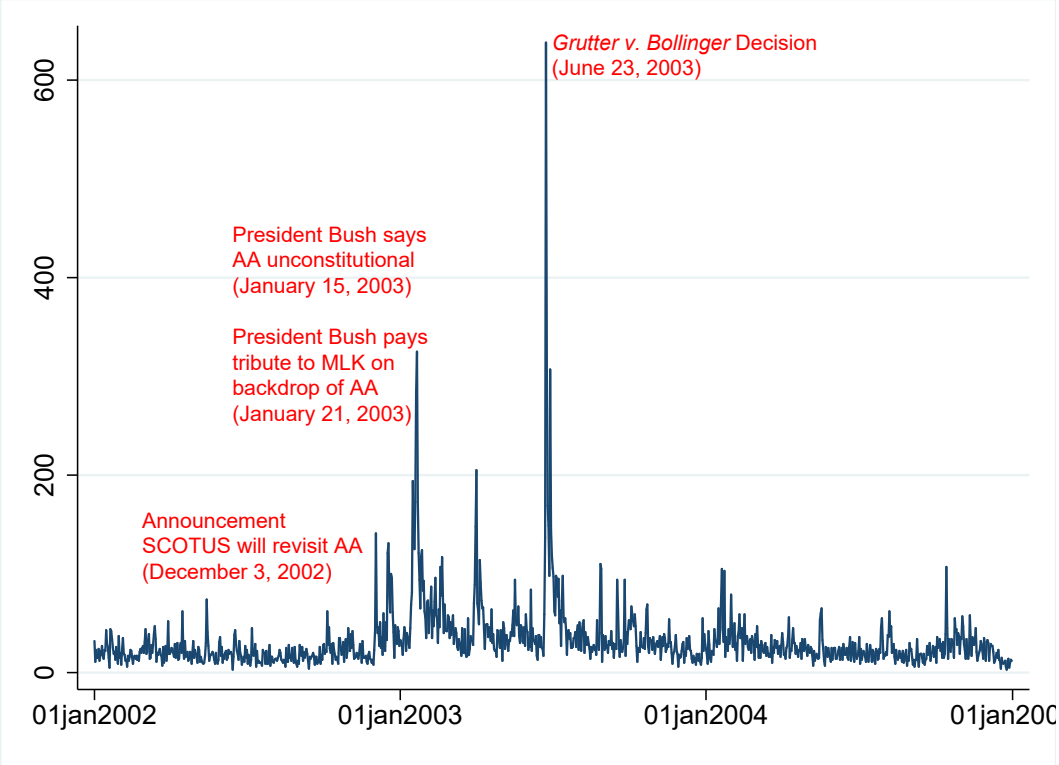
Online Appendix for "Affirmative Action and Pre-College Human Capital"

by

Mitra Akhtari, Natalie Bau, and Jean-William Laliberté

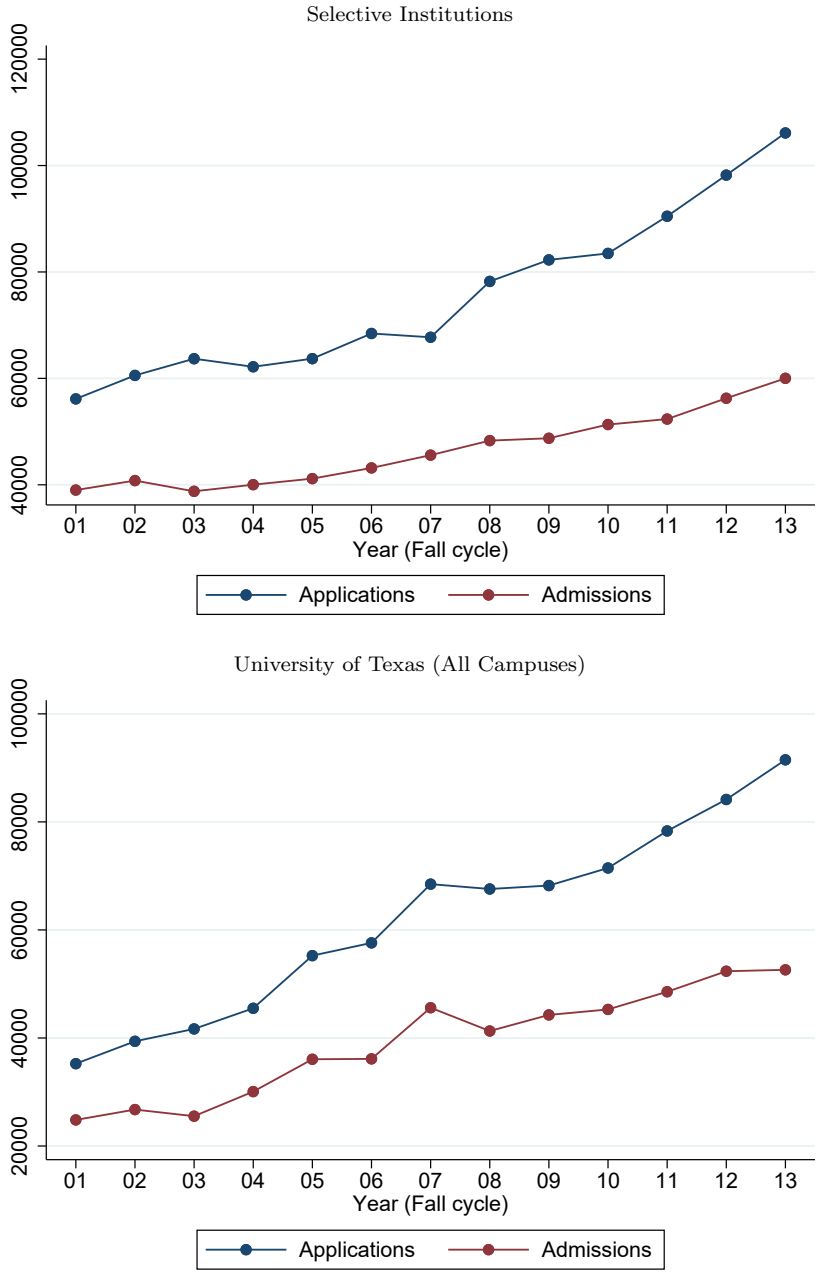
Appendix Figures

Figure A1. Number of Articles Mentioning Affirmative Action by Day, 2002-2004



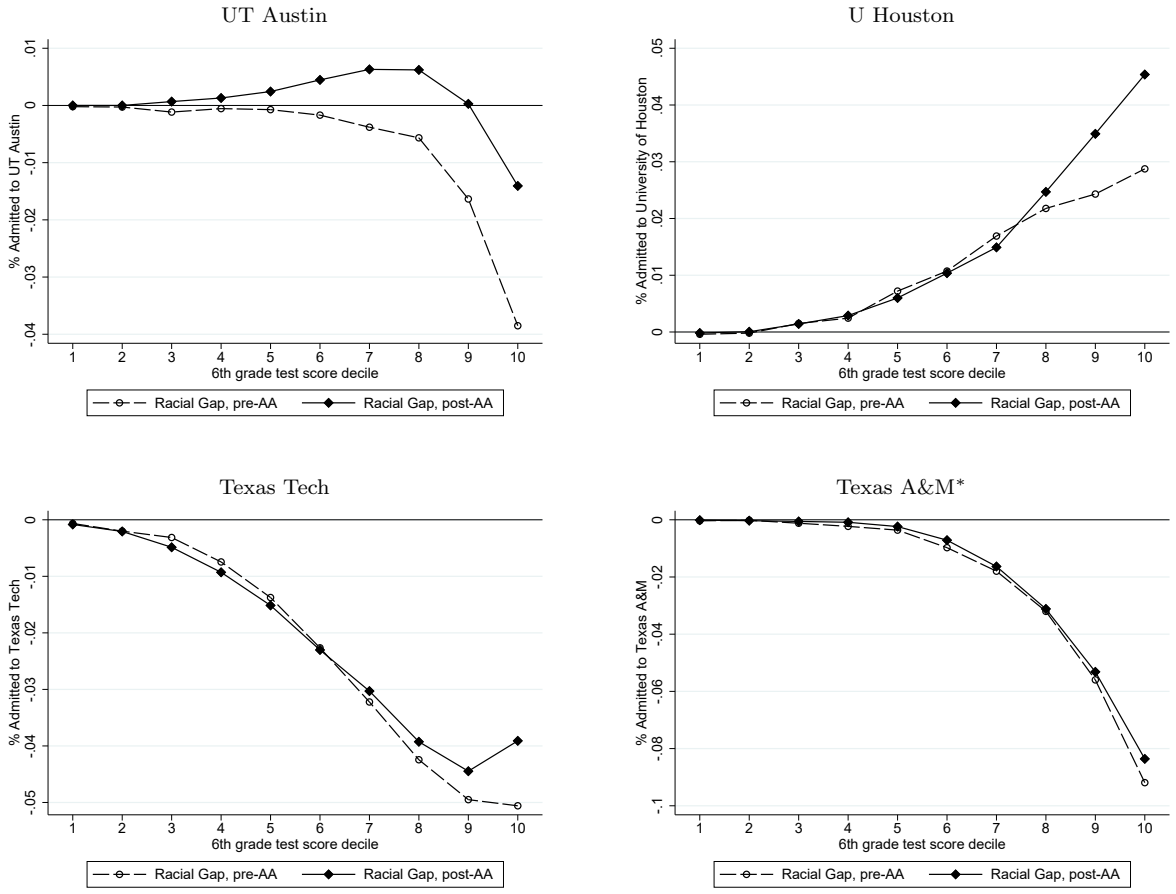
Note: This figure reports the number of US newspaper articles by day that contained the phrase "affirmative action" on newslibrary.com.

Figure A2. Admissions to Texas Institutions Over Time (IPEDS)



Note: This figure reports the number of applicants and admissions to selective institutions (UT Austin, U Houston, Texas A&M, and Texas Tech) and campuses of the University of Texas (UT Arlington, UT Austin, UT Dallas, UT El Paso, UT Permian Basin, UT Rio Grande, UT San Antonio, and UT Tyler). The data are from IPEDS and are available from 2001 onward.

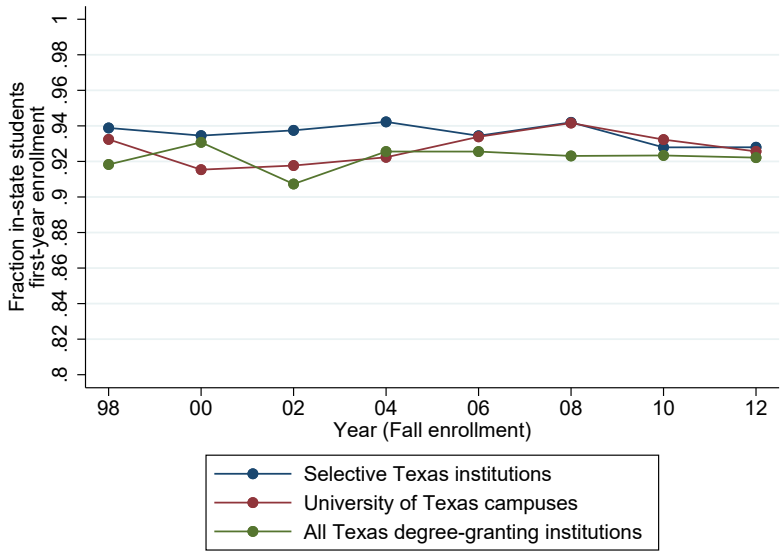
Figure A3. Racial Gaps in Admissions (URM - Whites) to Selective Texas Universities by Institution and Test Score Decile



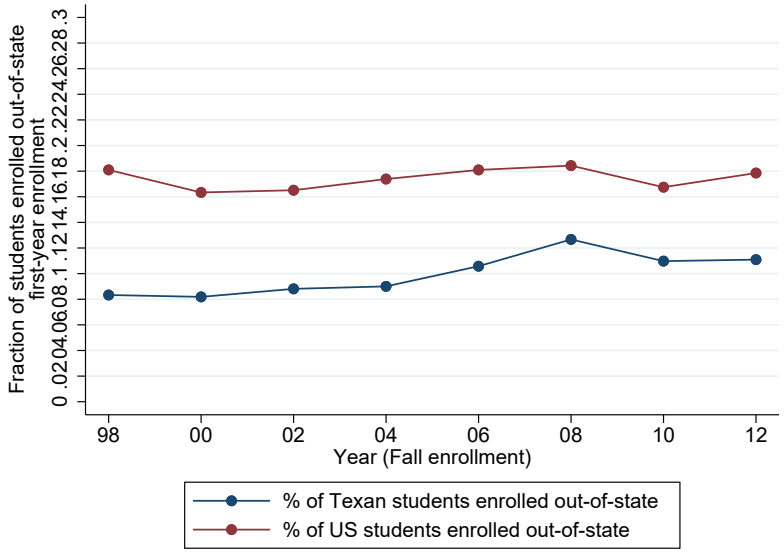
Note: This figure presents racial gaps (URMs relative to whites) in admission probabilities for each selective Texas university (UT Austin, U Houston, Texas Tech, and Texas A&M) by decile of the statewide distribution of 6th grade test scores. Statistics are shown separately for the pre-AA period (1997-2000 cohorts of 9th graders) and the post-AA period (2001-2010 cohorts of 9th graders). *Texas A&M publicly announced that it would *not* use race in admissions (Parker, 2018).

Figure A4. Migration, Residence, and First-Year Enrollment Over Time (IPEDS)

Panel A: In-State Enrollment at Texas Institutions

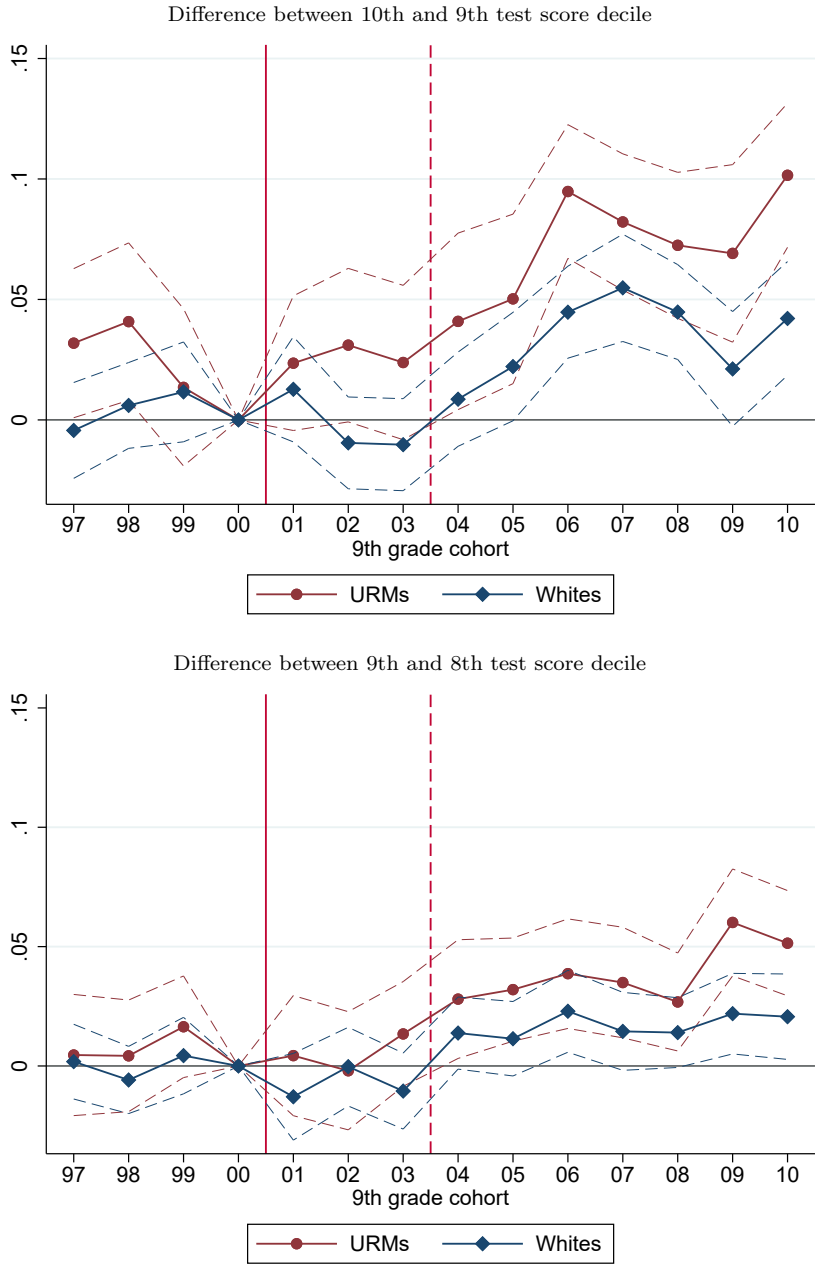


Panel B: Out-of-State Enrollment Among Texas and All US Students



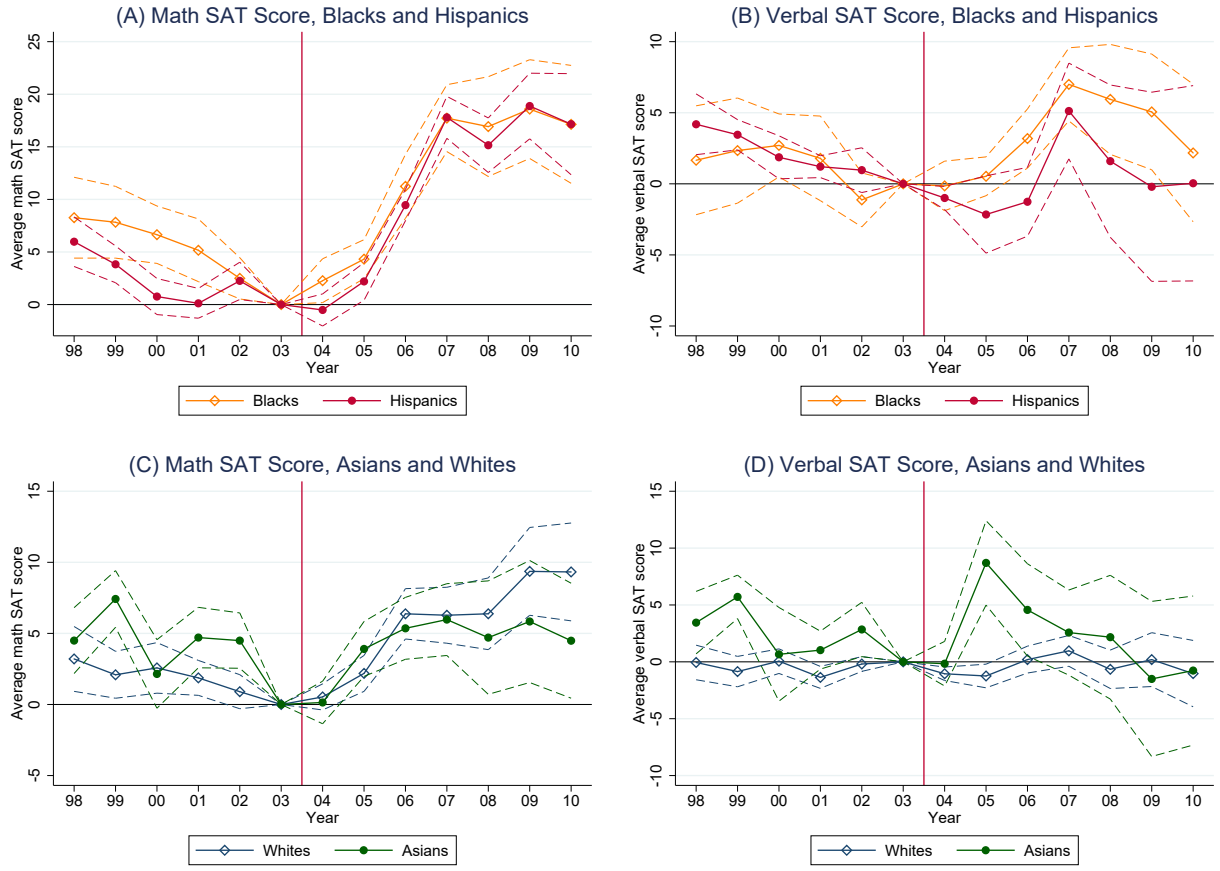
Note: Panel A reports the fraction of first-year students enrolled at different sets of Texas institutions that resided in Texas when they were admitted. Selective institutions include UT Austin, U Houston, Texas A&M, and Texas Tech. Campuses of the University of Texas include UT Arlington, UT Austin, UT Dallas, UT El Paso, UT Permian Basin, UT Rio Grande, UT San Antonio, and UT Tyler. Panel B reports the fraction of first-year students that either resided in Texas (blue line) or in any US state (red line) upon admission that are enrolled in out-of-state institutions. The data are from IPEDS, only include degree-granting institutions, and are only available every 2 years.

Figure A5. Evolution of Differences in Admissions at Selective Institutions



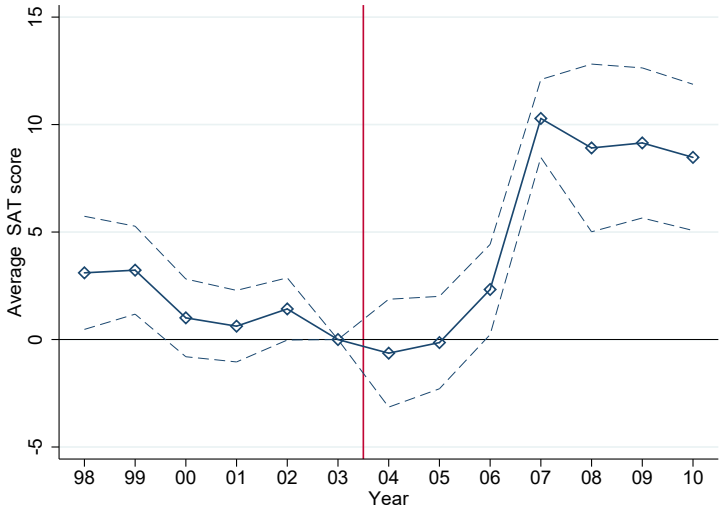
Note: This figure shows the evolution over time of differences in number of admissions at selective institutions between students at the 10th and 9th deciles of the 6th grade test score distribution (top panel) and between students at the 9th and 8th deciles (bottom panel), separately by race. Differences are normalized relative to that of the 2000 cohort of 9th graders. The solid vertical red line indicates cohorts that were already in high school at the time of the policy change, and the dashed vertical red line indicates cohorts that started high school after the policy change. The statistics are based on TEA data. Dashed lines report 95% confidence intervals with standard errors clustered at the district level.

Figure A6. Effect of AA on SAT Scores by Race



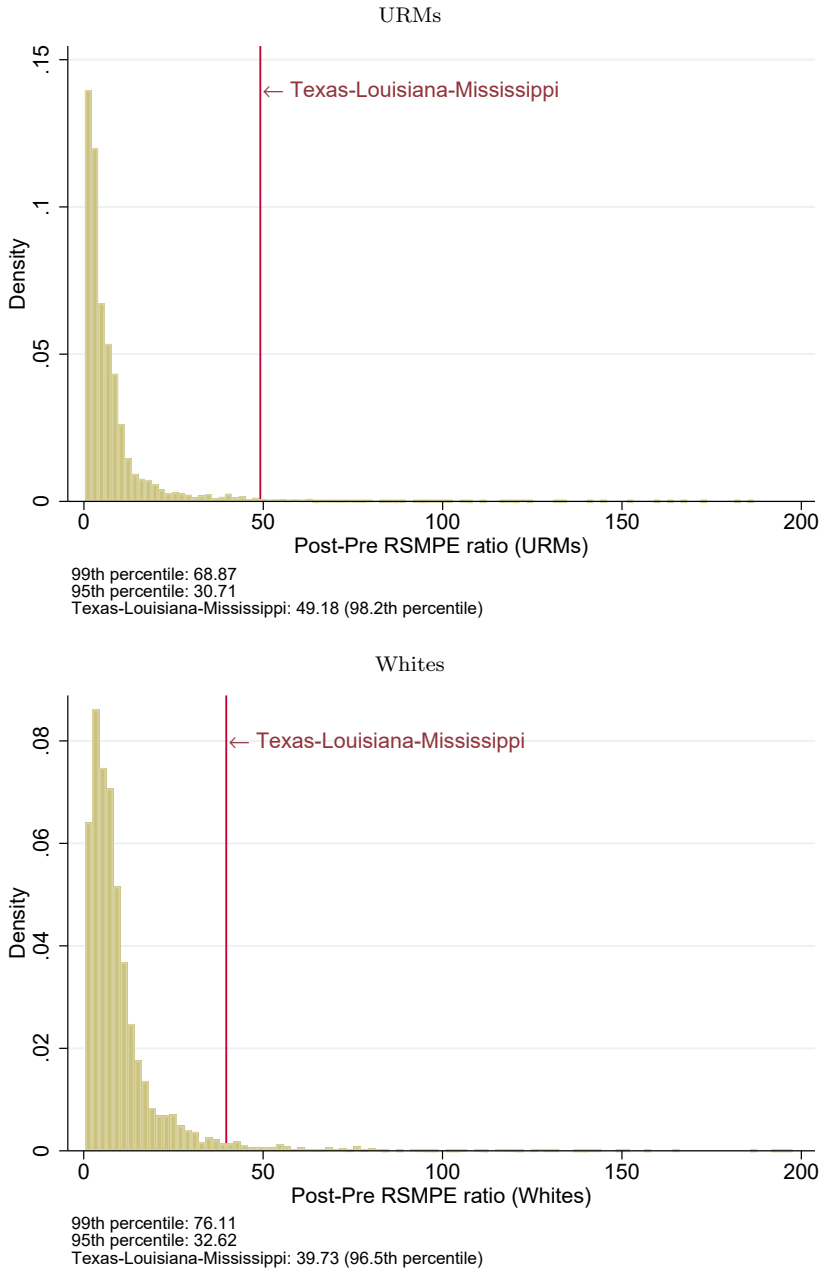
Note: The outcomes are state-year average SAT math and verbal scores. Dots indicate coefficients from a regression of the outcome on year indicator variables interacted with an indicator variable for the three treated states, estimated separately for Black and Hispanic students in Panels A and B and separately for Asian and white students in Panels C and D. Cells are weighted by the number of SAT test-takers. Dashed lines show 95% confidence intervals for standard errors clustered at the state level.

Figure A7. Effect of AA on SAT Scores of URM Students Relative to White Students (Triple-Difference)



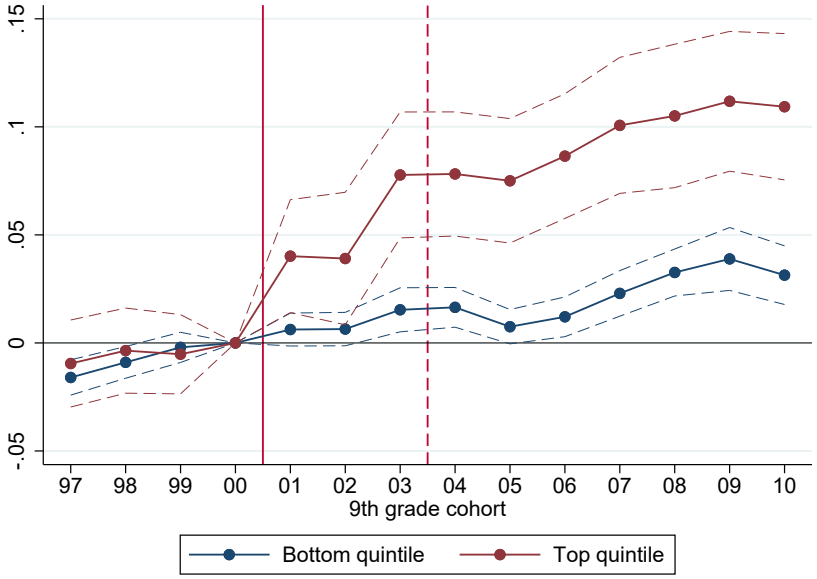
Note: The outcome is the state-year average SAT math scores. Dots indicate coefficients from a regression of the outcome on year indicator variables interacted with an indicator variable for the three treated states and an indicator for URM status. Cells are weighted by the number of SAT test-takers. Dashed lines show 95% confidence intervals for standard errors clustered at the state level.

Figure A8. Synthetic Control Permutation Tests



Note: This figure plots the distribution of post/pre RMSPE ratio statistics for placebo synthetic cohort estimates for whites (top panel) and URMs (bottom panel). The vertical red line indicates the post/pre RMSPE ratio for the treated states (Texas, Mississippi, and Louisiana).

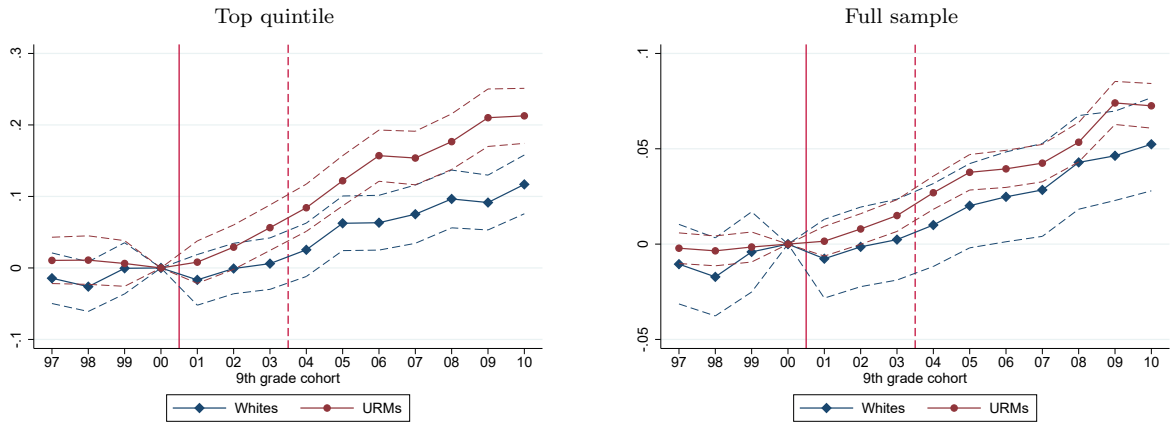
Figure A9. Effect of AA on Probability of Applying to Any Texas Public University for URMs Relative to Whites



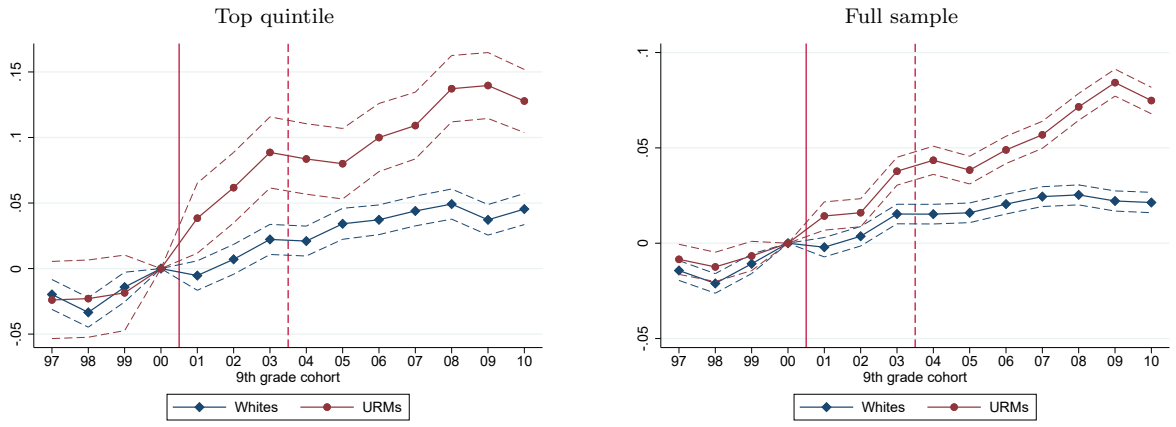
Note: The outcome is the probability of applying to any Texas 4-year public university within 4 years of starting 9th grade. Dots are coefficients from a regression of the outcome on year dummies interacted with URM status. All regressions condition on cohort-test score, race-test score, and district-test score fixed effects, where test score quintiles are from the cohort-specific distribution of 6th grade standard test scores. Dashed lines show 95% confidence intervals for standard errors clustered at the district level.

Figure A10. Raw Trends in College Applications by Race

Panel A: Number of Applications to Selective Universities

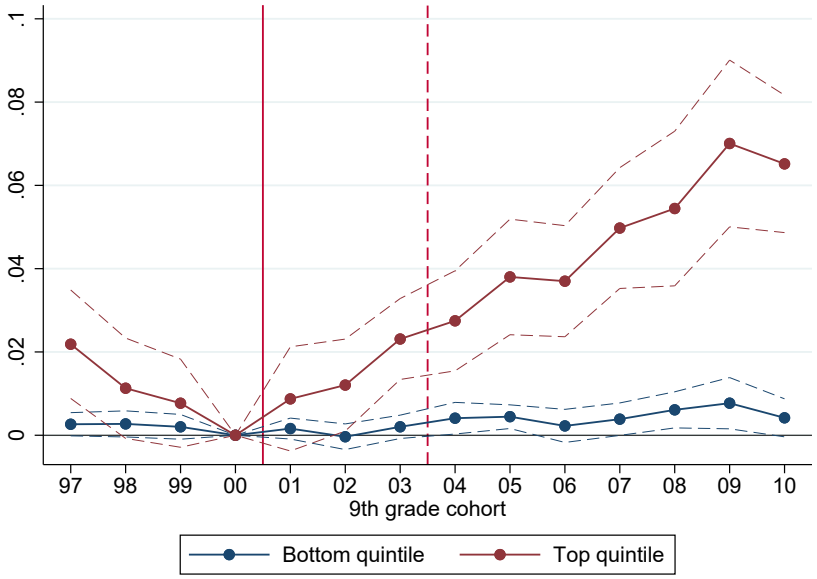


Panel B: Probability of Applying to Any UT Institution



Note: This figure reports raw average trends in college applications behavior in our analytical sample. Time series are normalized relative to the cohort that entered 9th grade in 2000. Dashed lines show 95% confidence intervals for standard errors clustered at the district level.

Figure A11. Effect of AA on Number of Applications to U Houston and Texas Tech (Combined) for URMs Relative to Whites



Note: The outcome is the average number of applications sent to the University of Houston and to Texas Tech within 4 years of starting 9th grade. Dots are coefficients from a regression of the outcome on year dummies interacted with URM status. All regressions condition on cohort-test score, race-test score, and district-test score fixed effects, where test score quintiles are from the cohort-specific distribution of 6th grade standard test scores. Dashed lines show 95% confidence intervals for standard errors clustered at the district level.

Appendix Tables

Table A1— Summary Statistics

	URMs		Whites	
SAT Data				
Years	1998-2003	2004-2010	1998-2003	2004-2010
Verbal scores	440.9	441.7	527.7	528.4
Math scores	438.7	443.4	530.1	534.7
Number of cells	878	1,026	306	357
Number of SAT takers	1,194,067	2,159,747	4,136,869	5,634,200
TEA Administrative Data				
Cohorts (grade 9)	1997-2000	2001-2010	1997-2000	2001-2010
Age (grade 9)	14.27	14.20	14.16	14.14
Limited English Proficiency (LEP)	0.067	0.048	0.000	0.000
Special Ed status	0.077	0.052	0.078	0.059
English as a Second Language (ESL)	0.042	0.038	0.000	0.000
Gifted	0.077	0.085	0.158	0.160
Immigrant	0.005	0.001	0.001	0.000
Poor	0.602	0.663	0.125	0.157
Female	0.508	0.508	0.499	0.496
6th grade test score decile	4.365	4.554	6.628	6.626
Attendance rate (grade 10)	0.934	0.941	0.954	0.955
Attendance rate (grade 11)	0.931	0.934	0.949	0.949
Applications to selective universities (within 4 years)	0.060	0.101	0.210	0.238
UT system application rate (within 4 years)	0.075	0.133	0.097	0.124
University application rate (within 4 years)	0.173	0.258	0.290	0.335
District-cohort-test score cells	12,492	36,462	17,414	41,614
Number of students	357,973	1,176,595	405,005	971,850
Number of districts	522	680	803	844
LUSD Administrative Data				
Cohorts (grade 11)	2001-2003	2004-2008	2001-2003	2004-2008
Age (grade 11)	16.39	16.41	16.21	16.22
Female	0.538	0.535	0.506	0.520
Mean school grades (grade 11)	77.34	78.17	82.24	83.45
Mean school grades (grade 8)	82.50	81.91	86.62	86.86
Attendance rate (grade 11)	0.929	0.927	0.943	0.948
Stanford test percentile rank (grade 11)	36.12	49.76	69.20	77.81
Number of students	17,620	34,107	3,623	5,779
Number of schools	42	49	36	42

Note: This table reports summary statistics (means) from the SAT data, the Texas Education Agency (TEA) administrative data, and the administrative data from a large, urban school district (LUSD). An observation in the SAT data is a race-year-state cell. An observation in the TEA data is a district-test score decile-cohort cell. The LUSD data consists of repeated cross sections of 11th graders, and an observation is a student.

Table A2— Summary Statistics for THEOP Survey Data

Panel A: Summary Statistics

	Full Sample		Whites		URMs	
	Mean	SD	Mean	SD	Mean	SD
Time (Minutes) Spent on Homework	64.54	56.69	56.06	53.60	70.74	56.22
Applied to First Choice College	0.65	0.48	0.70	0.46	0.60	0.49
Parental Involvement Index (0-15)	5.98	3.87	5.94	3.78	6.20	3.95
Discussed College App. w. Counselor	0.67	0.47	0.65	0.48	0.70	0.46

Panel B: Total Numbers

	N
Total Students	13,852
Whites	6,406
URMs	7,446
Students in 2002	11,025
Students in 2004	2,827

Note: This table reports summary statistics for the Texas Higher Education Opportunity Project (THEOP) survey data for two cohorts of seniors, one in 2002 and one in 2004. For the measure of how many minutes per day students spend on homework, students were asked how many hours per day they spent on their homework and were given the options zero hours, less than 1 hour, 1 to 2 hours, 3 to 4 hours, and 5+ hours. We convert these to minutes so that 0 hours is 0 minutes, less than 1 hour is 30 minutes, 1 to 2 hours is 90 minutes, and so on. The parental involvement index is constructed using several questions that ask "How often do your parents ... (i) give you special privileges because of good grades, (ii) try to make you work harder if you get bad grades, (iii) know when you are having difficulty in school, (iv) help with your schoolwork, and (v) talk with you about problems in school." Students' responses range from "very rarely" (1) to "almost all the time" (4). We sum across the answers to these questions to construct the "parental involvement index" so that a higher index corresponds to more involvement along these dimensions and renormalize the measure by subtracting 5 so that the minimum score is 0 rather than 5.

Table A3— Changes in Returns to HCI: Differences in Regression Coefficients for URMs Relative to Whites

Test score decile	Difference in regression coefficients	P-value
1	.	.
2	0.000	0.612
3	0.000	0.778
4	0.001	0.398
5	0.000	0.959
6	0.006	0.017
7	0.004	0.283
8	0.007	0.098
9	0.016	0.010
10	0.026	0.003

Note: This table reports the difference between the regression coefficients of URM and white students shown in Figure 3. It also reports the p-value on the null of equality between the regression coefficients.

Table A4— Effect of AA on College Applications Behavior – Black Students Relative to White Students

		Percentile of 6th grade test score distribution				
	All students	Bottom quintile	2nd quintile	3rd quintile	4th quintile	Top quintile
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: Applications to selective universities					
Partial treatment	0.018	0.004	0.009	0.010	0.021	0.071
	(0.004)	(0.003)	(0.004)	(0.007)	(0.010)	(0.014)
Full treatment	0.031	0.007	0.012	0.037	0.054	0.076
	(0.005)	(0.002)	(0.005)	(0.006)	(0.013)	(0.019)
Mean dependent variable	0.197	0.013	0.044	0.107	0.225	0.444
Test: quintile q = top quintile						
Partial treatment p-value		0.000	0.000	0.000	0.002	
Full treatment p-value		0.000	0.001	0.026	0.162	
	Panel B: Application to any UT institution					
Partial treatment	0.007	0.006	0.003	0.005	0.000	0.027
	(0.003)	(0.002)	(0.003)	(0.006)	(0.006)	(0.009)
Full treatment	0.014	0.007	0.010	0.017	0.013	0.032
	(0.004)	(0.002)	(0.003)	(0.005)	(0.008)	(0.014)
Mean dependent variable	0.102	0.011	0.032	0.062	0.114	0.220
Test: quintile q = top quintile						
Partial treatment p-value		0.015	0.003	0.017	0.007	
Full treatment p-value		0.059	0.113	0.278	0.138	
	Panel C: Application to any university					
Partial treatment	0.024	0.018	0.027	0.012	0.023	0.062
	(0.004)	(0.004)	(0.005)	(0.009)	(0.008)	(0.011)
Full treatment	0.031	0.015	0.026	0.028	0.040	0.072
	(0.005)	(0.004)	(0.007)	(0.008)	(0.009)	(0.014)
Mean dependent variable	0.300	0.074	0.149	0.238	0.357	0.519
Test: quintile q = top quintile						
Partial treatment p-value		0.000	0.003	0.000	0.008	
Full treatment p-value		0.000	0.004	0.003	0.019	
Observations (cells)	46,032	7,609	9,152	9,919	10,102	9,250

Note: This table reports difference-in-differences estimates of the effect of affirmative action on college application behavior. The sample is restricted to Black and white students. The regressions use the TEA data, an observation is at the district-cohort-race-test score quintile level, where test score quintile is assigned based on 6th grade (pre-AA) test scores on the state standardized test. The sample is restricted to students who were in 9th grade between 1997 and 2006. Cells are weighted by the number of student-years in a cell. Partial treatment is the interaction between an indicator for being a URM and an indicator variable for entering high school after 2001 and before 2003. Full treatment is the coefficient on the interaction between entering high school after 2003 and being a URM. The outcome variable in Panel A is the average number of selective universities to which students applied. The outcome variable in Panel B is the fraction of students in a cell that applied to any institution of the University of Texas system. In Panel C, the outcome is the probability of applying to any Texas 4-year public university. Standard errors are clustered at the district level.

Table A5— Effect of AA on College Applications Behavior – Hispanic Students Relative to White Students

		Percentile of 6th grade test score distribution				
	All students	Bottom quintile	2nd quintile	3rd quintile	4th quintile	Top quintile
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: Applications to selective universities					
Partial treatment	0.005	-0.000	-0.001	-0.002	0.010	0.015
	(0.003)	(0.002)	(0.002)	(0.003)	(0.007)	(0.010)
Full treatment	0.011	-0.002	0.001	0.006	0.019	0.034
	(0.005)	(0.002)	(0.003)	(0.005)	(0.008)	(0.013)
Mean dependent variable	0.157	0.006	0.029	0.083	0.198	0.418
Test: quintile q = top quintile						
Partial treatment p-value		0.113	0.107	0.097	0.668	
Full treatment p-value		0.004	0.012	0.026	0.181	
	Panel B: Application to any UT institution					
Partial treatment	0.001	0.002	-0.005	0.001	-0.002	0.011
	(0.002)	(0.002)	(0.002)	(0.003)	(0.005)	(0.007)
Full treatment	0.004	-0.003	-0.002	0.005	0.003	0.015
	(0.003)	(0.002)	(0.002)	(0.005)	(0.006)	(0.008)
Mean dependent variable	0.115	0.024	0.054	0.090	0.140	0.237
Test: quintile q = top quintile						
Partial treatment p-value		0.225	0.020	0.168	0.101	
Full treatment p-value		0.038	0.049	0.221	0.151	
	Panel C: Application to any university					
Partial treatment	0.000	0.004	-0.007	-0.003	0.002	0.011
	(0.003)	(0.003)	(0.004)	(0.004)	(0.006)	(0.009)
Full treatment	0.009	-0.006	-0.007	0.005	0.018	0.034
	(0.004)	(0.003)	(0.003)	(0.007)	(0.007)	(0.010)
Mean dependent variable	0.261	0.054	0.124	0.216	0.342	0.509
Test: quintile q = top quintile						
Partial treatment p-value		0.449	0.048	0.105	0.297	
Full treatment p-value		0.000	0.000	0.001	0.086	
Observations (cells)	58,024	9,946	11,917	12,668	12,470	11,023

Note: This table reports difference-in-differences estimates of the effect of affirmative action on college application behavior. The sample is restricted to Hispanic and white students. The regressions use the TEA data, an observation is at the district-cohort-race-test score quintile level, where test score quintile is assigned based on 6th grade (pre-AA) test scores on the state standardized test. The sample is restricted to students who were in 9th grade between 1997 and 2006. Cells are weighted by the number of student-years in a cell. Partial treatment is the interaction between an indicator for being a URM and an indicator variable for entering high school after 2001 and before 2003. Full treatment is the coefficient on the interaction between entering high school after 2003 and being a URM. The outcome variable in Panel A is the average number of selective universities to which students applied. The outcome variable in Panel B is the fraction of students in a cell that applied to any institution of the University of Texas system. In Panel C, the outcome is the probability of applying to any Texas 4-year public university. Standard errors are clustered at the district level.

Table A6— Effect of AA on Grades for URMs Relative to Whites by Achievement Terciles

	All students			Grades in 8th grade		
				Bottom tercile	Middle tercile	Top tercile
	(1)	(2)	(3)	(4)	(5)	(6)
Treated	0.877	1.002	0.955	0.872	0.406	1.386
	(0.309)	(0.298)	(0.311)	(0.510)	(0.390)	(0.421)
Lagged dep. var. (grade 8)		0.555				
		(0.009)				
Observations	61,089	46,346	92,847	15,881	15,614	14,778
R^2	0.226	0.345	0.784	0.189	0.224	0.208
Mean dependent variable	78.67	79.48	81.11	75.79	79.49	83.46
S.D. dependent variable	8.67	7.80	7.37	7.43	6.99	6.97
Test: tercile q = top tercile p-value				0.423	0.073	
School-year FE	X	X	X	X	X	X
Ethnicity FE	X	X		X	X	X
Demographic controls	X	X		X	X	X
Student FE			X			
Grade-year FE			X			
Grade-ethnicity FE			X			

Note: This table reports difference-in-differences estimates of the effect of affirmative action on grades in a large urban school district. An observation is a student, and the sample consists of repeated cross sections of 11th graders. "Treated" is the coefficient on the interaction between being a URM and being observed post 2003. Achievement terciles are assigned based on 8th grade average grades. Standard errors are clustered at the school-cohort level.

Table A7— Effect of AA on Stanford Test Scores for URMs Relative to Whites

		Grades in 8th grade				
	All students	Bottom quintile	2nd quintile	3rd quintile	4th quintile	Top quintile
	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent variable: Stanford Test Scores (grade 11)					
Treated	4.780	3.936	5.533	2.812	6.556	7.885
	(1.135)	(1.626)	(2.031)	(1.927)	(1.542)	(1.693)
Observations	58,096	9,287	9,282	9,180	9,083	8,532
R^2	0.444	0.471	0.470	0.498	0.492	0.481
Mean dependent variable	49.40	40.48	45.87	50.43	55.31	62.77
S.D. dependent variable	25.74	23.30	23.50	23.94	23.85	23.32
Test: quintile q = top quintile p-value		0.084	0.356	0.029	0.519	
School-year FE	X	X	X	X	X	X
Ethnicity FE	X	X	X	X	X	X
Demographic controls	X	X	X	X	X	X

Note: This table reports the difference-in-differences estimates of the effect of affirmative action on mean Stanford test scores (in national percentile ranks) in a large, urban school district. Mean Stanford test scores are the average across 5 subjects (reading, math, language, science, and social science). An observation is a student, and the sample consists of repeated cross sections of 11th graders. "Treated" is the coefficient on the interaction between being a URM and being observed post 2003. Achievement quintiles are assigned based on 8th grade average school grades, and the sample is restricted to years 2001 to 2008. Standard errors are clustered at the school-cohort level.

Table A8— Effect of AA on College Applications Behavior – Excluding Houston & Dallas

		Percentile of 6th grade test score distribution				
	All students	Bottom quintile	2nd quintile	3rd quintile	4th quintile	Top quintile
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: Applications to selective universities					
Partial treatment	0.010	0.003	0.001	0.002	0.015	0.031
	(0.003)	(0.002)	(0.002)	(0.004)	(0.007)	(0.009)
Full treatment	0.021	0.003	0.004	0.016	0.036	0.052
	(0.003)	(0.001)	(0.003)	(0.005)	(0.006)	(0.010)
Mean dependent variable	0.151	0.008	0.032	0.086	0.198	0.417
Test: quintile q = top quintile						
Partial treatment p-value		0.002	0.001	0.002	0.107	
Full treatment p-value		0.000	0.000	0.000	0.113	
	Panel B: Application to any UT Institution					
Partial treatment	0.003	0.004	-0.003	0.002	0.002	0.013
	(0.002)	(0.001)	(0.002)	(0.003)	(0.004)	(0.006)
Full treatment	0.009	0.001	0.002	0.009	0.010	0.024
	(0.003)	(0.002)	(0.002)	(0.004)	(0.005)	(0.007)
Mean dependent variable	0.111	0.023	0.054	0.090	0.141	0.237
Test: quintile q = top quintile						
Partial treatment p-value		0.134	0.007	0.075	0.070	
Full treatment p-value		0.002	0.004	0.038	0.055	
	Panel C: Application to any university					
Partial treatment	0.008	0.010	0.004	0.002	0.010	0.024
	(0.003)	(0.003)	(0.004)	(0.005)	(0.005)	(0.007)
Full treatment	0.018	0.004	0.004	0.014	0.029	0.048
	(0.003)	(0.003)	(0.004)	(0.006)	(0.006)	(0.008)
Mean dependent variable	0.264	0.064	0.139	0.230	0.350	0.513
Test: quintile q = top quintile						
Partial treatment p-value		0.088	0.009	0.007	0.071	
Full treatment p-value		0.000	0.000	0.000	0.016	
Observations (cells)	67,909	12,813	14,395	14,689	14,025	11,987

Note: This table reports difference-in-differences estimates of the effect of affirmative action on college applications behavior. The regressions use the TEA data, an observation is at the district-cohort-race-test score quintile level, where quintiles are assigned based on 6th grade (pre-AA) test scores on the state standardized test. The sample is restricted to students who were in 9th grade between 1997 and 2006. The sample excludes the Houston Independent School District and the Dallas Independent School District. Cells are weighted by the number of student-years in a cell. Partial treatment is the coefficient on the interaction between an indicator for being a URM and an indicator variable for entering high school after 2001 and before 2003. Full treatment is the coefficient on the interaction between entering high school after 2003 and being a URM. The outcome variable in Panel A is the average number of selective universities to which students applied. The outcome variable in Panel B is the fraction of students in a cell that applied to any institution of the University of Texas system. In Panel C, the outcome is the probability of applying to any Texas 4-year public university. Standard errors are clustered at the district level.

Appendix A: Robustness and Details of SAT Results

In this appendix, we evaluate the robustness of our difference-in-differences results (Part A) to

1) Alternative ways of accounting for AA bans in control states,
2) Accounting for potential non-compliance in Louisiana and Mississippi,
3) Accounting for group-specific pre-trends.

We also report the details of the synthetic control method and verify the robustness of these results (Part B). The subsections are as follows:

4. Robustness to using alternative sets of pre-treatment variables in the matching process,
5. Robustness to using alternative numbers of pre-treatment years in the matching process,
6. Robustness to dropping Louisiana and Mississippi from the treated unit,
7. Details on synthetic control inference procedures.

A. Difference-in-Differences Estimates

Baseline estimates for math and verbal test scores are presented in columns (1) and (2) of Table 1.

1. Accounting for AA Bans.. — Several states implemented affirmative action bans in university admissions during our study period. Washington, Michigan, and Nebraska passed affirmative action bans through ballot initiatives in November 1998, 2006, and 2008, respectively. Governor Jeb Bush issued an executive order banning affirmative action in Florida in November 1999.[35] In our main empirical specification, we control for affirmative action bans in control states.[36] In this subsection, we verify that our difference-in-differences estimates of the effect of *Grutter v. Bollinger* on students in Texas, Louisiana, and Mississippi are robust to alternative ways of accounting for these affirmative action bans. Robustness tests for math SAT scores are reported in Table A9 and, for completeness, corresponding results for verbal scores are reported in Table A10. Column (1) reports results for our baseline specification. In column (2), we omit the controls for the effect of affirmative action bans. Our estimates of the effect of *Grutter v. Bollinger* on whites (4 points) and on URMs (8 points) are unaffected by the exclusion of an indicator for AA bans as a control variable. Failing to control for AA bans yields a slightly smaller triple-differences estimate of 4 points. In column (3), we drop the four states that banned affirmative action between 1998 and 2010 from the estimating sample. Again, our estimates of the effect of the reinstatement of affirmative action are virtually unchanged.

Finally, we report results from a specification that leverages changes in AA ban status in all 7 switching states in Table A11 (Texas, Mississippi, and Louisiana terminating a ban, and Florida, Michigan, Washington, and Nebraska implementing one). This identifies not just the pure affirmative action effect but the effect of affirmative action relative to any compensatory policies that may be enacted in response to banning affirmative action, such as percent plans (Hinrichs, 2012). The point estimates are slightly smaller but qualitatively similar to the baseline estimates.

[35] Following our study period, Arizona (2010), New Hampshire (2011) and Oklahoma (2012) also banned affirmative action in college admissions.

[36] The indicator turns on after 1999 in Washington, after 2000 in Florida, after 2007 in Michigan, and after 2009 in Nebraska. It is zero in all years for all other states.

2. Accounting for Potential Non-Compliance.. — There is some evidence that Louisiana and Mississippi may have continued to use race in university admissions to some extent in 1998-2003 despite the *Hopwood v. Texas* ruling due to pre-existing rulings that required them to desegregate their institutions of higher education (Hinrichs, 2012). Thus, we also drop these two states from the sample and estimate the effects of the policy change on Texas alone relative to the control states in column (4) of Table A9. Dropping these two states has little effect on the estimates.

3. Accounting for Group-Specific Pre-Trends.. — It is apparent in Panel A, Figure 4, that Texas, Louisiana and Mississippi were falling behind the rest of the country prior to the reinstatement of AA. To account for these differential pre-trends, we estimate a linear trend term separately for each racial group and treatment group using only the pre-treatment years, and partial out this linear trend from the full panel. We use the resulting de-trended data as our outcome variable in column (5). The point estimates are generally twice as large as they are in our baseline specifications. This suggests that, if anything, our main estimates put a lower bound on the effect of affirmative action on SAT scores.

B. Synthetic Control Specification

We construct a synthetic control group of states by matching those states' pre-trends in test scores to the pre-trends of the treated unit (the weighted average of Texas, Mississippi, and Louisiana). When generating the synthetic control groups, we exclude Florida, Michigan, Washington, and Nebraska from the donor pool since these states implemented AA bans during the study period. We also exclude South Dakota, North Dakota, Wyoming, and Washington DC from the pool of potential controls because SAT scores are missing for some ethnic groups in some years in these states due to small samples. Then, we choose the control group by minimizing the mean squared prediction errors in 1998-2003. We consider the following potential predictors:

- average math SAT scores over years 1998-2000 and/or 2001-2003, measured separately for URMs and whites
- average verbal SAT scores over years 1998-2000 and/or 2001-2003, measured separately for URMs and whites
- average number of SAT test-takers over years 1998-2000 and/or 2001-2003, measured separately for URMs and whites
- SAT taking rates in years 2000-2001 and/or 2002-2003, measured separately for URMs and whites[37]
- average grade 8 NAEP math test scores in 2003[38]
- per capita state expenditures in education in 2000 and/or 2003
- the fraction of total state expenditure on education in 2000 and/or 2003.[39]

[37] To calculate SAT taking rates, we divide the number of test-takers by the number of 17-19 years olds in each state-by-race-by-year cell, which we obtain from ACS/Census data. Note that these population counts are not available in 1998 and 1999, and that population counts at such a disaggregated level are quite volatile, which may introduce measurement error.

[38] Data available at https://www.nationsreportcard.gov/ndecore/landing

[39] Both the per capita and fraction of state expenditures measures are obtained from https://www.census.gov/programs-surveys/state.html

Table A9— Robustness Checks for the Effect of AA on Math SAT Scores

	Baseline	No control for AA bans	Drop AA ban states	Drop Mississippi and Louisiana	De-trended data
	(1)	(2)	(3)	(4)	(5)
			Panel A: URMs		
DD coefficient	8.009	7.998	8.122	8.092	20.02
	(1.544)	(1.498)	(1.694)	(1.538)	(1.522)
Observations (cells)	1,904	1,904	1,748	1,830	1,904
R^2	0.844	0.844	0.839	0.844	0.693
State, year and ethnicity FE	X	X	X	X	X
			Panel B: Whites		
DD coefficient	4.048	4.145	3.835	4.455	7.208
	(0.984)	(0.995)	(1.021)	(0.849)	(0.984)
Observations (cells)	663	663	611	637	663
R^2	0.969	0.968	0.967	0.969	0.968
State, year and ethnicity FE	X	X	X	X	X
			Panel C: Asians		
DD coefficient	0.658	0.658	0.447	0.362	6.659
	(1.827)	(1.813)	(1.962)	(1.788)	(1.828)
Observations (cells)	663	663	611	637	663
R^2	0.944	0.944	0.939	0.944	0.939
State, year and ethnicity FE	X	X	X	X	X
			Panel D: Triple-Difference (URMs vs Whites)		
DDD coefficient	4.155	3.975	4.253	4.059	10.17
	(0.828)	(0.872)	(0.816)	(0.817)	(1.263)
Observations (cells)	2,555	2,555	2,347	2,455	2,555
R^2	0.998	0.998	0.998	0.998	0.989
State-year FE	X	X	X	X	X
State-ethnicity FE	X	X	X	X	X
Ethnicity-year FE	X	X	X	X	X

Note: This table reports difference-in-differences and triple-difference effects of affirmative action on SAT scores. Each observation is a state-race-year group. In all specifications, cells are weighted by the number of test-takers in a group. In Panels A, B and C, the DD coefficient reports the interaction of an indicator variable for belonging to a treated state (Texas, Louisiana, and Mississippi) and being tested after *Grutter v. Bollinger* (post 2003). In Panel D, the coefficient is on the interaction between being a URM, being tested post 2003, and belonging to a treated state. All specifications control for a AA ban indicator variable (and in the case of Panel D, its interaction with URM) except columns (2) and (3). Standard errors are clustered at the state level.

Table A12 reports the weights put on each control state as well as on the variables used in the matching algorithm for our preferred specification, which minimizes the pre-treatment RMSPE for URMs.

4. Alternative Sets of Predictors.Figure A12 shows results for 12 alternative sets of predictors and highlights our preferred specification, which minimizes the RMSPE in the pre-treatment period for URMs. The predictors included in the preferred model are math SAT scores, verbal SAT scores, SAT taking rates in 2002-2003, and the fraction of total state expenditure on education in 2000 and 2003. We use that specification for all other synthetic control analyses in the paper.

All 12 specifications produce results consistent with positive effects for both white and minority students. Our main (preferred) specification yields neither abnormally large nor abnormally small effects compared to other models.

5. Matching on Fewer Pre-treatment Cohorts.Here, we verify that our results are robust to using fewer pre-treatment years in the construction of the synthetic control group. That is, we ensure that the treatment and control groups do not just diverge in 2003 because this is the first year we do not choose weights to match the treatment and control groups' outcomes. Figure A13 shows time series of math SAT scores for synthetic control groups based on 4, 5, and 6 years of pre-treatment data.[40]

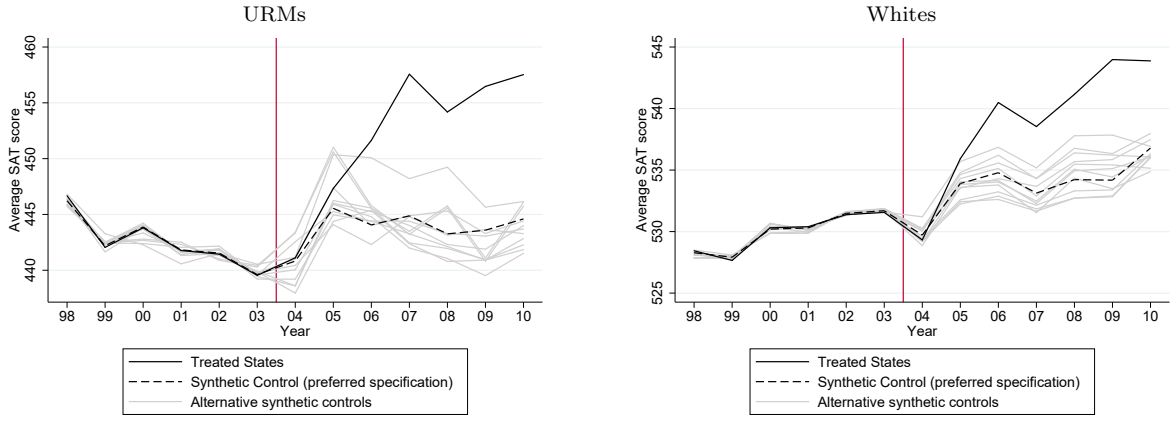
For both URMs and whites, the results are largely insensitive to using either 5 or 6 years of pre-treatment data. The pre-treatment fit is poorer if only 4 years are used, but the gap between treated and untreated states during pre-treatment years remains relatively small, even in years that were not used in the construction of the synthetic control group. The post-treatment outcomes of the three synthetic control groups are qualitatively similar, and in all cases, the control groups' SAT scores are significantly below those of the treated states.

6. Dropping Louisiana and Mississippi.As in the difference-in-differences robustness section, in Figure A14, we drop Louisiana and Mississippi from the sample and estimate the effects of the policy change on Texas alone relative to its synthetic control group. Because there are many more SAT test-takers in Texas than in Louisiana and Mississippi, looking at Texas in isolation produces very similar results.

7. Inference.Our first inference approach is based on permutation tests in which the treatment is randomly reassigned to control units. Our treatment unit is a weighted average of three states (Texas, Mississippi, and Louisiana). Our donor pool contains 40 untreated states. There are therefore 9,880 possible placebo combinations of three control states. For each of these combinations, we run the synthetic control algorithm using our main specification and store the post/pre-treatment ratio of root mean squared prediction errors (RMSPE). The distributions of RMSPE ratios for the 9,880 placebo estimates are shown in Figure A8. We obtain a p-value by finding the rank of the RMSPE ratio for the real treated unit in those distributions. The treatment unit's post-pre ratio of RMSPE is at the 96.5th percentile of the distribution for whites and at the 98.2th percentile for URMs. Note that this permutation test based on Abadie, Diamond and Hainmueller (2010) relies on random assignment of treatment across states. The two alternative methods of inference we examine below do not rely on random assignment.

[40]When minimizing the RMSPE over 4 years, we only use predictors measured in years 1998 to 2001. When minimizing the RMSPE over 5 years, we only use predictors measured in years 1998 to 2002.

Figure A12. Synthetic Control Estimates of the Effect of AA on Math SAT Scores, Alternative Sets of Matching Variables

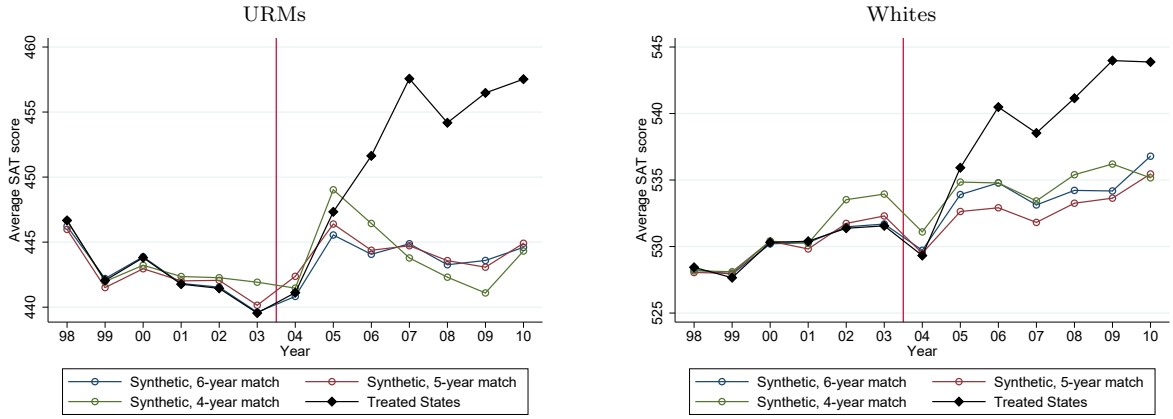


Note: This figure reports synthetic cohort analyses separately for whites and URMs. It shows SAT math scores for the treated states (Texas, Mississippi, and Louisiana) and for synthetic control groups based on 12 alternative sets of predictors.

As an alternative method of inference, we implement the t-test proposed by Chernozhukov, Wüthrich and Zhu (2022), which is a K-fold cross-fitting procedure. The procedure is as follows. We split the pre-treatment period into K mutually exclusive, consecutive blocks (subperiods). Then, for all $k = 1, ..., K$, we run the synthetic control algorithm (using the our main specification), excluding the corresponding pre-treatment subperiod from the data, and save the estimated synthetic control weights $\hat{w}_{i,k}$. For each k, we compute the average treatment effect $\hat{\tau}_k$ as the pre-post difference in average differences between the treated unit and the synthetic control. Note that when calculating the average differences over the pre-treatment period for block k, we only use the relevant subperiod. That is, in the spirit of cross-validation approaches, for each block k, weights are estimated *excluding* years from that block, whereas treatment effects are calculated by only including pre-treatment years from that block. The t-statistic is equal to $(\sqrt{K}(\hat{\tau} - \tau_0))/(\hat{\sigma}_{\hat{\tau}})$, where $\hat{\tau}$ is the average treatment effect across the K blocks, τ_0 is the null hypothesis, and $\hat{\sigma}_{\hat{\tau}}$ is the sample standard deviation of $\hat{\tau}_k$. We use $K = 2$, which allows us to average our predictors over the 1998-2000 and 2001-2003 subperiods to maintain consistency with our main specification. The null of no effect is rejected at at least the 10% level for both URMs and whites, with p-values equal to 0.048 and 0.072, respectively.

Finally, we consider the conformal inference method proposed by Chernozhukov, Wüthrich and Zhu (2021), which produces p-values by permuting residuals (differences in outcomes between the treated unit and the synthetic control) in the *time* dimension. Importantly, under this approach, synthetic control weights are estimated under the null hypothesis using *all* periods (i.e. including the post-treatment years). However, this approach requires a long pre-treatment period. Our data only includes 6 pre-treatment years, which makes this approach ill-suited for our setting. With this caveat in mind, we run the synthetic control algorithm using the same predictors as in our main specification for consistency, but minimizing the mean squared prediction error over the full 1998-2010 period, and save the associated residuals. We then produce p-values based on 10,000 i.i.d permutations of the residuals with replacement. The null of no effect is rejected at the 10% level

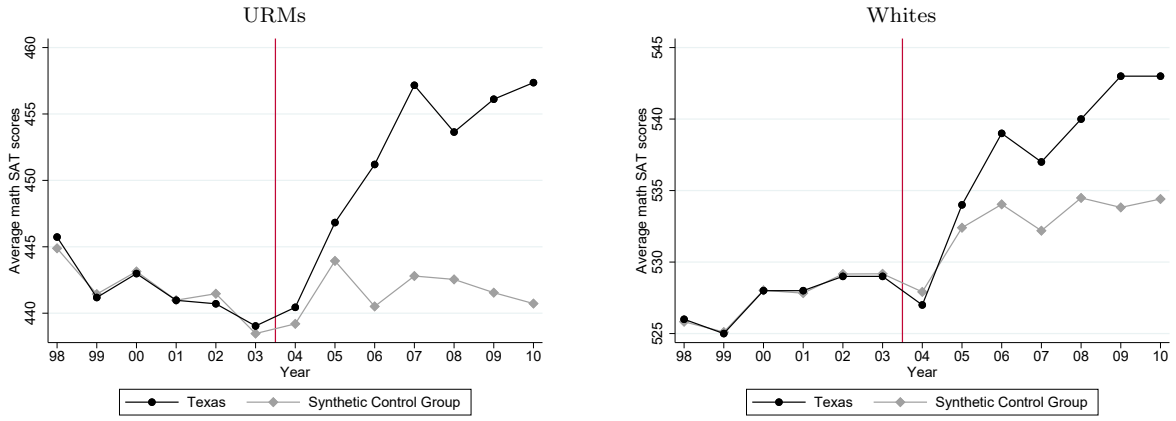
Figure A13. Synthetic Control Estimates of the Effect of AA on Math SAT Scores by Number of Pre-treatment Years Used in Match



Note: This figure reports synthetic cohort analyses separately for whites and URMs. It shows SAT math scores for the treated states (Texas, Mississippi, and Louisiana) and for the synthetic control group under alternative matching specifications. The control group "Synthetic, 6-year match" is obtained by minimizing the root mean squared prediction error (RMSPE) over the 1998-2003 period. For "Synthetic, 5-year match," the RMSPE is minimized over the 1998-2002 period, and for "Synthetic, 4-year match," it is minimized over the 1998-2001 period.

for both URMs and whites, with p-values equal to 0.085 and 0.054, respectively. Chernozhukov, Wüthrich and Zhu (2021) also propose using moving block permutations, which is valid under weaker assumptions on the distribution of residuals. However, with only 13 years of data, there are only 12 such permutations possible in our setting, making it difficult to test at low significance levels. Under this approach, the true treatment ranks 3/13 for URMs and 2/13 for whites.

Figure A14. Synthetic Control Estimates of the Effect of AA on Math SAT Scores Dropping Louisiana and Mississippi



Note: This figure reports synthetic cohort analyses separately for whites and URMs. It shows SAT math scores for the treated state (Texas) and for the synthetic control group. In constructing the control group, Louisiana and Mississippi were omitted from the donor pool.

Table A10— Robustness Checks for the Effect of AA on Verbal SAT Scores

	Baseline	No control for AA bans	Drop AA ban states	Drop Mississippi and Louisiana	De-trended data
	(1)	(2)	(3)	(4)	(5)
			Panel A: URMs		
DD coefficient	-0.634	-0.779	-0.170	-0.649	8.000
	(1.784)	(1.756)	(1.929)	(1.795)	(1.935)
Observations (cells)	1,901	1,901	1,745	1,828	1,901
R^2	0.796	0.795	0.788	0.793	0.717
State, year and ethnicity FE	X	X	X	X	X
			Panel B: Whites		
DD coefficient	0.034	0.025	0.001	-0.026	-0.179
	(0.888)	(0.878)	(0.955)	(0.888)	(0.888)
Observations (cells)	663	663	611	637	663
R^2	0.971	0.971	0.970	0.970	0.971
State, year and ethnicity FE	X	X	X	X	X
			Panel C: Asians		
DD coefficient	-0.176	-0.145	-0.456	-0.426	5.088
	(2.753)	(2.709)	(2.880)	(2.743)	(2.751)
Observations (cells)	663	663	611	637	663
R^2	0.929	0.929	0.925	0.928	0.920
State, year and ethnicity FE	X	X	X	X	X
			Panel D: Triple-Difference (URMs vs Whites)		
DDD coefficient	1.260	1.083	1.440	1.331	7.137
	(0.753)	(0.825)	(0.634)	(0.763)	(1.081)
Observations (cells)	2,552	2,552	2,344	2,453	2,552
R^2	0.998	0.998	0.998	0.998	0.990
State-year FE	X	X	X	X	X
State-ethnicity FE	X	X	X	X	X
Ethnicity-year FE	X	X	X	X	X

Note: This table reports difference-in-differences and triple-difference effects of affirmative action on SAT scores. Each observation is a state-race-year group. In all specifications, cells are weighted by the number of test-takers in a group. In Panels A, B and C, the DD coefficient reports the interaction of an indicator variable for belonging to a treated state (Texas, Louisiana, and Mississippi) and being tested after *Grutter v. Bollinger* (post 2003). In Panel D, the coefficient is on the interaction between being a URM, being tested post 2003, and belonging to a treated state. All specifications control for a AA ban indicator variable (and in the case of Panel D, its interaction with URM) except columns (2) and (3). Standard errors are clustered at the state level.

Table A11— Effect of AA on SAT Scores for URMs and Whites, Including Variation From AA Ban States

	Math	Verbal
	(1)	(2)
	Panel A: URMs	
AA effect	6.180	-1.402
	(1.760)	(1.385)
Observations (cells)	1,904	1,901
R^2	0.844	0.795
State, year, and race FE	X	X
	Panel B: Whites	
AA effect	3.792	-0.085
	(1.033)	(1.111)
Observations (cells)	663	663
R^2	0.969	0.971
State, year, and race FE	X	X
	Panel C: Asians	
AA effect	0.460	0.315
	(2.265)	(1.583)
Observations (cells)	663	663
R^2	0.944	0.929
State, year, and race FE	X	X

Note: This table reports difference-in-differences estimates of the effect of affirmative action on SAT scores. Each observation is a state-race-year group. Cells are weighted by the number of test-takers in a group. The DD coefficient reports the interaction of an indicator variable for belonging to a AA switching state (Texas, Louisiana, Mississippi, Florida, Washington, Michigan, or Nebraska) and being tested during AA years. Standard errors are clustered at the state level.

Table A12— Synthetic Control Weights

	Weight	
Group:	**URMs**	**Whites**
States in synthetic control group		
Alabama		0.022
Arizona		0.158
California		0.321
Indiana		0.225
Kentucky	0.035	
Minnesota	0.030	
Nevada	0.222	
New Jersey	0.248	
North Carolina		0.094
Ohio	0.192	0.071
Pennsylvania	0.232	0.003
West Virginia	0.042	0.107
Predictor variables		
Math SAT scores, URMs, 1998-2000	0.265	0.038
Math SAT scores, URMs, 2001-2003	0.182	0.014
Math SAT scores, whites 1998-2000	0.314	0.471
Math SAT scores, whites 2001-2003	0.005	0.400
Verbal SAT scores, URMs, 1998-2000	0.000	0.003
Verbal SAT scores, URMs, 2001-2003	0.000	0.003
Verbal SAT scores, whites 1998-2000	0.209	0.000
Verbal SAT scores, whites 2001-2003	0.025	0.009
SAT taking rate, URMs, 2002-2003	0.000	0.000
SAT taking rate, whites, 2002-2003	0.000	0.003
% state expenditures on education in 2000	0.000	0.046
% state expenditures on education in 2003	0.000	0.012

Note: States with zero weights for both URMs and whites are omitted from the table.