

# Representation versus Assimilation: How do Preferences in College Admissions Affect Social Interactions?

Peter Arcidiacono\* Shakeeb Khan† Jacob L. Vigdor‡

Duke University Duke University Duke University

October 17, 2007

## Abstract

Given the existence of non-selective universities, the question of whether to employ racial preferences in college admissions reduces to one of optimal allocation of a finite resource: students who are members of under-represented racial or ethnic groups. In this paper, we assess recent legal arguments that racial preferences at selective colleges promote meaningful on-campus interracial interaction. As such, we model such interaction as a function of minority representation and, in some cases, perceived social similarity between students of different races. We estimate a structural model to capture these effects and use the results to trace out the net effects of racial preferences on population rates of interracial contact. The results suggest that the interaction-maximizing degree of racial preference, while positive, is significantly weaker than that observed in practice.

**Key Words:** Affirmative Action, social stratification, diversity.

JEL: I2, J15, K0

---

\*Department of Economics, Box 90097, Durham, NC 27708. Phone: (919) 660-1816. Email: psarcidi@econ.duke.edu.

†Department of Economics, Box 90097, Durham, NC 27708.

‡Department of Public Policy Studies, Box 90245, Durham, NC, 27708. Phone: (919)613-7354. Email: jacob.vigdor@duke.edu. Gregory Besharov, William Darity, Kevin Lang, Tom Nechyba, Jesse Rothstein, Jeff Smith, Steven Stern, Sarah Turner, and Harriet Zuckerman as well as seminar participants at Duke University, Yale University, the NBER Higher Education Meetings, and the Winter Meetings of the American Economic Association provided helpful comments. We thank the Andrew W. Mellon Foundation for data access and the Smith Richardson Foundation for financial support. Any opinions expressed herein are those of the authors and not those of any affiliated institution.

# 1 Introduction

In 2003, two U.S. Supreme Court rulings (*Gratz v. Bollinger* and *Grutter v. Bollinger*) affirmed the constitutionality of certain types of racial preferences in college admissions. The main premise behind the Supreme Court’s decisions was that racial preferences in admissions benefit not only the minority students targeted by the policy, but majority students as well. One purported benefit to majority students is rooted in an increased likelihood of meaningful inter-racial contact associated with increased minority representation on campus.<sup>1</sup>

Given the existence of non-selective colleges, however, increasing the diversity of one campus necessarily entails reducing the diversity of another. Since only the most selective colleges employ racial admissions preferences, such preferences have little effect on the attendance rates of blacks (Arcidiacono, 2005). Bowen and Bok, in their seminal book *The Shape of the River*, write that:

“Many people are unaware of how few colleges and universities have enough applicants to be able to pick and choose among them. There is no single, unambiguous way of identifying the number of such schools, but we estimate that only about 20 to 30 percent of all four-year colleges and universities are in this category. Nationally, the vast majority of undergraduate institutions accept all qualified candidates and thus do not award special status to any group of applicants, defined by race or on the basis of any other criterion.” (Bowen and Bok 1998, pp. 16)

Thus, while racial admission preferences may increase diversity, and hence interracial interaction, at very selective colleges, there may be a countervailing loss of diversity and interaction at another campus. From a societal perspective, it is important to ask whether the net effect of this redistribution of minority students is to increase, or indeed to decrease, the population level of interracial interaction. This paper seeks to address this question.

To this end, we first put forth some stylized facts about representation of blacks at selective schools, with the average SAT score of the school used as our measure of selectivity. Whereas a simple comparison of SAT score distributions by race would predict the lowest levels of diversity at the most selective colleges, we find that diversity is actually *higher* at these

---

<sup>1</sup>It has also been argued that higher minority representation at selective colleges allows more minority students to ascend to positions of power or influence in society, which may benefit majority students as well to the extent that the overall productivity of an organization depends on the degree of racial similarity between management and the more general workforce. Our paper is not capable of evaluating this claim.

schools, relative to moderately selective universities.

We then develop a series of models of inter-racial interaction, where interaction depends on the degree of minority representation on campus, and in some cases the degree of similarity between minority and majority students. We expand this latter model to introduce the possibility of statistical discrimination, wherein students form judgements about similarity by using easily observable characteristics such as race. The models make varying predictions regarding interaction rates and interaction maximizing admissions policies. None suggest that the interaction-maximizing relationship between percent black and college selectivity should be upward sloping. The models do suggest that in the quest to maximize contact between members of different groups, there may arise a tradeoff between representation of the minority group and the degree of background similarity between members of different groups.

We test the empirical predictions of the models using data from the College & Beyond survey. The College & Beyond data set provides administrative records at thirty selective colleges in the United States as well as information from a follow-up survey conducted roughly seven years after matriculation.<sup>2</sup> This is the same data set used in Bowen and Bok's *The Shape of the River*. We analyze variation in self-reported measures of whether individuals who matriculated at a selective college in 1989 became well acquainted with two or more members of various racial groups. We then relate these measures to the racial mix and characteristics of the school.

Empirically, we find substantial support for a model where similarity in academic background affects contact rates. We estimate models where the probability of inter-racial interaction is a function not only of the racial composition of one's cohort, but also of the degree of similarity between one's own test scores and those of other racial groups on the same campus. We find that individuals are most likely to interact with individuals of other races if those individuals have test scores similar to their own. In other words, students appear to socially stratify themselves by characteristics that correlate strongly with their test scores: cross-race interaction is most likely to occur within these strata. Increasing minority representation within the group of individuals having SAT scores significantly below one's own SAT score has very little impact on the probability of inter-racial interaction.

Evidence that likes interact with likes corroborates existing results in the emerging eco-

---

<sup>2</sup>The sample restriction to selective colleges can be justified on the ground that the racial preference policies under examination here are only relevant at these schools (see Kane, 1998 and Arcidiacono, 2005).

nomics literature on friendship formation.<sup>3</sup> Foster (2005) and Marmaros and Sacerdote (2006) both find that race is extremely important in college friendship formation but find that similarity in academic background is also important. Similarly, Weinberg (2005) analyzes high school students and also finds evidence for similarity in characteristics matters for friendship formation.

Reduced form results also suggest some scope for statistical discrimination. Increasing racial diversity of the group of individuals with SAT scores significantly higher than one's own SAT score has close to the same effect on inter-racial contact as increasing diversity of the group with scores similar to one's own score. This is consistent with a model where individuals prefer to interact with students of similar academic backgrounds and make inferences about peers' academic background based on their race. Well prepared black students may be penalized by this statistical discrimination as the presence of those with poor backgrounds has negative effects on the expectations of non-blacks.

The reduced form results motivate a structural model that is designed to highlight the potentially offsetting effects of representation and similarity in academic background. The model has prospective friends of particular races arrive according to a process that is governed in part by the representation of those races at their school. Match quality is in part a function of the gap between one's own academic background and the background of the potential friend. Individuals receive noisy signals on the potential friend's academic backgrounds and use these signals as well as the school-specific distribution of academic background for members of the potential friend's race in assessing whether to form the friendship. Estimates from the structural model show that both representation and similarity in academic background are important factors in inter-racial friendship with some evidence of statistical discrimination.

With the structural estimates in hand, we perform a series of simulations to predict the amount of inter-racial interaction that would occur if the degree of racial preference in selective college admissions were adjusted to any of several points between currently observed levels and zero. Although race-blind assignments would reduce the representation of blacks at the most selective schools, the predicted population level of inter-racial interaction actually *increases*.<sup>4</sup> Decreases in contact at the most-selective colleges are more than offset by increases

---

<sup>3</sup>Likes interacting with likes has a much longer history in sociology under the heading of *homophily*. See Carli et al. (1991), Henderson and Furnham (1985), Kandel (1978), and McPherson and Smith-Lovin (1988).

<sup>4</sup>These simulations confine students to attend one of the institutions within the College & Beyond data set. If racial preferences were eliminated many of these students would not be admitted to any of these schools, reducing inter-racial contact. However, the students who did not attend one of these top schools would attend

at moderately-selective colleges, owing in part to the reduction in academic background disparities on each campus.

Do these results imply that society would be best served by reducing the degree of racial preferences in selective college admissions? While the aim of this paper is not to answer this question, our work illuminates a trade-off that was not clearly articulated in arguments before Federal courts earlier in this decade.<sup>5</sup> If the intention of the U.S. Supreme Court was to promote interracial interaction only at the small set of colleges with the most stringent admissions standards, continuation of existing policy might make sense. If the social value of interracial contact is not strongly related to selectivity, however, it does appear that current policies are suboptimal.

The rest of the paper proceeds as follows. Section 2 provides a brief background on the distribution of minority students across college campuses in the United States. Section 3 presents theoretical models of inter-racial interaction. Section 4 describes the data and methods used in the analysis. Section 5 examines the factors that influence inter-racial interaction. Section 6 describes and shows the estimates of the structural model. Section 7 shows the simulations of alternative assignment rules. Section 8 concludes.

## 2 Stylized Facts on Minority Representation in US Colleges

Given the existence of racial preferences in selective college admissions, one might expect the underlying relationship between minority representation and college selectivity to be strongly negative. After all, many professions or industries engage minority recruitment initiatives to address under-representation. Figure 1, which presents some basic information on the relationship between college quality and minority representation in the United States, shows that this intuition is at least partly wrong – over some range of the distribution, the relationship between selectivity and representation is positive. We use average SAT score as our measure of college selectivity, and percent black as our measure of minority representation. We collected data on these two variables from the Princeton Review Best 357 colleges.<sup>6</sup> Figure 1 fits a fractal

---

institutions further down the rankings and increase inter-racial contact at those schools. We discuss these effects in detail in section 7.

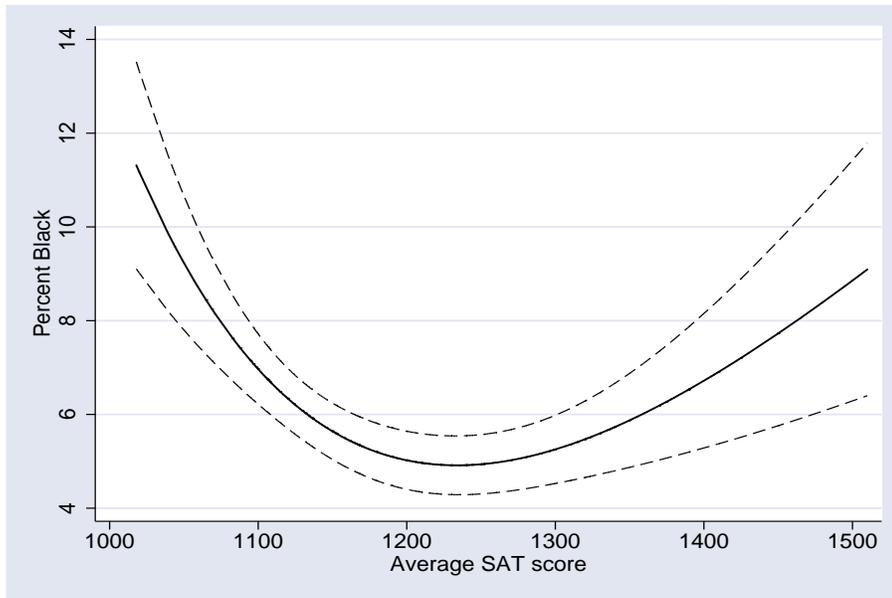
<sup>5</sup>It is also worthwhile noting that this paper is silent on the question of whether affirmative action is desirable from an equity perspective.

<sup>6</sup>Some schools—particularly those at the very top— did not report their average SAT score. For these schools we used data from America’s Best Colleges 2005 produced by U.S. News & World Report. The U.S. News & World Report data gives the 25th and 75th percentiles of the SAT distribution. For those schools not

polynomial of average SAT scores to percent black. The results are weighted by enrollment, but the unweighted results are similar and presented in the Appendix.

Plotting percent black as a function of average SAT score yields a U-shaped pattern: the most selective schools have a higher percent black than moderately selective schools. Figure 2 displays the same table by region and in all regions the U-shaped pattern emerges. This U-shaped pattern holds despite blacks scoring significantly lower on the SAT than their non-black counterparts.

Figure 1: College Percent Black as a Function of Average SAT Score<sup>†</sup>



Source: Princeton Review Best 357 colleges and America’s Best Colleges 2005 produced by U.S. News & World Report. Weighted by college enrollment.

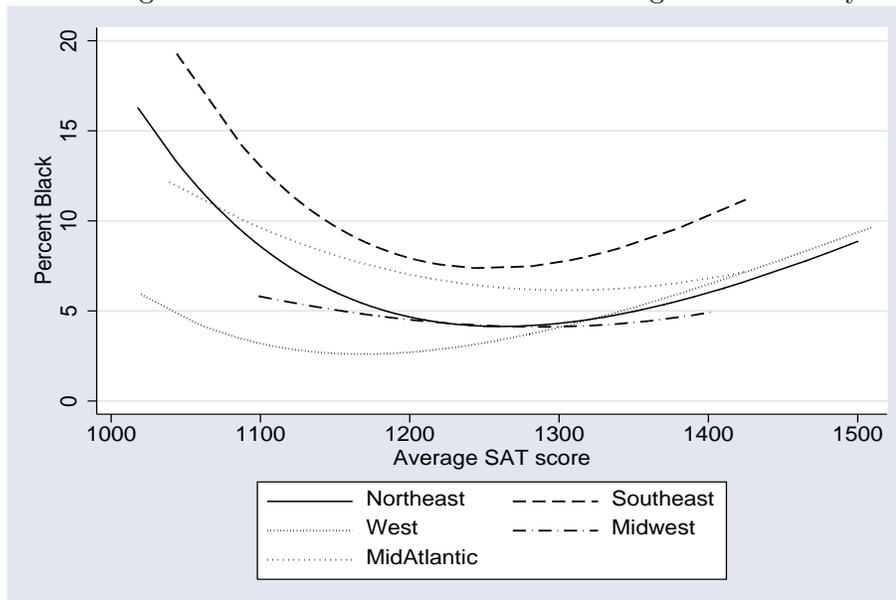
The U-shaped pattern suggests that preferences for blacks at the most selective institutions have come at the expense of diversity at moderately selective institutions. Indeed, racial preferences in admissions have been shown to have little effect on admissions at most schools (see Kane 1998 and Arcidiacono 2005) and in turn little effect on the overall attendance rates of blacks (Arcidiacono 2005). Race preferences in college admissions then primarily affect which colleges blacks attend conditional on attending at all.

Is the U-shaped pattern evident in Figures 1 and 2 consistent with some underlying model of optimizing behavior on the part of college administrators? In the next section we develop

---

reporting an average SAT score, we used the midpoint of the 25th and 75th percentiles.

Figure 2: College Percent Black as a Function of Average SAT Score by Region<sup>†</sup>



Source: Princeton Review Best 357 colleges and America’s Best Colleges 2005 produced by U.S. News & World Report. Weighted by college enrollment.

three models of inter-racial contact maximization. None of the three cases yield a U-shaped pattern as the optimal assignment rule. When interaction occurs randomly among individuals collected on an individual campus, inter-racial contact is maximized when minority representation is equalized across all schools. In models where similarity in academic background matters, however, the contact-maximizing relationship between selectivity and minority representation is monotonically downward sloping. Current policies appear to be designed to maximize the degree of inter-racial contact on the most selective campuses. From a social perspective, these policies are optimal only to the extent that contact on these campuses is valued much more highly than contact at moderately selective universities.

### 3 Models of Inter-racial Contact

#### 3.1 Model 1: Inter-racial contact does not depend upon individual characteristics

Our first model considers the simplest case where inter-racial interaction does not depend upon background characteristics. In particular, let there be two groups of individuals,  $W$ ’s

and  $B$ 's, at each of two schools, 1 and 2. For each of the models we take the number of  $W$  types at the two schools as given at  $N_{W1}$  and  $N_{W2}$ , though our results are consistent with a model where schools trade off enrollment between the two types.

Each individual is randomly assigned a match. For this base model, the probability of a successful match is exogenously given, and for simplicity we set the probability to one. The number of successful inter-racial contacts at school  $j$ ,  $IC_j$  can be written as the number of  $W$  types at the school times the probability of being matched with a  $B$  type:

$$IC_j = N_{Wj} \left( \frac{N_{Bj}}{N_{Wj} + N_{Bj}} \right)$$

A social planner assigning  $B$  types to schools to maximize inter-racial contact solves:

$$\begin{aligned} \max_{N_{B1}, N_{B2}} \quad & N_{W1} \left( \frac{N_{B1}}{N_{W1} + N_{B1}} \right) + N_{W2} \left( \frac{N_{B2}}{N_{W2} + N_{B2}} \right) \\ \text{s.t.} \quad & N_{B1} + N_{B2} = N_B \end{aligned}$$

Substituting in for  $N_{B2}$  from the constraint and taking first order conditions then gives the optimal choice of  $N_{B1}$ :

$$N_{B1} = N_B \left( \frac{N_{W1}}{N_{W1} + N_{W2}} \right)$$

The fraction of  $B$  types assigned to school 1 is then the same fraction of  $W$  types assigned to school 1. The optimal fraction of  $B$  types assigned to each school in this simple model is then constant across schools.

Having the group shares be equal across schools minimizes the probability that an individual will be matched with a member of his own race. Since inter-racial interaction has nothing to do with the characteristics of the individual, the optimal policy for maximizing inter-racial contact would yield no relationship between the share of a group at a school and the quality of the school. In the context of Figure 1, the optimal policy would yield a flat line at the share of blacks in the population that attended this set of schools. If this model is correct, differences in background characteristics should have little effect on interaction probabilities.

### 3.2 Model 2: Inter-racial contact depends upon similarity in academic backgrounds

While equal representation across schools will maximize inter-racial contact when interaction does not depend upon other factors, it is easy to write down situations when equal representation is not optimal once interaction depends upon the characteristics of the individuals.

We now allow individuals to be one of two academic background types,  $H$  and  $L$ . To keep the proofs simple, we make the following additional assumptions about the distribution of academic experience types across race and across schools as well as the relationship between academic experience and the probability of a successful contact:

1. The probability of being type  $H$  given race  $W$  is 0.5:  $N_{WH} = N_{WL}$ .
2. School 2 has relatively more  $W$  types with characteristic  $H$  than school 1:

$$\begin{aligned} N_{W1} &= \alpha N_{WL} + (1 - \alpha) N_{WH} \\ N_{W2} &= (1 - \alpha) N_{WL} + \alpha N_{WH} \end{aligned}$$

where  $0.5 \leq \alpha < 1$ .

3. The distribution of academic experience is such that  $W$  types are more likely to have academic experience  $H$ : the probability of a member of group  $B$  being the  $L$  type is  $\pi_{BL} > 0.5$ .
4. The social planner can assign at most  $\alpha$  of each academic experience type to each of the schools.<sup>7</sup>
5. Conditional on matching, the probability of a successful contact is 1 if the academic experiences are the same and  $\gamma < 1$  if the academic experiences are different.

This last assumption ensures that similarity in academic background matters at every school.

We then examine the optimal assignment policies for  $B$  types under these assumptions. First, note that it is optimal to assign as many members of group  $B$  with academic experience  $H$  as possible to school 2:  $N_{BH2} = \alpha N_{BH}$ . The issue is then how to distribute members of group  $B$  of type  $L$  across the two schools. Let  $\beta$  represent the fraction of  $L$  type blacks assigned to school 1. Inter-racial contact at each of the schools is given by:

$$IC_1 = \alpha N_{WL} \left( \frac{\beta N_{BL} + \gamma(1 - \alpha) N_{BH}}{N_{W1} + N_{B1}} \right) + (1 - \alpha) N_{WH} \left( \frac{\gamma \beta N_{BL} + (1 - \alpha) N_{BH}}{N_{W1} + N_{B1}} \right) \quad (1)$$

$$IC_2 = \alpha N_{WH} \left( \frac{\gamma(1 - \beta) N_{BL} + \alpha N_{BH}}{N_{W2} + N_{B2}} \right) + (1 - \alpha) N_{WL} \left( \frac{(1 - \beta) N_{BL} + \gamma \alpha N_{BH}}{N_{W2} + N_{B2}} \right) \quad (2)$$

To show that there should be a lower fraction of  $B$  types at school 2, we first find the  $\beta$  that leads to equal representation:  $\beta_{ER}$ . Next, we show that the derivative of inter-racial

---

<sup>7</sup>This assumption ensures that there is some diversity in academic background for each race at each school.

contact is positive with respect to  $\beta$  when evaluated at  $\beta_{ER}$ . The value of  $\beta$  that leads to equal to representation,  $\beta_{ER}$  is given by:

$$\beta_{ER} = \frac{1}{2} + \frac{(2\alpha - 1)(1 - \pi_{BL})}{2\pi_{BL}}$$

Note that at  $\pi_{BL} = 1/2$ ,  $\beta_{ER} = \alpha$ . Further, with  $\alpha > 1/2$  the expression is decreasing in  $\pi_{BL}$ . With  $\pi_{BL} > 1/2$ ,  $\beta_{ER} < \alpha$ .

Let  $IC = IC_1 + IC_2$ . Taking the partial derivative with respect to  $\beta$  yields:

$$\begin{aligned} \frac{\partial IC}{\partial \beta} &= \alpha N_{WL} \left( \frac{N_{BL}}{N_{W1} + N_{B1}} - \frac{(\beta N_{BL} + \gamma(1 - \alpha)N_{BH})N_{BL}}{(N_{W1} + N_{B1})^2} \right) \\ &+ (1 - \alpha)N_{WH} \left( \frac{\gamma N_{BL}}{N_{W1} + N_{B1}} - \frac{(\gamma\beta N_{BL} + (1 - \alpha)N_{BH})N_{BL}}{(N_{W1} + N_{B1})^2} \right) \\ &+ (1 - \alpha)N_{WL} \left( -\frac{N_{BL}}{N_{W2} + N_{B2}} + \frac{((1 - \beta)N_{BL} + \gamma\alpha N_{BH})N_{BL}}{(N_{W2} + N_{B2})^2} \right) \\ &+ \alpha N_{WH} \left( -\frac{\gamma N_{BL}}{N_{W2} + N_{B2}} + \frac{(\gamma(1 - \beta)N_{BL} + \alpha N_{BH})N_{BL}}{(N_{W2} + N_{B2})^2} \right) \end{aligned}$$

If  $\gamma = 1$  we are back to the case where similarity in academic background does not matter. We can write the derivative above evaluated at  $\beta_{ER}$  as a function of the first order condition from the maximization problem when academic background did not matter plus additional terms. Since equal representation involved setting this first order condition equal to zero, the sign of the additional terms determines the sign of the expression. Substituting in such that the partial derivative is only a function of  $N_W$ ,  $N_B$ ,  $\pi_{BL}$ ,  $\alpha$ , and  $\beta_{ER}$ , the partial derivative evaluated at  $\beta_{ER}$  is then given by:

$$\begin{aligned} \frac{\partial IC}{\partial \beta} \Big|_{\beta_{ER}} &= \frac{\alpha N_W}{N_W + N_B} \left[ \frac{(1 - \gamma)(1 - \alpha)(1 - \pi_B)\pi_B N_B^2}{N_W + N_B} \right] \\ &+ \frac{(1 - \alpha)N_W}{N_W + N_B} \left[ -(1 - \gamma)\pi_B N_B + \frac{(1 - \gamma)\beta\pi_B^2 N_B^2}{N_W + N_B} \right] \\ &+ \frac{(1 - \alpha)N_W}{N_W + N_B} \left[ -\frac{(1 - \gamma)\alpha(1 - \pi_B)\pi_B N_B^2}{N_W + N_B} \right] \\ &+ \frac{(1 - \alpha)N_W}{N_W + N_B} \left[ (1 - \gamma)\pi_B N_B - \frac{(1 - \gamma)(1 - \beta)\pi_B^2 N_B^2}{N_W + N_B} \right] \end{aligned}$$

implying that the sign evaluated at  $\beta_{ER}$  is given by:

$$\text{sign} \left( \frac{\partial IC}{\partial \beta} \Big|_{\beta_{ER}} \right) = \alpha \left[ \frac{N_W}{2} + \alpha(1 - \pi_B)N_B \right] - (1 - \alpha) \left[ \frac{N_W}{2} + (1 - \alpha)(1 - \pi_B)N_B \right]$$

Since  $\alpha > 0.5$ , the derivative is positive at  $\beta_{ER}$ . With  $\beta$  representing the fraction of group of  $B$  of type  $L$  assigned to the lower quality school, it is clear that the contact-maximizing allocation pattern will involve a disproportionate number of  $B$  types represented at the lower quality school.

In addition to equal representation not being optimal in this case, it is important to note that the observability of academic background produces a symmetry in conditional match probabilities. That is to say, a member of group  $W$  with type  $L$  who has been assigned a member of group  $B$  of type  $H$  has the same probability of matching as a member of group  $W$  of type  $H$  who has been assigned a member of group  $B$  of type  $L$ . This does not hold in the case of unobserved background below. In such a case, statistical discrimination compounds the effect of true differences in academic backgrounds. The presence or absence of symmetric effects is an important distinction that will allow us to empirically distinguish between the two models.

### 3.3 Model 3: Inter-racial contact depends upon similarly in previous academic experiences and individuals statistically discriminate.

The problem with matching based upon academic experience can be recast with  $\gamma$  serving as the benefit of a match between two individuals of different academic experiences. There is then a cost of forming a match which is borne by one of the individuals. Which individual bears the cost is exogenously assigned and costs are assumed to be distributed uniformly on the zero-one interval. If academic experience is perfectly observed before an individual decides to pay the cost of accepting the match with someone of a different academic experience type then the probability the individual will accept the match is  $\gamma$ .

The advantage of recasting the problem this way is that it can be easily extended to the case when individuals only receive signals of their potential partner's academic experience and use this information in forming expectations about the expected benefits of engaging in the friendship. We employ the same assumptions regarding the distributions of academic experience types across the two groups as well as the assumption of how members of group  $W$  are distributed across the two schools. Now, however, individuals emit signals—signals that they have no control over—regarding their academic experience type. In particular, an individual correctly signals their type  $\phi$  percent of the time where  $0.5 < \phi \leq 1$ . We assume that the  $W$  types always bear the cost though all the qualitative results hold when  $B$  types bear the cost instead.

As with the case when academic experience is observed, it will again be optimal to assign as many members of group  $B$  that are type  $H$  to school 2:  $N_{BH2} = \alpha(1 - \pi_{BL})N_B$ . Let  $\beta$  again represent the fraction of  $L$  types from group  $B$  that are assigned to school 1. Let  $P(M|A, S, j)$  be the probability a  $W$  type will accept a match given his academic experience

is  $A$ , given he has received a signal  $S$  from his potential partner, and given that he is at school  $j$ . This probability for someone of academic experience  $L$ , who has seen a signal  $H$  at school 1 is given by:

$$P(M|L, H, 1) = \gamma \left( \frac{\phi(1-\alpha)(1-\pi_{BL})}{\phi(1-\alpha)(1-\pi_{BL}) + (1-\phi)\beta\pi_{BL}} \right) + 1 \left( \frac{(1-\phi)\beta\pi_{BL}}{\phi(1-\alpha)(1-\pi_{BL}) + (1-\phi)\beta\pi_{BL}} \right)$$

where the first term gives the benefit of matching with an  $H$  type conditional on being an  $L$  type times the probability that the individual he is matched with is an  $H$  type given the  $H$  signal and given the distribution of types at school 1. The second term is similar except that now it is the benefit and the conditional probability of the individual being an  $L$  type.

We can now perform the same analysis as before: calculate the partial derivative of the inter-racial contact function with respect to  $\beta$  and evaluate it at  $\beta_{ER}$ . If the partial derivative evaluated at  $\beta_{ER}$  is positive, then it is optimal to assign relatively more  $B$  types to school 1 than school 2. However, the optimal assignment rule is exactly the same as in the case with academic experience matching but no statistical discrimination. To see this, consider the number of inter-racial contacts generated by from matches including  $L$  types who are members of group  $W$  at school 1. This number is given by:

$$IC_{L1} = \alpha N_{WL} \left[ \left( \frac{N_B(\phi(1-\alpha)(1-\pi_{BL}) + (1-\phi)\beta\pi_{BL})}{N_{W1} + N_{B1}} \right) \left( \frac{\gamma\phi(1-\alpha)(1-\pi_{BL}) + (1-\phi)\beta\pi_{BL}}{\phi(1-\alpha) + (1-\phi)\beta\pi_{BL}} \right) + \left( \frac{N_B((1-\phi)(1-\alpha)(1-\pi_{BL}) + \phi\beta\pi_{BL})}{N_{W1} + N_{B1}} \right) \left( \frac{\gamma(1-\phi)(1-\alpha)(1-\pi_{BL}) + \phi\beta\pi_{BL}}{(1-\phi)(1-\alpha) + \phi\beta\pi_{BL}} \right) \right]$$

The term on the first line gives the probability of meeting a  $B$  type and seeing the  $H$  signal times the probability of accepting a match given an  $H$  signal was seen. The second line then gives the similar expression for the  $H$  signal. Note that this expression simplifies to:

$$IC_{L1} = \alpha N_{WL} \left( \frac{\beta N_{BL} + \gamma(1-\alpha)N_{BH}}{N_{W1} + N_{B1}} \right)$$

which is exactly the same as the corresponding term when academic experience is observed without noise. The terms for  $IC_{H1}$ ,  $IC_{L2}$ , and  $IC_{H2}$  also match so that the  $\beta$  which maximizes inter-racial contact when academic experience is observable is the same as the  $\beta$  which maximizes inter-racial contact when academic experience is measured with noise. Hence, to maximize inter-racial contact  $B$  types will be disproportionately represented at school 1.

Note that although the best choice of  $\beta$  is the same across the two cases the symmetry present in the observable academic experience case is not present here. In particular, conditional on matching with someone of the opposite academic experience type, the probability the match is accepted is no longer the same for  $L$  types and  $H$  types and differs across schools

as well. Consider an  $L$  type at school 2 who is matched with an  $H$  type member of group  $B$ .

Label the probability of a successful match in this case as  $Q(M|L, H, 2)$  and is given by:<sup>8</sup>

$$Q(M|L, H, 2) = \phi \left[ \frac{\gamma\phi\alpha(1 - \pi_{BL}) + (1 - \phi)(1 - \beta)\pi_{BL}}{\phi\alpha(1 - \pi_{BL}) + (1 - \phi)(1 - \beta)\pi_{BL}} \right] + (1 - \phi) \left[ \frac{\gamma(1 - \phi)\alpha(1 - \pi_{BL}) + \phi(1 - \beta)\pi_{BL}}{(1 - \phi)\alpha(1 - \pi_{BL}) + \phi(1 - \beta)\pi_{BL}} \right]$$

The corresponding probability of an  $H$  type member of group  $W$  accepting a match with an  $L$  type member of group  $B$  conditional on assignment,  $Q(M|H, L, 2)$ , is:

$$Q(M|H, L, 2) = \phi \left[ \frac{\gamma\phi(1 - \beta)\pi_{BL} + (1 - \phi)\alpha(1 - \pi_{BL})}{(1 - \phi)\alpha(1 - \pi_{BL}) + \phi(1 - \beta)\pi_{BL}} \right] + (1 - \phi) \left[ \frac{\gamma(1 - \phi)(1 - \beta)\pi_{BL} + \phi\alpha(1 - \pi_{BL})}{\phi\alpha(1 - \pi_{BL}) + (1 - \phi)(1 - \beta)\pi_{BL}} \right]$$

Only special cases satisfy  $Q(M|H, L, 2) = Q(M|L, H, 2)$  such as when academic experience can be seen without error or when the members of group  $B$  at school 2 are evenly divided between  $H$  types and  $L$  types. Hence, asymmetries in the match probabilities are suggestive of statistical discrimination.

## 4 Data

To examine the impact of collegiate diversity on inter-racial interaction, we employ the College & Beyond Data set, made available by the Andrew W. Mellon Foundation.<sup>9</sup> This data set contains information from two sources: administrative information on a set of mostly selective undergraduate institutions, and survey responses collected from a sample of students who matriculated at those institutions in one of three cohorts. Our analysis focuses on the 1989 entering cohort.<sup>10</sup> The administrative data include information on each student's SAT scores, major, and their ultimate means of exit, whether graduation, transfer, or withdrawal. For most institutions, the administrative data represents the entire entering cohort. For the remainder, the data comprise a nonrandom sample of the student body. Weights are provided to adjust for this sampling. A complete list of institutions represented appears in Table A1 in the appendix. A subset of the administrative sample, approximately 40%, received and replied to

---

<sup>8</sup>Note that  $Q(M|L, H, 2)$  differs from  $P(M|L, H, 2)$  in that in the former case the individual is actually matched with an  $H$  type while in the latter case an individual is matched with someone who has given the  $H$  signal.

<sup>9</sup>We omit observations from historically black colleges as race preferences for under-represented minorities are not relevant at these schools. We also omit observations from women's colleges as the selection into and the environment of these colleges may be substantially different from other institutions. Including data from women's colleges does not affect our results.

<sup>10</sup>Other cohorts available in the C&B data set include the classes entering in 1951 and 1976. We focus on the 1989 cohort because it is the most recent and more detailed questions were asked about inter-racial interaction.

a follow-up survey in 1996. This survey provides information on many outcomes of interest such as collegiate satisfaction and inter-racial interaction.

#### 4.1 Evidence of admission preferences: disparities in academic background

As a proxy for academic background, we focus on SAT scores.<sup>11</sup> Table 1 shows the distribution of SAT scores relative to the school average for blacks and non-blacks. Relative to members of other racial groups, blacks in the College & Beyond received lower SAT scores.<sup>12</sup> Within a particular school, the seventy-fifth percentile of the black distribution roughly matches the twenty-fifth percentile of the non-black distribution. These large differences in SAT scores support the essential assumptions made in the models above.<sup>13</sup>

Table 2 examines the distribution of black SAT scores relative to the school average for three tiers of schools. Tier 1 includes those with average SAT scores above 1300, tier 2 between 1200 and 1300, and tier 3 less than 1200. The gap between black SAT scores and the average SAT score is fairly similar across tiers, though the bottom end of the black distribution is relatively worse at tier 3 schools.

#### 4.2 Measures of inter-racial contact

Our key outcome measures are the self-reported answers to questionnaire items inquiring about the friends that respondents made while enrolled in college. For a number of racial and social groups, the survey begins by asking whether respondents knew at least one member of the group. For those responding yes, a follow-up question asks whether the respondent knew two or more group members “well.” While these responses are inherently subjective, it is plausible to think that the correlate strongly with a latent variable indicating the degree of inter-racial

---

<sup>11</sup>We experimented with other measures such as private versus public school and found little evidence that similarity in these measures mattered besides through the SAT score.

<sup>12</sup>Note that the SAT has been accused of being biased against blacks. This paper takes no stand on this issue, simply noting that if this is true, it will reflect our empirical results in predictable ways. We return to this in the results section.

<sup>13</sup>Given these differences, our initial thought was that perhaps blacks would look similar to legacies in terms of their SAT scores due to preferential admissions policies for both groups. These preferential admissions policies would then serve as a device that would encourage interaction between blacks and legacies. However, the distribution of legacy SAT scores at particular schools is actually similar to those of non-blacks in general. Note that legacies may still receive preferential treatment in the admissions process. This is because legacies should on average have higher SAT scores than in the population as a whole. Hence, conditional on being in a particular SAT score range, we would expect legacies to be disproportionately represented at the top end.

Table 1: Distribution of SAT Scores Relative to the School Average<sup>†</sup>

SAT minus		
School Average SAT	Blacks	Non-Blacks
10th percentile	-338	-150
25th percentile	-260	-65
50th percentile	-167	22
75th percentile	-70	100
90th percentile	11	165
Mean	-164	14
Observations	1561	20566

<sup>†</sup>Sample is taken from the 17 coeducational institutions in the College & Beyond that had participants in the followup survey.

Table 2: Distribution of Black SAT Scores Relative to the School Average By Tier<sup>†</sup>

Black SAT minus			
School Average SAT	Tier 1	Tier 2	Tier 3
10th percentile	-300	-339	-358
25th percentile	-233	-248	-280
50th percentile	-143	-145	-190
75th percentile	-59	-61	-85
90th percentile	7	9	20
Mean	-146	-155	-179
School Average SAT	1330	1245	1127
Observations	5910	7929	8288

<sup>†</sup> Sample is taken from the 17 coeducational institutions in the College & Beyond that had participants in the followup survey.

contact respondents experienced while in college. We focus particularly on whether non-blacks knew two or more blacks well and whether whites knew two or more Hispanics or asians well. Table 3 shows the proportion of individuals answering these questions affirmatively as well as the corresponding shares of blacks, Hispanics, and asians in the sample. The last three columns disaggregate the numbers by the same tiers as in Table 2.

Representation does appear to correlate with our measure of inter-racial contact. Survey respondents who attended tier 1 schools have the highest probabilities of knowing members from all three groups well and each of the three groups is disproportionately represented at the tier 1 schools. As we move down the tiers, the fraction of both asians and hispanics fall. This may be related to the location of the college themselves as Hispanics may be more concentrated in particular regions of the country. For blacks, however, we see the same U-shape pattern as in the national data: the middle tier schools have on average the lowest percent black.

Table 3: Probability of Knowing Two of More Members of a Group Well<sup>†</sup>

	Group	Population	Tier 1	Tier 2	Tier 3
Know 2 or More	Blacks	59.8%	67.5%	56.8%	58.0%
	Asians	59.6%	81.4%	63.3%	48.1%
	Hispanics	30.1%	45.7%	38.1%	18.5%
Share	Blacks	6.4%	7.5%	5.7%	6.2%
	Asians	8.2%	12.9%	8.6%	5.6%
	Hispanics	3.3%	5.7%	3.5%	1.9%

<sup>†</sup> Sample is taken from the 17 coeducational institutions in the College & Beyond that had participants in the followup survey. Tier 1, Tier 2, and Tier 3 refer to institutions with average SAT scores above 1300, between 1200 and 1300, and below 1200 respectively. For knowing two or more blacks well, the sample includes all non-blacks. For knowing two or more Hispanics or Asians, the sample includes only whites. Results are not sensitive to this sample selection.

## 5 Reduced Form Methods and Results

### 5.1 Estimation Strategy

We relate the outcome measure, a respondent’s self-report of knowing two or more members of a racial or ethnic group well, to observable characteristics of the individual and the school. For our baseline model, we work within the latent dependent variable model framework. Analogously to the latent dependent variable in an ordered response model, the latent variable monotonically related to the number of members of a group  $k$  that are well known by an individual  $i$  at school  $j$  is given by<sup>14</sup>:

$$Y_{ijk}^* = \alpha_0 + X_{1i}\alpha_1 + SHR_{jk}\alpha_2 + (SHR_{jk}\alpha_2)^2\alpha_3 + \epsilon_{ijk} \quad (3)$$

where  $X_{1i}$  refers to individual-specific characteristics that may affect inter-racial interaction or an individual’s standard for knowing someone well. The  $SHR$  terms refer to the fraction of the group in question (for example, blacks) at the school. The squared term allows for decreasing (or potentially increasing) returns from increasing representation. Assuming that the error terms are normally distributed and independent from the other variables, we can estimate the probability of knowing two or more well in a probit model where the outcome measure is  $Y_{ijk} = 1$  if  $Y_{ijk}^* \geq \tau_2$ , an unknown threshold. All of the  $\alpha$ ’s are allowed to vary across the groups in question.

For a reduced form test of whether similarity in academic background matters, we add to the model the relative SAT score at each school. The underlying index function is then:

$$Y_{ijk}^* = \alpha_0 + X_{1i}\alpha_1 + SHR_{jk}\alpha_2 + (SHR_{jk}\alpha_2)^2\alpha_3 + (SAT_i - \overline{SAT}_j)\alpha_4 + \epsilon_{ijk} \quad (4)$$

If similarity in SAT scores is unimportant—or the SAT score is not related to the characteristic on which similarity matters— $\alpha_4$  will be zero. Note that it is possible to include school fixed effects in the above specification, with the within-school variation in SAT scores driving the estimate of  $\alpha_4$ . Further, because we are using the within-school variation, it is possible to estimate the specification for each school separately and obtain separate estimates of  $\alpha_4$  for each school to further ensure that our estimates not being driven by the features of particular

---

<sup>14</sup>In the models that follow we are assuming that the latent error term  $\epsilon_{ijk}$  is distributed independently of the covariates. This assumption may be most likely to be violated regarding the regressor  $SHR$  since for historical reasons share black may vary locationally with ways that are correlated with attitudes toward interracial friendship. We discuss below how to control for this potential source of endogeneity bias.

institutions.<sup>15</sup>

We can also explicitly introduce the possibility of a relationship between an individual's own characteristics and the distribution of those same characteristics among the other race. Hence, we develop a model that nests the relationship between the share of group  $k$  and interaction rates in equation (3) but allows for similarity in SAT scores across the distribution to matter. In particular, divide all members of group  $k$  into three subgroups: those with SAT scores 160 above their own (*HIGH*), 160 points below their own (*LOW*), and within 160 points of their own score (*MED*).<sup>16</sup> Dividing these numbers by the total member of classmates then gives the joint probability of being in the particular SAT group and in the racial group in question. We then allow increasing the shares of each of these groups to differ in their effect on inter-racial interaction. This leads to the following specification:

$$Y_{ijk}^* = \alpha_0 + X_{1i}\alpha_1 + \mathbf{SHR}_{jk}\alpha_2 + (\mathbf{SHR}_{jk}\alpha_2)^2\alpha_3 + \epsilon_{ijk} \quad (5)$$

where

$$\mathbf{SHR}_{jk}\alpha_2 = \frac{\alpha_{20}N_{jkHIGH} + \alpha_{21}N_{jkMED} + \alpha_{22}N_{jkLOW}}{N_j}$$

$N_{jkHIGH}$  refers to the number of students at school  $j$ , of race  $k$ , who have SAT scores 160 points above individual  $i$  while  $N_j$  refers to the total number of students at school  $j$ . Note that the coefficients on the linear term also enter into the squared term. In this way if one group has little direct benefit then that group will also have little effect on crowd out. That is, if a group has no first-order effect on interaction then the group should also have no second-order effect.

The specification in (5) nests all three of the models discussed in section 3. If SAT scores are irrelevant to inter-racial interaction, then the coefficient on the three share variables will be the same. If similarity in SAT scores matters but there is no statistical discrimination, we would expect the coefficient on the share of the group within the SAT range to be higher than

---

<sup>15</sup>In this specification, there may be omitted factors correlated with individual SAT scores that influence the rate of inter-personal contact. As a crude example, students at the high end of the SAT distribution for their institution may be “bookworms” who get to know very few other students. To address this concern, we present one specification below that controls simultaneously for relative and absolute SAT scores. Absolute SAT scores prove to be insignificant determinants of inter-racial contact conditional on relative SAT scores. We will also present specifications demonstrating that this is not a universal tendency. White students with high SAT scores are more likely to know members of a group with high average SAT scores – Asians.

<sup>16</sup>160 points corresponds to the standard deviation in SAT scores across the population of College & Beyond students.

the coefficients on the other groups,  $\alpha_{21} > \alpha_{20}, \alpha_{22}$ . However, we would also expect symmetry between the coefficients on the share of students above and the share of students below,  $\alpha_{20} = \alpha_{21}$ . If statistical discrimination is present and the group in question has significantly lower SAT scores than the population, the coefficient on the share of the group significantly above the individual's SAT score should be higher than the corresponding coefficient on the share of the group significantly below the individual's SAT score,  $\alpha_{20} > \alpha_{22}$ .

## 5.2 Is there evidence that similarity in academic background affects contact rates?

Estimates of a baseline probit of the probability of knowing two or more blacks well are presented in Table 4. The sample consists of all non-black College & Beyond respondents. Variation in percent black is solely at the level of the institution and standard errors are adjusted for clustering. The specification in column 1 gives the results when the only controls are female, percent black, and percent black squared. Females are less likely to report that they know two or more blacks well and this pattern continues throughout the rest of the regression results. However, as we will see later in the paper, this seems to be because females are less likely to report that they know anyone well— what is considered knowing someone well appears to differ across the sexes.

Both percent black and percent black squared have the expected signs. With the negative squared term, an interior interaction-maximizing percent black exists. The interior optimum is at 10%. The existence of an interior optimum could reflect several different mechanisms. Beyond this point, almost all non-blacks may know two or more blacks well. Thus further increases in representation could increase interaction along a margin we do not directly observe. It is also possible that minority groups begin to eschew cross-racial interaction once their group share reaches some critical mass. Finally, this point could be the threshold at which the effect of mismatch in backgrounds trumps the effect of increased representation.

The second column adds the individual's SAT score relative to the school average. Those with higher SAT scores at a particular school are substantially less likely to know two or more blacks well. Given that blacks tend to be at the lower end of the within school SAT distribution, this is suggestive that similarities in the characteristics associated with SAT scores may facilitate inter-racial relationships.

It may also be the case, however, that those with high SAT scores are more likely to have discriminatory preferences or are less likely to report that they know someone well. For

example, Fryer (2005) finds that those in junior high and high school who have high grades are listed as friends of more students than those who have low grades. Column 3 adds the individual's SAT score to the regression to test whether the result is driven by high SAT score individuals being less likely to interact with blacks. The coefficient on own SAT score was small and insignificant: we cannot reject the exclusion of the individual's SAT score once we control for relative SAT score.

Table 4: Probit Estimates of the Relationship Between Percent Black and Knowing Two or More Blacks Well<sup>†</sup>

Variable	Specification		
	(1)	(2)	(3)
Share Black	16.15	16.40	14.26
	(3.31)	(3.22)	(7.16)
(Share Black) <sup>2</sup>	-0.312	-0.302	-0.332
	(0.075)	(0.070)	(0.216)
Female	-0.083	-0.136	-0.136
	(0.055)	(0.059)	(0.059)
( $SAT - \overline{SAT}$ ) (00's)		-0.105	-0.124
		(0.016)	(0.059)
$SAT$			0.019
			(0.051)
Constant	-0.338	-0.309	-0.474
	(0.121)	(0.118)	(0.401)
Log likelihood	-4755	-4720	-4720

<sup>†</sup>N=7126. Sample is all non-blacks who had valid answers for the questions on knowing two or more blacks well in the 1996 followup survey.

To further test whether similarities in SAT score are important to inter-racial interaction, we perform the same analysis except using the probabilities of knowing two or more Hispanics or Asians well. Results of these specifications, estimated using only the sample of non-Hispanic white students, are reported in table 5. Consistent with the results for knowing blacks, women

are less likely to report that they know two or more Hispanics or asians well. The standards for knowing someone well appear to be different across the sexes.

More important are the effects of relative SAT score. Although about a third of the magnitude of the corresponding coefficient in the black regression, increasing one's SAT score relative to the school average makes interaction with Hispanics less likely. This is consistent with the fact the gap between Hispanic and white SAT scores at the school level is -58 points—about a third of the gap between black and white SAT scores. The second column displays the results for asians. Relative SAT score has the opposite effect here: higher relative scores make interaction with asians more likely. This is consistent with similarity in academic background mattering as asians on average scored 56 points higher than their white counterparts. Similarity in SAT scores seems to be important not only in inter-racial relationships with blacks, but also in inter-racial relationships with Hispanics and asians.

Although including school fixed effects makes it impossible to identify how representation affects interaction, we can see if the coefficient on relative SAT score changes once school fixed effects are included. Results of these specifications are in Table 6. Once again we see that those with high relative SAT scores are more likely to interact with asians and less likely to interact with blacks and Hispanics. In addition to the results in Table 6, we estimated the model with major effects and with school-major effects. In all cases the same results emerged with little change in the magnitudes and no change in the statistical significance of the relative SAT score.

Finally, since we are relying on within school variation in SAT scores, we can estimate the model separately for each institution. Again we see strong evidence that similarity in academic backgrounds matter. For twelve of the seventeen institutions the effect of an individual's SAT score on the probability of knowing two or more blacks well was negative and significant. In only one case was there a positive sign but even then the coefficient was not close to significant. In seven of the seventeen institutions an individual's SAT score had a positive and significant effect of knowing two or more asians well. While five of the institutions had negative signs, none of the five were significant. Results for Hispanics were generally insignificant.

### **5.3 Further Tests of Likes Interacting with Likes and Evidence of Statistical Discrimination**

With evidence to this point suggesting that those with high SAT scores at a particular school are less likely to know two or more blacks well, we now attempt to relate this back to the

Table 5: Probit Estimates of the Relationship Between Group Share and the Probability Knowing Two or More Members of the Group Well<sup>†</sup>

Variable	Know Two or More:	
	Hispanics	Asians
Group Share	28.70 (3.96)	17.43 (2.72)
(Group Share) <sup>2</sup>	-0.122 (0.060)	-0.123 (0.053)
Female	-0.136 (0.025)	-0.233 (0.047)
( $SAT - \overline{SAT}$ ) (00's)	-0.031 (0.013)	0.079 (0.020)
Constant	-1.250 (0.044)	-0.658 (0.087)
Log likelihood	-3021	-3244
Average Group SAT -		
Average White SAT	-58	56

<sup>†</sup> N=5459. Sample is all non-Hispanic whites who had valid answers for the questions on knowing two or more blacks well in the 1996 followup survey.

Table 6: Probit Estimates of the Relationship Between Group Share and the Probability Knowing Two or More Members of the Group Well with School Fixed Effects<sup>†</sup>

Variable	Know Two or More:		
	Blacks	Hispanics	Asians
Female	-0.137	-0.151	-0.246
	(0.060)	(0.024)	(0.044)
$(SAT - \overline{SAT})$ (00's)	-0.105	-0.039	0.078
	(0.015)	(0.012)	(0.021)
Observations	7126	5459	5459
School Fixed Effects	Yes	Yes	Yes

<sup>†</sup> For blacks, sample is all non-whites who had valid answers for the questions on knowing two or more blacks well in the 1996 followup survey. For Hispanics and Asians the corresponding sample is for whites only.

Results are not sensitive to the sample specification.

distributions of SAT scores of blacks and non-blacks within the school. In particular, we create three individual-specific peer groups: those with SAT scores 160 points above the individual's own SAT score, 160 points below, and those with SAT scores within plus or minus 160 points. We then examine how percent black within these groups affects the probability of knowing two or more blacks well. The shares are adjusted for the size of the group such that the initial model is nested: if the coefficients of percent black in the three groups are the same, we will obtain the results from the initial model. Results are displayed in Table 7.

The first column of Table 7 uses the institutional variation used in the previous tables. Similarity in SAT scores seems extremely important to inter-racial interaction. The coefficient on the black share of those of similar SAT scores is four times as large as the black share of those with SAT scores significantly below the individual's score.

The response to share black in the different peer groups, however, is not symmetric. Namely, the coefficient on black share for those with significantly higher SAT scores is actually higher than corresponding coefficient on similar SAT scores. To test whether this difference is statistically significant, the second column of the table restricts the effect of the higher SAT

group share to be the same at those with similar SAT scores. We cannot reject the restriction that the coefficients are the same.

That the coefficients are the same for higher SAT score share and similar SAT score share stands in contrast to the significantly smaller effects for the lower SAT score share. This is consistent with the model of statistical discrimination: non-blacks choose to engage with blacks who emit the right signals and these signals are more likely to be emitted by those with higher SAT scores.

The next set of columns expands the number of groups to six. Each of the SAT groups is split in two, one for within the individual's major and one outside. Again if the coefficient estimates are the same we are back to the restricted model. Although the estimates are much more noisy, the same patterns emerge both within and outside one's major. Regardless of whether it is outside or inside the major, the black share for those with lower SAT scores is much smaller than the corresponding effects through the black shares of higher SAT scores and similar SAT scores.

## 6 Structural Model and Results

In this section we specify a model of interactions from which we can perform the policy simulations. We first describe how potential friends arrive and how individuals decide whether or not to form a friendship. We next describe our simulated maximum likelihood procedure where simulation is necessary because we do not actually observe whether to form a friendship. The limited data on interactions means that some of our parameters are not identified. However, conditional on setting one of the parameters in the model we are able to identify the others. We estimate the model and show how the results change given different values for the unidentified parameter.

### 6.1 Model

With the reduced form evidence suggesting that similarity in academic background affects inter-racial contact and that asymmetries may be present, we now proceed to the formulation and estimation of the structural model. With the structural model we will be able to forecast how interactions will change both from the changing composition of the student body and through the corresponding updating of beliefs about the academic backgrounds of their peers. All students are assigned  $N$  potential friends from among their student bodies. The

Table 7: Probit Estimates of the Relationship Between Percent Black Both In and Out of Major and Knowing Two or More Blacks Well Allowing for Variation by SAT Gap<sup>†</sup>

		Specification					
	Variable	(7)	(8)	(9)	(10)	(11)	(12)
Black Share Overall	Higher SAT	17.66 (6.47)	—				
	Similar SAT	14.98 (2.55)	15.41 (2.35)				
	Lower SAT	3.65 (1.24)	3.63 (1.26)				
Black Share in Major	Higher SAT			7.40 (49.17)	—	34.91 (51.95)	—
	Similar SAT			26.66 (15.08)	23.8 (11.80)	37.82 (18.13)	36.57 (13.60)
	Lower SAT			1.91 (6.69)	1.91 (6.57)	-1.95 (7.38)	-1.81 (7.31)
Black Share Outside Major	Higher SAT			19.56 (9.60)	—	15.46 (9.23)	—
	Similar SAT			13.96 (2.71)	14.62 (2.49)	12.09 (2.78)	12.66 (2.50)
	Lower SAT			4.01 (1.57)	3.93 (1.54)	4.43 (1.70)	4.37 (1.67)
	Share Black <sup>2</sup>	-0.314 (0.032)	-0.315 (0.032)	-0.316 (0.031)	-0.315 (0.031)	-0.335 (0.035)	-0.336 (0.035)
	Female	-0.141 (0.031)	-0.140 (0.031)	-0.138 (0.031)	-0.138 (0.031)	-0.161 (0.032)	-0.161 (0.032)
	Constant	-0.090 (0.064)	-0.093 (0.064)	-0.100 (0.065)	-0.100 (0.064)	-0.220 (0.101)	-0.222 (0.100)
	Major Fixed Effects	No	No	No	No	Yes	Yes
	Log likelihood	4729.9	4730.0	4729.6	4729.7	4716.9	4717.0

<sup>†</sup>N=7126. Sample is all non-blacks who had valid answers for the questions on knowing two or more blacks

probability of an individual of group  $i$  being assigned  $n$  potential friends of group  $j$  follows a binomial distribution:

$$Pr(n) = \binom{N}{n} \lambda_{ij}^n (1 - \lambda_{ij})^{N-n} \quad n = 0, 1, \dots, N$$

where  $\lambda_{ij}$  is given by:

$$\lambda_{ij} = \lambda_{0i} + \lambda_{1i} SHR_j$$

One student among the pair is exogenously assigned to make the decision as to whether the friendship forms given the information the student has at the time. Match quality from the perspective of the decision maker is a function of  $A$ , which the decision-maker only observes with noise,  $X$ , the observed characteristics of the decision-maker, and  $\epsilon$ , an unobserved match-specific component. Decision-makers know the distribution of  $A$  conditional on  $j$  and observe a signal on  $A$  as well. This signal,  $S$ , is given by:

$$S = A + \zeta$$

where  $\zeta$  is distributed  $N(0, \sigma)$ .  $A$  is assumed to be discrete, taking on one of  $K$  values. The probability of  $A$  taking on the value  $A_k$  conditional on the partner being a member of  $j$  is  $\pi_{jk}$ . The decision-maker updates his beliefs about  $A$  using Bayes' Rule:

$$P(A_k | S, j) = \frac{\pi_{jk} \mathcal{L}(A_k | S)}{\sum_{k'=1}^K \pi_{jk'} \mathcal{L}(A_{k'} | S)}$$

The gains from matching with someone of match quality  $A_k$  depend upon how similar  $A_k$  is to the individual's own academic background. In practice, we measure this similarity by the absolute distance between the two academic backgrounds. The expected utility for student  $i$  accepting the  $n$ th match with a member of group  $j$  is then given by:

$$EU_{in} = \frac{\sum_{k=1}^K \pi_{jk} \mathcal{L}(A_k | S_n) (\alpha_0 |A_i - A_k|)}{\sum_{k'=1}^K \pi_{jk'} \mathcal{L}(A_{k'} | S_n)} + X_n \alpha_1 + \epsilon_{in} \quad (6)$$

Student  $i$  accepts the match when  $EU_{in} > 0$ . We assume that the  $\epsilon$ 's are iid Type I extreme value. With the  $\epsilon$ 's unobserved by the econometrician, the probability of a match being accepted follows:

$$p_{in} = \frac{\exp(EU_{in} - \epsilon_{in})}{\exp(EU_{in} - \epsilon_{in}) + 1} \quad (7)$$

In practice we use SAT scores as our approximation of  $A$ . The empirical distribution of SAT scores for group  $j$  is then used in formulating the  $\pi_{jk}$ 's. The parameters to be estimated are then the arrival rate parameters, the  $\lambda$ 's, the utility function parameters, the  $\alpha$ 's, and the variance on the signal of academic background,  $\sigma$ . The next section describes the procedure used to estimate these parameters.

## 6.2 Simulated Maximum Likelihood Procedure

In the College & Beyond dataset, we actually observe very little about the matching process. We do not know anything about same-race friendships nor do we know the number of friends—actual or potential— a student has. The only information we have is whether an student had two more friends of particular ethnic group that was not their own ethnic group. We know nothing about the characteristics of the friends from other races. This necessarily restricts the scope of estimation. We focus our attention on the decisions by non-blacks as to whether to accept black matches.

We use simulated maximum likelihood as a means of obtaining consistent parameter estimates given the limited available data. Our procedure involves taking random draws in the following manner:

1. We first randomly draw an integer between 0 and 8 from the discrete uniform distribution. This integer will serve as the potential number of black friends for (non-black) individual  $i$ . We will denote this integer  $n^{(r)}$ , where  $r = 1, \dots, R$  denotes the simulation draw.
2. From the population of blacks at the college attended by individual  $i$ , randomly draw  $n^{(r)}$  people.
3. For each person drawn in the previous stage, draw a signal error, as a function of  $\sigma$ , the standard deviation of  $\zeta$ , from the normal distribution. With this draw, we can compute  $P(A_k | S_n^{(r)}, j)$  using the expression previously derived from Bayes rule. With these probabilities and the observed characteristics ( $X$ ) of the individual drawn in the previous state,  $p_{in}^{(r)}$  is calculated using the logit probability of a match being accepted in (7).

This procedure provides us with a conditional probability of a match, where we are conditioning on both the number of potential friends, and the observed characteristics of the individuals drawn from the population. An unconditional probability can be derived by integrating with respect to both variables we are conditioning on. Since this integral has no analytic form, we approximate it using importance sampling, using the standard normal distribution as the weighting function.

Next note that from these match probabilities for draw  $r$ , we can compute the probability

of two or more matches,  $p_{2i}^{(r)}$ , using the binomial formula with  $N = 8$ :

$$p_{2i}^{(r)} = 1 - \prod_{n=1}^N (1 - \lambda p_{in}^{(r,n)}) - \sum_{n=1}^N \lambda p_{in}^{(r)} \prod_{n' \neq n}^N (1 - \lambda p_{in'}^{(r)})$$

The above is an expression for a conditional probability of two or more matches, where we are conditioning on a particular simulation draw in the 3 step simulation procedure described above. Our simulation estimator of the unconditional probability averages across  $R$  3-step draws.

$$\hat{p}_{2i} = \frac{1}{R} \sum_{r=1}^R p_{2i}^{(r)}$$

Noting that this simulated probability is a function of the unknown parameters, collected and denoted here by  $\theta$ , we can now construct a maximum likelihood estimator of  $\theta = (\sigma, \gamma, \alpha)$ . For individual  $i$ , let  $d_i$  denote the indicator that he had 2 or more matches (this is in the data). Pick the value of the unknown parameters to maximize the likelihood function:

$$\hat{\theta} = \arg \max_{\theta \in \Theta} \frac{1}{\mathcal{N}} \sum_{i=1}^{\mathcal{N}} d_i \log \hat{p}_{2i}(\theta) + (1 - d_i) \log(1 - \hat{p}_{2i}(\theta))$$

Letting the number of simulation draws increase at a rate faster than the square root of the sample size, the simulated MLE will be asymptotically equivalent to the regular MLE- see, e.g. Hajivassiliou and Ruud (1994).

The above discussion assumes that  $N$ , the number of potential friends, is known. Separate identification of  $N$  and  $\lambda$  are not possible except through functional form restrictions. We therefore estimate the model for different values of  $N$  and note how  $N$  affects the results. The policy simulations are also conducted at different values of  $N$ .

### 6.3 Estimates

Estimates of the model for  $N = 40$  are displayed in Table 8. the results for other values of  $N$  between 20 and 100 are given in the appendix. The first two rows give the arrival parameters. As can be seen from the appendix, these are the parameters that are sensitive to the choice of  $N$ . Increasing  $N$  leads to smaller values for both  $\lambda_0$  and  $\lambda_1$  as the probability of meeting black individuals is increasing in  $N$ ,  $\lambda_0$ , and  $\lambda_1$ .

While the arrival rate parameters vary with  $N$ , the utility function parameters remain virtually unchanged as  $N$  changes. The reason for this is that these coefficients are driven by differences in interaction rates across individuals of different academic backgrounds—changing

$N$  does not affect the relationship between academic background and interaction. Most important to understanding the tradeoff between representation and similarity in academic background is the estimate of the parameter on the difference between own SAT score and the SAT score of the partner. Here we see that similarity in academic background is significant in determining successful matches. Since the model is highly non-linear—particularly given individuals only have expectation on academic backgrounds—how big of an effect this is is difficult to see from the estimates themselves. In the next section we use different rules for assigning students to schools which will make the tradeoff explicit. The estimate of the variance on the signal of academic background suggests—consistent with the reduced form evidence—some scope for statistical discrimination.

Table 8: Estimates of the Structural Model<sup>†</sup>

	Coefficient	Standard Error
<i>Arrival Parameters</i>		
$\lambda_0$	0.0278	0.0013
$\lambda_1$	0.2071	0.0110
<i>Utility Parameters</i>		
Intercept	2.7314	0.6117
$\text{Sex}_i$	-0.4015	0.1332
$ A_n - A_i $ (000's)	-3.9506	1.0066
$\sigma$	0.3595	0.3721
Log likelihood	-4746	

<sup>†</sup> $N=7126$ . Dependent variable is the probability of knowing two or more blacks well.

## 7 Alternative Assignment Rules

The evidence presented in the previous sections suggests that rates of inter-racial contact are governed both by minority representation and the degree of background similarity between students of different races. Given the large differences in SAT scores across blacks and whites at the College & Beyond schools and the locally positive relationship between college quality

and percent black within highly selective colleges, racial preferences may actually have a negative effect on inter-racial interaction. While we do not have the data to examine this claim directly, we are in a position to analyze the potential impact of altering the assignment of blacks and non-blacks across the institutions represented by the College & Beyond.

To do this, we need to quantify the relationship between race and college assignment. We estimate a multinomial logit model of college assignment where the probability of being assigned to one of the  $j$  schools in the data set is a function of the individual’s race and SAT score. The probability of being assigned to school  $j$  is given by:

$$Pr(j|X_i) = \frac{\exp(\beta_{0j} + BLACK_i\beta_{1j} + SAT_i\beta_{2j})}{\sum_{j'=1}^J \exp(\beta_{0j'} + BLACK_i\beta_{1j'} + SAT_i\beta_{2j'})}$$

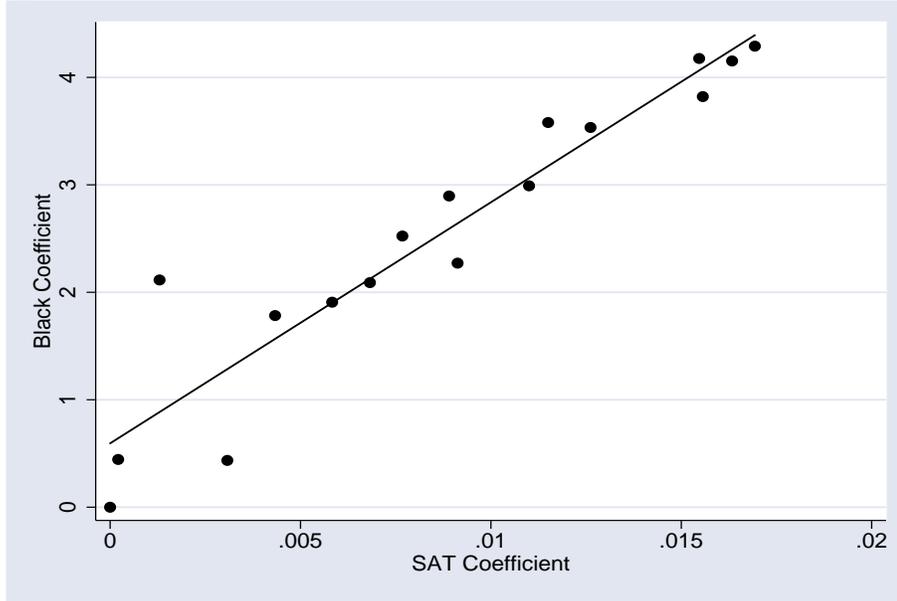
Figure 3 plots the relationship between the coefficient on SAT and the coefficient on black. The coefficient on black is highest when the coefficient on SAT score is highest. The  $R^2$  of the fitted line is 0.89. Consider two individuals, one of whom is black. The points on the fitted line imply that the non-black individual would need to have an SAT score 224 points higher than his black counterpart in order to face the same assignment rules.

We examine assignment policies that attempt to break the tie between SAT scores of the school and the race of the individual. Given the coefficients associated with the fitted line, we can simulate assignment rules that are less aggressive than those used in the College & Beyond. Namely, we can rotate this line around the mean of the SAT coefficients. A less aggressive assignment policy would then be associated with a flattening of this line.

The residuals from the fitted line represent the advantages particular schools have in attracting black students. However, there are not enough schools to be representative of the population of schools in each tier. Hence, all simulations are done using points on this line— purging the school-specific residuals from the black coefficient. Note further that some individuals who are in the College & Beyond data set would not have been able to attend any of those schools without race preferences – the data simply does not contain any completely unselective colleges. Our procedure will, admittedly unrealistically, assign the least qualified students of either race to the least selective college in the sample, even though in reality those students might attend non-sampled unselective colleges in the presence of a policy change. While we have no way of evaluating the full effects of removing race conscious admissions, we can partially address this issue by omitting the least selective sampled institution in the policy analysis. To the extent that our procedure leads minority students to “pool” at the least selective institution, this correction will improve our estimate of the overall impact. The

least selective sampled institution thus becomes the representative of all unselective colleges.

Figure 3: The Relationship Between the Coefficients on Black an SAT in the School Assignment Policies<sup>†</sup>



<sup>†</sup>Coefficients from a multinomial logit model of college assignment normalized with respect to the institution with the lowest coefficient on black. See text for details.

The estimated effects of dampening the relationship between the SAT coefficients and the black coefficients on the probability of knowing two or more blacks well are displayed in Table 9. The first column shows the aggregate effect while the next set of columns breaks out the effects by tier. Assigning blacks according to the white rules would actually yield a small increase in the probability of knowing two of more whites well. The slope of the black-SAT coefficient line that maximizes inter-racial contact is 80% of the actual slope. In other words, a small degree of racial admissions preferences appears to maximize the population rate of inter-racial contact, but the observed degree of preferences goes well beyond this maximum.

Clearly as racial preferences are weakened some blacks would be unable to obtain admission to any of the institutions in the College & Beyond. To take this into account, we repeated the assignment rules using the full set of schools and then removed the least selective sampled school when discussing the policy effects. The simulations would then be representative of what would actually happen with the removal of racial preferences if the shift of blacks to the least selective school is representative of the shift that would occur out of the College &

Beyond data set. Removing the least selective school from the policy analysis did show that completely removing racial preferences from the rest of the College & Beyond schools would lead to less inter-racial contact. However, the interaction maximizing policy would still involve a 50% reduction of the slope of the black-SAT coefficient line.

Table 9: Probabilities of Knowing Two or More Blacks Well Under Various Assignment Rules<sup>†</sup>

	Overall	Tier 1	Tier 2	Tier 3
10% weaker	0.34%	-1.39%	0.13%	1.39%
20% weaker	0.62%	-2.79%	0.15%	2.73%
30% weaker	0.85%	-4.18%	0.06%	4.04%
40% weaker	1.03%	-5.56%	-0.10%	5.29%
50% weaker	1.18%	-6.88%	-0.37%	6.51%
60% weaker	1.28%	-8.16%	-0.73%	7.64%
70% weaker	1.34%	-9.39%	-1.16%	8.71%
80% weaker	1.36%	-10.54%	-1.67%	9.72%
90% weaker	1.35%	-11.59%	-2.23%	10.65%
100% weaker	1.32%	-12.57%	-2.85%	11.50%

<sup>†</sup>See text for details on how the simulations were conducted.

The small changes at the aggregate level, however, are masking much larger changes across the tiers of institutions. Namely, top tier schools see large drops in interaction rates as race preferences are weakened with corresponding larger increases in interaction rates at tier 3 schools. A social welfare function which puts a higher premium on interactions at top tier schools may then find that the base assignment rules are optimal.

The effects of varying the assignment rules on the probability of knowing two or more blacks well are also not uniform within a school. In Table 10 we examine how this probability changes for those who have the highest test scores within their school and those who have the lowest. The standard deviation of SAT minus school average SAT for non-blacks is 125 in our data. We then examine the effects of changing the assignment rules for those non-blacks with scores 125 points above or below the school average.

Overall, those with the highest scores within a school see their interaction rates increase

the most as the assignment rules for blacks become more like the assignment rules for whites. Indeed, for this group interaction is maximized at the corner where blacks are assigned according to the white rules. Again, the benefits are coming primarily from tier 3 schools at the expense of the other schools. This is true for those with the lowest scores at each school as well. However, the blacks that are pushed from the top schools lower down have a bigger effect on those at the top SAT distribution within a school than on those at the bottom.

The results for Tier 2 schools illustrate the heterogeneity in the effects of weakening the tie between the black and SAT score coefficients. The lowest interaction rates for top SAT score students at tier 2 schools are observed under the current assignment policy. However, the current assignment policy leads to the highest interaction rates for those at tier 2 schools with the lowest SAT scores.

Table 10: Probabilities of Knowing Two or More Blacks Well Under Various Assignment Rules Conditional on SAT Rank<sup>†</sup>

	Overall		Tier 1		Tier 2		Tier 3	
	+ 1 std.	-1 std.						
10% weaker	0.92%	0.04%	-1.29%	-1.73%	0.58%	-0.30%	1.92%	0.93%
20% weaker	1.76%	0.04%	-2.59%	-3.47%	0.94%	-0.70%	3.79%	1.83%
30% weaker	2.55%	-0.02%	-3.87%	-5.20%	1.23%	-1.20%	5.63%	2.68%
40% weaker	3.27%	-0.09 %	-5.15%	-6.90%	1.47%	-1.75%	7.34%	3.52%
50% weaker	3.95%	-0.02%	-6.38%	-8.55%	1.52%	-2.42%	9.05%	4.34%
60% weaker	4.55%	-0.35%	-7.54%	-10.14%	1.46%	-3.14%	10.66%	5.08%
70% weaker	5.11%	-0.52%	-8.63%	-11.68%	1.37%	-3.93%	12.17%	5.80%
80% weaker	5.61%	-0.73%	-9.67%	-13.13%	1.11%	-4.80%	13.65%	6.44%
90% weaker	6.05%	-0.96%	-10.55%	-14.45%	0.83%	-5.69%	14.95%	7.01%
100% weaker	6.43%	-1.18%	-11.40%	-15.72%	0.47%	-6.62%	16.18%	7.59%

<sup>†</sup>See text for details on how the simulations were conducted.

## 8 Conclusion

It is commonly argued that increasing minority representation on selective college campuses will increase the likelihood, frequency, and intensity of inter-racial interaction at those institutions. While we offer no evidence here to directly contradict this presumption, this paper makes two counterpoints rooted in basic economic models of friendly interaction between agents. The first point is exceedingly straightforward. Policies that influence only the distribution of minority students, and not their total number, necessarily reduce representation on some campuses. In practice, this reduction tends to occur at moderately-selective institutions. Even in a simple model of random interaction between agents on a campus, the net impact of racial preferences in admissions can only be considered positive if interaction at these institutions is valued less than interaction at the most selective institutions.

The second point follows from a long history of social scientific research establishing that individuals tend to associate with those who are similar to them along a number of dimensions. This observation motivates a more compelling model of interpersonal interaction, where agents actively choose whether to invest in friendships. Empirically, we show that the probability of interaction between races on a campus is sensitive to the degree of mismatch between racial groups, as measured by SAT scores.<sup>17</sup> This sensitivity may lead to statistical discrimination or to behavioral norms within a school that are more race-specific than they otherwise would be.<sup>18</sup>

Because race preferences serves to exacerbate these mismatches, the increase in inter-racial contact associated with its practice is at best weakly positive. Our simulations indicate that racial admissions policies as currently practiced actually have a mild negative impact on population rates of inter-racial contact. More generally, results indicate that the use of racial preferences has only a very small impact on the population rate of inter-racial contact. Once again, in light of this evidence the racial preferences can be considered positive only if interaction at lower-tiered schools is valued less than that at the most selective universities.

We intend for these results to illustrate trade-offs present when admissions officers utilize racial preferences in admissions. Our results do suggest that racial preferences are a relatively weak tool for increasing the amount of between-race interaction that takes place on a particular campus, and, because diversity at one institution precludes diversity at another, an even

---

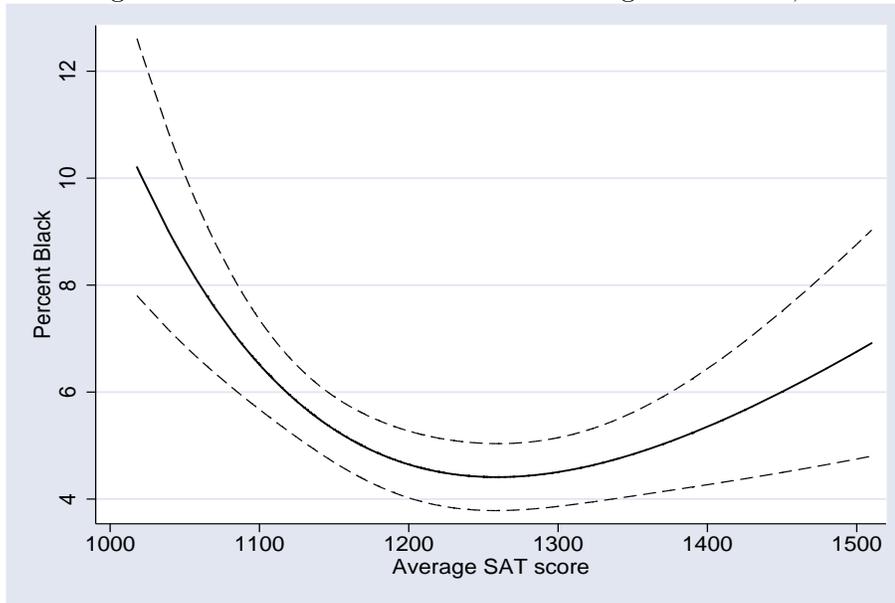
<sup>17</sup>Note that mismatch is different here from the standard argument that racial preferences put minorities in schools where they can not succeed. See Rothstein and Yoon (2006) for empirical tests of this argument.

<sup>18</sup>See Akerlof and Kranton (2000) for a discussion of identity in economics.

weaker tool for increasing the total amount of such interaction in the population.

## Appendix

Figure 4: College Percent Black as a Function of Average SAT Score, Not Weighted<sup>†</sup>



Source: Princeton Review Best 357 colleges and America's Best Colleges 2005 produced by U.S. News & World Report.

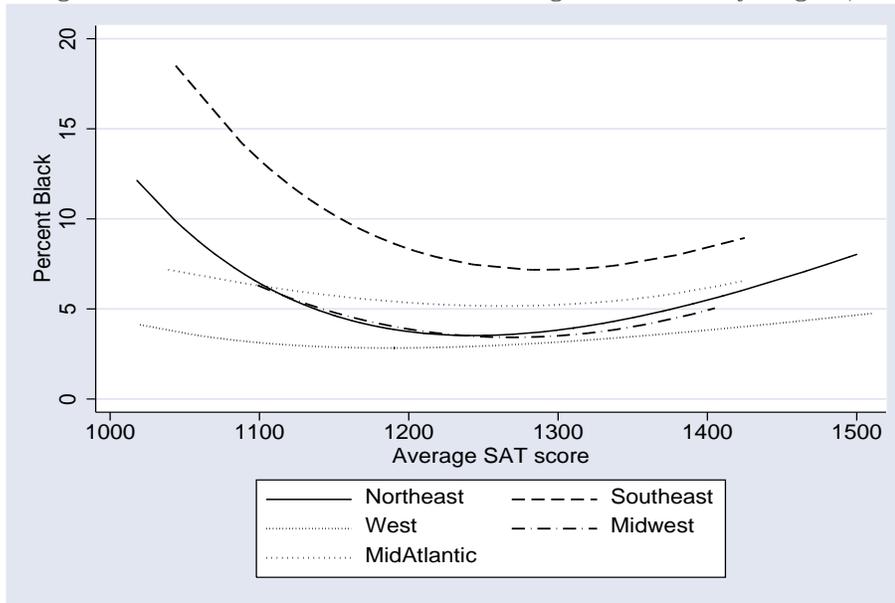
Table 11: Estimates of the Structural Model with Different Values for the Number of Potential Friends<sup>†</sup>

$R^\ddagger$	Specification						
	$\lambda_0$	$\lambda_1$	Intercept	$\text{Sex}_i$	$ A_n - A_i $	$\sigma$	Log like
100	0.0197	0.0737	2.7282	-0.4029	-3.9718	0.3589	4746.4
90	0.0219	0.0818	2.7284	-0.4028	-3.9700	0.3595	4746.4
80	0.0246	0.0918	2.7291	-0.4027	-3.9690	0.3591	4746.4
70	0.0281	0.1045	2.7291	-0.4025	-3.9658	0.3591	4746.4
60	0.0327	0.1214	2.7295	-0.4022	-3.9621	0.3593	4746.4
50	0.0392	0.1448	2.7308	-0.4020	-3.9584	0.3594	4746.4
40	0.0490	0.1794	2.7314	-0.4015	-3.9506	0.3595	4746.4
30	0.0651	0.2356	2.7332	-0.4008	-3.9386	0.3596	4746.3
20	0.0971	0.3427	2.7359	-0.3991	-3.9130	0.3608	4746.2

<sup>†</sup>See text for details on how the simulations were conducted.

<sup>‡</sup> $R$  refers to the number of simulated individuals.

Figure 5: College Percent Black as a Function of Average SAT Score by Region, Not Weighted<sup>†</sup>



Source: Princeton Review Best 357 colleges and America’s Best Colleges 2005 produced by U.S. News & World Report.

## References

- [1] Akerlof, G. and R. Kranton (2000) “The Economics of Identity” *Quarterly Journal of Economics*, v. 115, No. 3, pp. 715-753.
- [2] Arcidiacono, P. (2005) “Affirmative Action in Higher Education: How do Admission and Aid Rules Affect Future Earnings” *Econometrica*, v. 73, No. 5, pp. 1477-1524.
- [3] Arcidiacono, P. and J. Vigdor (2007) “Does the River Spill Over? Estimating the Economics Returns to Attending a Racially Diverse College” working paper.
- [4] Black, D., K. Daniel and J. Smith (1997) “College Quality and the Wages of Young Men.” Unpublished manuscript.
- [5] Black, D., K. Daniel and J. Smith (2001) “Racial Differences in the Effects of College Quality and Student Body Diversity on Wages” in *Diversity Challenged*, Harvard Educational Review.

- [6] Bowen, W.G. and D. Bok (1998) *The Shape of the River: Long-Term Consequences of Considering Race in College and University Admissions*. Princeton University Press.
- [7] Carli, L.L., R. Ganley and A. Pierce-Otay (1991) "Similarity and Satisfaction in Roommate Relationships." *Personality and Social Psychology Bulletin*, v. 17 pp. 419-426.
- [8] Duncan, G.J., J. Boisjoly, D.M. Levy, M. Kremer, and J. Eccles (forthcoming) "Empathy or Antipathy? The Consequences of Racially and Socially Diverse Peers on Attitudes and Behaviors." *American Economic Review*.
- [9] Hajivassiliou, V.A. and P.A. Ruud (1994) "Estimation Methods Using Simulation," in *Handbook of Econometrics, Vol. 4*, eds. R. Engle and D. McFadden, New York: Elsevier Science.
- [10] Foster, J.R. (2005) "Making Friends: A Non-experimental Analysis of Social Group Formation." *Human Relations*, v. 58 pp. 1443-1465.
- [11] Henderson, M. and A. Furnham (1985) "Similarity and Attraction: The Relationship between Personality, Beliefs, Skills, Needs, and Friendship Choice." *Journal of Adolescence* v. 5 pp. 111-123.
- [12] Kandel, D.B. (1978) "Homophily, Selection, and Socialization in Adolescent Friendships." *American Journal of Sociology* v. 84 pp. 427-436.
- [13] Kane, T. (1998) "Racial and Ethnic Preferences in College Admissions." in Christopher Jencks and Meredith Phillips (eds.) *The Black-White Test Score Gap*, Brookings Institution.
- [14] Marmaros, D. and B. Sacerdote (2006) "How do Friendships Form?" *Quarterly Journal of Economics*, v. 121, No.1, pp. 79-119.
- [15] McPherson, J.M. and L. Smith-Lovin (1988) "Homophily in Voluntary Organizations: Status Distance and the Composition of Face-to-Face Groups." *American Sociological Review* v. 52 pp. 370-379.
- [16] Rothstein, J. and A. Yoon (2006) "Mismatch in Law School." working paper.

- [17] Weinberg, B. (2005) “Social Interactions with Endogenous Association.” working paper.