Statistical Discrimination and Optimal Mismatch in College Major Selection

Mary Kate Batistich
Timothy N. Bond
Sebastian Linde
Kevin J. Mumford

University of Notre Dame
Purdue University
Texas A&M University

August 21, 2024

Mismatch and Major Choice

- College major decision is one of most important investment choices for high skill workers
- Wage difference between college graduates with high paying and low paying degrees nearly as large as gap between high school and college graduates (Altonji et al., 2012)
- Central argument against affirmative action in admissions preferences is potential effect on major choice (e.g., *Students for Fair Admissions v. Harvard*)
- Black students attempt lower paying majors at more prestigious institutions than they would have if they attended less prestigious institutions (e.g., Arcidiacono et al., 2012, 2016).

Information and Major Choice

- Focus on role of two types of information frictions in major choice and empirical implications for mismatch
- Mismatch literature (Arcidiacono et al. 2011)
  - Students have incomplete information on their own aptitude
  - Admissions decisions cause students to update their beliefs
  - Black students who are admitted due to diversity preferences will form overly positive beliefs about their aptitude, make too difficult human capital investments which lower welfare
- Statistical discrimination literature (Lang and Manove 2011)
  - Employers have less precise information on the productivity of black applicants than white applicants
  - Rely more heavily on observable indicators for black applicants
  - Incentivizes black students to overinvest in education

Our Model of Majors

- Students with incomplete information on their aptitude choose from a menu of majors that differ in their human capital production function
- Black students have less precise beliefs about their aptitude than white students
- Employers cannot initially observe accrued human capital but they do observe major, college grades, and a signal of productivity
- Following standard assumptions in statistical discrimination literature, signal is more precise for white students

Tensions of Information Frictions

- Student incomplete information
  - *Lowers* the value of black major choice as a signal to employers since black students are less informed of their aptitude when choosing their major
  - *Reduces* incentive for black students to choose more difficult majors
  - Causes black students in equilibrium to attempt *less* difficult majors than similar white students
- Statistical Discrimination
  - *Raises* value of black major choice as a signal to employer since labor market signal less reliable
  - *Increases* incentive for black students to choose a more difficult major leading black students to attempt less difficult majors than similar white students
  - Causes black students in equilibrium to attempt *more* difficult majors than similar white students

Empirical Results

- Test for which of these two frictions are more important using three different data sets
- Find support that statistical discrimination is dominant force
- Black students take higher paid and more STEM-focused majors than white students conditional on SAT, high school grades
- Disparity grows when moving up the SAT distribution
- Black students earn lower wages than white students in same major, both conditional and unconditional on SAT scores
- Largest racial wage disparity is among those in highest earning majors
- Find evidence that black students have less precise beliefs about their aptitude when choosing college major using racial differences in labor market return to college grades

Literature Review

- Affirmative Action and College Mismatch
  - Sander (2004)
  - Arcidiacono et al. (2011)
  - Mountjoy and Hickman (2021)
  - Bleemer and Mehta (2022)
  - Akhtari et al. (2024)
- Racial Differences in College Major Selection
  - Arcidiacono et al. (2012)
  - Arcidiacono et al. (2016)
  - Hill (2017)
  - Sovero et al. (2021)
  - Bleemer and Mehta (2021)
- Effect of market conditions on major choice
  - Ersoy (2020)
  - Han and Winters (2020)
  - Blom et al. (2021)
  - Weinstein (forthcoming)

Primitives

- Large number of (b)lack and (w)hite students possess normally distributed beliefs about their aptitude, with black students having a higher variance in their beliefs than white students
- Choose from continuum of investment technologies m which differ in complementary with aptitude (difficulty)
- Conditional on aptitude, human capital production function single-peaked, choosing too easy or too difficult major will lead to lower realized productivity
- Employers do not observe realized productivity, instead observe major choice, college grades, and an unbiased labor market signal
- Labor market signal more precise for white workers, unobservable to econometrician
- Grades equally precise across race, observable to econometrician

Equilibrium Major Selection

- In equilibrium, there is a race-specific one-to-one mapping of aptitude to major
- Lowest types choose the major which maximizes human capital (no incentive to deviate)
- All other workers choose more difficult majors than optimal (sheepskin incentives)
- Student information frictions dominate: White workers overcredentialize more than black workers, because employers view major choice as an imprecise measure of black worker productivity $\rightarrow$ more human capital, higher wages, larger observed return to major difficulty
- Statistical discrimination dominates: Black workers overcredentialize more than whites, because employers view major choices as a relatively more precise measure of black worker productivity $\rightarrow$ less human capital, lower wages, lower observed return to major difficulty

Grades

- Previous tests only differentiate between whether statistical discrimination is stronger or weaker than student information frictions
- Do not tell us if weaker force exists at all
- From econometrician's perspective regression of wages on grades and major choice is simply $E[w|m, g]$ (law of iterated expectations)
- That is, regression coefficients will tell us which is a stronger predictor of worker productivity: major or grades
- Grades are equally precise across race, but major less correlated with black student's productivity only if black students had worse information about their aptitude when making major choice
- Provides independent test of information friction hypothesis

Testable Predictions

- Black students should graduate in more (less) difficult majors conditional on measures of college preparation (SAT scores) if statistical discrimination (information frictions) dominates
- This gap should increase (decrease) as we move up the SAT score distribution if statistical discrimination (information frictions) dominates
- Black workers will earn less (more) than similar white workers within major if statistical discrimination (information frictions) dominates
- Black workers should have a lower (higher) observed return to major difficulty (i.e., this gap should grow as we move up the major difficulty distribution) if statistical discrimination (information frictions) dominates
- Black workers should have a higher observed return to college grades if they face stronger information frictions about their preparation than white workers

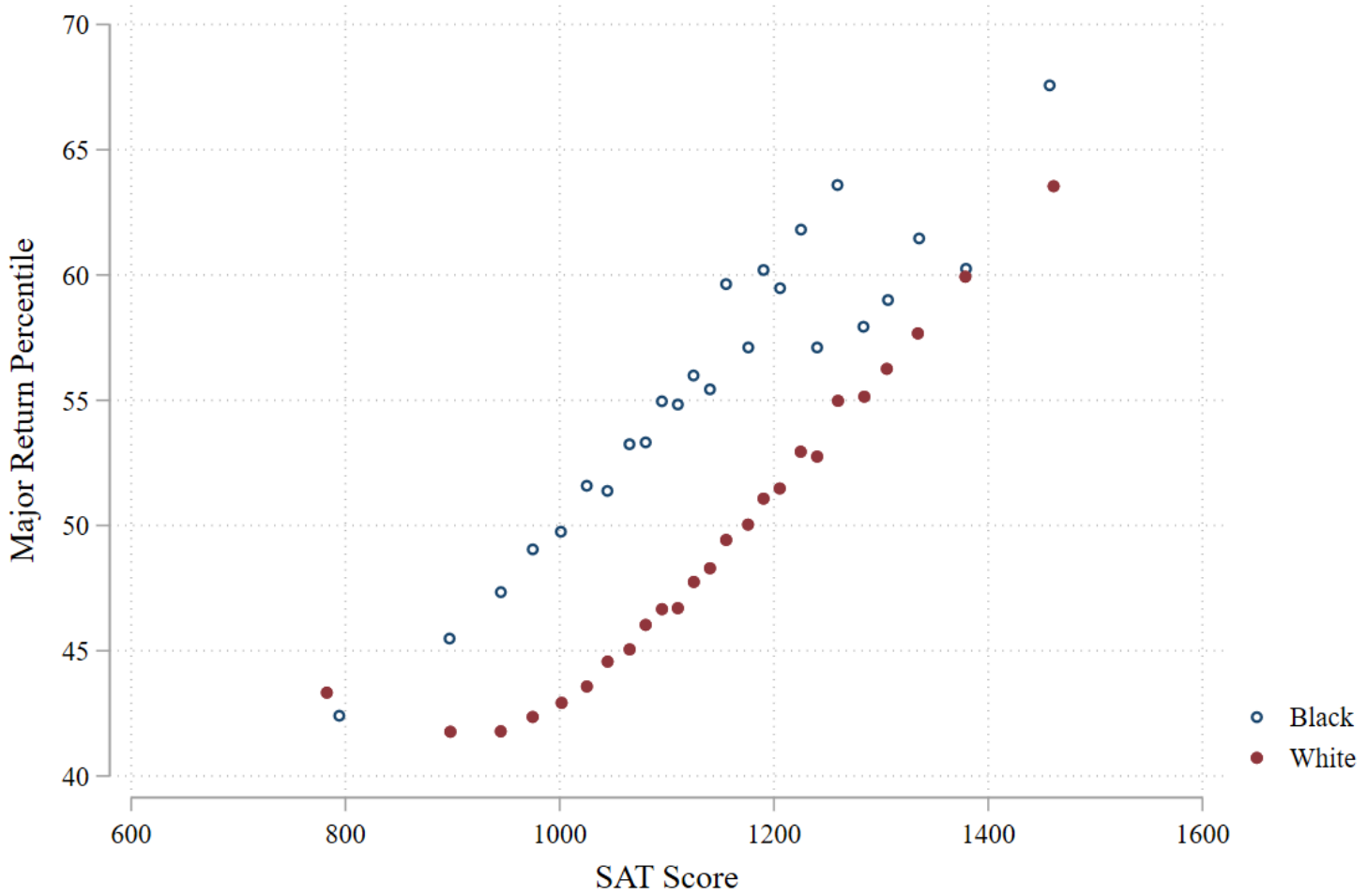
Data

- MIDFIELD State School Sample Data
  - Administrative data from 12 large public universities: Clemson, Colorado, Colorado State, Florida, Florida State, Georgia Tech, North Carolina State, North Carolina - Charlotte, Oklahoma, Purdue, Utah State, Virginia Tech
  - Include courses taken, majors, grades, test scores, GPAs for students between 1987-2018
- American Community Survey 2011-2021 (wages and college major)
- Baccalaureate and Beyond
  - Nationally representative longitudinal data of college students in 2007-2008 graduating class
  - Information on major, grades, test scores, and institution
  - Wage data for 2009, 2012, 2018

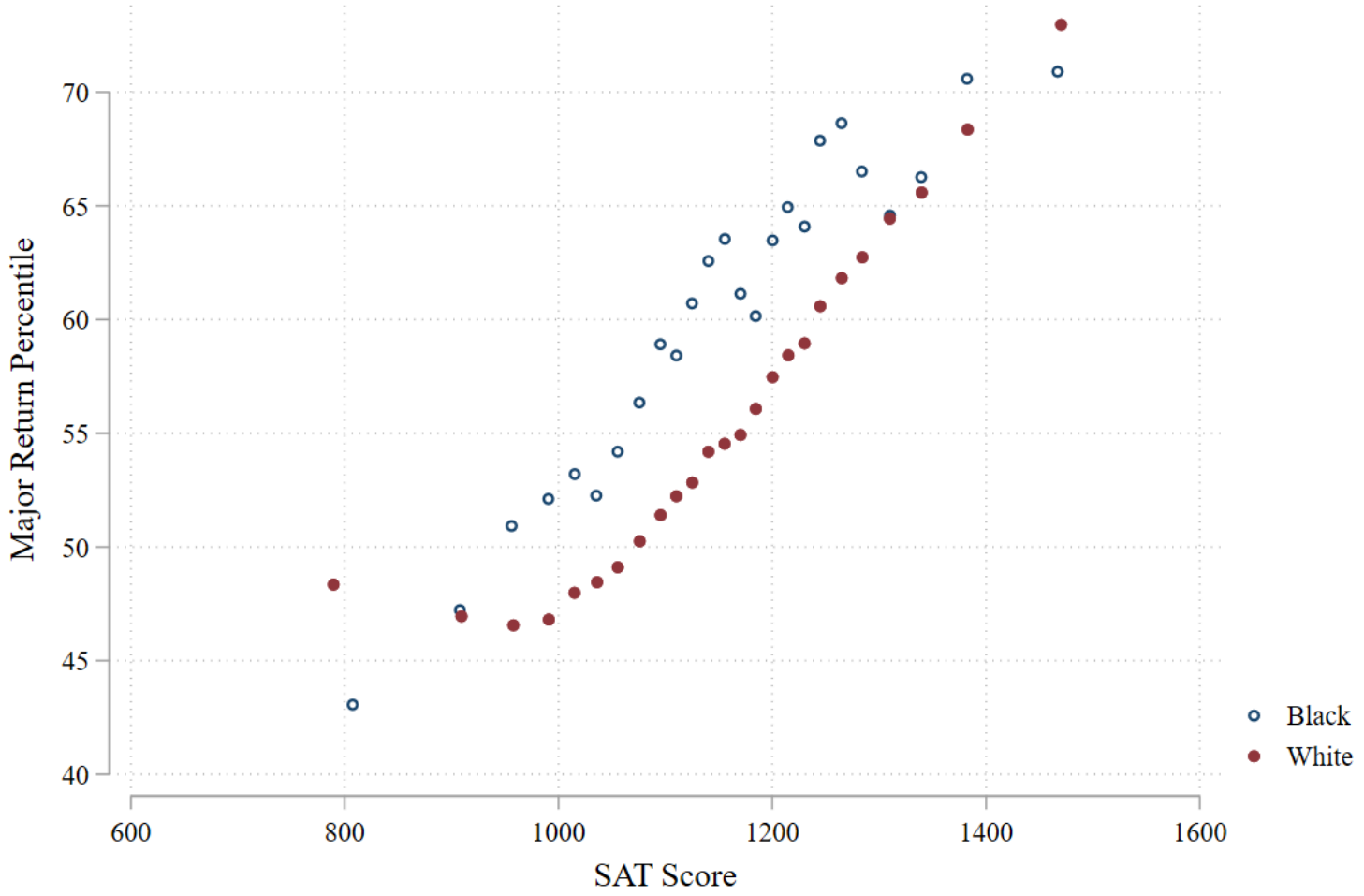
Major Difficulty

- Compute two wage-based metrics
  - Average residual from regression of log wage on age and year fixed effects for native-born full-time year-round employed 25-54 year old whites with at least a four-year college degree
  - Percentile ranking of majors from those residuals
- 5 Lowest Return: Early Childhood Education, Library Science, Studio Arts, Human Services and Community Organization, Teacher Education: Multiple Levels
- 5 Highest Return: Petroleum Engineering; Metallurgical Engineering; Mining and Mineral Engineering; Pharmacy, Pharmaceutical Sciences, and Administration; Naval Architecture and Marine Engineering
- Economics 18th, Computer Science 19th, Finance 23rd (out of 173)
- Compute one course-based metric: fraction of course credits in STEM for average graduate of each major

SAT Scores and First Major Percentile Return by Race: State School Sample



SAT Scores and Graduation Major Percentile Return by Race: State School Sample



SAT Scores and Major Percentile Return by Race: B&B Sample

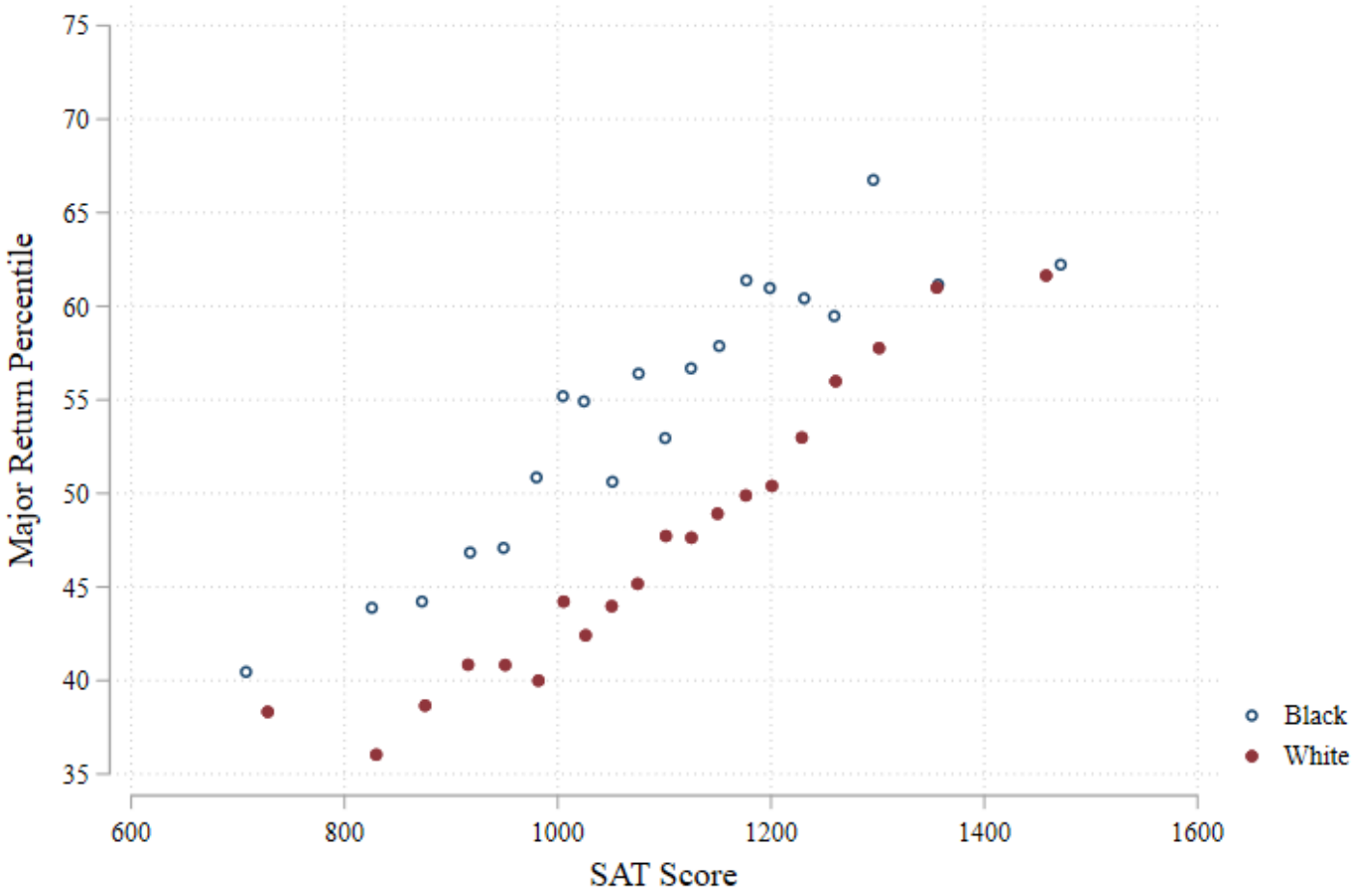


Table 1: Major Selection by Race and SAT Score

	State Schools				B&B	
	1st-Yr. Major		Grad. Major		Grad. Major	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Major Wage Return						
Black	0.032***	0.037***	0.030***	0.037***	0.053***	0.067***
	(0.002)	(0.002)	(0.003)	(0.003)	(0.006)	(0.009)
Black × SAT		0.005***		0.008***		0.008**
		(0.001)		(0.001)		(0.003)
Panel B: Major Percentile Return						
Black	0.042***	0.049***	0.037***	0.047***	0.076***	0.094***
	(0.003)	(0.003)	(0.004)	(0.004)	(0.009)	(0.012)
Black × SAT		0.007***		0.010***		0.011**
		(0.001)		(0.001)		(0.004)
Panel C: Major STEM Courses						
Black	0.029***	0.034***	0.018***	0.027***	0.042***	0.060***
	(0.002)	(0.002)	(0.003)	(0.003)	(0.008)	(0.013)
Black × SAT		0.005***		0.009***		0.010**
		(0.001)		(0.001)		(0.004)
Student Characteristics	X	X	X	X	X	X
SAT Fixed Effects	X	X	X	X	X	X
Institution x Start Year FE	X	X	X	X		
Carnegie Classificiation FE					X	X
Observations	934,448	934,448	450,987	450,987	11,530	11,530

Table 2: Adult Log Earnings by Graduation Major Selection and Race

	ACS			B&B		
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: Major = Wage Return					
Black	-0.220***	-0.229***		-0.083***	-0.074***	-0.055***
	(0.016)	(0.016)		(0.016)	(0.014)	(0.018)
Major	0.866***	0.832***	0.833***	0.594***	0.586***	0.590***
	(0.024)	(0.033)	(0.033)	(0.123)	(0.124)	(0.123)
Major × Black	-0.325***	-0.321***	-0.325***	-0.142**	-0.152**	-0.151**
	(0.052)	(0.051)	(0.052)	(0.071)	(0.072)	(0.069)
College GPA						0.069***
						(0.017)
College GPA × Black						0.105***
						(0.034)
	Panel B. Major = Percentile Return					
Black	-0.099***	-0.109***		-0.033	-0.019	-0.001
	(0.013)	(0.013)		(0.029)	(0.026)	(0.029)
Major	0.649***	0.625***	0.625***	0.424***	0.417***	0.421***
	(0.023)	(0.029)	(0.029)	(0.097)	(0.098)	(0.097)
Major × Black	-0.246***	-0.242***	-0.245***	-0.102*	-0.110**	-0.109**
	(0.038)	(0.037)	(0.038)	(0.052)	(0.053)	(0.051)
College GPA						0.069***
						(0.017)
College GPA × Black						0.104***
						(0.034)
	Panel C: Major = STEM Courses					
Black	-0.182***	-0.193***		-0.044	-0.032	-0.014
	(0.040)	(0.041)		(0.031)	(0.030)	(0.032)
Major	0.460***	0.452***	0.452***	0.380***	0.367***	0.379***
	(0.080)	(0.084)	(0.084)	(0.111)	(0.113)	(0.110)
Major × Black	-0.121**	-0.118**	-0.118**	-0.080	-0.079	-0.082
	(0.057)	(0.057)	(0.058)	(0.079)	(0.079)	(0.075)
College GPA						0.074***
						(0.017)
College GPA × Black						0.106***
						(0.035)
State FE		X		X	X	X
State × Race FE			X			
Carnegie Classification FE				X	X	X
SAT FE					X	
Observations	2,650,399	2,650,399	2,650,399	26,360	26,360	26,360

Race or SES?

- Alternative hypothesis is that results driven by low SES students having a stronger desire for monetary rewards
- Correlation between SES and race drives results
- Unlikely statistical discrimination mechanism holds for low SES white students
- Can compare effects on low SES white students to black students to test our mechanism
- While data on students own childhood SES is not available, both State School Sample and B&B data include home ZIP code
- Include ZIP code conventional SES measures, as well as intergenerational mobility statistics computed as part of Opportunity Insights (Chetty et al., 2018)

Table 3: Graduation Major Selection by Race, SAT Score, and Neighborhood Characteristics, State School Sample

	State Schools			
	(1)	(2)	(3)	(4)
Black	0.042***	0.043***	0.043***	0.042***
	(0.004)	(0.004)	(0.004)	(0.004)
Black × SAT	0.009***	0.008***	0.008***	0.007***
	(0.002)	(0.002)	(0.002)	(0.002)
Median Income (10,000s)		0.001***		
		(0.000)		
Median Income × SAT		-0.001***		
		(0.000)		
Median Education			0.002***	
			(0.001)	
Median Education × SAT			-0.001***	
			(0.000)	
Income Mobility				0.086***
				(0.018)
Income Mobility × SAT				-0.090***
				(0.013)
Student Characteristics	X	X	X	X
SAT FE	X	X	X	X
Institution x Start Year FE	X	X	X	X
Observations	311,520	311,520	311,520	311,520

Table 4: Graduation Major Selection by Race, SAT Score, and Neighborhood Characteristics, Baccalaureate and Beyond Sample

	B&B			
	(1)	(2)	(3)	(4)
Black	0.069***	0.068***	0.068***	0.069***
	(0.009)	(0.009)	(0.009)	(0.009)
Black × SAT	0.007*	0.007*	0.007*	0.007*
	(0.004)	(0.004)	(0.004)	(0.004)
Median Income (10,000s)		-0.001		
		(0.001)		
Median Income × SAT		-0.000		
		(0.000)		
Median Education			-0.003*	
			(0.001)	
Median Education × SAT			-0.000	
			(0.000)	
Income Mobility				-0.033
				(0.056)
Income Mobility × SAT				-0.005
				(0.026)
Student Characteristics	X	X	X	X
SAT FE	X	X	X	X
Carnegie Classification FE	X	X	X	X
Observations	8,500	8,500	8,500	8,500

Table 5: Log Earnings by Graduation Major Selection, Race, and Neighborhood Characteristics

	B&B				
	(1)	(2)	(3)	(4)	(5)
Black	-0.088***	-0.069***	-0.079***	-0.086***	-0.062***
	(0.022)	(0.022)	(0.022)	(0.021)	(0.019)
Major	0.626***	0.541***	0.552***	0.628***	0.562***
	(0.117)	(0.139)	(0.131)	(0.175)	(0.149)
Black × Major	-0.297***	-0.304***	-0.296***	-0.291***	-0.273***
	(0.084)	(0.080)	(0.082)	(0.079)	(0.084)
Median Income (10,000s)		0.017***			0.017***
		(0.002)			(0.002)
Median Income × Major		0.010			0.006
		(0.008)			(0.008)
Median Education			0.020***		
			(0.004)		
Median Education × Major			0.009		
			(0.010)		
Income Mobility				0.974***	
				(0.134)	
Income Mobility × Major				-0.039	
				(0.722)	
Student Characteristics	X	X	X	X	X
Year FE	X	X	X	X	X
Carnegie Classification FE	X	X	X	X	X
SAT FE					X
Observations	21,920	21,920	21,920	21,920	21,920

Other Results

- Results robust to other major measures as well as using first declared major in the State Schools sample
- No evidence of heterogeneity by age or gender
- Results robust to major difficulty measures including non-white workers in calculation, as well as including only white men
- Institution fixed effects reduce precision in B&B but have little impact on point estimates

Summary of Results

- Results strongly indicate that black students choose more difficult majors due to anticipated statistical discrimination
- This 'mismatch' is optimal behavior of students, not distorted behavior due to institutional factors
- Important implications for current methodologies that test mismatch hypothesis on university admissions

A Simple Extension

- Consider simple extension of model where black students face barriers to human capital investments, $c(m)$, that are increasing in difficulty
- Such barriers will reduce black student investment choices (potentially even beneficially)
- Policymakers concerned with equity can give black students an affirmative action subsidy $b(m)$ which will induce black students to attempt higher levels of m
- If $b(m)$ is too low, black students will "undermatch" and would see better outcomes if they attempted more difficult m
- If $b(m)$ is too high, black students will "overmatch" and lowering affirmative action subsidies will raise average black outcomes
- When $b(m) = c(m)$ black students will optimally mismatch as in our model, and a reduction in $b(m)$ is arguably beneficial

Some Definitions

- Weak mismatch: Lowering $b(m)$ on the margin would improve black outcomes
- Strong mismatch: $b(m) = 0$ (i.e., abolishing racial admissions preferences) would lead to better black outcomes than current $b(m)$
- Strong mismatch implies weak mismatch but not vice versa

Pseudo-Random Assignment

- Consider a natural experiment which leads to a small number of black students to randomly attend a more difficult institution (e.g., a RD around an admissions cutoff)
- If information is incomplete, these students will be paid a higher wage than those who attend a less difficult institution *even if* they are less productive
- Thus cannot reject weak or strong mismatch
- Signaling value of institution will lead to sharp increase in wages at discontinuity
- If information is complete (older workers) this provides test of weak mismatch because it compares outcomes from marginally changed students whose matriculation decisions depend on $b(m)$

Affirmative Action Ban

- Several states have banned affirmative action in admissions
- Frequent empirical strategy is to compare black outcomes before and after affirmative action ban
- Whether ban harms minority students seems to depend on state and minority group studied
- This provides a test of strong mismatch regardless of whether information is complete or incomplete, but cannot reject weak mismatch
- Natural that results could vary dramatically across studies, since each study is comparing a different level of affirmative action subsidies (state policy differences towards different classes) to same $b(m) = 0$ treatment

Summary

- Developed a new model of human capital investment when students have incomplete information about their aptitude and anticipate statistical discrimination
- Two different information channels have opposing effects on major selection
- Empirical evidence suggests statistical discrimination is stronger than student information frictions
- Find that black students enroll in more difficult majors, and have a lower return to majors in the labor market
- Because behavior is optimal, in equilibrium moving black students to "better matched" investments is harmful
- Researchers must think carefully about policy question of interest and level of information possessed by employers when evaluating empirical studies of racial admissions preferences