Academic Outcomes and Texas Top Ten Percent Law

Eric Furstenberg*
Comments welcomed.

January 30, 2009

Abstract

In this paper, I estimate the causal effect of attending a selective college on a student's academic performance. My results differ from previous studies, because I estimate a local effect, identified only for students who enroll in a selective college but would not have been able to without the guaranteed admissions granted to them by Texas' Top Ten Percent Law. Differing from many previous studies, I find significant negative effects of attending a selective college on first semester GPA, sixth semester GPA, and graduation probability. I interpret these results as evidence that students admitted to Texas' selective institutions via the Top Ten Percent Law, but would not have been admitted without the Law are "mismatched."

Keywords: Higher Education, Public Policy.
JEL Classifications: I23, I28.

*Department of Economics, University of Virginia, ekf4p@virginia.edu. I thank Mark Long, John Pepper, Marta Tienda, and Sarah Turner for helpful comments and suggestions. In addition, Dawn Koffman graciously answered all my questions about the THEOP dataset. All errors are my own.

1 Introduction

Following the 1996 *Hopwood*[1] decision ending affirmative action admissions programs, minority student enrollments fell precipitously at Texas' flagship public universities. As an alternative to affirmative action the Texas legislature passed the so-called Top Ten Percent Law (H.B.588), which guarantees to students graduating in the top decile of their high school class admission to any public college or university. The policy is designed to compensate for the prohibition of affirmative action and increase minority access to the flagship institutions by taking advantage of moderately segregated public high schools and granting admissions to 10 percent of the students in each high school. Even though many top decile high school students would earn admission to Texas' selective institutions without the Top Ten Law, the Law does open access to these institutions for many other students. A natural question to ask is how have these students performed academically while attending an institution more selective than they otherwise would have been able to attend.

The goal of this paper is to estimate the effect of attending a more selective institution on a variety of measures of collegiate academic performance for the students whose admissions outcomes (and enrollment decisions) were affected by the Top Ten Percent Law. To identify this effect, I rely on a two stage least squares estimation of a Local Average Treatment Effect (LATE). The counterfactual evaluated in this paper can be explained as follows. Consider a student who ranks in the top decile of her class and enrolls at a selective school (UT–Austin for example). Further, suppose she would not

[1]US Fifth Circuit Court of Appeals.

have been admitted without the Top Ten Percent Law. The LATE framework estimates the effect of attending UT–Austin on her collegiate academic performance relative to her academic performance if admission to UT–Austin was not available to her via the Top Ten Percent Law.

Many studies have show that Texas Top Ten Percent Law altered application and enrollment patterns of Texas' college bound high school students.[2] The goal of this paper is not to detail the effects of the Top Ten Percent Law on application and enrollment decisions, but it is important to note that the Top Ten Percent Law did have a significant impact on the application and enrollment decisions of Texas' high school students. Raw enrollments at UT-Austin in 1995 are 55, 350, 434, and 1501 for Blacks, Hispanics, Asians and Whites graduating in the top decile of their high school class. In 1999, the second year in which top decile students were eligible for admission under the Law, the corresponding numbers are 161, 513, 610, and 1632. There are similar declines for students graduating in the third decile of their high school class.[3] In addition to in-

[2] Niu and Tienda (2007) show that the Top Ten Percent Law does boost enrollment at Texas' flagship public universities for eligible students at predominantly minority high schools. However, Niu, Sullivan, and Tienda (2008) show that lack of information about the Top Ten Percent Law mitigates the effect of the Law on bringing more minority students to the flagship institutions. Niu, Tienda, and Cortes (2006) study students' preferences over college selectivity levels. Bucks (2002) concludes that the Top Ten Percent Law was unsuccessful at restoring minority enrollment levels at the flagship institutions to Pre-Hopwood levels. Long and Tienda (2008) show that average standardized test scores rose at less selective schools following the implementation of the Top Ten Percent Law, and that at UT–Austin the trend of increasing standardized test scores halted.

[3] Source: Author's calculations using the THEOP Administrative data files. See Table 4 in Section 2.

creasing minority enrollments at UT–Austin, the Top Ten Percent Law also dramatically increased non–minority enrollments. This paper evaluates the academic performance of all of these students.

Others have studied the effects of attending a more selective college. Most authors have focused on labor market returns to selectivity.[4] Others have focused on the effects of attending a selective college on academic outcomes. The goal for much of these papers has been to study effects of affirmative action programs, and specifically to assess the "mismatch hypothesis." In the context of race based admissions preferences, the mismatch hypothesis states that the beneficiaries of the policy are in fact harmed because they are allowed entry to schools where they are underqualified academically. The majority of authors have found minimal, if any, evidence in support of the mismatch hypothesis.[5] In a sense, the counterfactual assessed in this paper provides a test of the mismatch hypothesis with respect to the Top Ten Percent Plan.

Most similar to this paper, Cortes and McFarlin Jr. (2008) use a two pronged strategy to analyze the mismatch hypothesis in the context of the shift between the admissions regimes defined by the affirmative action and the Top Ten Percent Law. First, in a 2SLS model they show that attending a more selective college has significant positive effect on college graduation probability, for both minority and non–minority students. However, this is a global effect, not a local average effect specific to the individuals whose behavior is affected by the policy. Second, they use differences-in-differences to show that the change from affirmative action to Top Ten Percent Law

[4]Brewer, Eide, and Ehrenberg (1999) and Dale and Krueger (2002), for example.

[5]For example, Bowen and Bok (1998), Alon and Tienda (2005), and Rothstein and Yoon (2007).

reduced the graduation probability of second decile students when compared to first decile students, and that this effect is larger for non-minority students than minority students.

I find strong *negative* selectivity effects on first semester and sixth semester GPA for students admitted under the Top Ten Percent Law. Additionally, I find significant negative selectivity effects on completion of the sixth semester and on graduation probability. These results are robust to a number of alternative specifications and are strongest for White and Hispanic students.

The remainder of this paper is organized as follows: Section 2 describes the data, Section 3 describes the empirical specification, Section 4 presents the results and discussion, and Section 5 concludes.

2 Data

The data used in this paper are part of the Texas Higher Education Opportunity Project Administrative Dataset. The THEOP data contain applicant data for more than 500,000 applicants to nine Texas colleges and universities from 1992 to 2005.[6] The main dataset includes a wide variety of applicant information. Among the variables common to all of the reporting institutions are: gender, race and ethnicity, citizenship, Texas residency, SAT and ACT scores, high school class rank, geographic identifiers, and high school identifiers. The dataset also includes student records for all students who enrolled at

[6]See www.texastop10.princeton.edu/admin_overview.html for further information.

each institution. The student records include information about each enrolled student's performance for each semester. These data include semester GPA, cumulative GPA, credit hours completed, total credit hours completed, and choice of major.

Table 1 lists the main variables (and their descriptions) used in this paper.

Table 1: Variables and Descriptions

Variable	Description
testscore_c	Composite test score (constructed)
hspctrank_c	HS class rank, percentile (constructed)
topten	1 = top decile of HS class rank & year >= 1998
rank_l	= *topten* * (*hspctrank_c*)
rank_r	= (1 − *topten*) * (*hspctrank_c*)
gender	1 = Male
year	Time trend
black	1 = Black
hispanic	1 = Hispanic
nat_am	1 = Native American
asian	1 = Asian
other	Base category (does not include White)
sel	median of *testscore_c* of enrolled school

Notes: "_c" indicates variables constructed by THEOP staff.

testscore_c is constructed as a composite test score, based on both SAT and ACT scores, depending on which test each student takes. This allows me to use a single variable to represent each student's standardized test performance, and avoids issues of non-random selection into the groups of students taking the SAT or ACT. The composite variable is based on recentered SAT scores.[7]

[7]Because the SAT was recentered in 1996, the SAT scores before and after 1996 are not directly comparable. Comparison tables exist for SAT and ACT scores after 1996, but the pre-1996 SAT scores would need to be recentered in order to be compared to the ACT scores. In order to included data from these years in the analysis, I estimated a recentered SAT score for students taking the SAT in the years prior to 1996. The details of this procedure are discussed in Appendix A. Because the LATE method does not rely on before/after analysis, inclusion of these observations from these years is not critical to the empirical identification strategy. All of the results discussed are robust to exclusion of

Because the goal of this paper is to study the performance of the students affected by the Top Ten Percent Law, it is necessary to identify high school class rank. In some cases, the institutions report class rank as a percentile, and in other cases, it must be calculated as the ratio of numeric class rank to class size. The class rank information is contained in the *hspctrank_c* variable. Lower values correspond to higher class rank – for example, the valedictorian has $hspctrank\_c = 1$, and the lowest ranked student has $hspctrank\_c = 100$.[8]

Three variables are derived from *hspctrank_c*, and ultimately *hspctrank_c* is not used in the analysis. *topten* is an indicator variable, equal to one if a student graduated in the top decile of their high school class and the Top Ten Law was in place (years 1998 and forward). *rank_l* is equal to an interaction of *topten* and *hspctrank_c*, and *rank_r* is equal to an interaction of $1 - topten$ and *hspctrank_c*. The terms *rank_l* and *rank_r* allow the marginal effect of high school rank on college GPA to have different slopes on either side of the discontinuity.

Identifying the admissions selectivity of the colleges and universities is important for this paper. My conclusions are based on four alternative definitions of selectivity. First, I use the median value of *testscore_c* at the institution attended by each individual

these observations.

[8]Note that before 1997 and for certain high schools, UT-Austin reports class rank as 8 for all top decile students and 18 for all second decile students, etc. This is not a major issue for this paper because it only involves data from a few high schools, from the pre-1997 time period. All of my analysis is based on data from 1996 and forward, so therefore there are few observations that are subject to this discretization.

as a continuous measure of selectivity. Higher values of *sel* denote a more selective school. Second, I use a binary variable to indicate whether or not a student attends a selective post-secondary school. UT–Austin or Texas A&M–College Station (TAMU) are the two schools defined as selective. The non-selective schools are Texas Tech, A & M – Kingsville, and UT–San Antonio. Third, I use a binary variable for selectivity, for which only UT–Austin is defined as selective. Fourth, I use a categorical variable with 5 different selectivity categories, as defined by Barron's. With a few minor exceptions, all of my results are robust to these specification changes. Therefore I report only the results using the first selectivity variable described above.[9]

Table 2 reports the values of *sel* for the 5 schools included in the analysis:

Table 2: Variables and Descriptions

School	*sel* value
UT – Austin	1210
TAMU	1160
TT	1090
UT – SA	870
AM – Kingsville	950

Notes: Based on author's calculations from THEOP Administrative Dataset.

2.1 Descriptive Analysis

Table 3 shows summary statistics for selected variables:

Table 3: Summary Statistics

Variable	Mean	Std. Error
cgpa	2.87	.88
hspctrank_c	17.31	16.74
testscore_c	1153.87	162.15

continued on next page

[9]The complete results are available upon request.

Table 3: Summary Statistics

Variable	Mean	Std. Error
gender	.48	.49
black	.032	.17
hispanic	.14	.35
asian	.089	.28

Notes: Based on author's calculations from THEOP Administrative Dataset. Calculations are based on data from enrolled students at the 5 institutions included in the analysis (listed above), for the years 1990 – 2003. $n = 188,835$.

All of the sample means fall well within the range of acceptable values.

Table 4 shows the enrollments at UT–Austin for 1995 – 2003, by race and decile of high school class rank. Shortly following the implementation of the Top Ten Percent Law, UT–Austin enrollments of students from the top decile of their high school class increased dramatically, while the enrollments of students from outside the top decile of their high school class shrank. Additionally, Table 4 shows that in addition to increasing minority enrollments at UT–Austin, the Top Ten Percent Law also dramatically increased non–minority enrollments.

Table 4: UT-Austin Enrollments by Race and Decile of HS Class Rank: 1995 – 2003

	Fist Decile				Second Decile				Third & Lower Deciles			
	Black	Hispanic	Asian	White	Black	Hispanic	Asian	White	Black	Hispanic	Asian	White
1995	62	360	476	1556	41	208	193	916	103	272	179	1084
1996	55	350	434	1501	45	216	222	994	95	284	247	1177
1997	48	338	501	1406	51	212	246	1006	79	259	292	1467
1998	60	392	509	1503	47	214	258	956	78	222	276	1261
1999	161	513	610	1632	40	214	249	978	80	237	301	1258
2000	156	586	655	1925	60	220	282	1097	89	261	351	1394
2001	139	570	720	1947	51	227	294	949	43	149	237	781
2002	156	693	804	2203	54	231	287	1086	50	144	205	787

Notes: Based on author's calculations from THEOP Administrative Dataset.

It is important to note the relationship between high school class rank and the selectivity of the college attended. Similar to Niu and Tienda (2007), Figure 1, based on data from 1998 forward, plots the average value of *sel* as a function of high school class rank.

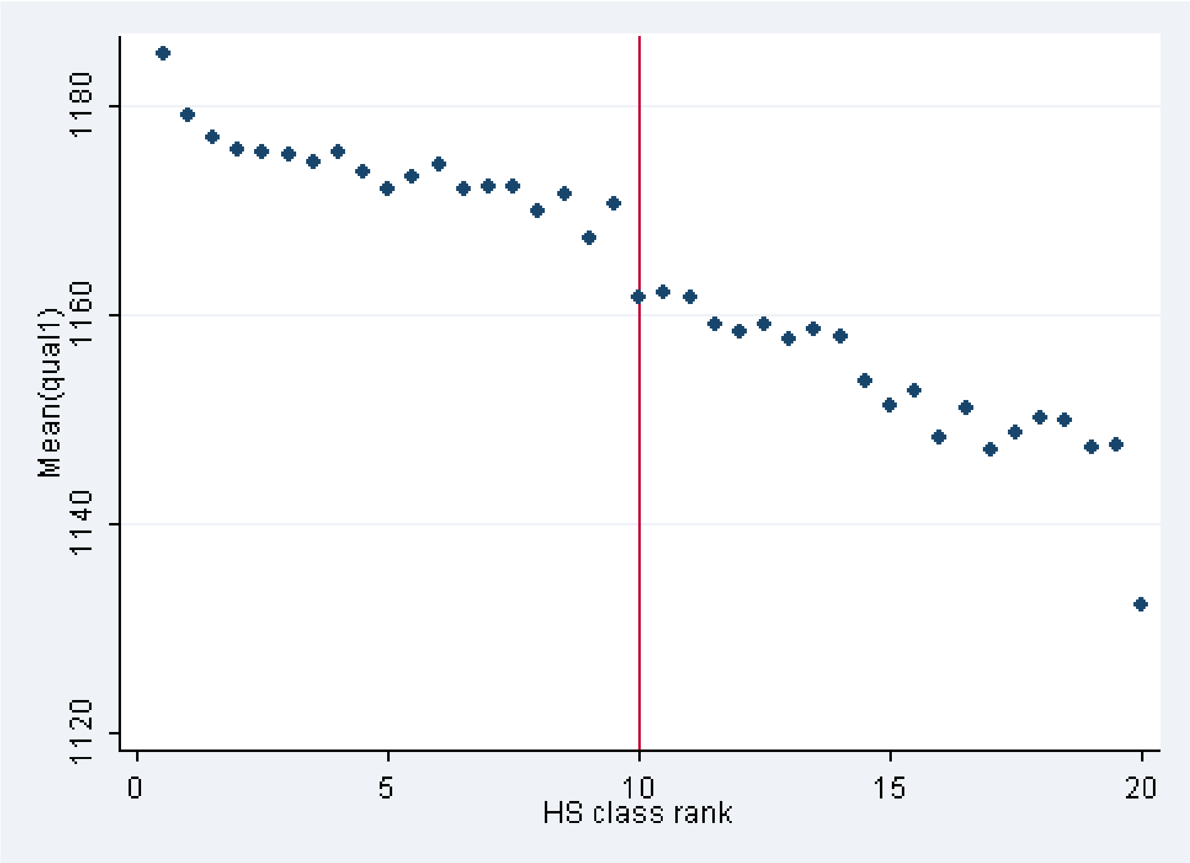

Figure 1: Average Selectivity and High School Class Rank

The effect of the Top Ten Percent Law is clear, as evidenced by the sharp discontinuity at class rank equal to 10. Students in the top decile of their graduating class are significantly more likely to attend a selective higher education institution.[10]

[10]Alternative definitions of selectivity produce a similar pattern. Additionally, data from the years

3 Empirical Methodology

The basic empirical model used is a simple linear model of a student's academic outcome as a function of the selectivity of the college attended and a set of control variables:

$$Y_i = \beta_0 + \beta_{1,i} sel_i + \gamma Z_i + \epsilon_i. \tag{1}$$

The coefficient of interest is $\beta_{1,i}$, the effect of attending a more selective institution on Y_i, the student's academic outcome.[11] The matrix X_i is a matrix of control variables containing test score, high school class rank, gender, race dummy variables, and a time trend. ϵ_i is a mean zero error term. Therefore, if larger values of Y_i indicate better academic performance, then positive values of $\beta_{1,i}$ indicate positive effects of attending a more selective school on Y_i – attending a more selective school improves academic performance.

When the treatment variable is uncorrelated with unobservable variables which affect the outcome, then equation (1) can be estimated with OLS and $\beta_{1,i}$ is identified as the Average Treatment Effect or ATE.[12] In this case, the treatment is attending a more selective post-secondary institution. Clearly, the ATE estimates are biased if students select into the treatment group based on unobservable characteristics which are related to academic performance.[13] Endogeneity problems with the OLS estimation immediately suggest use of Two Stage Least Squares estimation of the selectivity effects.

prior to the Top Ten Law show no such discontinuity.

[11] I index $\beta_{1,i}$ by i to emphasize that the parameter is allowed to vary.

[12] Note that the discussion of average treatment effects assumes the treatment effect is homogeneous.

[13] See Dale and Krueger (2002) for a discussion of selection on unobservables.

Before proceeding, additional discussion is necessary to clarify the interpretation of the parameters in the 2SLS models.

When the response to the treatment varies within the treated population, then different instruments identify the effect of the treatment for different subsets of the population. If the treatment effect is homogeneous throughout the population, then the choice of instruments does not matter for the estimation – any valid instrument will give a consistent estimate of the average treatment effect. If this were the case, then LATE and ATE are equivalent. Imbens and Angrist (1994) show that when the treatment effect is heterogeneous, then the choice of instruments in the 2SLS model is critical for the interpretation of the estimates. Then, the slope parameter is interpreted as the treatment effect for the group of eligible individuals who accept the treatment, where the instrument defines the subpopulation who is eligible for the treatment. This is the Local Average Treatment Effect (LATE).

The goal of this paper is to document the effects of the Top Ten Percent Law on Texas students' academic outcomes. In particular the population of interest is the group of students who attend more selective schools than they would have in the absence of the policy in question. The slope parameter identified in the 2SLS model is the LATE parameter, and is defined as the the effect of the treatment on the group of students whose behavior is affected by a change of the instrument. The instrument I use is the indicator variable *topten*, which is equal to one if the individual graduated in the top decile of their high school class, and zero otherwise. Thus, in this context, the slope parameter in the 2SLS model is the effect of attending a more selective school on the outcome of interest for the individuals who attend a more selective school than they would have in the absence of the Top Ten Percent Law.

In the 2SLS model, identification of the treatment effect ($\beta_{1,i}$) is generated from this discontinuity in the relationship between attending a selective institution and high school class rank. The discontinuity in Figure 1 can also be estimated by the first stage regression in the 2SLS model (See Table 5). The effect of the Top Ten Percent Law is clear in Figure 1 and this effect is statistically significant in the first stage regression. This confirms the validity of using *topten* as an instrument.[14]

Other authors have used similar methods. Cortes and McFarlin Jr. (2008) use a 2SLS framework to study the effect of attending a selective college on graduation probability. Following Card (1995) their solution to the endogenous selection problem is to use distance to the nearest selective college as the instrument for selectivity. LATE effects are identified as the effects brought about by a change of the instrument. Consequently, the results of Cortes and McFarlin Jr. (2008) are best interpreted as the effects of attending a selective college on graduation probability, but not as effects of the Top Ten Plan. That is, using distance to the nearest college identifies the effect of selectivity for

[14]It is worth mentioning that the 2SLS estimates are equivalent to the estimates derived from a Fuzzy Regression Design framework (FRD). For example, see Imbens and Lemieux (2008), Angrist and Lavy (1999), Lee and Card (2008), Card and Shore-Sheppard (2004), Hahn, Todd, and Van Der Klaauw (1999), or Hahn, Todd, and Van Der Klaauw (2001). The FRD design contrasts with the Sharp Regression Discontinuity design (SRD) used by Niu and Tienda (2007). In the FRD design not all eligible individuals accept the treatment. In fact, the first stage regression in the 2SLS is essentially the same as the SRD model used by Niu and Tienda (2007).

the whole population (as everyone is subject to a change of the distance between their home and the nearest selective college), not the effect of selectivity on the individuals whose behavior is affected by the Top Ten Percent Law.

4 Results

This section presents the empirical estimates of the models discussed in Section 3. The first set of results uses first semester grade point average (GPA) on a 4 point scale as the dependent variable. Throughout the analysis, I limit the sample to include students who graduated in the first or second decile of their high school class. The LATE methodology relies on comparing the academic outcomes of a student in the top decile of her high school class and an otherwise equivalent student outside the top decile, so therefore students outside the second decile are not relevant to the analysis.[15]

Table 5 displays the estimates of four regressions. The first column shows a baseline OLS model. The second shows the LATE estimates. The third and fourth columns show the estimated first stage and reduced form models corresponding to the 2SLS interpretation of the LATE design. The reference racial group is Whites.

Table 5: Effect of Selectivity on First Semester GPA

	OLS	LATE	First Stage	Reduced Form
testscore_c	0.002*	0.002*	0.089*	0.002*
	(0.000)	(0.000)	(0.001)	(0.000)
rank_l	-0.042*	-0.041*	0.185**	-0.042*
	(0.001)	(0.001)	(0.090)	(0.001)
rank_r	-0.031*	-0.031*	-0.084**	-0.031*

continued on next page

[15]See Imbens and Lemieux (2008) for a discussion of "bandwidth selection." The results are robust to limiting the sample to the top three deciles of high school students.

Table 5: Effect of Selectivity on First Semester GPA

	OLS	LATE	First Stage	Reduced Form
	(0.001)	(0.001)	(0.035)	(0.001)
gender	-0.147*	-0.164*	-2.708*	-0.146*
	(0.004)	(0.005)	(0.315)	(0.004)
year	0.025*	0.007*	-2.945*	0.026*
	(0.001)	(0.002)	(0.059)	(0.001)
black	-0.169*	-0.057*	18.620*	-0.177*
	(0.013)	(0.019)	(0.835)	(0.013)
hispanic	-0.124*	-0.161*	-6.023*	-0.122*
	(0.007)	(0.009)	(0.690)	(0.007)
nat_am	-0.189*	-0.156*	5.456**	-0.191*
	(0.038)	(0.041)	(2.313)	(0.038)
asian	0.085*	0.241*	25.763*	0.074*
	(0.007)	(0.018)	(0.310)	(0.007)
other	0.088*	0.119*	5.136*	0.086*
	(0.025)	(0.027)	(1.394)	(0.025)
sel	**-0.0004***	**-0.006***		
	(0.000)	(0.001)		
topten	-0.088*		14.576*	-0.094*
	(0.008)		(0.671)	(0.008)
Constant	-48.536*	-6.591***	6,943.349*	-51.478**
	(1.608)	(3.940)	(117.042)	(1.592)
n	121411	121411	121562	121411

Notes: Based on author's calculations from THEOP Administrative Dataset. *** significant at 10%; ** significant at 5%; * significant at 1%. Standard errors in parentheses. The first stage model uses *topten* as the dependent variable, and the reduced form uses GPA as the dependent variable.

Table 5 shows negative and statistically significant effect of attending a selective college. Consider a 100 point increase in the median SAT of the school attended.[16] According to the OLS model, one's expected first semester GPA would decrease by 0.04 GPA points. This corresponds to a 1/3 of a letter grade decrease in one of eight courses, or a 1/6 letter grade decrease in one of four courses (a typical course load in a semester). According to the LATE design, such a change in selectivity would lead to a decrease of first semester GPA of 0.6 GPA points. This is approximately 2/3 of a letter grade change of the first semester GPA. Note that the OLS estimates should be interpreted

[16]Based on the median values of *testscore_c* shown in Table 2, this is a reasonable difference between a selective and unselective school.

as a population level treatment effect, and the estimates from the LATE design should be interpreted as the causal effect of attending a more selective school on GPA for the students with class rank in a small neighborhood of 10.[17] All of the other coefficients have the predicted sign.

Based on the first stage model, the estimated discontinuity in the relationship between *sel* and class rank at class rank equal to 10 is statistically significant and equal to 14.57. This estimate correspond the discontinuity shown in Figures 1. Similarly, the reduced form shows a statistically significant discontinuity in the relationship between GPA and class rank, equal to -0.094.

4.1 LATE Estimates by Year and by Race

Next, I repeated the LATE regressions for each year covered by the THEOP data. As suggested by Table 4 the main "action" of the Top Ten Percent Law appears to take place after 1998. Table 6 shows the LATE estimates by year for the years 1996–2002.[18]

[17] 2SLS estimates are often larger (in magnitude) than OLS. As discussed by Card (2001), one explanation for why this is true is that OLS estimates report the average effect for the whole population to whom the treatment is available, but the 2SLS estimates report the average effect for only the individuals whose behavior is affected by the treatment (thus the instrument is an indicator for top decile of high school class rank). When not all of the individuals to whom the treatment is available elect to take the treatment, then the OLS estimates are in a sense 'diluted'. In this case, only a certain percentage of the population to whom the treatment is available take the treatment, but because OLS measures the effect of the treatment on the entire population, the estimated effect is attenuated in the OLS estimates.

[18] The full regression results are available upon request.

Table 6: LATE Selectivity Effects by Year

	1996	1997	1998	1999	2000	2001	2002
sel	0.003	0.013	0.022	0.001	-0.004	-0.013**	-0.006**
	(0.011)	(0.023)	(0.020)	(0.020)	(0.006)	(0.006)	(0.003)
n	9513	9629	9972	10179	11143	11249	12029

Notes: Based on author's calculations from THEOP Administrative Dataset. *** significant at 10%; ** significant at 5%; * significant at 1%. Standard errors in parentheses.

The LATE estimates by year follow a predictable pattern. As shown in Table 6, the LATE selectivity effects do not surface in the data until the Fall semester of 2001. The estimates from the years 1996 – 2000 are not statistically significant. As discussed in Section 1 and illustrated in Table 4, the enrollment of top ten ranked students at the selective schools grew consistently in the years following 1998, the first year the Top Ten Law was in place. The results above suggest that in each successive year, the additional students taking advantage of the Top Ten Law performed worse academically. Perhaps these students are the ones who, despite their top decile rankings, are the least prepared for the challenging academic environment of a selective institution.

To check if the LATE effects are consistent across races, I ran the regressions separately for the four major racial groups (White, Black, Hispanic, Asian). Table 7 shows the results.[19]

Table 7: LATE Selectivity Effects by Race

	Black	Hispanic	Asian	White
sel	-0.002	-0.003*	-0.007	-0.009*
	(0.002)	(0.001)	(0.005)	(0.001)
n	3722	16519	12298	87607

Notes: Based on author's calculations from THEOP Administrative Dataset. *** significant at 10%; ** significant at 5%; * significant at 1%. Standard errors in parentheses.

[19]The full regression results are available upon request.

Only the coefficients for Hispanics and Whites are significant. The estimates are negative and large in magnitude for both groups. The estimates show that Hispanics' first semester GPA decreases by about 1/3 of a letter grade and Whites' first semester GPA decreases by almost a full letter grade for every 100 point increase of selectivity. The estimates for Blacks and Asians are not statistically significant. For both groups, the estimates are negative and large, but imprecisely estimated. The likely explanations for the lack of precision are small sample size for Blacks and the small number of Asians attending schools in the lower selectivity categories.

4.2 Selectivity effects on Sixth Semester GPA and Graduation Probability

It is possible that the selectivity effects on first semester GPA are temporary and that the lasting and more appropriate measures of academic performance occur later in a student's academic career. Therefore, I repeated the LATE design using 6th semester GPA and graduation as the dependent variables. Table 8 displays the relevant estimates from the models using sixth semester GPA, completion of the 6th semester and graduation (completion of 100 credit hours) as the dependent variables.[20]

Table 8: Effect of Selectivity on Sixth Semester GPA and Graduation

	LATE	LATE Probit Model	
	Dep. Var. = Sixth Sem. GPA	Dep. Var. = Completion of 6th Sem.	Dep. Var. = Graduation
sel	-0.018*	-0.006**	-0.013*
	(0.001)	(0.002)	(0.004)
n	920856	85206	70006

[20]The full results are available upon request.

Notes: Based on author's calculations from THEOP Administrative Dataset. *** significant at 10%; ** significant at 5%; * significant at 1%. Standard errors in parentheses. The estimates reported for the Probit models are not marginal effects.

As a word of caution, note that the number of students eligible for admission via the Top Ten Law is very small for these regressions. Students who are eligible for admission under the Top Ten Percent Law began enrolling in the fall of 1998. At the earliest, these students would have completed their sixth semester in the spring of 2001 and would have graduated in the spring of 2002. For most institutions, the THEOP data do not contain complete administrative records for years later than 2002. Therefore, the estimates for the sixth semester models are based on two years of Top Ten eligible students, and the graduation models are based on only one year.

The first column of Table 8 shows the LATE estimates of the effect of attending a selective college on sixth semester GPA. The second and third columns show the LATE probit estimates of the selectivity effect on completion of the sixth semester and graduation. All of the estimates are negative and statistically significant. In fact, the effect of attending a more selective institution (as measured by a 100 point increase of $testscore\_c$) is a 1.8 point decrease of sixth semester GPA. These results show the potential for significant and long lasting effects for underqualified students who attend a selective institution.

4.3 Robustness Checks

As described in Section 2, I estimated all of the regressions described above using three alternative definitions of selectivity. In addition, I tested the following alternative specifications. First, because GPA is not directly comparable across universities, I estimated the models using rank of college GPA as the dependent variable. Second, I estimated the models using high school fixed effects, and choice of major fixed effects. Third, I estimated a number of alternative models using quadratic and cubic terms. Fourth, in the graduation equations, I used completion of 120 credit hours instead of completion of 100 credit hours as the dependent variable.

Few of the results are substantively different from those presented above. The most notable difference is that some of the specifications using the alternative definitions of selectivity fail to produce statistically significant results for the sixth semester GPA and graduation regressions. The complete results of the alternative specifications are available upon request.

5 Conclusions

In this paper, I use a two stage least squares to estimate the local average treatment effect (LATE) of attending a selective college. The population of interest is the group of students who enroll in a selective college, but would not have been able to without the guaranteed admissions granted to them by Texas' Top Ten Percent Law. Thus, I estimate the causal effect of attending a more selective college, but the results only apply to a subset of the college going population.

I find significant negative effects of attending a selective college on first and sixth semester GPA and graduation probability. It should not be surprising that there are negative effects. Constraining universities' admissions decisions in the manner dictated by the Top Ten Percent Law must have the results of selecting less qualified students. The results are strongest starting a few years after the implementation of the Law, and coincide with the large increases in the number of top decile students attending Texas' most selective institutions. The results are the strongest for Whites and Hispanics. Thus, the results do show that many top decile students are "mismatched" at selective colleges, and that the mismatch is not based on race.

My results are quite important given the scale of the Top Ten Percent Law. At UT–Austin roughly 40 percent of incoming students graduated in the top decile of their high school class in 1997, the last year before the Top Ten Law was in effect. In 2002, almost 75 percent of UT–Austin students graduated in the top decile of their high school class. While I do not consider the benefit of attending a selective institution such as UT–Austin, and I do not conduct a full cost benefit analysis of the Law, my results suggest that the Top Ten Law is doing a poor job of selecting the most qualified students for admission to the selective institutions. To the extent that administrators at selective institutions want to maintain their academic standards, policymakers should reconsider policies such as Top Ten Percent Law. Admissions policies without guarantees and admissions decisions based on individual evaluation of the applicants' qualifications are likely to avoid this problem. Alternatively, policymakers might consider a guaranteed admissions law with higher standards: for example, a policy under which only the top five percent of students from each high school are guaranteed admission might avoid much of the mismatch produced by the Top Ten Law.

References

Alon, S., and M. Tienda (2005): “Assessing the ”Mismatch” Hypothesis: Differences in College Graduation Rates by Institutional Selectivity,” *Sociology of Education*, 78, 294–315.

Angrist, J., and V. Lavy (1999): “Using Maimonides’ Rule to Estimate the Effect of Class size on Student Achievement,” *Quarterly Journal of Economics*, 114(2), 533–575.

Bowen, W. G., and D. Bok (1998): *The Shape of the River: Long Term Consequences of Considering Race in College and University Admissions*. Princeton University Press, Princeton, NJ.

Brewer, D., E. Eide, and R. Ehrenberg (1999): “Does it Pay to Attend an Elite College? Cross Cohort Evidence on the Effects of College Type on Earnings,” 34(1), 104–123.

Bucks, B. (2002): “The Effects of Texas’ Top Ten Percent Plan on College Choice,” *Working Paper*.

Card, D. (1995): “Using Geographic Variation in College Proximity to Estimate the Return to Schooling,” in *Aspects of Labour Market Behaviour: Essays in Honor of John Vanderkamp*, ed. by L. Christofides, E. Grant, and R. Swindisky.

——— (2001): “Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems,” *Econometrica*, 96(5), 1127–1160.

Card, D., and L. Shore-Sheppard (2004): “Using Discontinuous Eilgibility Rules to Identify the Effects of the Federal Medicaid Expansions on Low Income Children,” 86.

Cortes, K. E., and I. McFarlin Jr. (2008): “College Quality and the Texas Top Ten% Plan: Implications for Minortiy Students,” *THEOP Working Paper*.

Dale, S. B., and A. B. Krueger (2002): “Estimating the Payoff to Attending a More Selective College: An Application of Selection on Observables and Unobservables,” *Quarterly Journal of Economics*, 117(4), 1491–1527.

Hahn, J., P. Todd, and W. Van Der Klaauw (1999): “Evaluating the Effect of an Antidiscrimination Law Using Rergession-Discontinuity Design,” Working Paper 7131, National Bureau of Economic Research.

——— (2001): “Identification and Estimation of Treatment Effects with a Regression Discontinuity Design,” *Econometrica*, 69(1), 201–209.

Imbens, G., and T. Lemieux (2008): “Regression Discontinuity Designs: A Guide to Practice,” *Journal of Econometrics*, 142(2), 615–635.

Imbens, G. W., and J. D. Angrist (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62(2), 467–476.

Lee, D. S., and D. Card (2008): "Regression Discontinuity Inference with Specification Error," *Journal of Econometrics*, 142(2), 655–674.

Long, M. C., and M. Tienda (2008): "Changes in Texas Universities' Applicant Pools after the Hopwood Decision," *THEOP Working Paper*.

Niu, S., M. Tienda, and K. Cortes (2006): "College Selectivity and the Texas Top 10% Law," *Economics of Education Review*, 25, 259–272.

Niu, S. X., T. Sullivan, and M. Tienda (2008): "Minority talent Loss and the Texas Top 10% Law," *THEOP Working Paper*.

Niu, S. X., and M. Tienda (2007): "The Impact of the Texas Top 10% Law on College Enrollment: A Regression Discontinuity Approach," *THEOP Working Paper*.

Rothstein, J., and A. Yoon (2007): "Mismatch in Law School," *Working Paper*.

A Data Appendix

The SAT recentering procedure is as follows. Note that UT-Austin and TAMU data contain SATQ and SATV scores. Based on conversion tables available from The College Board, THEOP staff were able to convert pre-1996 SAT scores to the same scale as the post-9916 recentered SAT scores. Data from other institutions do not contain SATQ and SATV scores, so the conversion described above was not possible. As an alternative, I performed the following procedure:[21]

1. Using Data UT-Austin and TAMU for the pre-1996 period, the recentered SATQ and SATV scores contained in the THEOP data were converted back to the original un-recentered scale.

2. Both the un-recentered and recentered SATQ and SATV scores were summed to derive a un-recentered and a recentered total SAT score.

3. Regress the recentered total SAT score on a fourth order polynomial of the un-recentered SAT score.

4. Predict the recentered total SAT score based on the un-recentered SAT scores for the institutions and years for which the recentered scores were not previously available.

[21] I thank Mark Long for suggesting this procedure.