Statistical Discrimination and Optimal Mismatch in College Major Selection

Mary Kate Batistich
University of Notre Dame

Timothy N. Bond
Purdue University and IZA

Sebastian Linde
Medical College of Wisconsin

Kevin J. Mumford
Purdue University

May 27, 2022

Abstract

We develop a model of college major selection in a labor market with statistical discrimination and learning. Heterogeneous students choose among college majors that differ in their return to ability. Employers do not initially observe productivity but do observe student major and a signal of output, where informativeness of the signal differs by race. Our model predicts black students will choose higher return majors than white students conditional on their college preparation, but receive lower equilibrium labor market returns to major. We find empirical support for our learning model using administrative data from several large universities and wages from two nationally representative surveys.

JEL Codes: J71, J15, I26

We are grateful to conference participants at the Mid-Midwest Applied Microeconomics Workshop for their helpful comments. All remaining errors are our own.

1 Introduction

The well-documented disparities in preparation between black and white students in universities have led several scholars to argue that black students are "mismatched" (e.g., Sander, 2004; Arcidiacono and Lovenheim, 2016). That is, black students select human capital investments that are too difficult given their aptitude, and would have better labor market outcomes had they chosen less difficult investments. Such an argument requires that black students make systematic mistakes, perhaps due to biased information. Yet, Lang and Manove (2011) show that the presence of statistical discrimination by employers can also distort human capital investments, leading black workers to optimally "overinvest" in education when education is easy for employers to observe.

We apply this insight to a model of college major selection. Students choose a major which augments their initial stock of human capital. Condition on this, the education production function is single-peaked so that students who choose a major more difficult than optimal will be less productive at graduation than had they made the optimal choice. Employers do not observe the productivity of new labor market entrants, but do observe major choice and an unbiased signal of productivity. As is standard in the statistical discrimination literature, this signal is less precise for black students than for white. In equilibrium, black students optimally "mismatch"; they choose majors that are more difficult than similarly prepared white students, and graduate less productive than they would be had they chosen a less difficult major.

Our model generates several novel empirical predictions. First, black students enroll in majors with a higher labor market return than white students with similar academic backgrounds. Second, as an outcome of the signaling game black students have a lower observed return to college majors. Third, due to employer learning black workers should experience a wage return to unobserved productivity that is increasing at a greater rate than for white workers.

We find support for these predictions using administrative data from twelve large public universities, nationally representative data from the American Community Survey (ACS), and longitudinal data from the Baccalaureate and Beyond (B&B). Conditional on SAT and ACT scores, black students choose majors which have higher average wages for mid-career white men, and graduate in these majors at higher rates. This result holds across several institutions with varying admission standards. We find that the racial wage gap is larger for those who graduated in higher-return majors in both the ACS and B&B. Finally, the measured return to parental income, a variable that is plausibly unobservable to employers and correlated with productivity, increases for black workers relative to whites in labor market experience.

While we focus on major choice rather than college choice, our results can help explain seemingly contrasting findings in the admissions affirmative action literature. It is widely assumed by researchers that school quality and student aptitude are complements (e.g., Sallee et al., 2008; Arcidiacono et al., 2011). Dillon and Smith (2020) find evidence for such complementarities when considering four-year graduation rates and long-term earnings. Yet, Mountjoy and Hickman (2021) find little racial differences in the marginal returns to attending selective schools in Texas, despite the large preparation advantage held by the white marginal admit.[1] In our framework, the signaling value of each university would differ across race. All black students would attend a university which is more demanding than optimal from a human capital perspective, but in equilibrium the average net human capital that black students graduate with, and thus wages, grows in university selectivity.

Despite the dramatic differences in the labor market returns across college majors, racial differences in major selection has received surprisingly little attention.[2] Arcidiacono et al. (2012) show that black students at Duke University are more likely to begin schooling in a science major than whites students, but have lower rates of finishing a major in science. Arcidiacono et al. (2016) similarly find substantial gaps in preparation between minority students who finish a STEM degree and those who do not within the University of California system. Using data from the University of California, Los Angeles, Sovero et al. (2021) show minority students actually have higher rates of STEM persistence after controlling for preparation. Hill (2017) finds that statewide affirmative action bans reduce the number of minorities who graduate with STEM degrees. Using a similar measure to the one we implement, Bleemer and Mehta (2021) document that the raw gap between white and minority students in the selection of high return majors grew during the 1990s. While we do not study trends in major choice across time, we show that this gap is reversed once controlling for college preparation.

Our study differs from these papers in several important ways. First, rather than attempting to quantify the impact of affirmative action, we take the distribution of students at a university as given and analyze the effect of labor market statistical discrimination on major choice. Second, our theory yields racial differences in major choice even in the absence of affirmative action, and in an environment where students and universities have symmetric information on aptitude. Third, we document racial differences across a larger and more diverse set of universities, and for a fuller set of majors, than is typical for this literature. Fi-

[1] See also Angrist et al. (2019) who do not find evidence for mismatch effects for high school students at Chicago exam schools.

[2] See Altonji et al. (2016) for a recent review on the returns across college majors.

nally, our model generates predictions on early career outcomes and labor market dynamics, which we confirm using nationally representative data on wages for recent college graduates.

Our empirical results contribute to the growing body of evidence that student major selection responds to labor market incentives. Previous studies have found that students switched majors in response to cyclical fluctuations in energy prices, the dot-com bust, the fracking boom, and the 2007-2008 financial crisis (Ersoy, 2020; Han and Winters, 2020; Weinstein, forthcoming). Similarly, Aalto et al. (2022) find the COVID-19 pandemic caused a decrease in applications to hospitality vocational programs by high school students in Sweden, while Ganguli et al. (2022) find the pandemic increased the demand for online courses promoting telework skills in Saudi Arabia. Blom et al. (2021) find that students enroll in majors with better labor market prospects during recessions. Bičáková et al. (2021) find evidence consistent with students exerting higher effort in college when they enroll during worse economic times.

The rest of the paper is organized as follows. In section 2, we introduce our model where students select a college major taking into account the statistical discrimination behavior of future employers. In section 3, we test our model's predictions on major selection. We test our model's predictions on labor market outcomes in section 4. Section 5 concludes.

2 A Model Of College Major Selection with Statistical Discrimination

We develop a three period model similar in spirit to Lang and Manove (2011), but extended to allow for learning. There exists a large number of students who are either (b)lack or (w)hite. They differ in their initial stock of human capital. We partition this stock into two components. a_i, the portion of a human capital which is complementary with college major choice, is bounded over $[a_L, a_H]$. ζ_i, a permanent component of human capital which is not augmented by university education, is independent of a_i and distributed normally with mean 0 and standard deviation σ_ζ. For example a_i could be cognitive skills and ζ_i could be interpersonal skills. Alternatively a_i could represent general knowledge learnable in school while ζ_i is learning through non-schooling experiences. As we discuss below, optimal major choice will involve a one-to-one mapping of a_i to majors conditional on race. Thus ζ_i is necessary for there to be uncertainty about productivity about which employers can learn.[3]

[3]All of our arguments should follow if instead there was some factor, such as heterogeneous major costs, which generated uncertainty in ability conditional on major choice. However, such an approach would be much less tractable. In particular, we would need to define a cost function so that the distribution of a_i conditional on major choice is normal, so that we can apply the usual formulas for Bayesian updating. It is

In period 1, students select from a continuum of college majors m which differ in their human capital production function. A student who selects major m will produce p_i when they enter the labor market, where

$$\log p_i = f(a_i, m_i) + \zeta_i. \tag{1}$$

$f(a, m)$ is the major-specific human capital production function. It is strictly increasing in a. Further, m is indexed by its complementarity with a; $f(a, m)$ is single-peaked in m, with $\arg\max_m f(a, m)$ increasing in a. Finally, denoting $F(a) \equiv \max_m f(a, m)$, $\frac{\partial F}{\partial a} > 0$, so that we would expect the highest m majors to also be the highest paid majors in the labor market.

In period 2, students enter the labor market. Employers do not observe a p, but do observe m. The market also observes a signal s which is an unbiased measures of a student's productivity:

$$s_i = \log p_i + \epsilon_i, \tag{2}$$

where ϵ_i is normally distributed with mean 0 and standard deviation σ_k^2, and $k \in \{b, w\}$ is a student's race. This reflects information this is learned, for example, from an interview. Following the tradition in the statistical discrimination employers are better able to interpret this information for whites, so that $\sigma_w^2 < \sigma_b^2$.

In the final period, p is revealed to the market. All agents are risk neutral and do not discount future earnings. The labor market is perfectly competitive. Firms use all available information when making wage offers. Students choose majors which maximize their expected utility.

Denote π_k as the race-specific employer belief function, w_k as the race-specific wage function, M_k as the race-specific function which maps from ability to college major.

Definition. *An equilibrium is a set of functions π_k, w_k, and M_k such that*

1. *w_k generates zero expected profit for firms given π.*

2. *M_k maximizes expected utility given w_k.*

3. *π_k is defined by Bayes' rule whenever possible.*

As in Lang and Manove (2011) we will restrict attention to separating equilibria which are "well-behaved."

difficult to imagine what that cost function would look like with a_i being bounded.

Definition. *A well-behaved equilibrium is an equilibrium with the following properties:*

1. *M_k is smooth, continuous, differentiable, and monotonically increasing in a*

2. *For any m which is not utilized by any individuals of race k in equilibrium, $\pi_k = a_L$.*

We will propose the existence of a well-behaved equilibrium and analyze its properties. We will then prove its existence.

2.1 Employer Beliefs and Wages

In period 3, p is revealed. As the market is competitive, this is the period 3 wage for all workers.

For period 2, note that in a well-behaved equilibrium, college major selection reveals a student's a to the market. Denote $A_k(m)$ as the inverse of $M_k(a)$. The distribution of productivity for students of race k with major m is normally distributed with mean $f(A_k(m), m)$ and standard deviation σ_ζ^2. As s is normally distributed, we can apply Bayes' rule to find period 2 employer beliefs for all m that are used in equilibrium,

$$\pi_k(m, g, s) = \tau_k^{-1} \left[\sigma_\zeta^{-2} f(A_k(m), m) + \sigma_k^{-2} s \right], \tag{3}$$

where $\tau_k \equiv \sigma_\zeta^{-2} + \sigma_k^{-2}$ is the precision of the market's posterior beliefs for a worker of race k with major m. Equation 3 is simply an average of the prior and signal, weighted by each's relative precision. It then follows from the zero profit condition that period 2 wages are simply

$$w_k(m, s) = \pi_k(m, s). \tag{4}$$

2.2 Optimal Major Selection

Now, consider a student with initial human capital a and ζ who is choosing her major to maximize her expected utility. The student's optimization problem is,

$$\max_m E_k(w|m, a, \zeta) + f(a, m) + \zeta, \tag{5}$$

where the objective function is simply the expected sum of the period 2 and period 3 wages. $E_k(w|m, a, \zeta)$ is the expected period 2 wage for a student of race k with initial human capital parameters α and ζ who attempts major m,

$$E_k(w|m, a) = \tau_k^{-1}(m) \left[\sigma_\zeta^{-2} f(A_k(m), m) + \sigma_k^{-2} \left(f(a, m) + \zeta \right) \right]. \tag{6}$$

This follows from taking the expectation of (4), recognizing that s is equal to $f(a,m)+\zeta$ in expectation. The expected wage is a weighted average of the market's beliefs about a student with major m and the student's actual productivity, with more weight being placed on the market's beliefs when the signal has more variance. In other words, in choosing a higher return major students gain benefits from a "sheepskin" effect $f(A_k(m),m)$, but beyond a certain point, students bear a cost of lower actual human capital from being in a major that is more difficult than optimal for their a.

Proposition 1. *Denote $M^*(a)$ as $\arg\max_m f(a,m)$. In any well-behaved equilibrium, $M_k(a_L) = M^*(a_L)$, and $M_k(a') \geq M^*(a') \forall a' > a_L$*

Proof. A similar proof is provided in Lang and Manove (2011). For the first part of the proposition, suppose $M_k(a_L) < M^*(a_L)$. As $A_k(m)$ is monotonic in m, an increase in m will raise $f(A_k(m),m)$ and $f(a,m)$. Thus increasing m is strictly preferred.

Now, suppose that $M_k(a_L) > M^*(a_L)$. By definition, $M_k(a_L)$ can only provide higher expected utility than $M^*(a_L)$ if $f(A_k(M_k(a_L)), M_k(a_L)) > f(A_k(M^*(a_L)), M^*(a_L))$. But in equilibrium, beliefs must be correct, so $f(A_k(M_k(a_L), M_k(a_L)) = f(a_L, M_k(a_L))$. Since $\frac{\partial f(a,m)}{\partial m} < 0$ when $m \geq M^*(a)$ and in any well-behaved equilibrium employers believe that all students who choose $m < M_k(a_L)$ have ability a_L, students could deviate to $M^*(a_L)$ and increase their expected utility.

For the second part of the proposition, suppose $M_k(a') < M^*(a')$. As $A_k(m)$ is monotonic in m, an increase in m will raise $f(A_k(m),m)$ and $f(a,m)$. Thus increasing m is strictly preferred. □

Proposition 1 follows from employer belief structures in well-behaved equilibria. The lowest ability individuals do not receive a benefit from choosing a higher m than the full-information optimum because they receive no sheepskin effect. In equilibrium, employers believe the least competitive major that is utilized must be the lowest type, and therefore the lowest type will want to choose their full-information optimal major.

Proposition 2. *In equilibrium, $M_k(a, w_k)$ can be characterized by the differential equation*

$$\frac{\partial M_k(a,w_k)}{\partial a} = -\tau_k^{-1}\left[\sigma_\zeta^2 \frac{\partial f(a,m)}{\partial a}\right]^{-1}\left[\frac{2\partial f(a,m)}{\partial m}\right]^{-1}.$$

Proof. By applying the chain rule to (5) and recognizing that in equilibrium $f(A_k(m),m) = f(a,m)$, we arrive at a first order condition of

$$2\frac{\partial f(a,m)}{\partial m} + \tau_k^{-1}\left[\sigma_\zeta^{-2}\frac{\partial f(a,m)}{\partial a}\frac{\partial A_k(m)}{\partial m}\right] = 0.$$

Note that since $A_k(m) = M_k^{-1}(a)$, $\frac{\partial A_k(m)}{\partial m} = \frac{\partial M_k(a)}{\partial a}^{-1}$. Rearranging terms then proves the proposition. □

Proposition 3. *In equilibrium, black students attempt majors with a higher return than white students conditional on a.*

Proof. The proof follows from inspection of $\frac{\partial M_k(a,w_k)}{\partial a}$. Note that $\tau_w \geq \tau_b$. At a_L, we know from $M_b(a_L) = M_w(a_L)$ from Proposition 1. Therefore $\frac{\partial M_b(a_L,w_b)}{\partial a} > \frac{\partial M_w(a_L,w_w)}{\partial a}$ and an ε increase in a will lead to $M_b(a+\zeta) > M_w(a+\zeta)$.

Now suppose that there was some $a > a_L$ for which $M_w(a') \geq M_b(a')$. Since M is continuous, it then must be the case that in some ball around a' there is an $a'' < a'$ for which $\frac{\partial M_w(a,w_w)}{\partial a} > \frac{\partial M_b(a,w_b)}{\partial a}$. But as the major choices approach equality, $\frac{\partial M_b(a,w_b)}{\partial a} > \frac{\partial M_w(a,w_w)}{\partial a}$ which is a contradiction. □

In equilibrium black students take majors with a higher return than white students. As all students choose m higher than the full-information optimum, this means that black students are more "mismatched." The phenomena is driven by statistical discrimination in the labor market. Black students have a higher marginal return to observable information during period 2 than white students, which gives them larger incentives to increase their academic credentials by investing in higher return majors.

Proposition 4. *In equilibrium, black college graduates have lower productivity than white students with the same college major.*

Proof. This follows from Propositions 3. Since black students have lower a within major they have lower p when entering the labor market. □

Proposition 4 follows from Proposition 3. Since black students are choosing more difficult majors, and human capital decreases on the margin when students attempt a more challenging major than optimal, black students will accumulate less human capital than white students with the same major.

2.3 Labor Market Outcomes

Our model predicts that black students will ceteris paribus choose relatively higher return majors than white students due to labor market statistical discrimination. We now turn analyze the impact of this on labor market outcomes.

Proposition 5. *Black students earn lower wages than white students conditional on m.*

Proof. From Proposition 4 we know that black students will earn lower human capital than white students conditional on m. From equation (6) we can see that wages are simply employer's beliefs about student's human capital. $\square$

Proposition 5 follows directly from Proposition 1. Since black students graduates are less productive than whites with the same major, they earn lower wages.

Proposition 6. *The observed labor market return to m with respect for black college graduates is lower than for white college graduates.*

Proof. A similar proof is provided in Lang and Manove (2011). First note that $M_w * (a) = M_b^*(a) \forall a$. Thus by Proposition 1, $M_b(a_L) = M_w(a_L) \equiv M^*(a_L)$. Now consider the equilibrium observed return to human capital from major $m' > M^*(a_L)$,

$$\frac{f\left(A_k(m'), m'\right) - f\left(a_L, M^*(a_L)\right)}{m' - M^*(a_L)}.$$

It follows from Proposition 4 that $f(A_w(m'), m') > f(A_b(m'), m')$. This holds for any $m' > M^*(a_L)$.

$\square$

Under statistical discrimination the observed return to m will be lower for black students than white students because the equilibrium of the signaling game is non-distortionary for the lowest abilities and majors, but induces a racial productivity gap for higher m.

2.4 Learning and Wage Dynamics

Now, suppose we have access to some variable z_i that is unobservable to employers and that

$$z_i = \log p_i + \nu_i, \tag{7}$$

where ν_i is distributed mean 0 with standard deviation σ_z^2 and is orthogonal to ϵ_i. Further, suppose we can observe $f(A_k(m), m)$, that is the average productivity of workers of race k with major m in equilibrium.

Proposition 7. *Conditional on $f(A_k(m), m)$, black workers have a lower labor market return to z than white workers in period 2. Conditional on z, black workers have a higher labor market return to $f(A_k(m), m)$.*

Proof. Applying Bayes' rule, the expected productivity of a worker given all information observed by the econometrician is

$$E_k\left[p_i|m,z\right]=\gamma^{-1}\left[\sigma_{\zeta}^{-2}f\left(A_k\left(m\right),m\right)+\sigma_z^{-2}z\right], \tag{8}$$

where $\gamma\equiv\sigma_{\zeta}^{-2}+\sigma_z^{-2}$. The expected wage is then

$$\begin{aligned}E\left[w|m,z\right]&=E\left[\pi_k\left(m,s\right)|m,z\right]\\&=\tau_k^{-1}\left[\left(1+\gamma^{-1}\sigma_k^{-2}\right)\sigma_{\zeta}^{-2}f\left(A_k\left(m\right),m\right)+\gamma^{-1}\sigma_k^{-2}\sigma_z^{-2}z\right],\end{aligned} \tag{9}$$

where the second expression follows since $E[s_i|m,z]=E[p_i|m,z]$.

Taking the derivative of (9) with respect to z,

$$\frac{\partial E\left[w|m,z\right]}{\partial z}=\tau_k^{-1}\gamma^{-1}\sigma_k^{-2}\sigma_z^{-2}z, \tag{10}$$

which is decreasing in σ_k^2 and thus lower for blacks than whites.

Taking the derivative of (9) with respect to $f(A_k(m),m)$,

$$\frac{\partial E\left[w|m,z\right]}{\partial f(A_k(m),m)}=\tau_k^{-1}\sigma_{\zeta}^{-2}\left(1+\gamma^{-1}\sigma_k^{-2}\right), \tag{11}$$

which is increasing in σ_k^2. □

Proposition 7 follows standard results from the statistical discrimination and learning literature. Because black workers have noisier signals than white workers, employers initial place more weight on observable information in evaluating their expected productivity.

Proposition 8. *Conditional on $f(A_k(m),m)$, the return to z for black workers will increase relative to whites with experience. Conditional on z, the observed return to $f(A_k(m),m)$ for black workers will decrease relative to whites with experience.*

Proof. Taking the difference between (8) and (9) yields the change in wages with experience:

$$\sigma_{\zeta}^{-2}\left[\gamma^{-1}-\tau_k^{-1}\left(1+\gamma^{-1}\sigma_k^{-2}\right)\right]f(A_k(m),m)+\gamma^{-1}\sigma_z^{-2}\left[1-\tau_k^{-1}\sigma_k^{-2}\right]z \tag{12}$$

It is straightforward to see that the derivative of this expression with respect to z is increasing in σ_k, and the derivative with respect to $f(A_k(m),m)$ is decreasing in σ_k. □

In period 3, employers learn the productivity of their workers. This increases the correlation between z and wages, since z is a measure of productivity that employers cannot observe. Since z is correlated with s, and employers place more weight on s in period 2 for whites, the increase in the observed return to z is larger for black workers. Likewise since $f(A_k(m), m)$ is correlated with s and employers put more weight on $f(A_k(m), m)$ for black workers in period 2, the decrease in the observed return to $f(A_k(m), m)$ is larger for blacks.

Note that at is important for these predictions that we observe $f(A_k(m), m)$ rather than simply m. This is because we know from Proposition 4 that the market has different beliefs on the average productivity for black workers with a given m than for white. Fortunately, $f(A_k(m), m)$ is readily observable in data; it corresponds to the race-specific average wage for individuals with major m.

3 Testing for Mismatch in Major Selection

We use administrative data from several large public universities which we refer to as our State School Sample, data from the American Community Survey (ACS) on college major and labor market outcomes, as well as longitudinal data from the Baccalaureate and Beyond (B&B) to empirically test our model's main prediction. In this section we focus on the model predictions related to major selection and find that black students select majors with a higher return than observationally equivalent white students.

3.1 Major Return Measure

As described in our theoretical model of major selection, our empirical analysis requires a measure of the average labor market return for each major. To do this, we draw on the American Community Survey (ACS). The ACS is an annual survey of people in the United States conducted by the U.S. Census Bureau. Importantly for our purposes, the ACS includes information on field of degree (the college major) aggregated to 173 categories for all individuals who hold a bachelor's degree or above.

To calculate our measure of major return, we compute the residuals from a regression on 35-45 year old native-born full-time year-round employed workers with exactly a bachelor's degree log annual wages with age and year fixed effects. This approach to measuring major return is similar to Bleemer and Mehta (2021), though we perform this regression separately for whites and blacks. The estimated black and white return for each major is reported in Table A1 along with the percentile return for white graduates. We refer to the average residual by major from this regression as the "Wage Return", which will mirror $f(A_k(m), m)$ in our model. We construct an alternative measure which is the percentile ranking of these residuals from the white regression that we refer to as "Percentile Return". We view this as a measure of m which is not directly related to $f(A_k(m), m)$.

3.2 State School Sample

We use administrative student data from 12 large public universities which we call the State School Sample. These data were obtained from school registrars through the MIDFIELD partnership. Institutions that participate in the MIDFIELD partnership share de-identified longitudinal student records for all degree-seeking undergraduate students. The data do not include job placement or any post-graduation information, but do include demographic characteristics and admissions data as well as course grades, major, and degree earned. The universities included in our sample are Clemson, Colorado, Colorado State, Florida, Florida State, Georgia Tech, North Carolina State, North Carolina – Charlotte, Oklahoma, Purdue, Utah State, and Virginia Tech.[4] While these universities are not nationally representative, Denning et al. (forthcoming) show that these students are quite similar to those from the nationally-representative NELS:88 and ELS:2002 sample of top-50 public universities in race, gender, and the distribution of SAT scores.

Table 1 reports summary statistics for the State School Sample. The primary advantage of these data is the large sample size, nearly 600,000 white or black students who entered college at one of these institutions between 1987 and 2018.[5] Black students have lower graduation rates, are disproportionally female, and are less likely to be transfer students. Strikingly, despite having on average 56 point lower SAT math scores and 50 point lower SAT verbal scores, black students enroll in majors that have a higher return on each of our measures. We also see black students are more likely to be enrolled in engineering, information technology, chemistry, or biology, while less likely to be enrolled in history, English, or agriculture.

3.3 Major Selection, Academic Preparation, and Educational Outcomes

For students with equivalent academic preparation, our model predicts that black students will select majors with higher average returns than white students. To show this in the raw data, we first plot the relationship between SAT scores (in 45 equal sized bins) and the white percentile return in Figure 1. We use the major return percentile for white workers rather than a race-specific measure since we want m to be measured the same across race.

[4]The MIDFIELD partnership does not allow us to report any results separately by institution that would enable readers to identify the institution.

[5]Students without a reported SAT or ACT test score are excluded from the analysis.

Consistent with our model, there is little difference in major selection by race by the worst performers on the SAT. However, as we move up the SAT distribution blacks students rapidly overtake white students in percentile return, and the gap appears to increase in SAT score. There is possibly some convergence at the top of the SAT distribution but we caution that there are fewer black students in these upper SAT bins.

To formally test our prediction, we estimate:

$$\text{Ret}_{ijt} = \alpha_1 \text{Black}_i + X_i \beta_r + \gamma_t + \delta_j + \epsilon_{ijt}. \tag{13}$$

The subscript i indicates the individual, j the educational institution, r is the individual's race, and t the year of enrollment. Black_i is an indicator for being black. X_i is a set of individual characteristics. γ_t is a vector of enrollment fixed effects and δ_j is a vector of institution fixed effects. Ret_{ijt} is our measure of major return.

The first two columns of Table 2 use our wage return measure. In addition to the fixed effects discussed above, column (1) includes controls for gender, transfer status, home zip code fixed effects (a proxy for socioeconomic status), high school GPA fixed effects, and SAT fixed effects. Consistent with with Table 1, we find that black students enroll in majors with a 3 log point higher residual wage than white students with the same SAT score. In column (2) we replace our SAT fixed effects with a linear variable in SAT and an interaction between SAT and race. Again as predicted by our model, the difference in major selection appears to grow in college preparation.

In columns (3) and (4) we repeat this analysis using instead the major percentile return for whites. We again see evidence supportive of our model. Black students enroll in a major 3.5 percentiles higher on average than whites with the same SAT score, and this difference is increasing in the SAT score.

4 Testing for Labor Market Impacts

Our model generates a stark prediction on the equilibrium labor market impacts of statistical discrimination. We use data from the ACS and from the B&B to empirically test these predictions and find that the empirical results support the predictions from our model, that black college graduates have lower wages than white graduates from the same major and that the return to a higher value major is lower for black college graduates than for white college graduates.

4.1 American Community Survey Data

For our analysis, we restrict our ACS sample to prime age (25 to 54) native non-Hispanic white and black workers who were full-time year round employed in the previous year. We display summary statistics for our ACS sample in Table 3. The black workforce is more female, in line with well-known racial differences in labor force participation (Neal, 2004). There is a substantial racial wage gap of 0.26 log points in the ACS sample. Note that there is little racial difference in our major return measure, though this is not in conflict with our theoretical results. Our model predicts that black students take higher return majors than white students conditional on their preparation and aptitude and we confirm this in our findings in Section 3. Given the large racial differences in preparation and aptitude, we would not necessarily expect to find black students enrolling in higher return majors on average in a nationally representative sample.

4.2 The Baccalaureate and Beyond Data

The biggest weakness for the ACS sample is a lack of controls for college quality. One concern then is that any racial differences we find in the labor market returns to major choice would be due to differences in university enrollment patterns between black and white students.[6] We therefore provide additional evidence from the Baccalaureate and Beyond 2008/18 (B&B). The B&B is a longitudinal study of 2007-2008 college graduates collected by the National Center for Education Statistics and designed to be nationally representative. Demographic characteristics, college admissions measures, detailed financial aid information, and college academic records are combined with follow-up surveys focused on employment, post-baccalaureate education, and other outcomes. The first follow-up was conducted in 2009, one year after graduation; the second follow-up was conducted in 2012, four years after graduation; and the third and final follow-up was conducted in 2018, ten years after graduation.

We display descriptive statistics for our B&B sample in Table 4. Similar to what we observe in the State School Sample data, black students are more likely to be female, and graduate with a lower GPA than white students. There is a greater than three point racial gap in average ACT scores. In raw terms, the the racial wage gap in each year is much

[6] Note that our results would not be biased if black students attended worse colleges than white students, but black major selection was uncorrelated with college quality, as the college quality effect would load onto the black indicator. Instead, we are concerned that black students may be more or less likely to enroll in, for example, STEM majors, when admitted to a selective institution. A central concern of the affirmative action and mismatch literature is that affirmative action in admissions leads black students to graduate in lower return majors than they would have had they attended a less selective college (e.g., Arcidiacono et al., 2016).

smaller than the unconditional racial gap in the United States. This reflects both the youth of the sample, as well as the fact that the racial gap is generally thought to be lower in more educated individuals (Lang and Lehmann, 2012). On both our metrics, black students are enrolled in majors with higher labor market returns than whites.

The major selection results reported in the previous section also hold in the B&B sample. Figure 2 reports the raw relationship between SAT score (in 25 equal sized bins) and the white percentile return in the B&B data. The relationship is similar to that found in the State School Sample in that black students select majors with an average percentile return that is higher than that of white students across most of the SAT distribution. Figure 2 for the B&B sample is more noisy than Figure 1 for the State School Sample, due to the small B&B sample size. However, the black-white difference in major selection does appear to be growing in SAT score with little difference in major selection for those with the lowest SAT scores. Table 5 reports regression results on major selection for the B&B sample and the results are similar to those found in Table 2 for the State School Sample. The first two columns use the wage return measure and find that black students enroll in majors with a more than 2 log point higher residual wage than white students with the same SAT score. The second two columns use the percentile return measure and show that black student enroll in majors that are 3 to 4 percentiles higher on average. The coefficient estimates for the Black-by-SAT terms have the predicted sign and are similar in magnitude to those found in Table 2, though not statistically significant.

4.3 Major Selection and Career Outcomes

Our model predicts the black students will have a lower observed return to major than whites. To test this, we estimate:

$$Y_{ijrst} = \alpha_r \text{Ret}_j + X_{ir}\beta + \gamma_{rs} + \delta_{rt} + \epsilon_{ijrst}. \tag{14}$$

The subscript i is for the individual, j is for the major, r indicates race, s indicates state of residence, and t indicates time. X_{ir} is a set of individual controls. γ_{rs} is a set of possibly race-specific state dummies. δ_{rt} is a set of possibly race-specific time dummies. Our model predicts that $\alpha_b < \alpha_w$; i.e., whites have a larger return to major difficulty than blacks.

In column (1) of Table 6 we estimate equation (14) using our ACS data. We cluster our standard errors at the field of degree level. With only our basic set of controls (gender, age, and age-squared) we find strong evidence for our model. Blacks have an observed return to major nearly 50% lower than whites. This result is unchanged with the addition of state and year fixed effects in column (2) and race-specific state and year fixed effects in column

(3). Columns (4)-(6) repeat the analysis using the percentile return instead as our measure of major. We find similar results.

In Table 7 we turn to the B&B data. These data allow us to use institution fixed effects rather than state fixed effects, to better account for quality of education.[7] The cost is a much smaller sample which is limited to early career outcomes. Column (1) estimates our modified version of equation (14) using our wage return measure calculated for whites. We find very similar results to that in the ACS. Recent black college graduates have an observed return to major roughly one-third as large as whites. Column (2) includes a control for college GPA and our results are essentially unchanged. In columns (3) and (4) we repeat this analysis using our percentile return measure for whites and find similar effects.

The summary statistics in Table 4 show that black students are more likely to come from a household where neither parent graduated from college, commonly called a first generation college student. In a robustness exercise reported in Table A2 we show that the results in Table 7 are not being driven by first generation students. First generation students have a lower wage return to major than other students, but the magnitude of this difference small. The estimated black return to major reported in columns (2) and (4) of Table A2 are very similar to those reported in Table 7.

4.4 Learning and Wage Dynamics

To test the dynamic predictions of our learning model, we adopt a similar specification to Altonji and Pierret (2001):

$$Y_{ijrut} = \alpha_{rt}\text{Ret}_{jr} + \delta_{rt} z_i + X_{ir}\beta + \gamma_{rs} + \delta_{rt} + \epsilon_{ijrst}, \tag{15}$$

where the subscript u now refers to the university of graduation for worker i and the other variables and subscripts are defined as before. Note the subscripts on Ret_{jr} indicating we now use our race-specific measure of major return. This follows from our theory as the analogue to $f(A_k(m), m)$. For z we use log parental income as reported on the senior year FAFSA, which we residualize of black-by-major fixed effects, graduation age fixed effects, and university fixed effects to account for factors that would be absorbed in a.[8] Our model makes four predictions. First, $\alpha_{b0} > \alpha_{w0}$ as employers put more weight on the prior for black workers when they enter the labor market. Second, $\delta_{b0} < \delta_{w0}$ as employers put less weight

[7] We do not include institution-by-race fixed effects as institution-specific features that improve black student outcomes could influence major selection.

[8] Note that we lose roughly half of our sample here as parents income is observable only for students who are not financially independent.

on the signal for black workers which is correlated with z. Finally, $\alpha_{bt} - \alpha_{wt}$ is decreasing in t and $\delta_{bt} - \delta_{wt}$ is increasing in t. Black wages see a sharper change in their observed correlation with observable and unobservable information across time, as employers based their initial beliefs about productivity for black workers on their prior.

We display our estimates in Table 8. The first three columns provide results from separate regressions for each of the three years of wage data from the B&B. Visibly these are consistent with our predictions. Though not statistically significant, our point estimates suggest that black students see a larger return to major in 2009, while having a sharply lower return to parental income. Relative to 2009, in 2012 the magnitudes of both coefficients decrease. By 2018, there is a large negative and statistically significant coefficient on the black-by-major return interaction, while we cannot reject the observed return to parental income is the same across race.

In column (4) we stack our three years of data and estimate a fairly standard Mincer specification allowing for the racial wage gap to vary with experience. Consistent with prior work (e.g., Altonji and Pierret, 2001) our results suggest if anything the racial wage gap grows with experience, which provides confidence that sample which is not representative in terms of educational attainment still provides similar wage dynamics to those in nationally representative populations. Column (5) implements our learning specification. All of the estimates match our empirical predictions, and all but the black-by-wage return interaction (i.e., $\alpha_{b0} - \alpha_{w0}$) are statistically significant.

5 Conclusion

In this paper we integrated a canonical statistical discrimination and learning framework into a model of major choice. Doing so revealed a new tension common with that in education choice models: black students are incentivized to overinvest in observable measures. The equilibrium outcome of the signaling game leads black students to attempt majors with a higher return to aptitude than similarly prepared whites. Yet they receive in equilibrium a lower wage return to these majors because the market correctly incorporates the incentives to overcredentialize that black students face. We find broad support for our predictions using administrative data from a set of widely differing universities, the ACS, and the B&B.

Our paper provides a novel contribution to the literature on academic mismatch and affirmative action. In equilibrium, black students are "overmatched" in their major choices, but not due to information asymmetries or deficiencies, and not due to affirmative action incentives provided by universities. Instead, it is the rational response to anticipated statistical discrimination on the labor market. This suggests an important policy role for universities would be to improve access of information on minority students to employers. For example, universities could provide additional interview training for black students, so that they are better able to convey information on their productivity to employers, overcoming the information disadvantage at the core of statistical discrimination. Universities could also work to provide better opportunities for black students in lower return majors to reveal their aptitude to employers. This could include research opportunities that produce tangible results, or academic competitions.

Our work also provides a valuable lesson on the interpretation of regression discontinuity approaches when the measured outcome is determined by a market with incomplete information. In fact, we should expect a discontinuity in wage outcomes between individuals just below and just above a, for example, university admissions cutoff, independent of any human capital effect of that university itself, because there is a sharp change in employer beliefs at this cutoff. Particularly in the context of the affirmative action literature, large policy changes that change equilibrium beliefs may provide a more useful way of testing for mismatch than narrow policies that incrementally change student university choices.

References

Aalto, Aino-Maija, Dagmar M uller, and J. Lucas Tilley (2022) "From epidemic to pandemic: Effects of the COVID-19 outbreak ou high school program choices in Sweden," IZA Discussion Paper No. 15107.

Altonji, Joseph G., Peter Arcidiacono, and Arnaud Maurel (2016) "The analysis of field choice in college and graduate school: Determinants and wage effects," in Hanushek, Eric A., Stephen Manchin, and Ludger Wößmann eds. *Handbook of the Economics of Education*, 5, 305–396: North Holland.

Altonji, Joseph G. and Charles R. Pierret (2001) "Employer learning and statistical discrimination," *Quarterly Journal of Economics*, 116 (1), 313–350.

Angrist, Joshua D., Parag A. Pathak, and Andrés Zárate Román (2019) "Choice and consequence: Assessing mismatch at Chicago exam schools," NBER Working Paper No. 26137.

Arcidiacono, Peter, Esteban M. Acuejo, and Ken Spenner (2012) "What happens after enrollment? An analysis of the time path of racial differences in GPA and major choice," *IZA Journal of Labor Economics*, 1.

Arcidiacono, Peter, Esteban M. Aucejo, Hanmin Fang, and Kenneth I. Spenner (2011) "Does affirmative action lead to mismatch? A new test and evidence?" *Quantitative Economics*, 2 (3), 303–333.

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

Arcidiacono, Peter and Michael Lovenheim (2016) "Affirmative Action and the quality-fit trade-off," *Journal of Economic Literature*, 54 (1), 3–51.

Bičáková, Alena, Guido Matias Cortes, and Jacopo Mazza (2021) "Caught in the cycle: Economic conditions at enrolment and labour market outcomes of college graduates," *Economic Journal*, 131 (638), 2383–2412.

Bleemer, Zachary and Aashish Mehta (2021) "College major restrictions and student stratification," UC Berkeley Research and Occasional Paper Series CSHE.14.2021.

Blom, Erica, Brian C. Cadena, and Benjamin K. Keys (2021) "Investment over the business cycle: Insights from college major choice," 39 (4), 1043–1082.

Denning, Jeffrey T., Eric R. Eide, Kevin J. Mumford, Richard W. Patterson, and Merril Warnick (forthcoming) "Why have college completion rates increaseed?" *American Economic Journal: Applied Economics*.

Dillon, Eleanor Wiske and Jeffrey Andrew Smith (2020) "The consequences of academic match between students and colleges," *Journal of Human Resources*, 55 (3), 767–808.

Ersoy, Fulya Y. (2020) "The effects of the Great Recession on college majors," *Economics of Education Review*, 77.

Ganguli, Ina, Jamal I. Haidar, Asim Ijaz Khwaja, Samuel W. Stemper, and Basit Zafar (2022) "Economic shocks and skill acquisition: Evidence from a a national online learning program at the onset of COVID-19," NBER Working Paper No. 29921.

Han, Luyi and John V. Winters (2020) "Industry fluctations and college major choices: Evidence from an energy boom and bust," *Economics of Education Review*, 77.

Hill, Andrew J. (2017) "State affirmative action bans and STEM degree completions," *Economics of Education Review*, 57, 31–40.

Lang, Kevin and Jee-Yeon K. Lehmann (2012) "Racial discrimination in the labor market: Theory and empirics," *Journal of Economic Litarature*, 50 (4), 959–1006.

Lang, Kevin and Michael Manove (2011) "Education and labor market discrimination," *American Economic Review*, 101 (4), 1467–1496.

Mountjoy, Jack and Brent R. Hickman (2021) "The returns to college(s): Value-added and match effects in higher education," NBER Working Paper No. 29276.

Neal, Derek (2004) "The measured black-white wage gap among women is too small," *Journal of Political Economy*, 112 (S1), S1–S28.

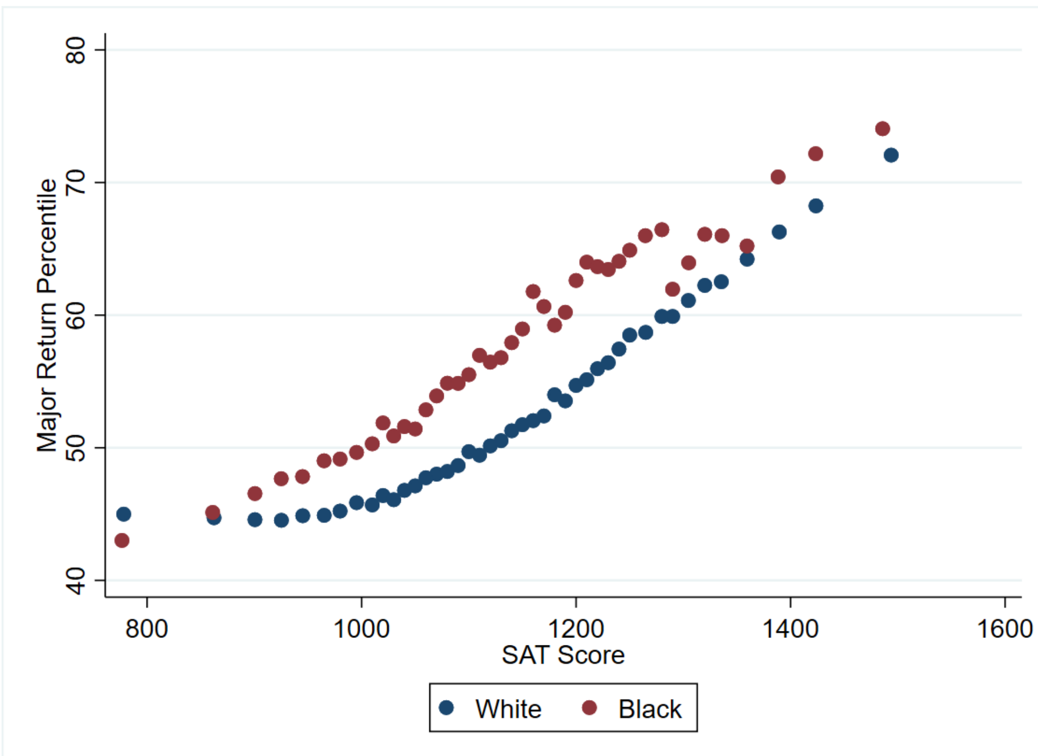
Sallee, James M., Alexandra M. Resch, and Paul N. Courant (2008) "On the optimal allocation of students and resources in a system of higher education," *B.E. Journal of Economic Analysis & Policy*, 8 (1).

Sander, Richard H. (2004) "A systematic analysis of affirmative action in American law schools," *Stanford Law Review*, 57 (2), 367–483.

Sovero, Veronica, Moshe Buchinsky, and Matthew D. Baird (2021) "Playing catch up: A term-level investigation of teh racial gap in STEM retention," *Economics of Education Review*, 83.

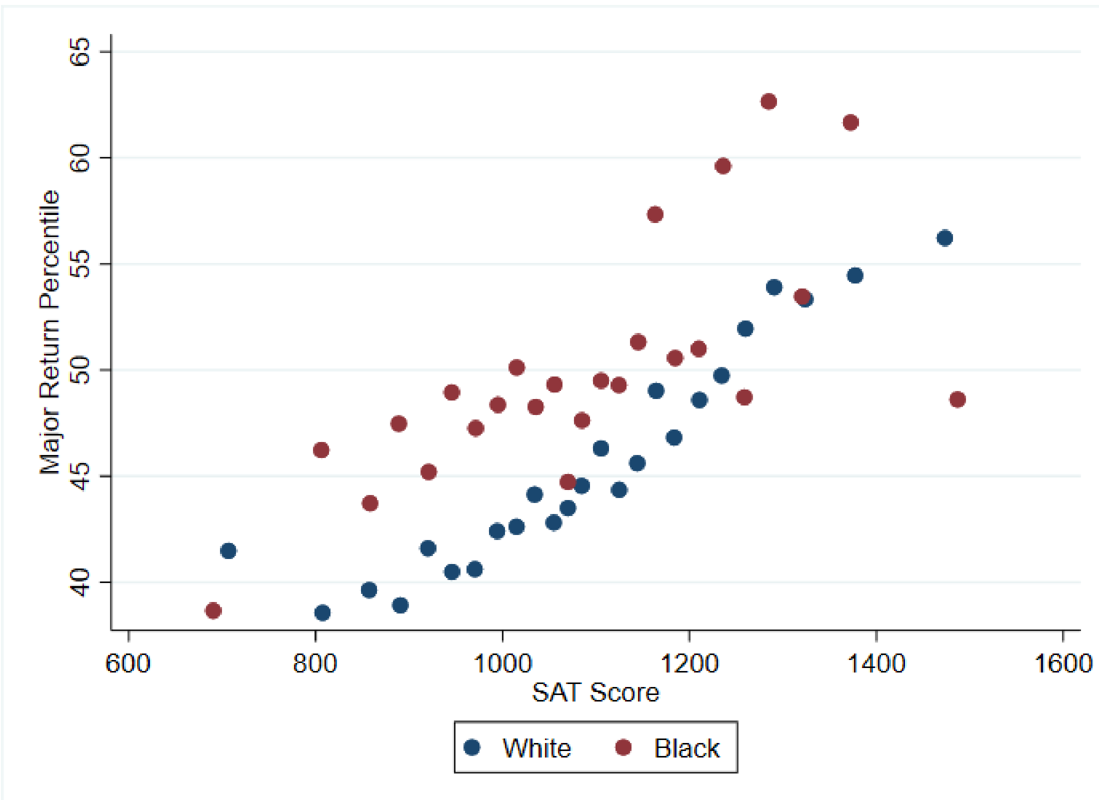
Weinstein, Russell (forthcoming) “Local labor markets and human capital investments,” *Journal of Human Resources*.

Figure 1: SAT Scores and Major Percentile Return by Race: State School Sample



Source - State School Sample includes students at Clemson, Colorado, Colorado State, Florida, Florida State, Georgia Tech, North Carolina State, North Carolina – Charlotte, Oklahoma, Purdue, Utah State, and Virginia Tech with observed SAT scores.

Figure 2: SAT Scores and Major Percentile Return by Race: B&B Sample



Source - Baccalaureate and Beyond longitudinal study of 2007-2008 college graduates. Students older than 30 when they graduate from college are excluded from the sample as are all student not identified as either Black or White

Table 1: State School Summary Statistics

	Black	White
	(1)	(2)
Wage Return to Major (White)	-0.053	-0.067
Percentile Return to Major (White)	55.23	53.62
College Graduate	0.436	0.510
Female	0.521	0.474
Transfer	0.094	0.135
Year Entered College	2000.0	2002.5
SAT Math Score	524.1	580.3
SAT Verbal Score	523.5	573.3
High School GPA	3.35	3.43
College GPA	2.50	2.96
Engineering Major	0.155	0.149
Information Technology Major	0.039	0.030
Chemistry Major	0.012	0.010
Biology Major	0.087	0.077
Social Science Major	0.068	0.051
Communications Major	0.049	0.044
Business Major	0.144	0.141
Liberal Arts Major	0.143	0.139
History Major	0.008	0.014
English Major	0.016	0.022
Education Major	0.036	0.038
Agriculture Major	0.019	0.038
Observations	29,647	565,428

Source - State school sample includes Black and White students at Clemson, Colorado, Colorado State, Florida, Florida State, Georgia Tech, North Carolina State, North Carolina – Charlotte, Oklahoma, Purdue, Utah State, and Virginia Tech with observed SAT scores.

Table 2: Major Selection by SAT Score and Race: State School Sample

	Wage Return (White)		Percentile Return (White)	
	(1)	(2)	(3)	(4)
Black	0.030***	0.034***	3.570***	3.925***
	(0.001)	(0.002)	(0.165)	(0.191)
SAT		0.041***		4.712***
		(0.000)		(0.042)
Black × SAT		0.003***		0.260***
		(0.001)		(0.101)
Female	-0.123***	-0.127***	-15.694***	-16.104***
	(0.001)	(0.001)	(0.065)	(0.065)
Transfer Student	-0.001	-0.003***	-0.308**	-0.536***
	(0.001)	(0.001)	(0.123)	(0.123)
Institution Fixed Effects	YES	YES	YES	YES
College Start Year Fixed Effects	YES	YES	YES	YES
High School GPA Fixed Effects	YES	YES	YES	YES
Zip Code Fixed Effects	YES	YES	YES	YES
SAT Fixed Effects	YES	NO	YES	NO
Observations	589,865	589,865	589,865	589,865
R^2	0.293	0.284	0.274	0.267

Source - State school sample includes students at Clemson, Colorado, Colorado State, Florida, Florida State, Georgia Tech, North Carolina State, North Carolina – Charlotte, Oklahoma, Purdue, Utah State, and Virginia Tech.
Notes - Robust standard errors reported in parenthesis. All specifications include an indicator for transfer student status. Time fixed effects include indicators for the year the student started college. $*p < .1, **p < .05, ***p < .01$

Table 3: ACS Summary Statistics

	Black	White
	(1)	(2)
Female	0.581	0.448
	(0.49)	(0.50)
Age	38.541	38.813
	(8.60)	(8.85)
Log Wage and Salary Income (2020 $)	10.868	11.123
	(0.59)	(0.66)
Wage Return	-0.089	-0.063
	(0.19)	(0.21)
Observations	77,538	973,388

Source - ACS, 2011-2018 waves

Notes - Summary statistics weighted using IPUMS perwt

Table 4: Baccalaureate and Beyond, 2007-2008 College Graduates

	Black	White
	(1)	(2)
Wage Return (White)	-0.129	-0.134
Wage Return (Black)	0.122	0.117
Percentile Return (White)	46.887	45.877
Female	0.674	0.572
SAT	950.355	1099.977
GPA at graduation	3.097	3.344
GPA (major centered)	-0.202	0.024
Age at Graduation	23.308	22.961
First Generation Student	0.605	0.421
Independent Student	0.369	0.302
Log Parent Income	10.497	11.061
Log Salary 2009	10.116	10.159
Log Salary 2012	10.499	10.578
Log Salary 2018	10.928	11.077
Observations	1,320	11,240

Source - Baccalaureate and Beyond longitudinal study of 2007-2008 college graduates. Students older than 30 when they graduate from college are excluded from the sample as are all student not identified as either Black or White

Table 5: Major Selection by SAT Score and Race: B&B Sample

	Wage Return (White)		Percentile Return (White)	
	(1)	(2)	(3)	(4)
Black	0.023***	0.024***	3.465***	3.919***
	(0.006)	(0.008)	(0.875)	(1.077)
SAT		0.012***		1.686***
		(0.001)		(0.158)
Black × SAT		0.001		0.294
		(0.003)		(0.407)
Female	-0.121***	-0.121***	-16.658***	-16.650***
	(0.003)	(0.003)	(0.481)	(0.480)
Institution Fixed Effects	YES	YES	YES	YES
SAT Fixed Effects	YES	NO	YES	NO
Observations	11,740	11,740	11,740	11,740
R^2	0.293	0.284	0.274	0.267

Source - Sample includes white and black college graduate in the 2008/18 Baccalaureate and Beyond Longitudinal Study, a nationally representative sample of individuals who earned a bachelor's degree in the 2007-08 academic year. We exclude those who are over the age of 30 at the time they graduate.

Notes - Wage Return is calculated as the average of the residuals by major from a regression of log annual earnings on age and year fixed effects for white college graduates in the ACS. Robust standard errors in parentheses.

Table 6: Major Selection on Adult Wages: ACS Sample, OLS Estimates

	LHS Variable: Log Wage					
	(1)	(2)	(3)	(4)	(5)	(6)
Female	-0.195***	-0.191***	-0.191***	-0.199***	-0.195***	-0.195***
	(0.024)	(0.024)	(0.024)	(0.023)	(0.023)	(0.023)
Black	-0.235***	-0.247***		-0.078***	-0.089***	
	(0.017)	(0.017)		(0.011)	(0.010)	
Wage Return	0.789***	0.775***	0.775***			
	(0.026)	(0.020)	(0.020)			
Wage Return × Black	-0.311***	-0.314***	-0.315***			
	(0.050)	(0.047)	(0.047)			
Percentile Return				0.006***	0.006***	0.006***
				(0.000)	(0.000)	(0.000)
Percentile Return × Black				-0.003***	-0.003***	-0.003***
				(0.000)	(0.000)	(0.000)
Age	0.098***	0.100***	0.100***	0.097***	0.100***	0.100***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Age^2	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
State FE	NO	YES	YES	NO	YES	YES
State × Race FE	NO	NO	YES	NO	NO	YES
Year FE	NO	YES	YES	NO	YES	YES
Year × Race FE	NO	NO	YES	NO	NO	YES
Observations	1,339,058	1,339,058	1,339,058	1,339,058	1,339,058	1,339,058

Notes - Robust standard errors clustered by field of degree in parenthesis. All models weighted using IPUMS perwt. $*p < .1, **p < .05, ***p < .01$

Table 7: Major Selection on Early Career Earnings: BB Sample, OLS Estimates

	LHS Variable: Log Wage			
	(1)	(2)	(3)	(4)
Black	-0.070***	-0.055***	-0.072***	-0.057***
	(0.015)	(0.015)	(0.015)	(0.015)
Wage Return	0.817***	0.817***		
	(0.023)	(0.023)		
Black × Wage Return	-0.256***	-0.263***		
	(0.074)	(0.074)		
Percentile Return			0.006***	0.006***
			(0.000)	(0.000)
Black × Percentile Return			-0.002***	-0.002***
			(0.001)	(0.001)
GPA (major centered)		0.061***		0.062***
		(0.010)		(0.010)
Year FE	YES	YES	YES	YES
University FE	YES	YES	YES	YES
Age FE	YES	YES	YES	YES
Observations	28,450	28,450	28,450	28,450
R-squared	0.351	0.351	0.349	0.350

Source - Sample includes white and black college graduate in the 2008/18 Baccalaureate and Beyond Longitudinal Study, a nationally representative sample of individuals who earned a bachelor's degree in the 2007-08 academic year. We exclude those who are over the age of 30 at the time they graduate.

Notes - To help with interpretation, college GPA is re-centered at 2.0. The major difficulty measure is calculated as the average of the residuals by major from a regression of log annual earnings on age and year fixed effects for white college graduates in the ACS. Robust standard errors reported in parenthesis: $*p < .1, **p < .05, ***p < .01$

Table 8: Labor Market Returns to Major Difficulty by Race with Parent Income

	2009	2012	2018	Overall
	(1)	(2)	(3)	(4)
Black	-0.043	0.008	-0.075**	-0.019
	(0.040)	(0.039)	(0.035)	(0.030)
Wage Return	0.733***	0.743***	0.982***	0.688***
	(0.055)	(0.051)	(0.049)	(0.042)
Black × Wage Return	0.118	-0.063	-0.612***	-0.009
	(0.190)	(0.207)	(0.182)	(0.155)
GPA (major centered)	-0.009	0.044*	0.117***	0.054***
	(0.025)	(0.024)	(0.023)	(0.014)
Parent Income	0.006	0.039***	0.063***	0.014
	(0.014)	(0.014)	(0.013)	(0.011)
Black × Parent Income	-0.131***	-0.047	-0.036	-0.101***
	(0.040)	(0.043)	(0.037)	(0.033)
Experience				0.104***
				(0.001)
Black × Experience				-0.005
				(0.004)
Wage Return × Experience				0.032***
				(0.007)
Black × Wage Return × Experience				-0.044*
				(0.025)
Parent Income × Experience				0.005***
				(0.002)
Black × Parent Income × Experience				0.010*
				(0.006)
University FE	YES	YES	YES	YES
Age FE	YES	YES	YES	YES
Observations	5,860	5,460	6,100	17,420

Source - Sample includes white and black college graduate in the 2008/18 Baccalaureate and Beyond Longitudinal Study who were not financially independent from their parents in college (in order to observe parent income). We exclude those who are over the age of 30 at the time they graduate.

Notes - Parent income is measured as the residuals from a regression of log parent income on Black-by-Major FE, Graduation Age FE, and University FE. Robust standard errors reported in parenthesis: $*p < .1, **p < .05, ***p < .01$

Table A1: Estimated Major Return

Major Code and Name		Wage Return White	Black	Prct	Major Code and Name		Wage Return White	Black	Prct
1100	General Agriculture	-0.171	0.078	45	3608	Physiology	-0.217	0.037	33
1101	Agriculture Production and Management	-0.108	0.144	53	3609	Zoology	-0.150	0.099	47
1102	Agricultural Economics	0.080	0.328	77	3611	Neuroscience	-0.380	-0.130	10
1103	Animal Sciences	-0.295	-0.044	20	3699	Miscellaneous Biology	-0.206	0.046	37
1104	Food Science	0.020	0.270	70	3700	Mathematics	0.110	0.360	79
1105	Plant Science and Agronomy	-0.180	0.071	43	3701	Applied Mathematics	0.185	0.436	85
1106	Soil Science	-0.105	0.142	55	3702	Statistics and Decision Science	0.147	0.398	83
1199	Miscellaneous Agriculture	-0.265	-0.013	25	3801	Military Technologies	0.162	0.420	84
1301	Environmental Science	-0.208	0.048	36	4000	Interdisciplinary and Multi-Disciplinary Studies	-0.348	-0.093	13
1302	Forestry	-0.056	0.195	59	4001	Intercultural and International Studies	-0.261	-0.010	26
1303	Natural Resources Management	-0.187	0.066	41	4002	Nutrition Sciences	-0.259	-0.009	27
1401	Architecture	0.055	0.306	74	4005	Mathematics and Computer Science	0.266	0.514	92
1501	Area, Ethnic, and Civilization Studies	-0.180	0.070	43	4006	Cognitive Science and Biopsychology	-0.085	0.167	56
1901	Communications	-0.107	0.146	53	4007	Interdisciplinary Social Sciences	-0.276	-0.023	24
1902	Journalism	-0.105	0.145	54	4101	Physical Fitness, Parks, Recreation, and Leisure	-0.342	-0.089	14
1903	Mass Media	-0.231	0.022	31	4801	Philosophy and Religious Studies	-0.207	0.044	37
1904	Advertising and Public Relations	-0.122	0.129	50	4901	Theology and Religious Vocations	-0.411	-0.161	6
2001	Communication Technologies	-0.256	-0.005	27	5000	Physical Sciences	0.042	0.292	71
2100	Computer and Information Systems	0.047	0.303	72	5001	Astronomy and Astrophysics	0.064	0.315	76
2101	Computer Programming and Data Processing	0.005	0.258	66	5002	Atmospheric Sciences and Meteorology	-0.026	0.225	62
2102	Computer Science	0.234	0.486	90	5003	Chemistry	-0.006	0.243	64
2105	Information Sciences	0.111	0.364	79	5004	Geology and Earth Science	0.003	0.253	66
2106	Computer Information Management and Security	-0.038	0.218	61	5005	Geosciences	-0.017	0.232	63
2107	Computer Networking and Telecommunications	-0.079	0.175	58	5006	Oceanography	-0.108	0.143	52
2201	Cosmetology Services and Culinary Arts	-0.411	-0.158	7	5007	Physics	0.117	0.367	80
2300	General Education	-0.336	-0.086	14	5008	Materials Science	0.057	0.308	74
2301	Educational Administration and Supervision	-0.191	0.061	40	5098	Multi-disciplinary or General Science	-0.081	0.170	57
2303	School Student Counseling	-0.351	-0.096	13	5102	Nuclear, Radiology, and Biological Technologies	-0.133	0.120	49
2304	Elementary Education	-0.446	-0.196	3	5200	Psychology	-0.284	-0.032	21
2305	Mathematics Teacher Education	-0.305	-0.057	18	5201	Educational Psychology	-0.318	-0.069	17
2306	Physical and Health Education Teaching	-0.238	0.009	29	5202	Clinical Psychology	-0.425	-0.175	5
2307	Early Childhood Education	-0.573	-0.322	1	5203	Counseling Psychology	-0.438	-0.186	5
2308	Science and Computer Teacher Education	-0.280	-0.032	22	5205	Industrial and Organizational Psychology	-0.071	0.181	58
2309	Secondary Teacher Education	-0.282	-0.036	22	5206	Social Psychology	-0.204	0.043	38
2310	Special Needs Education	-0.388	-0.140	9	5299	Miscellaneous Psychology	-0.301	-0.048	18
2311	Social Science or History Teacher Education	-0.321	-0.073	16	5301	Criminal Justice and Fire Protection	-0.202	0.053	39
2312	Teacher Education: Multiple Levels	-0.459	-0.210	3	5401	Public Administration	-0.008	0.243	64
2313	Language and Drama Education	-0.394	-0.144	8	5402	Public Policy	0.017	0.270	69
2314	Art and Music Education	-0.382	-0.135	9	5403	Human Services and Community Organization	-0.463	-0.209	2
2399	Miscellaneous Education	-0.168	0.076	46	5404	Social Work	-0.440	-0.188	4
2400	General Engineering	0.221	0.472	89	5500	General Social Sciences	-0.211	0.040	35
2401	Aerospace Engineering	0.295	0.543	94	5501	Economics	0.243	0.493	90
2402	Biological Engineering	0.104	0.354	77	5502	Anthropology and Archeology	-0.298	-0.046	19
2403	Architectural Engineering	0.250	0.502	91	5503	Criminology	-0.213	0.040	33
2404	Biomedical Engineering	0.062	0.313	75	5504	Geography	-0.113	0.141	51
2405	Chemical Engineering	0.372	0.624	96	5505	International Relations	-0.004	0.247	65
2406	Civil Engineering	0.266	0.517	92	5506	Political Science and Government	0.022	0.273	70
2407	Computer Engineering	0.259	0.513	91	5507	Sociology	-0.223	0.028	32
2408	Electrical Engineering	0.373	0.624	97	5599	Miscellaneous Social Sciences	-0.085	0.165	57
2409	Engineering Mechanics, Physics, and Science	0.163	0.410	84	5601	Construction Services	0.126	0.378	81
2410	Environmental Engineering	0.125	0.380	80	5701	Electrical and Mechanic Repairs and Technologies	-0.192	0.059	40
2411	Geological and Geophysical Engineering	0.358	0.606	95	5901	Transportation Sciences and Technologies	0.133	0.386	82
2412	Industrial and Manufacturing Engineering	0.293	0.544	94	6000	Fine Arts	-0.277	-0.026	24
2413	Materials Engineering and Materials Science	0.193	0.444	86	6001	Drama and Theater Arts	-0.363	-0.112	11
2414	Mechanical Engineering	0.308	0.559	95	6002	Music	-0.310	-0.060	17
2415	Metallurgical Engineering	0.493	0.740	99	6003	Visual and Performing Arts	-0.406	-0.157	7
2416	Mining and Mineral Engineering	0.462	0.705	99	6004	Commercial Art and Graphic Design	-0.248	0.004	28
2417	Naval Architecture and Marine Engineering	0.387	0.635	98	6005	Film, Video and Photographic Arts	-0.272	-0.021	25
2418	Nuclear Engineering	0.359	0.608	96	6006	Art History and Criticism	-0.234	0.016	29
2419	Petroleum Engineering	0.799	1.043	100	6007	Studio Arts	-0.464	-0.212	2
2499	Miscellaneous Engineering	0.183	0.433	85	6099	Miscellaneous Fine Arts	-0.364	-0.111	11
2500	Engineering Technologies	0.011	0.263	69	6100	General Medical and Health Services	-0.285	-0.034	21
2501	Engineering and Industrial Management	0.220	0.469	88	6102	Communication Disorders Sciences and Services	-0.413	-0.163	6
2502	Electrical Engineering Technology	0.107	0.358	78	6103	Health and Medical Administrative Services	-0.210	0.046	36
2503	Industrial Production Technologies	0.132	0.380	81	6104	Medical Assisting Services	-0.154	0.098	46
2504	Mechanical Engineering Related Technologies	0.068	0.319	76	6105	Medical Technologies Technicians	-0.057	0.192	59
2599	Miscellaneous Engineering Technologies	0.050	0.303	73	6106	Health and Medical Preparatory Programs	-0.230	0.019	31
2601	Linguistics and Comparative Language and Literature	-0.232	0.018	30	6107	Nursing	-0.042	0.209	60
2602	Common Foreign Language Studies	-0.204	0.046	39	6108	Pharmaceutical Sciences, and Administration	0.452	0.701	98
2603	Other Foreign Languages	-0.107	0.144	54	6109	Treatment Therapy Professions	-0.136	0.117	48
2901	Family and Consumer Sciences	-0.379	-0.129	10	6110	Community and Public Health	-0.289	-0.034	20
3201	Court Reporting	-0.244	0.006	28	6199	Miscellaneous Health Medical Professions	-0.332	-0.081	16
3202	Pre-Law and Legal Studies	-0.212	0.041	34	6200	General Business	0.051	0.302	73
3301	English Language and Literature	-0.183	0.069	42	6201	Accounting	0.143	0.394	83
3302	Composition and Speech	-0.353	-0.101	12	6202	Actuarial Science	0.276	0.529	93
3401	Liberal Arts	-0.181	0.071	42	6203	Business Management and Administration	0.001	0.251	65
3402	Humanities	-0.229	0.023	32	6204	Operations, Logistics and E-Commerce	0.048	0.301	72
3501	Library Science	-0.468	-0.224	1	6205	Business Economics	0.212	0.462	88
3600	Biology	-0.180	0.072	44	6206	Marketing and Marketing Research	0.006	0.257	68
3601	Biochemical Sciences	-0.139	0.110	48	6207	Finance	0.199	0.450	87
3602	Botany	-0.141	0.110	47	6209	Human Resources and Personnel Management	-0.102	0.150	55
3603	Molecular Biology	-0.176	0.076	44	6210	International Business	0.006	0.260	67
3604	Ecology	-0.333	-0.078	15	6211	Hospitality Management	-0.212	0.040	35
3605	Genetics	-0.277	-0.027	23	6212	Management Information Systems and Statistics	0.193	0.448	87
3606	Microbiology	-0.027	0.222	62	6299	Miscellaneous Business and Medical Administration	-0.032	0.220	61
3607	Pharmacology	0.007	0.256	68	6402	History	-0.129	0.122	50
					6403	United States History	-0.112	0.140	51

Source - ACS, 2011-2018 waves, sample of 35-45 year old native-born full-time year-round employed workers with exactly a bachelor's degree

Notes - The wage return is the average residual by major from a regression of log annual wages on age and year fixed effects separately for white and black workers.

Table A2: Major Selection on Early Career Earnings: BB Sample, OLS Estimates

	LHS Variable: Log Wage			
	(1)	(2)	(3)	(4)
First Generation Student	-0.011	-0.009	-0.012	-0.010
	(0.008)	(0.008)	(0.008)	(0.008)
Wage Return	0.840***	0.856***		
	(0.029)	(0.029)		
First Gen × Wage Return	-0.106***	-0.095**		
	(0.041)	(0.041)		
Percentile Return			0.006***	0.006***
			(0.000)	(0.000)
First Gen × Percentile Return			-0.001***	-0.001**
			(0.000)	(0.000)
Black		-0.055***		-0.056***
		(0.015)		(0.015)
Black × Wage Return		-0.249***		
		(0.074)		
Black × Percentile Return				-0.002***
				(0.001)
GPA (major centered)	0.066***	0.061***	0.067***	0.062***
	(0.010)	(0.010)	(0.010)	(0.010)
Year FE	YES	YES	YES	YES
University FE	YES	YES	YES	YES
Age FE	YES	YES	YES	YES
Observations	28,450	28,450	28,450	28,450
R-squared	0.351	0.352	0.350	0.350

Source - Sample includes white and black college graduate in the 2008/18 Baccalaureate and Beyond Longitudinal Study, a nationally representative sample of individuals who earned a bachelor's degree in the 2007-08 academic year. We exclude those who are over the age of 30 at the time they graduate.

Notes - To help with interpretation, college GPA is re-centered at 2.0. The major difficulty measure is calculated as the average of the residuals by major from a regression of log annual earnings on age and year fixed effects for white college graduates in the ACS. Robust standard errors reported in parenthesis: $*p < .1, **p < .05, ***p < .01$