^a and c_u^S in order to maximize a function that values average the ability within their STEM program $\bar{A}_u^S$ and total enrollment M_u:

$$\left(\bar{A}_u^S\right)^{\beta_1} + \left(M_u\right)^{\beta_2}$$

where the β terms determine the university's preferences.