Admitting Students to Selective Education Programs: Merit, Profiling, and Affirmative Action*

Dario Cestau
Carnegie Mellon University

Dennis Epple
Carnegie Mellon University

Holger Sieg
University of Pennsylvania

October 6, 2012

*We would like to thank Peter Arcidiacono, Pat Bayer, Flavio Cunha, Glen Ellison, John Engberg, Hanming Fang, Eric Hanushek, Joe Hotz, Sunny Ladd, Kjell Salvanes, Petra Todd, and seminar participants at numerous workshops and seminars for comments and suggestions. Financial support for this research is provided by the Institute of Education Sciences. Email: D. Cestau: dcestau@andrew.cmu.edu, D. Epple: epple@cmu.edu, H. Sieg: holgers@econ.upenn.edu.

Abstract

For decades, colleges and universities have struggled to increase participation of minority and disadvantaged students. Urban school districts confront a parallel challenge; minority and disadvantaged students are underrepresented in selective programs that use merit-based admission. We analyze optimal school district policy and develop an econometric framework providing a unified treatment of affirmative action and profiling. Implementing the model for a central-city district, we find profiling by race and income, affirmative action for low-income students, and no affirmative action with respect race. Counterfactual analysis reveals that these policies achieve 80% of African American enrollment that would be could be attained by race-based affirmative action.

KEYWORDS: Gifted and talented education, profiling, affirmative action, merit-based admission, equilibrium analysis, estimation.

1 Introduction

For decades, colleges and universities have grappled with the challenge of increasing participation of minority and disadvantaged students. Differential admissions criteria designed to increase participation of minority and disadvantaged students inevitably collide with concerns for fairness toward those who are displaced. Courts, legislatures, and electorates have weighed in to define the acceptable limits of affirmative action in higher education.The Supreme Court has also provided some hints about the extent to which profiling might be used in lieu of affirmative action. The Supreme Court is revisiting these issues during its current term, raising the possibility of major change in permissible use of race-conscious admission policies. While the spotlight has largely focused on higher education, school districts in large metropolitan areas confront the same challenge; minority and disadvantaged students are underrepresented in selective programs for gifted and talented students.[1] Efforts to increase participation of minority and disadvantaged students in these programs are no less controversial than their higher-education counterparts.[2]

School districts nationwide make a large investment in providing programs to serve gifted and talented students. The National Association of Gifted Children estimates that there are approximately 3 million academically gifted children in grades K-12 in

[1] To cite a recent example, on September 27, 2012, a complaint was filed with the U.S. Department of education claiming that the admissions policies of eight elite public high schools in New York City violate the 1964 Civil Rights Act, causing extreme under-representation of black and Hispanic students. (Baker, 2012).

[2] The New York Times dubbed the director of the city's gifted and talented program a "lightning rod for fury" when she announced changes to the admissions policies for the city's gifted programs (NYT, March 22, 2006).Approximately, two years later, the times reported: "When New York City set a uniform threshold for admission to public school gifted programs last fall, it was a crucial step in a prolonged effort to equalize access to programs that critics complained were dominated by white middle-class children whose parents knew how to navigate the system." The new policy has also met with intense criticism, and the director of the city's gifted and talented program has concluded: "We implemented the eligibility criteria, it didn't shake out that way and now we have to take another look at it." (NYT, June 19, 2008)

the U.S - almost 7 percent of the student population. As a proportion of the student population, this is comparable to the percentage of high school students admitted to selective colleges and universities.[3]

In this paper we investigate these issues in the context of a gifted and talented program of a mid-sized urban school district. We document the divergence between the demographics of the district's student body relative to the demographics of those participating in the district's gifted program. We delineate the constraints under which the district operates and the policy tools available to the district. We then study properties of optimal district policy. We formulate and estimate a model to characterize the district's use of profiling and affirmative action policies to enhance participation of minority and disadvantaged students while respecting the merit-based guidelines promulgated by the state government. We believe this to be the first econometric model providing a unified treatment of affirmative action[4] and profiling [5] in an equilibrium framework that permits counterfactual analysis of policy.

Most large public school districts in the U.S. operate selective programs to serve

[3]Reardon, et. al. (2012) identify the 171 most selective colleges and universities and report that approximately 7% of high school students attend one of these selective institutions.

[4]Becker (1957) introduced the analysis of taste based discrimination into economics. Phelps (1972) and Arrow (1973) developed a theory of statistical discrimination. The effects of affirmative action in employment have been studied by Lundberg (1991), Coate & Loury (1993), and Moro & Norman (2003, 2004). Chung (2000) considers the relationship between role models and affirmative action. Affirmative action in higher education is considered by Chan & Eyster (2003) , Loury, Fryer, & Yuret (2008), and Epple, Romano, & Sieg (2006, 2008). Arcidiacono (2005) estimates a structural model to determine how affirmative action in admission and aid policies affects future earnings. Hickman (2010a, 2010b) develops and estimates a structural model of college admissions and compares the effects of alternative admissions policies on incentives for academic achievement, the racial achievement gap, and the racial college enrollment gap. Long (2007) provides a valuable summary of the legal status of affirmative action and a review of the evidence regarding the effects of affirmative action.

[5]Knowles, Persico, & Todd (2001) stimulated a large body of work on profiling in law enforcement. In higher education, research on profiling includes Loury et al. (2008), Epple et al. (2008), and Epple, Romano, Sarpca, and Sieg (2012). State policies mandating admission of a specified fraction of each high school to state universities, motivated in part to achieve racial balance through profiling, are studied by Long (2004a, 2004b) and Cullen, Long, & Reback (2012).

gifted and talented students.[6] Admission to these programs is competitive and merit-based. School districts must promulgate admission policies that establish eligibility criteria. A district must also set up referral or testing policies that determine which students are evaluated as potential candidates for the program. Finally, the district must determine whether criteria will include consideration of disadvantaged or minority status.

The goal of gifted and talented programs is to provide education to help highly able students reach their potential.[7] Districts typically utilize achievement or intelligence tests in their admissions decisions; our district uses an IQ test. Ability is difficult to observe and, more importantly, non-verifiable, while IQ can be observed and verified by parents and teachers. Testing is costly. For the district we study, an IQ test is administered orally on a one-to-one basis between a district psychologist and each student considered for gifted admission. Testing is also stressful for parents and students, and can be viewed as an unwelcome burden in cases where students have a low probability of gaining admission to the selective program. Hence, a referral procedure is needed to determine which students take the IQ test.

In designing a referral procedure, the district must make a judgment of the extent to which the benefits of the program vary with observed student characteristics such as income, race or gender. This judgment may reflect an assessment of potential gains that students experience from participating in the program. Since ability is not fully observed, there is uncertainty about these gains. The referral procedure is then designed to determine which students are referred for testing, balancing the potential benefits against the costs of testing. We characterize the optimal referral

[6]International Baccalaureate programs also use stringent admission procedures. The most highly ranked public high schools in New York, Boston, and Philadelphia all use merit-based admission procedures.

[7]Neal & Johnson (1996) show that black-white differences in premarket skill do account for a significant portion of the black- white earnings gap in the early 1990's.

and admission policy for the district. We characterize the way in which referral and admission policies depend on the weights that the district assigns to the potential gains to different students.

One key property of our model is that optimal policies are cut-off strategies. As a consequence, optimal testing and admission policies give rise to a sharp regression discontinuity design.[8] This design is a popular, quasi-experimental method to evaluate the effectiveness of interventions in education. Our model implies an optimal admission rule for which the marginal treatment effect for a student at the cut-off is determined by the shadow value of the capacity constraint. Hence, a finding of a small or even zero marginal treatment effect using an RD design would not necessarily be evidence that the gifted program is ineffective. Broadly speaking, an RD design is primarily informative for outcomes that are of limited decision relevance for the district.

While use of an IQ test facilitates transparency, use of such tests is controversial because evidence suggests that, ceteris paribus, minority students tend to perform worse on the IQ tests than majority students. A student who has experienced "hardship" may also underperform on achievement tests relative to his or her capability. These observations motivate policymakers to consider affirmative action and profiling.[9] Hence, we study referral policies when districts are permitted to give preferential treatment to disadvantaged groups, i.e., when affirmative action policies can be designed to increase program participation of underrepresented minorities. We also study profiling policies that may be employed, either as an alternative or a supplement to affirmative action. Such policies may incorporate adjustments for differential

[8]Regression discontinuity design was used by Thistlethwaite & D.Cambell (1960). For a rigorous treatment econometric treatment see Hahn, Todd, & van der Klaauw (2001).

[9]The state guidelines require that districts consider "intervening factors" such as "gender or race bias, or socio/cultural deprivation" that may "mask gifted abilities."

performance on achievement tests that correlate with observed characteristics such race or free and reduced lunch status.

To gain additional insights into the quantitative properties of our model and demonstrate its practical relevance, we parametrize our model and estimate its parameters. We develop a Maximum Likelihood Estimator that exploits the fact that important variables are either latent (ability), partially latent (IQ scores), or measured with error (income and achievement). We conduct a Monte Carlo study that establishes that our identification strategy is valid and that we can recover the parameters of the model with reasonable sample sizes.

Our empirical analysis is based on a sample of elementary school students that entered first grade in the academic year 2004-05. We find that our model fits the data well and that estimated referral and admission thresholds are broadly consistent with policies articulated by the district. We find that the district adopts profiling with respect to both race and subsidized lunch status. The district also uses affirmative action in admission with respect to subsidized lunch status. We do not find any evidence that the district engages in affirmative action based on race.

A large proportion of African American students are eligible for free or reduced lunch. As a consequence, our estimated model implies that applying the FRL admission policies to all African American students, regardless of FRL status, would have a modest effect on increasing African American enrollment in the gifted program. Hence, in contrast to higher education[10], we find that profiling that exploits FRL status can have a quite substantial effect in increasing minority student participation. Our model also implies that eliminating preferential policies for FRL students and standardizing referral and admission decisions at the level of non-FRL students

[10] In his review of affirmative action in colleges and universities, Long (2007) concludes that "...correlates of race are unlikely to substitute successfully for the consideration of race itself."

would reduce the size of the gifted program by up to 50 percent. We thus conclude that adopting strict merit based referral and admission policies, would significantly alter the size and composition of the program.

The rest of the paper is organized as follows. Section 2 develops the theoretical model of optimal referral and admission for a merit-based selective program. Section 3 discusses the implications of affirmative action and profiling within our model. Section 4 introduces a parameterization of the model and develops a Maximum Likelihood Estimator. It also reports some evidence from a Monte Carlo study. Section 5 discusses our data set. Section 6 reports the empirical results of this study. Section 7 discusses the policy implications that can be drawn from this analysis.

2 A Model of Merit-Based Admission

We consider the decision problem of a public school district that operates a selective educational program. Admission is competitive and merit-based. Ability is inherently difficult to observe and cannot be verified. Admission criteria are, therefore, based on a standardized achievement or aptitude test. Testing is costly. Hence, not all students are referred for testing. We model the choice of policy for referral of students for testing and the choice of policy for admission to the selective program.

2.1 Preferences and Heterogeneity

There is a continuum of students that differ by ability, b. Let q denote the student's score on an IQ test and a performance on a prior achievement test. Observed socio-economic status is denoted y, which, for simplicity of exposition, we term household income.[11] Let $f(a,b,q,y)$ be the joint density of achievement, academic ability, IQ score, and income in the population. Students may also differ by discrete characteristics such as race. We first present the analysis for a single discrete type and then extend to consideration of more than one such type.

Assumption 1 *The density $f(a,b,q,y)$ is continuous and positive on its support $(0,\bar{a}) \times (0,\bar{b}) \times (0,\bar{q}) \times (0,\bar{y})$.*[12]

Students benefit from attending a program that is targeted at high-ability students. The value that the district attaches to having a student of type (b,y) participate in the selective program is denoted by $v(b,y)$.

Assumption 2 *The value function $v(b,y)$ is continuous and differentiable in both arguments. Moreover, it is monotonically increasing in b, i.e. $v_b(b,y) \geq 0$.*

This assumption presumes that the district places weight on student ability and household income. The mission of the program is to serve more able students. Hence, the sign of $v_b(b,y)$ is straightforward. The sign of $v_y(b,y)$ may be either positive or negative, depending on whether a district places greater weight on program participation of low or high income students.

Let c be the cost of testing a student.[13] The district initially observes (a,y) and chooses a referral policy denoted by $\alpha(a,y)$. The testing procedure provides the student IQ score.

[11] Unlike colleges and universities, school districts do not generally obtain financial information for individual families. Hence, y in our model denotes SES as inferred by district administrators from student eligibility for free or reduced lunch and, possibility, information such as a household's neighborhood of residence.

[12] For notational simplicity, we assume in the following that the upper bounds are infinity.

[13] An IQ test is an interactive test administered by a psychologist to an individual student. State law mandates that gifted status be determined by a certified district psychologist. Cost is typically cited as the reason for limiting the number of students who are tested. (See NYT, June 19, 2008)

Finally, there may exist an exogenously given capacity constraint that limits the size of the program.

Assumption 3 *The district faces an exogenous capacity constraint for the program given by k.*

2.2 Optimal Admission and Referral

We start by studying the case in which the district can adopt referral and admission rules that are not constrained by state rules. In the absence of state guidelines, referral policies can potentially depend on observables, such as household income and prior achievement.

Assumption 4 *The decision rule that determines who is referred for testing is a function of prior achievement and income:*

$$0 \leq \alpha(a, y) \leq 1 \tag{1}$$

While it is possible for admission to be conditioned on a number of observed characteristics, we consider a district that bases admission decisions on IQ. [14]

Assumption 5 *The decision rule that determines admission to the selective program is restricted to be a monotonic function of the IQ score, i.e.*

$$0 \leq \beta(q) \leq 1 \tag{2}$$

with $\beta_q \geq 0$.

[14] We relax this assumption in our application and allow the district to use a comprehensive "portfolio review" of borderline cases.

This assumption specifies that the admission rule is based on the score on an IQ test, q, which is observed by both district administrators and parents. In contrast b cannot be verified. As a consequence parents and teachers may disagree about b.[15] Hence, the key part of Assumption 3 is that admission is based on an observable, the IQ test score, and not on ability.

The district chooses optimal referral and admission policies. The optimization problem of the district can be written as:

$$\max_{\alpha(a,y),\beta(q)} \int_0^\infty \int_0^\infty \int_0^\infty \int_0^\infty \alpha(a,y) \Big(v(b,y)\ \beta(q)\ -\ c\Big)\ f(a,b,q,y)\ da\ db\ dq\ dy \quad (3)$$

subject to the capacity constraint:

$$k = \int_0^\infty \int_0^\infty \int_0^\infty \int_0^\infty \alpha(a,y)\ \beta(q)\ f(a,b,q,y)\ da\ db\ dq\ dy \quad (4)$$

and the feasibility constraints for $\alpha(a,y)$ and $\beta(q)$. Proposition 1 characterizes the optimal admission and referral policies.

Proposition 1 *Optimal admission and referral policies satisfy the following first order conditions:*

$$\alpha(a,y) \begin{pmatrix} = 1 \\ \in [0,1] \\ = 0 \end{pmatrix} \ as \int_0^\infty \int_0^\infty \Big(\beta(q)\ [v(b,y) - \lambda]\ -\ c\Big)\ f(b,q|\ a,y)\ db\ dq \begin{pmatrix} > 0 \\ = 0 \\ < 0 \end{pmatrix} \quad (5)$$

[15] If q is only a noisy measure of b, it may be useful to allow for retesting and multiple referrals. We do not consider this extension of our model in this paper.

and

$$\beta(q) \begin{pmatrix} = 1 \\ \in [0,1] \\ = 0 \end{pmatrix} \textit{ as } \int_0^\infty \int_0^\infty \int_0^\infty [v(b,y) - \lambda] \, \alpha(a,y) f(a,y,b|q) \, da \, dy \, db \begin{pmatrix} > 0 \\ = 0 \\ < 0 \end{pmatrix} \quad (6)$$

where λ is the Lagrange multiplier associated with the capacity constraint.

The expected value of referring a student type (a, y) for testing is given by:

$$V(a,y) = \int_0^\infty \int_0^\infty \beta(q) \, [v(b,y) - \lambda] \, f(b,q|a,y) \, db \, dq \quad (7)$$

The cost of testing is c. The first condition in equation (5) states that a student of type (a, y) is referred for testing if the expected value from testing that student type exceeds the cost of testing.

The second condition in equation (6) characterizes the optional admission policy. To get some additional insight define

$$W(q) = \frac{\int_0^\infty \int_0^\infty \int_0^\infty \; v(b,y) \, \alpha(a,y) f(a,y,b|q) \, da \, dy \, db}{\int_0^\infty \int_0^\infty \int_0^\infty \; \alpha(a,y) f(a,y,b|q) \, da \, dy \, db} \quad (8)$$

as the expected value to the district of a student of type q.[16] A student with IQ score q is admitted to the program if and only if $W(q)$ exceeds the shadow price λ associated with the capacity constraint.

[16] If the district also conditions the admission policy on achievement and income, then the FOC for optimal admission is given by:

$$\beta(a,y,q) \begin{pmatrix} = 1 \\ \in [0,1] \\ = 0 \end{pmatrix} \text{ as } \int_0^\infty [v(b,y) - \lambda] \, \alpha(a,y) f(b|a,y,q) \, db \begin{pmatrix} > 0 \\ = 0 \\ < 0 \end{pmatrix} \quad (9)$$

In particular, we have $\beta(a,q,y) = 0$ if $\alpha(a,y) = 0$.

Proposition 2 *There exists a threshold level $\bar{q}$ such that students above the threshold are admitted to the program while students below the threshold are not admitted to the program. The marginal expected treatment effect is determined by the capacity constraint and satisfies $W(\bar{q}) = \lambda$.*

Proposition 2 implies that the optimal design of the admission policy gives rise to a sharp regression discontinuity design. While a researcher may not know a district's objective function, increasing the achievement of students assigned to its program will be a prominent concern of any the district. Hence, it is natural to consider using regression discontinuity to investigate the impact of a gifted program on student achievement.[17]

It is well known that the RD estimator identifies the expected treatment effect of a marginal student, i.e. a student at the threshold $\bar{q}$. As Proposition 2 demonstrates, the expected treatment depends on the capacity constraint. The shadow value of the capacity constraint may be near zero in some districts, especially in urban districts that are experiencing declining enrollment and have excess capacity. Hence, an RD analysis may be useful to a district in assessing whether to undertake a marginal increase or decrease in enrollment, but it will be of limited usefulness in assessing the overall value of a district's gifted program.

Assumptions 1-3 imply that $V(a, y)$ is continuous and differentiable. To get additional insights, consider the joint density of (b, q) conditional on achievement and income, $f(b, q|a, y)$.

Assumption 6 *$f(b, q|a, y)$ is increasing in both a and y for each value of (b, q).*

[17] Recall that a is achievement prior to admission to a gifted program. In an RD analysis, the value added of the gifted program, $v(b, y)$, may be approximated by the gain in achievement due to the gifted program.

The former is a natural consequence of the positive association of IQ and achievement; the latter reflects the well-known tendency of students from higher income households to exhibit higher IQ. It then follows that

$$V_a(a, y) \ > \ 0 \tag{10}$$

The effect of y on V is less obvious. While $f(b, q|a, y)$ is increasing in y, the district valuation of participation in the program may be decreasing in y. We assume that the former dominates the latter, implying

$$V_y(a, y) \ > \ 0, \tag{11}$$

Summarizing the discussion above, we have:

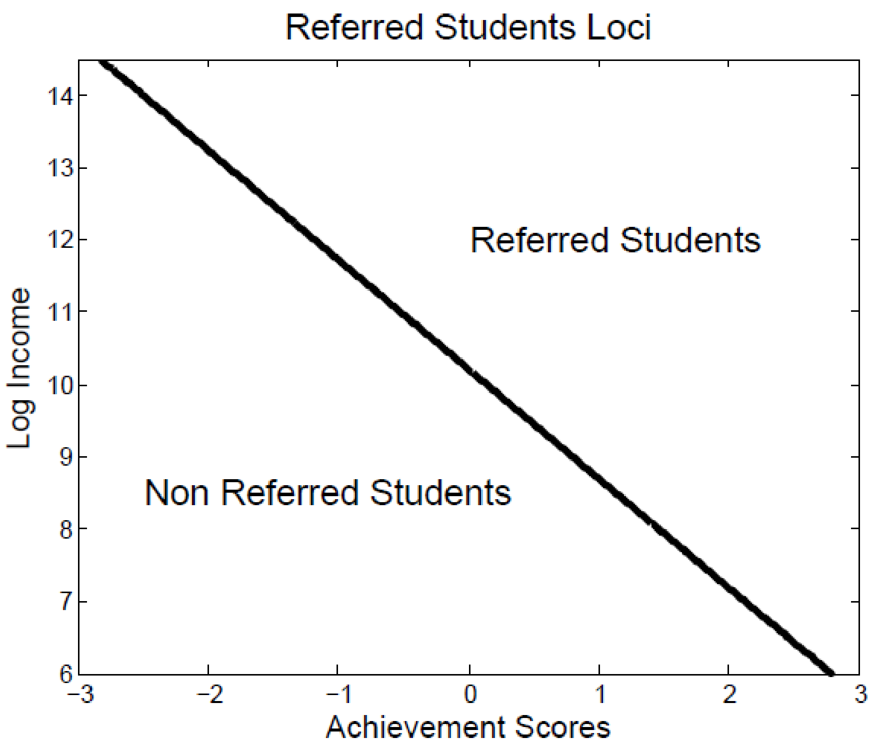
Assumption 7 *The function $V(a, y)$ is continuous and differentiable with $V_a(a, y) > 0$ and $V_y(a, y) > 0$ everywhere.*

Given Assumptions 6 and 7, the next proposition follows directly from the Implicit Function Theorem.

Proposition 3 *The equation $V(a, y) = c$ defines a downward sloping locus $a \ = \ a(y)$ in the (a, y) plane. Students on or above this locus are referred for testing. Those below the locus are not referred for testing.*

A downward sloping referral locus implies that students with higher incomes need lower achievement scores to be referred for testing since they are expected to perform better on the IQ test. Figure 1 illustrates the shape of this indifference locus for a numerical example.

Figure 1:



Referred Students Loci
Referred Students
Non Referred Students
Log Income
Achievement Scores
14
13
12
11
10
9
8
7
6
−3
−2
−1
0
1
2
3

We should point out that our model is sufficiently general to generate an upward sloping referral locus. A necessary condition for that is that $v_y(b, y) < 0$. i.e that the district assigns a negative value to higher income households. But this condition is clearly not sufficient if income is positively correlated with performance on IQ tests. As a consequence, the district has an incentive to adopt profiling techniques. We discuss these issues in more detail in the next section.

Finally, the district that we study chooses admission thresholds for gifted admission following guidelines set by the state. Policy for referral of students for testing is set at the district level. Hence, in our econometric model, we take admission thresholds as being set exogenously and we take referral policy as set by the district.[18]

3 Profiling

Students from disadvantaged backgrounds tend to be underrepresented in selective programs. A potential strategy to reduce this imbalance is via profiling to take account of factors, such as "hardship," that may cause underperformance relative to ability on achievement tests. Being a minority student or a student with FRL status are observable indicators that may be correlated with having experienced hardship. Different criteria for such students can potentially be adopted in either admission or referral policies. In this section, we undertake analysis to provide an economic foundation for understanding profiling in referral when affirmative action is not permitted.

To study the implications of profiling, we extend our model to consider two discrete types. Let m denote a student of a disadvantaged or minority type and M denote

[18] While the state defines gifted to be an IQ of 130, the state gives the district considerable latitude in setting lower thresholds for some groups of students, and latitude as well in conducting portfolio review for students near the thresholds. Our treatment of these additional thresholds is detailed in extensions of the model discussed below.

the advantaged or majority type.[19] Let $\beta_j(q)$ now denote the admission probability for type j. Affirmative action in admission arises if $\beta_M(q) < \beta_m(q)$.

To provide some insight into the nature of profiling, we develop analytic results using the following three simplifying assumptions:

Assumption 8

a) The admission rules cannot depend on j and must satisfy $\beta_M(q) = \beta_m(q)\ \forall q$.

b) Values are constant for each type and equal across types: $v_j(b, y) = v$ for $j = m, M$

c) There is no capacity constraint, i.e. $\lambda = 0$.

Assumption a) rules out that differences in referrals are driven by affirmative action in admission rules. Assumption b) implies that differences in referrals cannot be driven by differences in valuations across types. Assumption c) is made for expositional convenience and can be easily relaxed.

Definition 1 *The district employs profiling if referral for testing conditional on (a, y), depends on type, i.e. if $\alpha_m(a, y) \neq \alpha_M(a, y)$ for some (a, y).*

With type-blind admission standards (Assumption 8a), the district may set different referral criteria for different groups. Assumption 7 implies that type (j, a, y) is referred for testing if:

$$p_j(a, y) \ = \ \int_0^\infty \int_0^\infty \beta(q)\ f_j(b, q|a, y)\ db\ dq \geq \frac{c}{v} \tag{12}$$

where $p_j(a, y)$ is the probability that type (j, a, y) passes the test and is granted access to the program. The achievement of the marginal student of type j who is referred

[19] In our empirical application we focus on African Americans and students that are eligible for free or reduced lunch.

for testing is then obtained by solving for the value of a at which the preceding holds as an equality. The referral margin for type j is then given by the function $a_j(y)$.

Hardship may cause an individual to underachieve relative to potential. In particular, consider two students with the same observed a and y, one who has experienced hardship and the other who has not.[20] The one who has experienced hardship is more likely to have higher ability that the student who has not experienced hardship, i.e., the one who has experienced hardship has a higher probability of gaining admission to the gifted program. To formalize the notion of hardship, we use the following definition:

Definition 2 *Hardship of m relative to M is defined if for every (a, y), $p_M(a, y) < p_m(a, y)$.*

Conditional on measured (a, y) the disadvantaged group has a higher a probability of success on the admission test. The next assumption imposes some plausible monotonicity restrictions on the admission probabilities.

Assumption 9 *The probability of test success conditional on referral for a given demographic type, $p_j(a, y)$, is increasing in both of its arguments for $j \in \{M, m\}$*

Let $a_m(y)$ and $a_M(y)$ denote the referral loci for types m and M when profiling is permitted, and let, and $a_C(y)$ denote the common referral locus when profiling is prohibited. We then have the following result:

Proposition 4

a) If there is no type-contingent preference, i.e., if v is the same for both types, and

[20]This implies that income is not a sufficient statistic for hardship, e.g., for individuals of different races. For individuals of the same race, income may be a more accurate indicator of hardship.

admission criteria are the same for both types, there is profiling in referral for testing, with the referral locus for m lying below the referral locus for M.
b) If profiling is prohibited and admission criteria are the same for both types, then, for all y, $a_m(y) < a_C(y) < a_M(y)$.

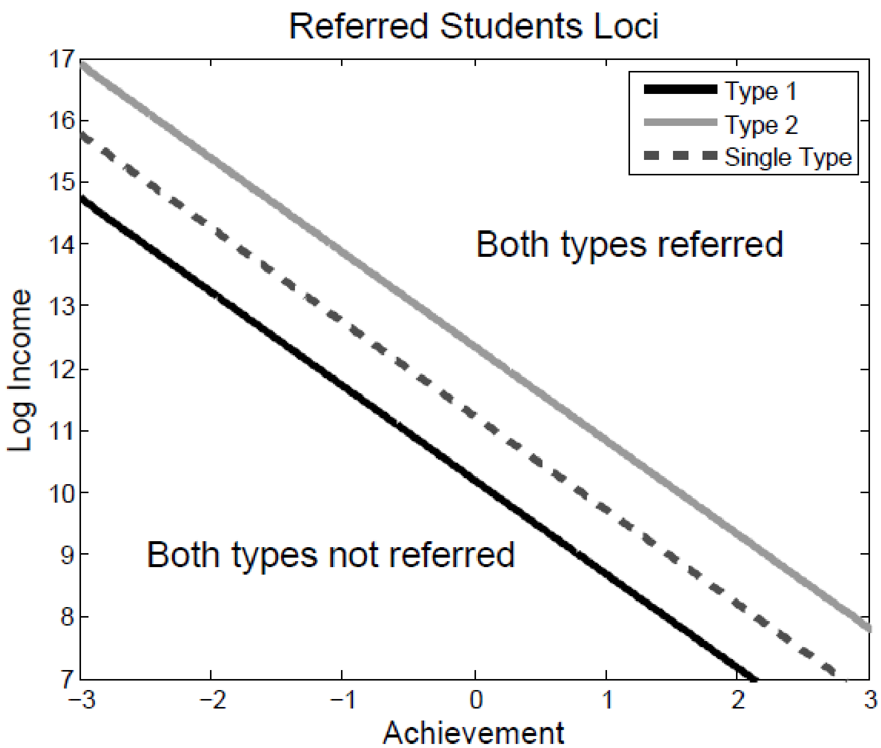
Proof:
a) Let (a', y') be a point on the referral threshold for type M: $p_M(a', y') = \frac{c}{v}$. Then $p_M(a, y) < p_m(a, y)$, Assumption 9, and the continuity of the distributions of (a, y) for each type imply that there exists (a'', y''), $a'' < a'$ and $y\prime < y''$, such that $p_m(a'', y'') = \frac{c}{v}$.
b) Let ρ be the proportion of type m in the student population. Let $\widehat{a}(y)$ be the referral locus when the proportion m is $\widehat{\rho}$. For each y, the latter referral locus satisfies $\widehat{\rho}\, p_m(\widehat{a}(y), y)) + (1 - \widehat{\rho}) p_M(\widehat{a}(y), y)) = \frac{c}{v}$. When $\widehat{\rho} = 0$, $\widehat{a}(y) = \alpha_M(y)$, and when $\widehat{\rho} = 1$, $\widehat{a}(y) = a_m(y)$. When $\widehat{\rho} = \rho$, $\widehat{a}(y) = a_C(y)$. Then the result follows from continuity and $a_m(y) < a_M(y)$. Q.E.D.

Figure 2 illustrates this result using a numerical example. Type 1 refers to minority students that have experienced hardship while type 2 refers to students that did not experience hardship.

Proposition 4a implies that there are (a, y) individuals from the group experiencing hardship, m, who are referred for testing when their advantaged counterparts, M, with the same (a, y) are not referred for testing. Between the two referral loci, a minority student is referred for testing when a majority student is not because the minority student is expected to perform better on the IQ test than a majority student with the same characteristics.

Proposition 4b demonstrates that prohibition of profiling results in a referral locus lies between the referral loci for types M and m. Relative to the case when profiling

Figure 2:



Referred Students Loci
Type 1
Type 2
Single Type
Both types referred
Both types not referred
Log Income
Achievement
17
16
15
14
13
12
11
10
9
8
7
−3
−2
−1
0
1
2
3

is permitted, prohibition of profiling decreases the proportion of type m referred for testing and increases the admission success rate of type m. Similarly prohibition of profiling increases the proportion of type M referred and decreases the success rate of type M. If profiling is not permitted, the proportion referred for testing will be higher for type M than type m. The success rate conditional on referral will be higher for m than for M.

In Proposition 4, we assume that the value added that the district attaches to admission is the same for both student types, i.e., $v_m = v_M$. Note that this equality and Assumption 8a are conditions sufficient for there to be no affirmative action. In particular, if we retain Assumption 8a but replace Assumption 8b with $v_m > v_M$, then differential referral criteria would be chosen even if type m experienced no hardship relative to type M .

4 Estimation

4.1 Observed Outcomes

Let R denote a discrete random variable that is equal to one if the student is referred for testing and zero otherwise. Let A denote a discrete random variable that is equal to one if the student is accepted into the selective program and zero otherwise. We characterize the information set of the econometrician.

Assumption 10 *In a random sample of student, we observe the following outcomes.*

1. *We observe R and A for all students in the sample.*

2. *For students who are referred for testing $R = 1$, and we observe q.*

3. *We observe income and achievement with error, denoted by* $\tilde{y}$ *and* $\tilde{a}$*.*

4. *We do not observe* b *for any student.*

5. *We observe race and subsidized lunch status.*

The assumption that income and achievement are measured with error is realistic and helps us generate a non-degenerate likelihood function for our model as described in detail below.

4.2 A Parametrization

In our application the district adheres to a predetermined admission policy that can be characterized by the following two equations:

$$\begin{aligned} \beta_m(b,q) &= 1(Q_m \leq q) + 1(q_m \leq q < Q_m)\,1(b_m \leq b) \\ \beta_M(b,q) &= 1(Q_M \leq q) + 1(q_M \leq q < Q_M)\,1(b_M \leq b) \end{aligned} \tag{13}$$

This specification relax our assumption in Section 2 and allow the district to use ability in admission as part of a comprehensive "portfolio review" of borderline cases. We assume the portfolio review reveals b. For each type, there are three thresholds in the admission function. Students whose IQ score, q, is above the upper IQ threshold, Q_j, gain admission, and students with IQ score below the lower IQ threshold, q_j, do not gain admission. Students between these IQ thresholds gain admission if ability, b, exceeds the ability threshold, b_j. We can estimate q_j to be the smallest value of q such that there exists at least one student in the sample who is admitted to the program with that level of IQ. Q_j can be estimated by the minimum of the IQ score such that the admission probability is one. The parameters b_m and b_M are estimated as part of the maximization of the likelihood function as detailed in Section 4.3.

The monetary cost of testing, c, is, for simplicity, treated as known to the econometrician and is set equal to \$100.

We assume that $(a, b, \ln y)$ are jointly normally distributed. Measured IQ equals ability plus a normally distributed error that is independent of $(a, b, \ln y)$:[21]

$$q = b \ + \ \epsilon_q$$

Errors for observed income and achievement are classical measurement errors with variances the same across types.

The district's value function for type j is given by:

$$v_j(b, y) = e^{\zeta_{b,j} \ b \ + \ \zeta_{y,j} \ \ln(y)}. \tag{14}$$

This functional form for $v_j(b, y)$ simplifies the computational burden by permitting the terms multiplying $\beta_j(b, q)$ in the integrand for $V_j(a, y)$ to written as a multivariate normal multiplied by a constant that is a function of the parameters of the joint distribution of $(a, b, \ln y)$.

The admission rules define two loci, $y_m(a)$ and $y_M(a)$, for minority and majority students respectively:

$$y_j(a) : V_j(a, y) = \int_0^\infty \int_0^\infty \beta_j(b, q) \ v(b, y) \ f_j(b, q|a, y) \ db \ dq \ = \ c$$

[21]Heckman (1995) presents evidence that the predictive power of IQ as a measure of academic ability differs very little by race or gender.

4.3 The Likelihood Function

In our application, we consider four types using race and free or reduced lunch status to defined each type. We begin by specifying an element of the likelihood for a given type, j. With measurement error in income and achievement, the probability of observing $R = 0$ conditional on $(j, \tilde{a}, \tilde{y})$ is given by:

$$\begin{aligned} Pr\{R = 0 \,|j, \tilde{a}, \tilde{y}\} &= \int (1 - \alpha_j(a, y))\, f_j(a, y|\tilde{a}, \tilde{y})\, dy\, da \\ &= \int \int_0^{y_j(a)} f_j(a, y|\tilde{a}, \tilde{y})\, dy\, da \end{aligned} \tag{15}$$

Similarly, the probability of observing $R = 1$ and $A = 1$:

$$\begin{aligned} & Pr\{R = 1, A = 1|\; j, q, \tilde{a}, \tilde{y}\} \\ = & \int \int \int \alpha_j(a, y)\, \beta_j(b, q)\, f_j(a, y, b|q, \tilde{a}, \tilde{y})\, dy\, da\, db \\ = & \begin{cases} \int \int_{y_j(a)} f_j(a, y|q, \tilde{a}, \tilde{y})\, dy\, da & \text{if } q \geq Q_j \\ \int \int_{y_j(a)} \int_{b_j} f_j(a, y, b|q, \tilde{a}, \tilde{y})\, db\, dy\, da & \text{if } Q_j > q \geq q_j \\ 0 & \text{if } q_j > q \end{cases} \end{aligned} \tag{16}$$

Finally, the probability of observing $R = 1$ and $A = 0$ is given by:

$$\begin{aligned} & Pr\{R = 1, A = 0|\; j, q, a, \tilde{y}\} \\ = & \int \int \int \alpha_j(a, y)\, (1 - \beta_j(q, b))\, f_j(a, y, b|\tilde{a}, \tilde{y}, q)\, db\, dy\, da \\ = & \begin{cases} 0 & \text{if } q \geq Q_j \\ \int \int_{y_j(a)} \int_0^{b_j} f_j(a, y, b|q, \tilde{a}, \tilde{y})\, db\, dy\, da & \text{if } Q_j > q \geq q_j \\ \int \int_{y_j(a)} f_j(a, y|q, \tilde{a}, \tilde{y})\, dy\, da & \text{if } q_j > q \end{cases} \end{aligned} \tag{17}$$

The likelihood for a single observation is then given by:

$$
\begin{aligned}
L \;=\;& [Pr\{R=0|\; j,\tilde{a},\tilde{y}\}\; f_j(\tilde{a},\tilde{y})]^{(1-R)} \\
& [Pr\{R=1, A=1|\; j,q,\tilde{a},\tilde{y}\}\; f_j(\tilde{a},\tilde{y},q)]^{R\,A} \\
& [Pr\{R=1, A=0|\; j,q,\tilde{a},\tilde{y}\}\; f_j(\tilde{a},\tilde{y},q)]^{R\,(1-A)}
\end{aligned}
\tag{18}
$$

The likelihood for a sample of N students is then a straightforward product of the above for the N observations.

Summarizing, we have introduced a parameterization of the model and shown how to consistently estimate the parameters of the model using a Maximum Likelihood estimator. We rely on measurement error in income and prior achievement, both quite natural in this application, to generate a well-defined likelihood function.[22]

4.4 A Monte Carlo Study

We conducted a Monte Carlo study to get some additional insights into the small sample properties of our ML estimator. We generated 50 samples with sizes equal to 5000 observations. Table 1 summarizes the findings of the Monte Carlo exercise. It reports the true parameter values under which the data were generated, the mean of the estimates, the standard deviation of the estimates and the sample mean of the estimated asymptotic standard errors.

We find that the small sample bias of most parameters is negligible. The only

[22] An alternative approach for estimating the model is to treat ability as observed by the district and unobserved by the econometrician. One could then model then treat the observed achievement scores as different measurements of ability and derive a likelihood function of the model using the techniques in Carneiro, Hansen, & Heckman (2003) and Cunha & Heckman (2008). As in our model, identification can be achieved without imposing strong linearity or normality assumption as discussed in Schenach (2004) and Cunha, Heckman, & Schenach (2010).

Table 1: Monte Carlo Results

	True Value	Mean	Std Dev.	Asymp. Errors
Ability: mean	100	100	0.75	0.65
Ability: std. dev.	13	13	0.54	0.51
IQ error: std.dev.	7	6.9	0.52	0.51
Achievement error: std. dev.	0.3	0.3	0.01	0.01
Income error: Std. dev.	0.05	0.07	0.05	0.08
Ability income: correlation	0.5	0.5	0.03	0.03
Achievement income: correlation	0.3	0.3	0.03	0.03
Ability: utility parameter	0.055	0.055	0.01	0.01
Income: utility parameter	-0.25	-0.26	0.09	0.08
Ability threshold FRL	115	115	0.40	0.40
Ability threshold no FRL	120	120	0.53	0.56

Simulations: 50

parameters that are slightly more difficult to estimate in small samples are the parameters of the value function.

5 Data

Our application focuses on a selective gifted program that is operated by a mid-sized urban district that prefers to stay anonymous. The district operates a pull-out program that is administered by a Gifted Center that serves students in elementary and middle school. Gifted students in grades 1 through 8 participate in a one-day-per-week program at a designated location away from the student's home school. Students participate in programs designed to enhance creative problem solving and leadership skills and are offered specially designed instruction in math, science, literature, and a variety of other fields.

The district adheres to state regulations concerning gifted students and services. The state regulations outline a multifaceted approach used to identify whether a student is gifted and whether gifted education is needed. The state requires gifted status to be determined by a certified district psychologist. A mentally gifted student who is not on free and reduced lunch is defined as someone with an IQ score of at least 130. The regulation specifies that a student with IQ score below 130 may be admitted "... when other educational criteria in the *profile* of the person strongly indicates gifted ability (emphasis added)." The state guidelines provide for consideration of factors that may "mask" giftedness including "... gender or race bias, or socio/cultural deprivation..." We treat these admission guidelines as exogenous to the district and focus on modeling and analysis of the district's implementation as reflected in referral and admissions data.

Most referrals and admissions for the gifted program occur during elementary school in grades 1-4. We focus on this population in this paper. The sample consists of the cohort of students that were in 1^{st} grade in 2004/05. We follow these students until the end of $4th$ grade. Pre-referral achievement scores are constructed based on the Oral Reading Fluency (ORF) test. We standardize ORF scores for every period in every grade and compute the average score for each student. In addition we have access to Pennsylvania System of School Assessment (PSSA) scores that measure reading and math skills in 3rd grade. We construct our prior achievement measures by averaging over both types of scores. The ORF measures have the advantage that they are available for young students. However, they tend to be more noisy measures of achievement than PSSA scores.[23]

Income is measured by median neighborhood (Census tract) income and taken

[23] We experimented with different measures of achievement scores in the empirical analysis reported below. Overall, the main findings are quite robust. Results are available upon request from the authors.

from the U.S. Census. Recall that district policymakers do not have measures of income of individual households. Hence, neighborhood income is likely to be a key part of the information set used by policymakers to infer household income. FRL refers to free or reduced lunch status. To assign an income level to the FRL cut-off we use the average threshold for the years 2006, 2007 and 2008, for 3.5 household members.

We start with a sample of 3,306 students. We keep those students who were at the school district during the first four grades of school. This reduces our sample by roughly one half, to 1,592 students. We drop 55 observations without lunch status data, and 185 that changed their lunch status over the four school grades. We eliminate 9 observations without income data, and 13 other observations without prior achievement data. Finally, we eliminate duplicate observations that correspond to students who were referred more than once. The final sample size is 1,326 observations.

Table 2 reports descriptive statistics for the full sample as well as the subsamples by FRL status and race. We find that the sample is predominantly poor and the

Table 2: Sample Statistics by FRL status and Race

		Lunch status		Race	
	All types	Non-FRL	FRL	Non Black	Black
Sample size	1326	345	981	585	741
Free or Reduced lunch	0.74			0.51	0.92
Black in sample	0.56	0.17	0.70		
Achievement avg.	-0.02	0.74	-0.29	0.37	-0.33
N. Income avg.	28,363	36,786	25,401	33,962	23,944
Fraction Referred	0.20	0.50	0.09	0.31	0.11
Fraction Gifted	0.10	0.30	0.03	0.19	0.03
Ratio Gifted/Referred	0.51	0.60	0.36	0.61	0.32

majority of students are black which is typical for many urban districts. Achievement and income are negatively correlated with FRL status and race. Table 2 also shows that 92 percent of all African American students in our sample are eligible for free or reduced lunch. Only 51 percent of non-black students in our sample qualify for subsidized lunch.

Table 3 reports additional descriptive statistics by referral and admission status. We find that gifted students have higher achievement scores, income and IQ than non-gifted students. Similarly, referred students have higher achievement scores and income than non-referred students.

Recall that income is measured as mean neighborhood income. For non-black students, neighborhood income is $38,296 and $30,028 depending on whether they are on FRL or not. For black students, the neighborhood income levels are $29,932 and $23,439, respectively.

As noted above, the state guidelines encourage consideration of measures of hardship in determining gifted eligibility. To comply with these guidelines, the district adopts different admission policies for different groups of students. We test for affirmative action in admission for students that come from low-income households, indicated by FRL status, and for affirmative action with respect to race.

Table 4 divides the sample of tested students into six IQ scores intervals. It reports the percentage of students admitted in each interval for each type of student. The upper panel of Table 4 breaks down the sample by free or reduced lunch status, the lower panel by race. We find that the the admission probabilities are increasing in IQ score. For students who are not on FRL, the probability of acceptance is, for all practical purposes, equal to one if the IQ score is larger than 125. It drops to 12 percent if the IQ is below 115. For students on FRL, the admission probability is

Table 3: Statistics for Non Referred, Referred and Gifted Students

All Students				
			Referred students only	
	Non referred	Referred	Non Gifted	Gifted
Free or Reduced lunch	0.84	0.34	0.45	0.23
Black in sample	0.62	0.31	0.43	0.19
Achievement avg.	-0.32	1.21	0.94	1.46
N. Income avg.	26,620	35,545	31,181	39,616
IQ avg.			109.6	125.3
Non-FRL Lunch lunch Students				
			Referred students only	
	Non referred	Referred	Non Gifted	Gifted
Black in sample	0.21	0.13	0.22	0.07
Achievement avg.	0.07	1.41	1.18	1.56
N. Income avg.	32,995	40,600	36,138	43,589
IQ avg.			111.9	127.0
Free or Reduced lunch Students				
			Referred students only	
	Non referred	Referred	Non Gifted	Gifted
Black in sample	0.70	0.66	0.70	0.58
Achievement avg.	-0.40	0.83	0.66	1.14
N. Income avg.	25,387	25,552	25,073	26,417
IQ avg.			106.7	119.7
Non Black Students				
			Referred students only	
	Non referred	Referred	Non Gifted	Gifted
Free or Reduced lunch	0.66	0.17	0.24	0.12
Achievement avg.	-0.08	1.38	1.09	1.56
N. Income avg.	31,444	39,626	35,847	42,087
IQ avg.			111.3	126.4
Black Students				
			Referred students only	
	Non referred	Referred	Non Gifted	Gifted
Free or Reduced lunch	0.95	0.72	0.72	0.72
Achievement avg.	-0.47	0.84	0.75	1.02
N. Income avg.	23,669	26,246	25,045	28,841
IQ avg.			107.3	120.7

Table 4: Ratios of Gifted Students by IQ bins, by FRL status

	Non-FRL Lunch		FRL	
IQ score	total	% Admitted	total	% Admitted
<110	30	0.13	40	0.08
110≤ <115	17	0.12	12	0.17
115≤ <120	28	0.50	19	0.53
120≤ <125	28	0.50	8	1.00
125≤ <130	30	1.00	5	1.00
130≤	39	1.00	3	1.00
	Non Black		Black	
IQ score	total	% Admitted	total	% Admitted
<110	33	0.12	37	0.08
110≤ <115	20	0.15	9	0.11
115≤ <120	30	0.57	17	0.41
120≤ <125	27	0.56	9	0.78
125≤ <130	33	1.00	2	1.00
130≤	37	1.00	5	1.00

one for all students with score larger than 120. It drops to 17 percent for students with IQ between 110 and 115. We see evidence suggesting that students on FRL face lower admission standards.

Next we consider admission standards by race shown in the lower panel of Table 4. There are two jumps in admission probabilities: the first jump is at 115 and the second at 125. Black students have lower admittance ratios at the range between 110 and 115, but higher admission probabilities in the range between 120 and 125. Overall, in the 115 to 125 interval, the admission rate for black students is slightly lower than for non-black students. Hence, there is little evidence that suggests affirmative action based on race.

Recall from the discussion following equation (12) that the upper thresholds, Q_j, are estimated by observing the threshold above which all students are admitted. From Table 4, we see that all students on FRL with IQ score above 120 obtain admission while students not on on FRL face a higher cut-off level of 125 for guaranteed admission. Hence, Q_M and Q_m are 120 for FRL and 125 for non-FRL. The IQ threshold for portfolio review is the level below which admission to the gifted program is zero. From We also find that, regardless of FRL status, this lower cutoff is 103. Hence, $q_M = q_m = 103$. Students of type j scoring between q_j and Q_j obtain portfolio review. We use these admission bounds when we estimate the structural parameters of the model.

Table 5 reports income and prior achievement by gifted status. We find that gifted students typically have higher income levels, higher prior achievement scores, and, by construction, higher IQ scores than students that are not gifted. The only notable exception is that we find little differences in neighborhood incomes and prior test scores for black students on free and reduced lunch. These students only differ by their performance on the IQ test.

Table 5: Gifted Status by FRL and Race

		Non Gifted			Gifted	
		Non-FRL	FRL		Non-FRL	FRL
Achievement avg.	Non Black	1.19	0.79	Non Black	1.57	1.45
N. Income avg.		37,665	30,072		43,968	28,204
IQ avg		113	108		127	121
Achievement avg.	Black	1.15	0.60	Black	1.33	0.91
N. Income avg.		30,638	22,894		38,391	25,127
IQ avg.		110	106		125	119

Finally, we provide some evidence of differential treatment in referral. Table 6 reports income and prior achievement by referral status. We find that there are large differences in prior achievement among students who are referred for testing and those who are not. Differences in income are only notable for students who are not on FRL.

Table 6: Referral Status by FRL and Race

		Non Referred			Referred	
		Non-FRL	FRL		Non-FRL	FRL
Achievement	Non Black	0.11	-0.18	Non Black	1.44	1.07
N. Income		34,256	30,007		41,699	29,262
Achievement	Black	-0.07	-0.49	Black	1.20	0.70
N. Income		28,199	23,408		33,105	23,599

6 Empirical Results

6.1 Overview of Main Results

We estimate and compare six different model specifications to explore the extent to which the district engages in profiling and/or affirmative action in referral and admission decisions. It is useful to begin by summarizing the set of parameters to be estimated. As detailed below, we find no role for estimated income after allowing for four race-by-FRL types. Anticipating this, we simplify the exposition by providing a summary of number of parameters in the model when income does not appear.

In principle, the district might engage in affirmative action in admission based on race or FRL status. Admission thresholds q_j and Q_j vary by type as does the portfolio review threshold b_j. With four types, there are then 12 parameters in the admission function.

The district may also have differential preferences with respect to participation by race or FRL. This is captured by four utility function parameters. The remaining parameters are those appearing in the distributions of types. For each type, we have means and covariances for ability and achievement. This yields five parameters per type for a total of 20 parameters. In addition, there are two measurement error variances–for achievement and IQ. Hence, the most general specification of the model has 38 parameters.

With the potential for profiling by type, affirmative action by type with respect to thresholds, and affirmative by type with respect to district preferences, there are a large number of potential policy combinations. Fortunately, we are able to isolate the policies used in practice by estimating six different models. Table 7 summarizes how the six different models estimated in this paper differ and reports the likelihood function values (LnL) associated with the models.

It proves to be quite instructive to begin by permitting no variation in policy by race. This is done in Models I through III. These models permit variation in treatment only with respect to FRL status. Model II permits profiling with respect to FRL and affirmative action with FRL thresholds but not preferences. It is the clear choice by LnL among these three. A more restrictive policy (Model I) that does not permit profiling substantially reduces LnL. On the other hand, extending Model II to permit preference variation by FRL status yields a negligible gain in fit as shown by LnL for Model III. Hence, among models that allow no racial preference, the best fitting alternative, Model II, permits profiling with respect to FRL and affirmative action with respect to the FRL admission threshold.

Next, consider Models IV through VI. All three permit profiling with respect to race while exploring variation in profiling and affirmative action thresholds. Among these, Model V is the most general, permitting differential treatment by both race and FRL on all policies except affirmative action with respect to preferences in FRL. Relative to Model V, Model IV eliminates both forms of affirmative action with respect to FRL. This yields a very significant reduction in fit, establishing the dominance of V over IV. Model VI restricts Model V by eliminating affirmative with respect to the race threshold, resulting in a negligible reduction in fit. Hence, model VI is preferred among this group.

Finally, we compare the dominant models from the two groups, II and VI. We see that II yields a significant reduction in fit relative to Model VI. Hence, we conclude that Model VI is the preferred alternative among this group of six. In short, no generalization of VI improves fit, and every restriction of VI significantly reduces the fit. We conclude that the district engages in profiling based on FRL status and race. It also engages in affirmative action in admission based on FRL status. There is no

Table 7: Model Summary

Model	Profiling		Affirmative Action Admission Threshold		Affirmative Action Preferences		Likelihood
	FRL	Race	FRL	Race	FRL	Race	
I	No	No	Yes	No	No	No	-3290
II	Yes	No	Yes	No	No	No	-3101
III	Yes	No	Yes	No	Yes	No	-3100
IV	No	Yes	No	Yes	No	Yes	-3179
V	Yes	Yes	Yes	Yes	No	Yes	-3072
VI	Yes	Yes	Yes	No	No	No	-3073

evidence that the district engages in affirmative action based on race. There is also no evidence suggesting preference based affirmative action with respect either to race or FRL.

Finally, we have explored specifications of our model in which referral decisions depend on income in addition to FRL status. In one model the school district observes income and achievement and makes referral decisions based on both variables while we observe individual achievement with error as well as the mean and standard deviation of income for the school district. Overall, we find that, conditioning on FRL status, adding a continuous income measure to our model does not improve the model fit. The finding that income plays no distinct role is potentially attributable to the following considerations. The district does not directly observe household income and would need to infer income informally from neighborhood of residence, interactions with household members, and similar observables. The lack of transparency associated with such inferences is clearly problematic and introduces potential for allegations of favoritism. By contrast FRL status is determined by criteria set externally to the district and therefore provides a transparent indicator for application of differential admission standards.

6.2 Parameter Estimates and Model Fit

Next we discuss the estimation results in further detail. We have seen from Table 7 that Model III provides no gain over II, and Model V provides no gain over VI. Hence, in Tables 8 and 9, we focus in on goodness of fit statistics by FRL and race respectively for the remaining four models. We see from Table 8 that the Model I fits the characteristics of FRL students well, but fails to explain the admission and referral outcomes for non-FRL students. Model II fits both subsamples well. A likelihood ratio of test of Model II against Model I shows the restrictions of the latter to be rejected with $p = .08$. These observations reinforce our conclusion above that that Model II is strictly preferred to Model I.

Table 9 reports goodness of fit statistics conditional on race. Recall that race does not play a role in any of the first three models that we have estimated. While Model II fits the conditional distributions by FRL status well, it does less well explaining the sorting by race. In particular, Table 9 suggests that Model II over-predicts the fraction of non-black students who are referred for testing and under-predicts the fraction of non-black students that are admitted to the program. Similarly, Model II under-predicts the faction of black students that are referred for testing and over-predicts the fraction of blacks students that are admitted to the program.

Recall that Model IV is similar to Model II but controls for race, instead of FRL status. Table 8 shows that Model IV is inconsistent with the observed sorting by

FRL status. Overall, Model IV fits the data much worse than Model II, supporting the judgment that FRL status is more important in explaining referral and admission policies than race.

Turning to Model VI, we see that the goodness of fit statistics reported in Tables 8 and 9 for Model VI accord well with the observed sorting by both FRL status and race. Hence, these goodness-of-fit comparisons support the conclusion that Model VI is the preferred alternative.

Our finding that Model VI fits the data well carries the implication that the district does not apply different criteria in admission decisions for African American students– either with respect to admission thresholds or preferences. Profiling with respect to race enhances prediction of performance by race, but does not imply preferential treatment by race. Hence, we conclude that the district provides no preferential treatment by race. By contrast, the district does give preferential treatment with respect to FRL status in setting admission thresholds.

We now turn to the parameter estimates. We report our preferred specification, Model VI, in Table 10. We also report the estimates for Model II to illustrate the importance of allowing for profiling based on race. Beginning with mean achievement estimates in the top panel, we see that students on FRL have lower achievement than their non-FRL, same-race counterparts. We see as well that non-black students have higher achievement than black students. The standard deviations of achievement exhibit some variation across race and FRL, with variance of non-white, non-FRL students being moderately higher than for the other three student types.

Our estimates of the parameters of ability are shown in the second panel. The means of ability exhibit a striking contrast to the means of achievement. Mean ability estimates for those on FRL are virtually the same for black and non-black students.

Table 8: Goodness of Fit by FRL Status

	Data	Model I	Model II	Model IV	Model VI
Non-FRL lunch Students					
Non Referred	50.1	80.5	50.9	72.5	51.2
$IQ \geq Q_{sup}$	20	4.6	16.8	8.5	17.0
$Q_{sup} > IQ \geq Q_{inf}, b \geq B_M$	9.9	5.6	11.1	6.8	11.1
$Q_{sup} > IQ \geq Q_{inf}, b < B_M$	18	7.4	18.8	10.5	18.3
$IQ < Q_{inf}$	2	2	2.3	1.7	2.4
Avg. Achieve. non Referred	0.1	-0.3	0.1	-0.2	0.1
Avg. IQ non Referred		93.1	101.6	97.4	102.9
Avg. Achieve. Referred	1.4	1.2	1.4	1.3	1.4
Avg. IQ Referred	121	117.1	120.9	119.8	121.0
Avg. Achieve. Referred non Gifted	1.2	1.0	1.1	1.0	1.1
Avg. IQ Referred non Gifted	111.9	109.5	112.4	111.6	112.3
Avg. Achieve. Referred Gifted	1.6	1.4	1.6	1.6	1.7
Avg. IQ Referred Gifted	127	124.2	127.4	126.3	127.4
FRL Students					
Non Referred	91.1	80.3	90.9	83.2	90.9
$IQ \geq Q_{sup}$	1.6	7.7	1.7	3.7	1.8
$Q_{sup} > IQ \geq Q_{inf}, b \geq B_M$	1.5	2.4	1.6	3.8	1.6
$Q_{sup} > IQ \geq Q_{inf}, b < B_M$	4.1	7.6	4.0	7.3	4.0
$IQ < Q_{inf}$	1.6	1.9	1.7	2.0	1.8
Avg. Achieve. non Referred	-0.4	-0.3	-0.4	-0.4	-0.4
Avg. IQ non Referred		93.1	90.5	91.1	89.1
Avg. Achieve. Referred	0.8	1.2	0.9	1.1	0.9
Avg. IQ Referred	111.3	117.2	111.6	116.4	111.6
Avg. Achieve. Referred non Gifted	0.7	1.0	0.8	0.9	0.7
Avg. IQ Referred non Gifted	106.7	108.8	106.7	109.5	106.6
Avg. Achieve. Referred Gifted	1.1	1.3	1.0	1.4	1.1
Avg. IQ Referred Gifted	119.7	125.2	119.8	124.9	120.1

Table 9: Goodness of Fit by Race

Non Black Students Students					
	Data	Model I	Model II	Model IV	Model VI
Non Referred	69.2	80.4	71.3	69.2	69.6
$IQ \geq Q_{sup}$	12.3	6.2	9.1	10.0	10.8
$Q_{sup} > IQ \geq Q_{inf}, b \geq B_M$	6.3	4.0	6.3	7.7	7.2
$Q_{sup} > IQ \geq Q_{inf}, b < B_M$	10.9	7.5	11.3	11.5	10.9
$IQ < Q_{inf}$	1.2	2.0	2.0	1.6	1.6
Avg. Achieve. non Referred	-0.1	-0.3	-0.2	-0.1	-0.1
Avg. IQ non Referred		93.1	94.4	99.8	97.2
Avg. Achieve. Referred	1.4	1.2	1.3	1.4	1.4
Avg. IQ Referred	120.5	117.2	119.4	120.4	120.5
Avg. Achieve. Referred non Gifted	1.1	1.0	1.0	1.1	1.1
Avg. IQ Referred non Gifted	111.3	109.1	111.2	112.0	111.8
Avg. Achieve. Referred Gifted	1.6	1.4	1.5	1.6	1.6
Avg. IQ Referred Gifted	126.4	124.7	126.6	126.5	126.6
Black Students					
Non Referred	89.3	80.3	87.8	89.3	89.3
$IQ \geq Q_{sup}$	1.8	7.5	2.9	1.0	1.7
$Q_{sup} > IQ \geq Q_{inf}, b \geq B_M$	1.6	2.7	2.3	2.1	1.6
$Q_{sup} > IQ \geq Q_{inf}, b < B_M$	5.7	7.6	5.2	5.4	5.2
$IQ < Q_{inf}$	1.6	1.9	1.7	2.1	2.2
Avg. Achieve. non Referred	-0.5	-0.3	-0.4	-0.5	-0.5
Avg. IQ non Referred		93.1	91.0	88.2	87.8
Avg. Achieve. Referred	0.8	1.2	1.1	0.8	0.8
Avg. IQ Referred	111.5	117.2	114.5	111.5	111.5
Avg. Achieve. Referred non Gifted	0.8	1.0	0.9	0.7	0.7
Avg. IQ Referred non Gifted	107.3	108.8	108.1	107.6	107.1
Avg. Achieve. Referred Gifted	1.0	1.4	1.3	1.1	1.1
Avg. IQ Referred Gifted	120.7	125.1	123.0	121.1	121.1

Table 10: Parameter Estimates

	MODEL II	MODEL VI
Mean Achievement	0.74	0.80
	(0.05)	(0.06)
Mean Achievement FRL	-0.29	-0.05
	(0.03)	(0.05)
Mean Achievement Black	0.74	0.41
	(0.05)	(0.12)
Mean Achievement FRL Black	-0.29	-0.39
	(0.03)	(0.03)
Std Dev Achievement	0.99	0.99
	(0.04)	(0.04)
Std Dev Achievement FRL	0.85	0.87
	(0.05)	(0.04)
Std Dev Achievement Black	0.99	0.88
	(0.04)	(0.08)
Std Dev Achievement FRL Black	0.85	0.83
	(0.05)	(0.02)
Mean Ability	111.0	112.8
	(1.5)	(1.5)
Mean Ability FRL	92.3	93.7
	(5.0)	(6.0)
Mean Ability Black	111.0	106.3
	(1.5)	(4.1)
Mean Ability FRL Black	92.3	94.9
	(5.0)	(6.2)
Std Dev Ability	13.0	12.5
	(1.5)	(1.5)
Std Dev Ability FRL	11.8	12.2
	(2.5)	(3.5)
Std Dev Ability Black	13.0	10.7
	(1.5)	(3.6)
Std Dev Ability FRL Black	11.8	10.2
	(2.5)	(2.9)
Std Dev Ability Error	7.3	7.2
	(0.6)	(0.6)
Std Dev Achievement Error	0.57	0.55
	(0.02)	(0.02)

Table 10: Parameter Estimates (cont.)

	MODEL II	MODEL VI
Correlation Ability Achievement	0.94	0.94
	(0.03)	(0.03)
Correlation Ability Achievement FRL	0.91	0.93
	(0.06)	(0.05)
Correlation Ability Achievement Black	0.94	0.78
	(0.03)	(0.15)
Correlation Ability Achievement FRL Black	0.91	0.84
	(0.06)	(0.12)
Utility Ability	0.060	0.058
	(0.006)	(0.005)
Utility Ability Black	0.060	0.058
	(0.006)	(0.005)
Ability Threshold Non-FRL	119.3	119.5
	(0.9)	(0.9)
Ability Threshold FRL	114.1	114.2
	(1.2)	(1.1)
Ability Threshold Non-FRL Black	119.3	119.5
	(0.9)	(0.9)
Ability Threshold FRL Black	114.1	114.2
	(1.2)	(1.1)
Achieve. Threshold Non-FRL	0.76	0.78
Achieve. Threshold FRL	0.57	0.80
Achieve. Threshold Non-FRL Black	0.76	0.64
Achieve. Threshold FRL Black	0.57	0.45
Log likelihood	-3101.2	-3072.6

Sample size: 1326

For students not on FRL, non-black students have a higher mean than black students, but the relative difference is substantially smaller than for achievement. The contrast between the variation in means of achievement by race and FRL to the variation in means of ability by race and FRL provides insight into our finding (Table 7) of significant scope for profiling by race and FRL. The standard deviations of ability are modestly higher for non-FRL than for FRL students.

Panel four provides further evidence of the basis for profiling. The correlation between achievement and ability is substantially higher for non-black than for black students. Differences in correlation across FRL status within race are much smaller, especially for non-black students. The fourth panel reveals, not surprisingly, that there is substantial error in measurement of both ability and achievement.

From the fifth panel, we see that the coefficient of utility for ability is .058. From panel two of Table 10, we see that the standard deviation of ability is on the order of 10. Hence, given two students differing in ability by one standard deviation, the district planner attaches approximately 75% higher monetary value to to having the higher-ability student participate in the gifted program.

From the sixth panel, we see the magnitude of affirmative action with respect to FRL students; the ability threshold for FRL students is approximately one half standard deviation below the ability threshold for non-FRL students.

The seventh panel in Table 10 presents profiling thresholds. While our estimated model is more general than the model underlying Proposition 4, that proposition gives insight into our findings with respect to profiling. As discussed above, our findings in Table 10 reveal that mean estimated ability is virtually the same for black and non-black students on FRL, but achievement of black students on FRL is substantially lower than achievement of non-Black students on FRL. Hence, in accordance with Proposition 4, the referral threshold for black students on FRL (.45) is lower than for non-black students on FRL (.80). For students not on FRL, black students have a lower referral threshold (.64) than non-black students (.78). This too is consistent with lower achievement relative to ability for black than non-black students. Interestingly, the referral threshold for non-black students is virtually the same for FRL and non-FRL students. This suggests that achievement relative to ability is roughly comparable for non-black students regardless of FRL status.

7 Policy Analysis

In 2003, the US Supreme Court considered two affirmative action cases against the University of Michigan, naming then-president Lee Bollinger, as defendant. One, Grutter v. Bollinger (2003), challenged the admission policy of the law school. The other, Gratz v. Bollinger (2003), challenged the undergraduate admission policy. The court upheld the former, seeing it as narrowly construed and employing race only as a possible favorable factor, while striking down the latter as being too closely akin to a quota system. The court decision with respect to Michigan's undergraduate admission policy prompted many colleges and universities to change their admission policies to reduce or eliminate race as a distinct admission criterion.

The Gratz v. Bollinger decision prompted research investigating the extent to which race-blind policies, coupled with selection employing correlates of race, might preserve racial diversity in higher education.[24] Here we use race-blind to mean that criteria for student admission do not vary by race. These analyses suggest that, at the college level, banning affirmative action while permitting of use of correlates of race results in a very significant reduction in minority attendance, especially at top-tier

[24]See, for example, Chan & Eyster (2003), Epple et al. (2008) and Loury et al. (2008).

institutions.

It is likely that the Gratz v. Bollinger decision also induced many school districts, possibly including the one that we study, to adopt race-blind referral and admission policies. Subsequent Supreme Court decisions directly focused on K-12 education reinforce the need for school districts develop race-blind policies. In 2007, the Court decided two cases involving K-12 education, one in Seattle (Parents Involved in Community Schools v. Seattle School District No. 1) and the other in Louisville (Meredith v. Jefferson County Board of Education). In those cases, the Court invalidated the districts' racial desegregation plans, ruling that students could not be assigned to schools solely to achieve racial balance. Not surprisingly, the Court was closely divided. Casting the pivotal vote, Justice Kennedy sided with the conservative wing of the Court to form a majority to ban use of race in determining school assignments. In his opinion, however, Justice Kennedy pointedly left the door open to use of correlates of race to achieve race-conscious objectives.

The following important policy question then arises: How closely can race-conscious but race-blind policies achieve the objectives of racial integration of gifted K-12 academic programs? Our model permits us to provide an analysis of the effect on racial diversity of affirmative action based on race. We can compare this approach to the effect of affirmative action based on a correlate of race—FRL status. We, therefore, consider two policy alternatives. Policy 1 employs affirmative action with respect to both FRL and race. This is accomplished by extending the admission threshold for FRL students to all African American Students. Policy 2 eliminates affirmative action entirely.

Our model predicts that adoption of Policy 1 would increase gifted enrollment of African American students in this cohort from 18.7 to 22.3 percent. Under Policy 2, enrollment of African American students would drop to 12.9 percent. Thus, elimi-

Table 11: Policy Analysis

	Data	Model VI	Policy 1	Policy 2
	Non Black Students			
Not Referred	69.2	69.6	69.6	72.4
Admitted IQ	12.3	10.8	10.8	9.9
Admitted Portfolio	6.3	7.2	7.2	6.5
Avg. IQ Referred	120.5	120.5	120.5	121.5
Avg. IQ Gifted	126.4	126.6	126.4	127.4
	Black Students			
Not Referred	89.3	89.3	88.2	93.1
Admitted IQ	1.8	1.7	2.2	1.0
Admitted Portfolio	1.6	1.6	1.9	1.0
Avg. IQ Referred	111.5	111.5	110.9	114.0
Avg. IQ Gifted	120.7	121.1	120.3	124.2
	Non-FRL Lunch Students			
Not Referred	50.1	51.2	48.7	51.2
Admitted IQ	20	17.0	17.9	17.0
Admitted Portfolio	9.9	11.1	11.7	11.1
Avg. IQ Referred	121	121.0	120.2	121.0
Avg. IQ Referred Gifted	127	127.4	126.8	127.4
	FRL Students			
Not Referred	91.1	90.9	90.9	95.5
Admitted IQ	1.6	1.8	1.8	0.7
Admitted Portfolio	1.5	1.6	1.6	0.7
Avg. IQ Referred	111.3	111.6	111.6	115.0
Avg. IQ Gifted	119.7	120.1	120.1	124.3
	All Students			
Referred	80.5	80.6	79.9	84.0
Admitted IQ	6.4	5.7	6.0	4.9
Admitted Portfolio	3.7	4.1	4.2	3.4
Total No Admitted	134	130	135	110
Avg. IQ Referred	117.7	117.7	117.3	119.7
Avg. IQ Gifted	125.3	125.5	125.2	127.0
Black gifted	18.7	19.1	22.3	12.9
FRL gifted	23.1	25.4	24.4	11.9

nating affirmative action entirely would reduce African American student enrollment by approximately 50% relative to a policy of affirmative action for African American students. Eliminating affirmative action for African American students but retaining affirmative action for FRL students reduces African American student enrollment by 20%. Thus, affirmative action with respect to FRL is extremely important in preserving enrollment of African American students in the district's gifted program.

8 Conclusions

U.S. educational institutions at all levels invest extensive resources in providing advanced programs to help high-ability students reach their potential. The challenge that arises in deciding admission to these programs is that ability is unobserved. Obtaining measures of ability that are not affected by family resources and vagaries of circumstance has proven to be extraordinarily challenging. Despite extensive research and design efforts, measures of ability such as IQ exhibit systematic variation across demographic groups. Educational institutions confront the challenge of using available evidence to assess suitability of students for admission to selective programs, distinguishing ability from influences attributable to family resources and circumstance. While not uncontroversial, differential treatment based on income has gained a considerable measure of acceptance. Race-based criteria have proven to be much more controversial.

The Supreme Court has endorsed a holistic approach, one that considers a broad set of factors in decisions about the admission of a student to selective programs and institutions, i.e., the Court encourages profiling. Going somewhat farther in his decision for the majority in Gratz v Bollinger, Justice Kennedy explicitly declined to reject use of non-race factors to achieve race-conscious ends, a view he reaffirmed in a concurring opinion in the Seattle and Louisville cases. It is natural to wonder whether it is possible to formalize the distinctions being made by the court, and, if so, whether it is possible to test empirically whether a policy does or does not conform to court guidelines. In this light, we have pursued a four objectives in this paper. One objective is to develop a framework sufficiently broad not only to encompasses merit, profiling, and affirmative action, but also to permit statistical testing of the roles that these factors play in practice. A second is to demonstrate, via econometric analysis, the practical relevance of this framework in application. A third is to extend the analysis of these policy issues to an important domain that has been understudied, selective programs for high-ability students in primary and secondary education. A fourth objective is to make a substantive contribution to policy by statistical testing of the extent of use of profiling and affirmative action in an operational setting, and by counterfactuals that investigate the consequences, either of curtailing or extending, the use of profiling or affirmative action.

We believe our application succeeds in these objectives. Of particular interest are the findings of our counterfactual analysis of race-based affirmative action. Long (2007), in his review of the evidence with respect to race-based affirmative action in colleges and universities, concludes, and we agree, that: "The pressures—and in some cases, legal requirements—to replace affirmative action are very real. All the same, the stark evidence of the inefficacy of replacement programs remains." By contrast, for the central school district we study, we find that profiling by race and income, coupled with affirmative action by income, can achieve 80 percent of the level of African American enrollment that could be achieved by race-based affirmative action.

Two questions naturally arise: Why the sharp contrast to findings with respect to elimination of race-based affirmative action in our setting as compared to findings for colleges and universities? To what extent is it likely to generalize to other central city school districts? The answer to the first question is that middle- and upper-income households with children tend to locate in suburban school districts, leaving relatively low-income households in the central city districts. African American households are disproportionately represented among low-income households. Hence, in central city districts, policies that courts have thus far found admissible (profiling by race and income and affirmative action by income) can go a long way toward achieving outcomes that would be achieved by adding race-based affirmative action to the district's decision criteria. The answer to the second question is that our findings are indeed likely to generalize to other central city school districts; the propensity of middle- and high-income households with children to locate in suburban districts is a phenomenon observed in every large metropolitan area in the United States.

References

Anwar, S. & Fang, H. (2006). An alternative test for racial prejudice in motor vehicle searches: theory and evidence. *American Economic Review*, *96 (1)*, 127–151.

Arcidiacono, P. (2005). Affirmative action in higher education: how do admission and financial aid rules affect future earnings?. *Econometrica*, *73 (5)*, 1477–1524.

Arrow, K. (1973). The theory of discrimination. In Ashenfelter, O. (Ed.), *Discrimination in Labor Markets*. Princeton University Press.

Baker, A. (2012). Charges of bias in admission test policy at eight elite public high schools. *New York Times*, *Sept 22*.

Becker, G. (1957). *The Economics of Discrimination*. Chicago University Press.

Carneiro, P., Hansen, K., & Heckman, J. (2003). Estimating distributions of treatment effects with an application to the returns of schooling and measurement of the effects of uncertainty on college choice. *International Economic Review*, *44(2)*, 361–422.

Chan, J. & Eyster, E. (2003). Does banning affirmative action harm college quality?. *American Economic Review*, *93 (3)*, 858–873.

Chung, K. (2000). Role models and arguments for affirmative action. *American Economic Review*, *90*, 640–48.

Coate, S. & Loury, G. (1993). Will affirmative action policies eliminate negative stereotypes?. *American Economic Review*, *83*, 1220–40.

Court, U. S. (2003a). Gratz and versus and bollinger, 539 u.s. 244..

Court, U. S. (2003b). Grutter versus bollinger, 539 u.s. 306..

Court, U. S. (2003c). Meredith v. jefferson county board of education, 548, us 938,..

Court, U. S. (2007). Parents involved in community schools v. seattle school district no. 1, 551 u.s. 701..

Cullen, J., Long, M., & Reback, R. (2012). Jockeying for position: strategic high school choice under texas' top ten percent plan. *Journal of Public Economics*, *forthcoming*.

Cunha, F. & Heckman, J. (2008). Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *Journal of Human Resources*, *43(4)*, 738–782.

Cunha, F., Heckman, J., & Schenach, S. (2010). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica*, *78 (3)*, 883–931.

Epple, D., Romano, R., & Sieg, H. (2006). Admission, tuition, and financial aid policies in the market for higher education. *Econometrica*, *74(4)*, 885–928.

Epple, D., Romano, R., & Sieg, H. (2008). Diversity and affirmative action in higher education. *Journal of Public Economic Theory*, *10 (4)*, 475–501.

Hahn, J., Todd, P., & van der Klaauw, W. (2001). Identification and estimation of treatment effects with regression discontinuity design. *Econometrica*, *69 (1)*, 201–09.

Heckman, J. (1995). Review: lessons from the bell curve. *Journal of Political Economy*, *103 (5)*, 1091–1120.

Hickman, B. (2010a). Effort, race gaps, and affirmative action: a game-theortic analysis of college admissions. Working Paper.

Hickman, B. (2010b). Effort, race gaps, and affirmative action: a structural policy analysis of u.s. college admissions. Working Paper.

Knowles, J., Persico, N., & Todd, P. (2001). Racial bias in motor vehicle searches: theory and evidence. *Journal of Political Economy*, *109*, 203–229.

Long, M. (2004a). College applications and the effect of affirmative action. *Journal of Econometrics*, *121*, 319–42.

Long, M. (2004b). Race and college admissions: an alternative to affirmative action. *Review of Economics and Statistics*, *86 (4)*, 1020–33.

Long, M. (2007). Affirmative action and its alternatives in public universities: what do we know?. *Public Administration Review*, *67 (2)*, 315–30.

Loury, G., Fryer, R., & Yuret, T. (2008). An economic analysis of color-blind affirmative action. *Journal of Law, Economics, and Organization*, *24(2)*, 319–55.

Lundberg, S. (1991). The enforcement of equal opportunity laws under imperfect information: affirmative action and alternatives. *The Quarterly Journal of Economics*, *106*(1), 309–26.

Moro, A. & Norman, P. (2003). Affirmative action in a competitive economy. *Journal of Public Economics*, *87*, 567–94.

Moro, A. & Norman, P. (2004). A general equilibrium model of statistical discrimination. *Journal of Economic Theory*, *114*, 1–30.

Neal, D. & Johnson, W. (1996). The role of pre-market factors in black-white wage differences. *Journal of Political Economy*, *104(5)*, 869–895.

Phelps, E. (1972). The statistical theory of racism and sexism. *American Economic Review*, *62*, 659–661.

Reardon, S., Baker, R., & Klasik, D. (2012). Race, income, and enrollment patterns in highly selective colleges, 1982-2004. Stanford University Working Paper.

Schenach, S. (2004). Estimation of nonlinear models with measurement error. *Econometrica*, *72 (1)*, 33–75.

Thistlethwaite, D. & D.Cambell (1960). Regression-discontinuity analysis: an alternative to the ex post factor experiment. *Journal of Educational Psychology*, *51*, 309–17.