Sorting and Statistical Discrimination in Schools: An Analysis Using the National Longitudinal Study of Adolescent Health

By

Anil Nathan

May 2009

COLLEGE OF THE HOLY CROSS, DEPARTMENT OF ECONOMICS
FACULTY RESEARCH SERIES, PAPER NO. 09-05*

Department of Economics
College of the Holy Cross
Box 45A
Worcester, Massachusetts 01610
(508) 793-3362 (phone)
(508) 793-3710 (fax)

http://www.holycross.edu/departments/economics/website

*All papers in the Holy Cross Working Paper Series should be considered draft versions subject to future revision. Comments and suggestions are welcome.

Sorting and Statistical Discrimination in Schools: An Analysis Using the National Longitudinal Study of Adolescent Health

By
Anil Nathan[†]
College of the Holy Cross

May 2009

Abstract

The rigorous economic analysis of peer group formation is a burgeoning subject. Much has been written about how peers influence an individual's behavior, and these effects are quite prevalent. However, less has been written on how exactly these peer groups begin and the resulting consequences of their formation. A reason for the dearth of knowledge on peer group formation is the lack of quality data sets that clearly define one's peers. To resolve this issue, this paper explores data which allows a peer group to be defined openly through self nominations. Using these nominations as well as characteristics of the students and their friends, it is possible to see on what dimensions these individuals are sorting into friendships. The data suggests that there is heavy sorting within race and academic ability. Additionally, tests for statistical discrimination on race and academics show that it is exhibited towards blacks and Hispanics. There is also weak evidence of statistical discrimination against whites. Empirical analysis also shows that the degree of statistical discrimination decreases for blacks and Hispanics over a year; however, there is little change for whites over the same period. This result suggests a process of learning about a noisy signal on academic characteristics. Future work includes models describing the benefit of having various friends and the probability of forming those friendships, which can be used to simulate redistribution policies.

JEL Classification Codes: J15, I2

Keywords: National Longitudinal Study of Adolescent Health, Add Health, friendship formation, statistical discrimination, school redistribution

[*]This research uses data from Add Health, a program project designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris, and funded by a grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 17 other agencies. Special acknowledgment is due Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Persons interested in obtaining data files from Add Health should contact Add Health, Carolina Population Center, 123 W. Franklin Street, Chapel Hill, NC 27516-2524 (addhealth@unc.edu). No direct support was received from grant P01-HD31921 for this analysis.

[†]Department of Economics, College of the Holy Cross, Worcester, MA 01610-2395, 508-793-2680 (phone), 508-793-3710 (fax), anathan@holycross.edu

Keywords: National Longitudinal Study of Adolescent Health, Add Health, friendship formation, statistical discrimination, school redistribution

JEL: J15, I2

1 Introduction

While the analysis of peer group formation and peer effects is well established in sociology and psychology, it has only recently been broached using rigorous economic analysis. Generally, the effect of one's peer group on one's behavior in many facets is very strong and profound. Knowing how one's peers actually come into being, however, is a subject that is not as well understood as what happens after the group is formed. Nevertheless, the process of peer group formation can be important in many ways.

Specifically, an understanding of sorting into peer groups is important whenever the distribution of characteristics on which individuals sort affects an outcome. Examples of such characteristics are race, academics, and attitudes (Clotfelter 2004). Two of the mechanisms that can affect sorting along these lines are the following.

- Homophily: A sociological term where individuals associate with others of similar characteristics (Kandel 1978).
- Statistical Discrimination: The perception and treatment of individuals based on the discernment of group characteristics; this is also known as stereotyping.

These two different but related processes may affect, for example, how students who are redistributed into schools based on race are received by the original members of the school, how the original members are received by those redistributed, and if any economic returns can be gained by the distributional changes. For example, Arcidiacono and Vigdor (2005) find that there are weak effects between racial diversity in college and post-graduation outcomes for white and Asian students. Perhaps the reason that the effects are not stronger are that white and Asian students remain entrenched in their racial groups due to preference (homophily) or that they only associate with members of other races of whom they perceive give a signal that is both of the following:

- Different from a mean perceived signal from the other races
- Similar to their own characteristics

Both of the above characteristics occurring together are hallmarks of statistical discrimination. If, for example, the signal is on intelligence/academics, both homophily and statistical discrimination may dampen any sort of gains to diversity that are attempted to be exploited by administrators and policy makers.

Knowing if homophily and statistical discrimination exist and the magnitudes of such effects based on a change in distributions is an important part of designing any sort of redistribution policy, such as school redistricting and affirmative action.[1] It is also important to know if repeated contact with peers over time results in any behavioral change with regards to homophily and statistical discrimination. With repeated exposure to signals and characteristics of potential peers, homophily and statistical discrimination may change. For example, after a school redistribution program is first implemented, homophily and statistical discrimination may decrease in magnitude. This result may further the goals of administrators and policy makers with regards to the policy's initial intentions of having more interracial contact; this may in turn allow for a better scholastic experience and better economic outcomes outside of school.

In order to address the above issues, this chapter will conduct the following analysis.

- Determine patterns of sorting across racial, academic, and attitudinal lines using two waves of the the National Longitudinal Study of Adolescent Health (Add Health).
- Develop and estimate a model of homophily.
- Alter the homophily model to develop and estimate a model of statistical discrimination.

Section 2 provides an in-depth background of the research and issues involved in peer effects analysis, peer group formation, and policies of integration and redistribution, as well as explaining the value added by this chapter to the literature. Section 3 describes the Add Health data and explains the key features of Add Health that are exploited. Section 4 shows descriptive evidence of sorting along racial, academic, and attitudinal lines. Section 5 develops a model of homophily and shows the results that homophily does at least weakly exist along certain lines and changes over time. Section 6 alters the model of homophily to describe a measure of statistical discrimination and presents results which show that statistical discrimination exists towards blacks and Hispanics, which changes over time. Section 7 revisits the

[1] A good example is Boston's Metropolitan Council for Educational Opportunity (Metco) Program, which redistributes minorities that must meet a certain academic standard across schools in the Greater Boston area. The main aims of the program are to help desegregate Boston area schools as well as provide opportunities to certain minorities by transferring them to advantageous school districts (Drews 2006).

motivation of the chapter presented above with results in hand, and outlines implications and future work.

2 Literature Review

A large portion of the literature on peer groups focuses on their effects rather than their formation. Peer groups have been shown to be very important in many facets of life, and can influence behavior immensely. Although this chapter does not analyze any peer effects per se, it is important to know that many papers have attempted to analyze these effects. Case and Katz (1991) explain that there exist significant neighborhood peer effects on drug and alcohol usage, church attendance, and unemployment from youth through adulthood. Duncan, Boisjoly, Kremer, Levy, and Eccles (2006) show that the racial composition of freshman housing assignments has an effect on student attitudes towards their peers. Even the potential biases of peer groups are explored in depth, such as the reflection problem in Manski (1993) where reference (or peer) groups may exhibit an endogenous social effect on individuals, as well as potential corrections using non-linearities outlined in Brock and Durlauf (2001). Weinberg (2003) goes even further by suggesting a model of social interaction with endogenous association using the National Longitudinal Study of Adolescent Health (Add Health), the same set of data that is used in this chapter. In many of these papers, however, peer groups are very anomalous. Add Health has clearly defined peer groups, so there can be more confidence that the effects calculated are from actual peer groups instead of assumed peer groups. The next chapter of this document will explore actual peer effects.

Discerning how a peer group forms and on what dimensions they sort has not been attempted as much due to the lack of proper data on peer groups. However, the theory of group formation has been explored in detail, especially by psychologists. Raino (1966) and Tuma and Hallinan (1978) believe that similarity and status are two important precursors to friendship. Blau (1964) offers a model where an agent calculates the expected benefit and cost of forming a friendship before making a decision on the friend. Akerlof and Kranton (2002) form a theoretical framework of group formation amongst students. They suggest that students match their characteristics to a set of pre-existing social categories. Students receive greater utility by matching to a group that is most similar to their observed characteristics. After first choosing the group, they then choose how much effort they put into schooling (an example of

a peer effect), which is conditional on group choice. However, the authors do not empirically test their premises. Marmaros and Sacerdote (2006) suggest a model where the expected benefit of a friendship is dependent on information gathered and any shared experiences, while the cost is the time used to develop the friendship. They do not assume that the individual can predict with a reasonable degree of certainty who would be a good friend. Arcidiacono, Khan, and Vigdor (2008) develop a model of interracial contact where individuals want to match with a friend who is similar academically, but where the signal of academic quality is noisy. This results in individuals statistically discriminating over any potential friends.

Marmaros and Sacerdote (2006) use a unique dataset from Dartmouth that measures the level of social interaction between any two individuals as the amount of e-mail sent between them. They find that the greatest dimensions of sorting are along racial lines and geographic boundaries by estimating poisson regressions of the the number of e-mails sent between any two people on various characteristics. Although e-mails may be a reasonable proxy for friendships, this chapter aims to use the more concrete friendship nomination data in Add Health. Foster (2005) and Arcidiacono, Khan, and Vigdor (2008) also use datasets that list characteristics of respondents and how many friends that they have across different lines.[2] However, it is not possible to identify the actual friendship nominations using these sets of data. Therefore, demographic data, apart from the line that is being matched, on friendship nominations is unavailable. Add Health has complete demographic data on friendship nominations within a particular school, so it is possible observe matches across multiple lines. This feature of Add Health is exploited, which is the true value added by this chapter.

Racial diversity is a very important topic that pertains to schools; programs such as school desegregation and busing have been implemented with the intention of forming new peer groups and fostering better educational and cultural outcomes. One such outcome is the elimination of the black/white achievement gap. Bowen and Bok (2001) argue that learning across races takes place and is quite useful; while Clotfelter (2004) chronicles how important a topic such as school desegregation in America is to both whites and blacks alike, and show how desegregation programs may have led to "white flight." On the other hand, Bifulco and Ladd (2007) show that black students choose charter schools in North Carolina with a higher portion of black peers, despite poorer results than private schools. In essence, racial

[2]Foster (2005) uses data from the University of Maryland registrar, while Arcidiacono, Khan, and Vigdor (2008) use the College and Beyond data.

homogeneity is chosen over academic excellence. However, there is some dispute as to the importance of racial composition on academic outcomes. Rumberger and Palardy (2001) argue that the socioeconomic level of students' schools as well as students' own socioeconomic status have about the same impact on achievement growth for both advantaged and disadvantaged students as well as for both white and non-white students. These findings question whether integration policies have any impact at all.

It is also argued that integration policies may actually lead to more segregation if there is a small minority present. Moody (2001) argues that segregation through clubs and sports can result in the appearance of segregation based on race. The same process could happen if students are on academic "tracks," such as honors classes. In these cases, racial sorting patterns like homophily and behaviors such as statistical discrimination can be confounded through other factors. However, Xie and Zeng (2002) model the selection of friendship based on "choice" (which includes dimensions such as race) and "opportunity" (which includes scholastic institutions that segregate) using a conditional logit framework. They find that race is the most important factor in choosing friendships. Regardless, this chapter checks for robustness of results by controlling for these issues through a random effects framework, with individuals as the group variable.

Finally, an important reason to model how peer groups form is to be able to perform simulations in order to see if and how new peer groups begin after a redistribution program. Marmaros and Sacerdote (2006) use their poisson model coefficients to simulate different housing programs at Dartmouth by moving students around, which they find to be very small because the geographic effect on peer group formation is only relevant over small distances, so it is difficult to get students close to a large number of other students conditional on the dormitory structure of Dartmouth. For this chapter and the data in general, schools are generally large, so any simulated redistribution program would not run into proximity constraints. However, Marmaros and Sacerdote do find that changing the entire composition of the Dartmouth class does change peer groups along many different dimensions. The results of this chapter can be used in the construction of a structural model to attempt similar redistribution and composition changing simulations. For example, school redistribution can be simulated by redistributing disadvantaged students into a high achieving school district. The peer groups formed after the redistribution can then be analyzed.

3 Data-The National Longitudinal Study of Adolescent Health (Add Health)

Add Health is a nationally representative study that explores the causes of health related behaviors of adolescents in grades 7 through 12 and their outcomes into young adulthood (Udry 2003). It seeks to examine how social contexts (families, friends, peers, schools, neighborhoods, and communities) influence adolescents' health and risk behaviors.

The study was initiated in 1994 with an In-School survey administered to a nationally representative sample of about 90,000 students in grades 7 through 12 in 132 schools. These schools were selected to ensure that region, population density, size, type, and ethnicity were representative of the national population. It was administered in one day in a 45-60 minute class period. Questions asked include those about social and demographic characteristics of respondents, self-reported grades, education and occupation of parents, household structure, risk behaviors, expectations for the future, self-esteem, health status, friendships, and school-year extracurricular activities.

The friendship nominations will be exploited in this chapter. Each individual in the surveys could nominate up to five male friends and five female friends, and they were asked to rank friends in order of preference. The friends can either be from the individual's current school, a sister school, or from neither the current nor sister schools. About 15% of the friendship nominations in the In-School survey are not in the current or sister school of the individual, while about 8% of the nominations are from the individual's current school but not on the school's roster.

All students who completed the In-School survey plus those who did not complete a survey but were listed on a school roster were eligible for selection into the core In-Home sample. It is a sample of adolescents in grades 7 through 12 in the US during the 1994-95 school year. The survey is clustered around schools (although students took the survey at home). Students in each school were stratified by grade and gender. About 17 students were randomly chosen from each stratum so that a total of approximately 200 adolescents were selected from each of the 80 pairs of schools. A total core sample of about 12,000 adolescents was interviewed. A special oversample of well educated blacks, Chinese, Cuban, Puerto Rican, disabled, and

some sibling pairs brought the total number of those who completed the In-Home survey in Wave I to about 20,000. In addition to the questions asked in the In-School survey, the In-Home survey asks additional questions on health status, health-facility utilization, nutrition, decision-making processes, family composition and dynamics, educational aspirations and expectations, employment experience, romantic experiences, substance use, and criminal activities. An aptitude test called the Add Health Picture Vocabulary Test (AHPVT) was administered before the survey. The AHPVT employs a series of images and words that describe these images in order to measure aptitude. The student must pick the word that best describes the picture. For example, a picture of a furry dog could be displayed to the user, with the words "furry," "greasy," "slimy," or "smooth."[3] Wave II of the In-Home Survey took place in 1996, and about 15,000 of the original respondents were retained. Wave III took place in 2001-2002, at which time the original respondents were asked about their current life situations. Topics included were current friends, drug usage, romantic relationships, and other activities. Additionally, they completed another AHPVT test.

4 Descriptive Statistics, Measures and Sorting Patterns

4.1 Descriptive Statistics

Sorting into friendships can potentially occur along many different dimensions. Race is often assumed to be a primary dimension, but others such as school performance and attitudes can affect friendship formation as well. Tables 1 and 2 describe the observations of the sample along race, academic, and attitudinal lines for the entire population using the In-School survey. Although sampling weights, stratification rules, and clustering procedures are provided in the data set, they were not used to construct these summary statistics. Therefore, these sample statistics should only be interpreted as accurate for the sample only, and not the general population.

Asians have the highest Grade Point Average (GPA) in English, math, sciences, and social studies, followed closely by whites.[4] Blacks and Hispanics subsequently have lower self-

[3]The AHPVT standardized score ranges from 13-146.
[4]GPA's are standard, where A=4.0 and F=0.0.

Table 1: Sample Statistics, In-School Survey

Race	Percentage
White Only	52.03%
Black Only	15.42%
Asian Only	4.63%
Hispanic Only	17.22%
American Indian Only	0.93%
Other Only	1.66%
Mixed (more than one race)	5.92%
Nothing (no race specified)	2.20%
Number of Observations	89940

Table 2: Sample Statistics, In-School Survey, Part 2

		Total	White	Black	Asian	Hispanic
GPA	> 3.3	24.53%	29.32%	13.53%	39.09%	14.2%
	2.3-3.3	37.92%	37.6%	40.56%	38.82%	36.42%
	1.3-2.3	28.36%	25.1%	35.9%	18.08%	35.79%
	0.3-1.3	8.82%	7.65%	9.71%	3.8%	12.99%
	0-0.3	0.38%	0.33%	0.29%	0.21%	0.61%
Likely to	0-Unlikely	4.25%	3.72%	3.00%	2.27%	6.71%
Graduate College	1	1.29%	1.30%	0.95%	0.32%	1.65%
	2	5.64%	4.46%	6.88%	3.25%	9.54%
	3	1.23%	1.29%	0.94%	0.57%	1.49%
	4	8.89%	8.37%	8.48%	6.06%	11.65%
	5	2.25%	2.44%	1.45%	1.55%	2.18%
	6	14.76%	15.27%	13.26%	13.33%	15.29%
	7	10.43%	12.59%	6.20%	10.25%	7.38%
	8-Certainly	51.07%	50.56%	58.84%	62.41%	44.13%
Happy to	1-Strongly Agree	24.76%	25.54%	22.70%	23.54%	25.54%
Attend School	2	32.90%	33.96%	29.34%	37.03%	33.75%
	3	23.59%	22.87%	25.49%	25.47%	22.77%
	4	9.46%	9.16%	11.40%	7.82%	8.68%
	5-Strongly Disagree	9.29%	8.47%	10.99%	6.15%	9.26%

Table 3: Sample Statistics for the Representative Population, In-Home Survey

Race	Percentage
White Only	65.18%
Black Only	14.52%
Asian Only	0.57%
Hispanic Only	12.27%
American Indian Only	3.33%
Other Only	0.77%
Mixed (more than one race)	3.34%
Nothing (no race specified)	0.00%
Number of Observations	13568

reported GPA's. It is interesting to note that blacks have a higher confidence in their future success in college, despite the grades. Asians are the most confident in their future collegiate success, while Hispanics have the least amount of confidence. The students' happiness at their school is reasonably even across races. The percentages of those in each race who responded positively to their school are close. However, there is a disparity between blacks and Asians who strongly disagree that they are happy at the school, with about 10% of blacks admitting that they are strongly unhappy as opposed to only about 6% of Asians having similar feelings.

Tables 3 and 4 describe the observations of only those in both wave 1 and wave 2 of the In-Home survey. The means reported in the table are population weighted to reflect sampling procedures (including the oversample). While most statistics seem to reflect the uncorrected values in the In-Home survey, there are a couple of exceptions.

The biggest exception is the severe undersampling of Asians, which even after correcting for the population weights, stratification, and clustering, does not reflect the uncorrected values for Asians in the In-School survey. Also, the descriptive statistics for Asians (not shown in Tables 3 and 4) do not reflect the values in the In-School survey. In fact, they are much lower achievement-wise. For these and other reasons, Asians are dropped from this analysis.[5]

[5] Arcidiacono and Nathan (2007) are working on describing and structurally modeling sorting and statistical discrimination patterns using the wave 1 In-School survey, in which Asians are used in the analysis.

Table 4: Sample Statistics for the Representative Population, In-Home Survey, Part 2

		Total	White	Black	Asian	Hispanic
GPA-Wave 1	> 3.3	22.95%	27.22%	11.30%	7.95%	12.88%
	2.3-3.3	41.01%	39.91%	45.07%	29.11%	40.19%
	1.3-2.3	28.53%	25.53%	36.73%	53.66%	36.97%
	0.3-1.3	7.08%	6.88%	6.35%	9.29%	9.45%
	0	0.45%	0.46%	0.55%	0.00%	0.50%
GPA-Wave 2	> 3.3	21.60%	25.45%	9.56%	3.73%	12.17%
	2.3-3.3	37.82%	37.47%	41.29%	37.25%	36.75%
	1.3-2.3	30.77%	27.98%	38.48%	45.98%	36.70%
	0.3-1.3	9.38%	8.64%	10.15%	13.02%	14.17%
	0	0.44%	0.47%	0.52%	0.00%	0.21%
Mean AHPVT Standardized Score		100.77	104.62	91.42	98.38	92.71
Likely to Graduate College-Wave 1		51.19%	49.76%	58.83%	32.29%	46.08%
Likely to Graduate College-Wave 2		38.61%	40.90%	37.27%	14.29%	27.18%
Happy to Attend School-Wave 1		26.37%	27.04%	22.00%	18.23%	29.37%
Happy to Attend School-Wave 2		22.75%	23.79%	19.59%	20.05%	22.27%

4.2 Measures- Friends

The In-School and all three waves of the In-Home surveys ask for friendship nominations. There is no data for those friends who are not on the school roster, those friends who are from a sister school, or those friends who are not in the school. Therefore, these friends are excluded from the analysis. Each male and female friend must be from the same school as the respondent and must have been surveyed as well. Of the 44,811 males in the In-School survey, 31,535 have at least one male friend who is on the individual's current school roster, and of the 44,401 females in the In-School sample, 35,688 have at least one female friend who is listed. Of the 7,190 males in the In-Home survey who are listed in both wave 1 and wave 2, 4,254 have at least one male friend listed, and of the 7,546 females listed in wave 1 and wave 2, 4,529 have at least one female friend listed.

From the friendship nominations, a binary variable on whether an individual has a friend of a certain characteristic may be constructed. The analysis is limited to same-gender friends for simplicity and to avoid some confounding factors such as romantic relationships. If romantic and platonic relationships across races and achievement do not follow the same patterns, then platonic friendship results can be biased by including romantic partners among friends.[6] Concentrating on same-gender friendships eliminates most romantic relationship possibilities.[7]

The In-Home survey allows respondents' friendship nominations to include those who have taken the In-School survey, but have not taken the In-Home Survey. Since respondent identifiers are consistent throughout all the surveys and waves, it is possible to back out relevant characteristics from the In-School survey that otherwise would not be available in just the In-Home survey.

4.3 Measures- Self-reported GPA vs. AHPVT as an Achievement Measure

Two potential measures for achievement include self-reported GPA and AHPVT score. There are advantages and disadvantages to using each measure. The advantage to using GPA is due to the number of people who report their grades in the In-School survey. As explained

[6] There are patterns in interracial romances that belie the general population (Foeman and Nance 1999).

[7] Another assumption is that male-female platonic friendships follow similar patterns to same-gender platonic friendships.

above, it is possible to back out the GPA of friends nominated in the In-Home survey, but not a respondent in the In-Home survey itself, from the In-School survey. The disadvantages to using GPA are that it can potentially be a noisy measure of achievement and that wave 2 GPA's for those friends nominated in the In-Home survey are unavailable.

AHPVT, on the other hand, has the advantage of generally being constant over time, since, like an IQ test, it attempts to appraise innate intelligence. The AHPVT is only taken by those in the In-Home survey in wave 1. However, it is plausible to suggest that an AHPVT score in wave 2 would be very similar to the AHPVT in wave 1. The big disadvantage to using the AHPVT is that the In-School survey cannot be used to back out the scores of friends nominated in the In-Home survey who were not respondents in the In-Home survey. Therefore, the number of friendship nominations who have AHPVT scores is significantly less.

As a result of these issues, this chapter uses GPA as the main measure for achievement. For those friendship nominees who did not take the In-Home survey in wave 2 (and thus have no GPA in wave 2), the GPA from wave 1 is carried over to wave 2.

4.4 Measures- Probabilities

In order to see if individuals in the sample are sorting across racial, academic, and attitudinal lines, it is important to compare the friendships that are actually formed with a random assignment of friendships in each school, where the friendship nominations in the sample originate. For example, group A may not have much interaction with group B due to the fact that they are not often in the same setting, so the probability that they form a random friendship is remote. However, if the probability that an individual from group A actually forms a friendship with an individual from group B is different from the probability of a random friendship conditional on the setting and characteristics, it may signal a sorting pattern. Sorting into a certain category is implied by a higher actual probability of friendship formation than the corresponding random probability of friendship formation, while sorting away from a certain category is implied by the opposite relationship.

- Actual $Pr(B|A)$ = Friendship probability from the data
- Random $Pr(B|A)$ = Indiscriminate matching probability within a school

Table 5: Sorting Along Racial Lines (Wave I)

Respondent	Probability	Have Same Gender Friend		
		White	Black	Hispanic
White	actual	92.77%	1.24%	4.93%
	random	80.65%	11.02%	7.90%
	ratio	1.15	0.11	0.62
Black	actual	5.80%	83.09%	10.54%
	random	28.89%	60.98%	9.73%
	ratio	0.20	1.36	1.08
Hispanic	actual	26.03%	4.42%	65.86%
	random	34.94%	16.11%	47.64%
	ratio	0.75	0.27	1.38
GPA>3.3	actual	83.23%	6.38%	6.34%
	random	73.46%	15.52%	10.45%
	ratio	1.13	0.41	0.61
Good College Prospects	actual	75.63%	12.65%	8.38%
	random	67.33%	20.70%	11.40%
	ratio	1.12	0.61	0.74
Happy at School	actual	77.94%	9.31%	9.77%
	random	69.55%	18.16%	11.75%
	ratio	1.12	0.51	0.83

Tables 5-8 list the actual and chance probabilities for respondents in the In-Home sample, divided on various lines, having listed same-gender friends along those same lines. The actual probabilities of friendship are calculated straight from the sample conditional on the individual's characteristics in the table. The random friendship probabilities are calculated by taking the mean of the relevant characteristic by school (since all friendships are contained within the school in the sample), conditional on an individual's characteristics in the table. The ratio of the actual probability to the random probability is reported. Any difference in the number of observations between the actual and random probabilities comes from individuals who are in the school population who do not list any valid friends.

In Tables 5 and 6, the actual probabilities that a respondent has a same-gender friend of the same race are higher for all races than the probability of randomly having a same-gender friend of the same race. Whites self-segregate less with regards to race than both blacks and Hispanics. In both waves, blacks are heavily sorting away from whites, due to the fact that

the actual probability of a black student naming a white student as a same-gender friend is much lower than the probability of random interactions. Whites have a higher ratio of actual probability to random probability with regards to having Hispanic friends when compared to the ratio with respect to black friends. There is not a large difference between waves when comparing racial sorting. Academic and attitudinal sorting follow expected patterns, where higher achievers and those with good attitudes about college and their happiness in school sort into having white friends, and sort away from blacks and Hispanics.[8] As a quick test to see along what lines (racial, academic, and attitudinal) sorting dominates, similar probabilities were calculated using the In-School survey, but with the columns labeled as having a same-gender friend who has a high GPA, has good college prospects, and is happy at the school. Whites sort slightly into high academic achievement friendships, while blacks and Hispanics sort away from these types of friendships. Also, whites seem to sort into friendships with positive attitudes about school, while blacks and Hispanics sort away from friendships with positive attitudes about school. The ratios between the actual and random probabilities of interaction in those cases are much closer to 1 than when friends are separated by race. In fact, the sorting patterns observed in these descriptive tables imply that racial sorting may drive academic and attitudinal sorting. Overall, the patterns of sorting illustrated imply that racial sorting occurs more than academic sorting, which in turn is more prevalent than attitudinal sorting.

Tables 7 and 8 take the strongest patterns of sorting (racial and academic lines) and make each item both racially and academically dependent in wave 1 and wave 2 of the In-Home survey, respectively. Same-gender friends are categorized by their race and their achievement based on GPA together. Once again, there is sorting within an individual's own category. For whites and blacks, high achievers tend to self-segregate themselves at a higher rate than low black and white achievers in both wave 1 and wave 2. In fact, the degree of self segregation that occurs for high achieving blacks is almost triple that of the degree of self-segregation for low-achieving blacks, as measured by comparing ratios between the groups. This result is similar to one in Bayer, Fang and MacMillan (2005), where highly educated blacks self-segregate

[8] In this particular context, an individual with a GPA of 3.3 (B+) or above is considered a "high" achiever, while an individual with GPA below 3.3 is considered a "low" achiever.

Table 6: Sorting Along Racial Lines (Wave II)

Respondent	Probability	Have Same Gender Friend		
		White	Black	Hispanic
White	actual	92.40%	1.17%	4.87%
	random	80.42%	11.19%	7.94%
	ratio	1.15	0.16	0.61
Black	actual	4.66%	85.15%	9.68%
	random	28.60%	61.33%	9.64%
	ratio	0.16	1.39	1.00
Hispanic	actual	29.72%	4.33%	62.26%
	random	34.67%	16.52%	47.47%
	ratio	0.86	0.26	1.31
GPA>3.3	actual	83.07%	5.97%	7.27%
	random	74.35%	14.57%	10.63%
	ratio	1.12	0.41	0.68
Good College Prospects	actual	74.63%	13.73%	8.47%
	random	67.95%	19.73%	11.84%
	ratio	1.10	0.70	0.72
Happy at School	actual	76.05%	11.56%	9.75%
	random	66.95%	18.95%	13.57%
	ratio	1.14	0.61	0.72

more than blacks who are less educated. Perhaps the small sample size of high-achieving blacks who actually have legitimate friends for analysis (61 in wave 1, 46 in wave 2) may be skewing results. A check was instituted using the In-School survey, where there are no sample size problems.[9] The degree of self-segregation among high-achieving blacks was confirmed to be higher than the degree of self-segregation among low-achieving blacks in that sample as well. High-achieving Hispanics who have actual legitimate friendships also have a sample size problem here, as there are instances where there are no friends who are high-achieving Hispanics. However, when comparing similar statistics with the In-School survey, it is true that it is very unlikely, for example, for a low-achieving white to have a high-achieving Hispanic as a friend. The same is true regarding other cells that are empty in the In-Home waves (i.e. high-achieving whites and low-achieving Hispanics are very unlikely to actually have a high-achieving black friend). Therefore, the sample size problem does not affect how actual and random probabilities are compared. Finally, when looking across waves, it seems that high-achieving whites are integrating more with other racial/academic groups (actual probability of 52.69% of self-segregation in wave 1 compared to 44.26% in wave 2). The opposite is happening for high-achieving blacks (actual probability of 28.68% of self-segregation in wave 1 compared to 44.22% in wave 2), although this result may be driven by small sample sizes. There is not much change in low achievers across all racial groups with regards to self-segregation over time. Overall, descriptive statistics lend credence to the fact that homophily actually does exist and that it transcends both race and academics. Now, formal models of both homophily and statistical discrimination can be developed.

5 Homophily

5.1 Model

According to tables 2 and 4, it is apparent that whites on average have stronger GPA's than their black and Hispanic counterparts. Tables 5-8 show that there are heavy interactions within race and within academic achievement. Assume that there is a large influx of one race into another school, for example. Will the subsequent changing of the distribution of academic

[9] Results in this chapter are compared to the In-School survey that is not corrected for survey design issues.

Table 7: Sorting Along Racial and Academic Lines (Wave I)

Respondent	Probability	Have Same Gender Friend					
		White >3.3	White <=3.3	Black >3.3	Black <=3.3	Hispanic >3.3	Hispanic <=3.3
White >3.3	actual (N=473)	52.69%	42.82%	0.06%	0.68%	1.78%	0.95%
	random (N=1805)	26.31%	54.34%	1.43%	9.74%	1.40%	6.37%
	ratio	2.00	0.79	0.04	0.07	1.27	0.15
White <=3.3	actual (N=1170)	23.16%	72.26%	0.00%	0.62%	0.19%	3.45%
	random (N=5043)	20.28%	60.82%	1.27%	9.45%	1.10%	6.66%
	ratio	1.14	1.19	0.00	0.07	0.17	0.52
Black >3.3	actual (N=61)	0.84%	2.26%	28.68%	63.68%	0.00%	3.36%
	random (N=347)	8.27%	20.42%	12.52%	49.03%	1.39%	8.10%
	ratio	0.10	0.11	2.29	1.30	0.00	0.42
Black <=3.3	actual (N=358)	1.05%	6.56%	7.51%	75.77%	0.09%	8.57%
	random (N=2213)	7.75%	21.14%	7.49%	53.72%	8.14%	1.41%
	ratio	0.14	0.31	1.00	1.41	0.01	6.09
Hispanic >3.3	actual (N=48)	13.00%	14.37%	0.00%	10.60%	15.45%	34.03%
	random (N=264)	12.71%	25.58%	2.56%	15.83%	7.35%	33.96%
	ratio	1.02	0.56	0.00	0.67	2.10	1.00
Hispanic <=3.3	actual (N=428)	6.54%	12.15%	0.15%	6.93%	8.01%	61.82%
	random (N=1899)	9.99%	25.45%	2.23%	13.97%	6.17%	40.94%
	ratio	0.65	0.48	0.07	0.50	1.30	1.51

Table 8: Sorting Along Racial and Academic Lines (Wave II)

Respondent	Probability	Have Same Gender Friend					
		White >3.3	White <=3.3	Black >3.3	Black <=3.3	Hispanic >3.3	Hispanic <=3.3
White >3.3	actual (N=355)	44.26%	48.48%	0.14%	0.62%	0.79%	3.71%
	random (N=1530)	24.51%	57.11%	1.08%	9.11%	1.31%	6.54%
	ratio	1.81	0.85	0.13	0.07	0.60	0.57
White <=3.3	actual (N=1040)	21.62%	71.81%	0.00%	1.64%	0.00%	4.20%
	random (N=4789)	18.82%	61.96%	1.15%	10.06%	1.00%	6.58%
	ratio	1.15	1.16	0.00	0.16	0.00	0.64
Black >3.3	actual (N=46)	0.00%	8.19%	44.22%	46.09%	0.00%	1.50%
	random (N=280)	6.83%	19.55%	9.99%	53.30%	1.40%	8.61%
	ratio	0.10	0.00	4.43	4.43	0.00	0.17
Black <=3.3	actual (N=357)	0.13%	2.37%	7.33%	82.19%	0.24%	7.11%
	random (N=2125)	6.72%	22.01%	7.07%	54.32%	1.12%	8.36%
	ratio	0.02	0.11	1.04	1.51	0.21	0.85
Hispanic >3.3	actual (N=29)	1.67%	20.02%	0.00%	2.62%	9.90%	65.79%
	random (N=207)	12.06%	28.56%	1.73%	16.60%	7.47%	33.85%
	ratio	0.14	0.70	0.00	0.16	1.32	1.94
Hispanic <=3.3	actual (N=328)	0.00%	28.11%	0.41%	5.49%	4.97%	57.66%
	random (N=1749)	8.98%	26.09%	2.43%	14.38%	5.57%	41.16%
	ratio	0.00	1.08	0.17	0.38	0.89	1.40

achievement within the school, in addition the the change in racial composition, cause changes in friendship formation within a school?

Let the probability of an individual in a certain school that has a same-gender friend of a certain racial group ($Prob(Y_{ijk})$) be represented by the following equation.[10]

$$Prob(Y_{ijk}) = \alpha_0 + X_i\alpha_1 + SHARE_{jk}\alpha_2 + (SHARE_{jk})^2\alpha_3 + \epsilon_{ijk} \quad (1)$$

- i = Individual respondent, j = School, k = Relevant racial group of friends
- X_i = Personal characteristics
- $SHARE_{jk}$ = Share of the relevant racial group in a school

There could potentially be nonlinear (most likely decreasing) returns to having more of a particular group at a school, which can be measured by including the squared term on the group shares at a school variable. So, it is expected that α_2 should be positive, while α_3 should be negative in order to confirm the decreasing returns to scale hypothesis. In order to test whether academics matters when sorting into friendship groups, the following addition can be made to the above equation.

$$Prob(Y_{ijk}) = \alpha_0 + X_i\alpha_1 + SHARE_{jk}\alpha_2 + (SHARE_{jk})^2\alpha_3 + (GRADE_i - \overline{GRADE}_j)\alpha_4 + \epsilon_{ijk} \quad (2)$$

- $GRADE_i$ = Academic metric for an individual i
- $\overline{GRADE}_j$ = Average of the academic metric at the school j, where i attends

$(GRADE_i - \overline{GRADE}_j)$ is a measure of academic achievement for the individual relative to the school. It can measure if the student is an above or below average student relative to the school that the individual is enrolled. If it is the case that this measure does affect friendship formation, α_4 should be significantly different from zero. If α_4 is positive, than higher-achieving students are sorting into the friendship group of the race in question (k). If α_4 is negative, then higher-achieving students are sorting away from the friendship group

[10] This model is outlined in Arcidiacono, Khan, and Vigdor (2008)

of the race in question. If α_4 is zero, then homophily along academic achievement lines is insignificant in facilitating cross-race relationships.

5.2 Results

Tables 9 and 10 provide probit estimates of the above equation in waves 1 and 2, respectively. The dependent variable is an indicator of whether an individual in a racial group that is not the race in question ($-k$) has a same-gender friend of the race in question (k), and that race in question is either white, black, or Hispanic. X_i is represented in this case by gender, race, and attitudinal variables such as how the individual views his prospects for college and to what degree the individual is happy with experiences at school. The academic metric analyzed here is the individual's GPA. The coefficient on group shares (α_2) is positive and significant on all groups in both waves, which is expected. The coefficient on the square of group shares (α_3) is not significant for any groups across the two waves, so there is no evidence in these samples that there are decreasing returns to scale.[11] The individual characteristic (X_i) attitudinal variables are insignificant. In both waves, blacks have a negative and significant coefficient compared to other races on the probability of having a white friend. In wave 1, blacks have a positive, although insignificant, coefficient compared to other races on the probability of having a Hispanic friend. However, this coefficient is negative in wave 2. This lends some credence to the fact that blacks are self-segregating more in wave 2 than wave 1, and the difference can be weakly attributed to the switching of Hispanic friends to black friends. The coefficient on relative GPA also follows expected patterns regarding signs. It is positive and significant for those who have white friends in wave 1. Therefore, if a student is above-average relative to schoolmates academically, this student is more likely to have a same-gender friend who is white. The opposite effect is true when analyzing the coefficient on relative grades when the relevant friendship racial group is black or Hispanic. The coefficients are negative, meaning that if a student is above-average relative to schoolmates academically, then the student is less likely to have a same-gender friend who is black or Hispanic. Since the individuals in these regressions exclude the racial group of the friends in question, and through racial dummy variables which takes away any homophily effects of across races, homophily based

[11] There is evidence of decreasing returns to scale in the In-School survey.

on GPA can be isolated. So, on average, increasing the relative GPA of non-white students in a school has a positive effect on the probability of having a white friend, while increasing the relative GPA of non-black or non-Hispanic students in a school has a weakly negative effect on the probability of having a black or Hispanic friend respectively. Since whites in general have higher GPA's amongst these races, followed by Hispanics and blacks, it seems like homophily along GPA lines can facilitate cross-race friendships. All coefficients in wave 2 are not significant, but do have the expected signs. This result may be attributable to the GPA noise and carryover that is mentioned previously. A robustness check using a random effects probit with the "group" variable being the individual shows that coefficients on relative GPA follow the expected patterns. This method eliminates any factors that the individuals and schools may have that affect homophily over the two waves, such as any sort of academic tracking (most plausibly) and other institutions such as clubs (less plausibly).

These tables, along with the above descriptive tables with the actual and random probabilities, all suggest that similarities in characteristics associated with academic achievement (as well as attitudes to a lesser extent) seem to at least have a weak effect on friendships within and across races.

6 Statistical Discrimination

6.1 Model

A test of statistical discrimination can be constructed as follows (Arcidiacono, Khan, and Vigdor 2008). Consider the share and share-squared variables in equation 1. In that equation, the share includes everyone in the group k (the race in question). These individuals in k can be split into those who have a better measure of achievement than individual i, those who have a similar measure of achievement to individual i, and those who have a worse measure of achievement than individual i.

$$SHARE_{jk}\alpha_2 = SHARE_{jkB}\alpha_{2B} + SHARE_{jkS}\alpha_{2S} + SHARE_{jkW}\alpha_{2W} \tag{3}$$

- $SHARE_{jkB}$ = Share of students in school j and group k who have a better academic

Table 9: Estimates on Having Friends from Various Groups
Homophily (Wave I)

	Non-Blacks with Black Same-Gender Friend††	Non-Hispanics with Hispanic Same-Gender Friend††	Non-Whites with White Same-Gender Friend††
$SHARE^{\dagger}_{jk}$	1.605^{*}	2.504^{***}	2.071^{*}
	(0.725)	(0.470)	(0.718)
$(SHARE_{jk})^2$	-0.554	-1.066	0.549
	(1.094)	(0.651)	(0.736)
$(GPA_i - \overline{GPA}_j)$	-0.0591	-0.0702	0.193^{*}
	(0.0670)	(0.0446)	(0.0939)
Good College Prospects	-0.0120	-0.117	0.148
	(0.118)	(0.0800)	(0.112)
Happy at School	0.0691	-0.0135	0.0404
	(0.123)	(0.0804)	(0.112)
Male	0.115	0.00495	0.273^{*}
	(0.108)	(0.0613)	(0.113)
Black		0.143	-1.084^{***}
		(0.262)	(0.261)
Hispanic	-0.796^{*}		-0.251
	(0.378)		(0.267)
White	-1.249^{**}	-0.147	
	(0.388)	(0.254)	
Constant	-1.336^{**}	-1.576^{***}	-1.813
	(0.404)	(0.260)	(0.266)

Standard errors in parenthesis
* $p < .05$, ** $p < .01$, *** $p < .001$
†: Race corresponds to column
††: All races except race in column are used in estimation

Table 10: Estimates on Having Friends from Various Groups
Homophily (Wave II)

	Non-Blacks with Black Same-Gender Friend††	Non-Hispanics with Hispanic Same-Gender Friend††	Non-Whites with White Same-Gender Friend††
$SHARE^{\dagger}_{jk}$	1.703^{**}	2.172^{***}	2.487^{**}
	(0.617)	(0.578)	(0.806)
$(SHARE_{jk})^2$	0.934	-0.702	-0.410
	(1.003)	(0.816)	(0.815)
$(GPA_i - \overline{GPA}_j)$	-0.108	-0.0393	0.00529
	(0.0719)	(0.0499)	(0.0811)
Good College Prospects	0.0771	-0.0499	0.244
	(0.145)	(0.0726)	(0.132)
Happy at School	0.112	-0.0368	0.235
	(0.115)	(0.0962)	(0.147)
Male	0.0973	0.170	-0.0917
	(0.141)	(0.0946)	(0.137)
Black		-0.112	-0.941^{**}
		(0.181)	(0.357)
Hispanic	0.117		0.0482
	(0.360)		(0.342)
White	-0.293	-0.323	
	(0.353)	(0.164)	
Constant	-2.416^{***}	-1.544^{***}	-1.873^{***}
	(0.356)	(0.196)	(0.395)

Standard errors in parenthesis
* $p < .05$, ** $p < .01$, *** $p < .001$
†: Race corresponds to column
††: All races except race in column are used in estimation

achievement metric than individual i

- $SHARE_{jkS}$ = Share of students in school j and group k who have a similar academic achievement metric than individual i
- $SHARE_{jkW}$ = Share of students in school j and group k who have a worse academic achievement metric than individual i

Equation 1.3 is simply splitting the $SHARE_{jk}\alpha_2$ variable and coefficient into three tiers of academic achievement relative to the individual. The share is still relative to the entire population of the school, not just of the race in question. If the coefficients α_{2B}, α_{2S}, and α_{2W} are carried into the squared term as well, equation 1.1 becomes the following.

$$Prob(Y_{ijk}) = \alpha_0 + X_i\alpha_1 + SHARE_{jkB}\alpha_{2B} + SHARE_{jkS}\alpha_{2S} + SHARE_{jkW}\alpha_{2W} \quad (4)$$
$$+(SHARE_{jkB}\alpha_{2B} + SHARE_{jkS}\alpha_{2S} + SHARE_{jkW}\alpha_{2W})^2\alpha_3 + \epsilon_{ijk}$$

The reason that the linear share coefficients enter into the squared term is to make sure that tiers with minimal first order effects (linear term) on the probability of having a friend in group k will also have minimal second order effects on the same probability. For example, if a certain tier does not have a large effect on the probability of having a friend in k, then an increase in the share of that tier should also be ensured not to have any effect on returns to scale, which is now purely measured by α_3.

If the tiering based on academic achievement is not important, then the coefficients on all three share variables should be the same. If the coefficient on the share of students in k who are better than the individual is higher than the coefficient on the share of students in k who are worse than the individual, and k has a measure of achievement that is lower than other races not in k, then the following is clear. Those individuals not in k are much more likely to have a friend in k if they are surrounded by high achieving members of k. In essence, individuals not in the group in question ($-k$) happen to project the academic characteristics of k in their school (j) onto those students who could be potential friends. In this case, those individuals who are not in k are statistically discriminating on the basis of academic achievement against group k. Now, if the coefficient on the share of students in k who are worse than the individual is higher than the coefficient on the share of students in k who are better than the individual,

and the measure of academic achievement for those not in k is lower than those in k, the opposite effect happens than mentioned above. However, once again, individuals not in the group in question $(-k)$ project characteristics of the $k's$ in their school (j) onto potential friends. This phenomenon also is an example of statistical discrimination against k by those not in k. Finally, the coefficient on the share of k that is similar in academic achievement to i can be used to measure the degree of homophily on academic achievement, since a projection of similar achievement to those not in k is placed on potential friends who happen to be in k. In summary, the estimation results of equation 4 can lead to the following.

- $\alpha_{2B} > \alpha_{2W}$, and $\overline{GRADE}_{-k} > \overline{GRADE}_k \Rightarrow$ Statistical Discrimination
- $\alpha_{2W} > \alpha_{2B}$, and $\overline{GRADE}_k > \overline{GRADE}_{-k} \Rightarrow$ Statistical Discrimination
- α_{2S} is a measure of homophily

6.2 Results

Tables 11 and 12 provide non-linear probit estimates for waves 1 and 2 of equation 4 and the marginal effects of a change in one standard deviation of the individual share variables of racial group k on the probability of having a same-gender friend in k, if the respondents are not in k.[12] Whites, blacks, and Hispanics are the racial groups that are analyzed. Achievement is once again measure by GPA. The tiers are defined as follows.

- Better: More than 0.5 GPA points above individual i
- Similar: Within 0.5 GPA points of individual i
- Worse: More than 0.5 GPA points below individual i

As implied in tables 2 and 4, blacks and Hispanics have lower GPA's in general than the population average, and whites have higher GPA's in general than the population average. Table 11 shows that there does exist statistical discrimination against blacks and Hispanics by non-blacks and non-Hispanics respectively in wave 1. A one standard deviation increase in the share of high-achieving blacks will result in an increase in the probability of having a black friend by 1.95% for non-blacks, while the corresponding probability increase that results from

[12] A maximum likelihood estimation procedure outlined in Gould, Pitblado, and Sribney (2006) was used. The marginal effects reported are the average of all individual marginal effects.

a one standard deviation increase in low-achieving blacks is 0.97%. With regards to having a Hispanic friend, the probability increases by 3.63% with a one standard deviation increase in the share of high-achieving Hispanics. The probability increases by 1.04% with a one standard deviation increase in the share of low-achieving Hispanics, but the estimate is insignificant. There is no evidence of statistical discrimination against whites, as the coefficients on the share of high-achieving whites and low-achieving whites are about the same, and the probabilities of having a white friend for non-whites change between 7% and 9% . In wave 1, there are decreasing returns to scale on the probabilities of having a Hispanic or white friend, but the coefficient on the share-squared coefficient is insignificant for blacks.

Table 12 shows similar patterns exhibited in wave 2 as in wave 1, but magnitudes of statistical discrimination have lessened somewhat against blacks and Hispanics. The range in the probability of having a black friend for non-blacks goes from a 1.69% increase with a one standard deviation increase in high-achieving blacks to a 0.88% increase with a one standard deviation change in low-achieving blacks. The range in the probability of having a Hispanic friend for non-Hispanics goes from a 3.09% increase with a one standard deviation increase in high-achieving Hispanics to a 1.57% increase with a one standard deviation increase in low-achieving Hispanics. For whites, there is a slight shift towards being weakly statistically discriminated against by non-whites. The probability of having a white friend for non-whites increases by 9.61% with a one standard deviation increase in the share of low-achieving whites, while the probability increases by 8.37% with a one standard deviation increase in the share of high-achieving whites. The difference is slight. Homophily across both waves 1 and 2 seem to be prevalent, since the coefficients on shares that are similar to the GPA's of individuals are significant. However, the estimates of homophily here may be inflated due to the generally normal distribution of GPA's across the population.[13] To once again control for factors such as academic tracking and clubs, a robustness check using a random effects probit model with the "group" variable as the individual supports results that there exists statistical discrimination against blacks and Hispanics, but not against whites.[14]

[13] One way to correct for the inflation is to change the boundaries of better/similar/worse to a non-fixed number, such as deciles, instead. This option can be explored in future work.

[14] A model similar to equation 1.4 was estimated, except without the embedded coefficients in the squared term.

Table 11: Estimates on Having Friends from Various Groups
Statistical Discrimination (Wave I)

	Non-Blacks with Black Same-Gender Friend††	Non-Hispanics with Hispanic Same-Gender Friend††	Non-Whites with White Same-Gender Friend††
$SHARE^{\dagger}_{jkB}$	2.606*	5.382*	3.383*
	(1.001)	(1.995)	(1.795)
Marg. Effect (1 sd change)	1.95%	3.63%	9.45%
$SHARE^{\dagger}_{jkS}$	3.282*	4.744*	6.478*
	(1.071)	(1.088)	(0.909)
Marg. Effect (1 sd change)	2.46%	3.20%	15.89%
$SHARE^{\dagger}_{jkW}$	1.292*	1.534	3.208*
	(0.449)	(0.924)	(1.549)
Marg. Effect (1 sd change)	0.97%	1.04%	7.87%
$SHARE^2$	-0.0177	-0.216***	-0.113***
	(0.203)	(0.0367)	(0.0128)
Male	0.0376	0.000511	0.251*
	(0.108)	(0.0561)	(0.0925)
Black		-0.00848	-1.047***
		(0.0937)	(0.123)
Hispanic	-0.299*		-0.215*
	(0.121)		(0.100)
White	-0.789***	-0.311**	
	(0.138)	(0.103)	
Constant	-1.777***	-1.392***	-1.777***
	(0.140)	(0.113)	(0.179)

Standard errors in parenthesis
* $p < .05$, ** $p < .01$, *** $p < .001$
†: Race corresponds to column
††: All races except race in column are used in estimation

Table 12: Estimates on Having Friends from Various Groups
Statistical Discrimination (Wave II)

	Non-Blacks with Black Same-Gender Friend††	Non-Hispanics with Hispanic Same-Gender Friend††	Non-Whites with White Same-Gender Friend††
$SHARE^{\dagger}_{jkB}$	3.214*	4.236*	3.961*
	(1.304)	(1.875)	(1.176)
Marg. Effect (1 sd change)	1.69%	3.09%	8.37%
$SHARE^{\dagger}_{jkS}$	6.215*	3.962*	6.390*
	(1.353)	(1.267)	(1.093)
Marg. Effect (1 sd change)	3.28%	2.89%	13.50%
$SHARE^{\dagger}_{jkW}$	1.676*	2.148*	4.549*
	(0.666)	(0.954)	(1.570)
Marg. Effect (1 sd change)	0.88%	1.57%	9.61%
$SHARE^2$	-0.0822	-0.141	−0.132*
	(0.0496)	(0.109)	(0.0496)
$SHARE^2$	-0.0822	-0.141	-0.132***
	(0.0496)	(0.109)	(0.0128)
Male	-0.00590	0.152	-0.0114
	(0.127)	(0.0883)	(0.112)
Black		-0.0196	-1.018***
		(0.130)	(0.151)
Hispanic	-0.287		0.00307
	(0.161)		(0.132)
White	-0.653***	-0.207	
	(0.132)	(0.131)	
Constant	-2.033***	-1.626***	-1.734***
	(0.157)	(0.131)	(0.182)

Standard errors in parenthesis
* $p < .05$, ** $p < .01$, *** $p < .001$
†: Race corresponds to column
††: All races except race in column are used in estimation

7 Conclusion

This chapter has shown that homophily is very prevalent along racial lines and somewhat prevalent along academic lines. It has also shown that statistical discrimination along academic lines exists against blacks and Hispanics by non-blacks and non-Hispanics and has developed a measure of the magnitude of the statistical discrimination incurred. Finally, it has shown that the degree of statistical discrimination is also decreasing between wave 1 and wave 2, suggesting that the signal that is sent out by potential friends regarding academic achievement is becoming clearer, and each individual who chooses a friend stereotypes less as this signal becomes clearer.

The major policies that involve redistribution along racial lines are school redistricting and affirmative action. In principle, these policies assume a certain randomness in interracial contact based on the sheer number of students of a racial group in a certain institution. Therefore, any peer effect benefits that may be garnered from the interracial contact is often analyzed based on this often assumed randomness in peer group formation.

The results of this chapter show how non-randomness in peer group formation can be explained, which in turn can influence any peer effects from these groups. For example, take a policy which redistricts high-achieving minorities (in the case of this chapter, high-achieving black and Hispanic students) from poor school districts into better school districts with relatively few minorities.[15] The results above show that these minorities would experience statistical discrimination if the minorities who are already in the advantageous school district are low achievers. Therefore, the redistricted minority students would not integrate very well with the majority. My results, though, show that the signal of achievement put forth by the new minority students can become clearer after some time, and the degree of statistical discrimination based on academic achievement can decrease.

Future work would involve adding some structure to the above reduced form analysis. For example, individuals can get a random utility based on how "close" they are to potential friends based on racial/academic variables. The signal that is received by individuals from the potential friends is a mix of the true signal from the friends and some noise. Simulations

[15] See footnote 1 regarding the Metco program.

can then be run based on various racial/academic assignment policies that assign friends to individuals, and the composition of friendship networks can then be analyzed.[16] Other future work involves, as mentioned previously, redefining tiers on a relative scale instead of an absolute scale to potentially remove bias when estimating homophily in equation 4.

Racial integration is both an end and a means to an end with regards to redistribution policies. This chapter analyzes how redistribution affects peer group composition, but the group composition's relation to the actual peer effects of the policy, such as future labor market outcomes or happiness due to the increase in diversity of these programs, is not analyzed. However, the composition of groups will certainly have an effect. This chapter shows that group composition forms in complex ways.

References

[1] **Akerlof, G. and R. Kranton** (2002), "Identity and Schooling: Some Lessons Learned for the Economics of Education ", *Journal of Economic Literature* 40: 1167-1201.

[2] **Arcidiacono, P. , S. Khan, and J. Vigdor** (2008), "Representation versus Assimilation: How do Preferences in College Admissions Affect Social Interactions?", Manuscript.

[3] **Arcidiacono, P. and A. Nathan** (2007) "Descriptive and Structural Analysis of Racial and Academic Statistical Discrimination in Schools", Working Paper.

[4] **Arcidiacono, P. and J. Vigdor** (2007) "Does the River Spill Over? Estimating the Economic Returns to Attending a Racially Diverse College", Working Paper.

[5] **Bayer, P., H. Fang and R. McMillan** (2005), "Separate When Equal? Racial Inequality and Residential Segregation", *National Bureau of Economic Research Working Paper # 11507.*

[6] **Bifulco, R., H. Ladd** (2007), "School Choice, Racial Segregation, and Test Score Gaps: Evidence from North Carolina's Charter School Program", *Journal of Policy Analysis and Management* 26(1): 31-56.

[7] **Bowen, W. and D. Bok** (1998), *The Shape of the River: Long Term Consequences of Considering Race in College and University Admissions* Princeton University Press.

[8] **Brock, W. and S. Durlauf** (2001), "Discrete Choice with Social Interactions", *Review of Economic Studies* 68: 235-259.

[16] A similar model is being thought about with regards to Arcidiacono and Nathan (2007), and something similar has been attempted in Arcidiacono, Khan, and Vigdor (2008).

[9] **Case, A. and A. Katz** (1991), "The Company You Keep: The Effect of Family and Neighborhood on Disadvantaged Youths", *National Bureau of Economic Research Working Paper # 3705.*

[10] **Clotfelter, C.** (2004) *After Brown: The Rise and Retreat of School Desegregation* Princeton University Press.

[11] **Drews, P.** (2006) "The Effects of Desegregation and School Choice Programs on Student Achievement: An Examination of the Metco Program", Unpublished Manuscript.

[12] **Duncan, G., J. Boisjoly, M. Kremer, D. Levy and J. Eccles** (2006), "Empathy or Antipathy? The Impact of Diversity", *The American Economic Review* 96(5): 1890-1905.

[13] **Foeman A. and T. Nance** (1999), "From Miscegenation to Multiculturalism: Perceptions and Stages of Interracial Relationship Development", *Journal of Black Studies* 29: 540-557.

[14] **Foster, J.R.** (2005), "Making Friends: A Non-Experimental Analysis of Social Group Formation", *Human Relations* 58: 1443-1465.

[15] **Gould, W., J. Pitblado and W. Sribney** (2006) *Maximum Likelihood Estimation with Stata, third edition* Stata Press.

[16] **Hoxby, C.** (2000), "Peer Effects in the Classroom: Learning from Gender and Race Variation", *National Bureau of Economic Research Working Paper # 7866.*

[17] **Kandel, D.B.** (1978), "Homophily, Selection, and Socialization in Adolescent Friendships", *American Journal of Sociology* 84: 427-436.

[18] **Manski, C.** (1993), "Identification of Endogenous Social Effects: The Reflection Problem", *Review of Economic Studies* 40: 531-542.

[19] **Marmaros, D. and B. Sacerdote** (2006) "How Do Friendships Form?", *Quarterly Journal of Economics* 121(1): 79-119.

[20] **Moody, J.** (2001), "Race, School Integration, and Friendship Segregation in America", *American Journal of Sociology* 107: 679-716.

[21] **Raino, K.** (1966) "A Study of Sociometric Group Structure: An Application of a Stochastic Theory of Social Interaction" in J. Berger, M. Zelditch, and B. Anderson, *Sociological Theories in Progress*, Vol.1, Houghton Mifflin.

[22] **Rumberger, R. and G. Palardy** (2001), "Does Segregation Still Matter? The Impact of Student Composition on Academic Achievement in High School", *Teachers College Record* 107: 1999-2045.

[23] **Tuma, N.B. and M. Hallinan** (1979), "The Effects of Sex, Race and Achievement on Schoolchildren's Friendships", *Social Forces* 57: 1265-1285.

[24] **Udry, J. R.** (2003), The National Longitudinal Study of Adolescent Health (Add Health), Waves I & II, 1994-1996; Wave III, 2001-2002 [machine-readable data file and documentation], *Carolina Population Center, University of North Carolina at Chapel Hill*, Chapel Hill, NC.

[25] **Weinberg, B.** (2006) "Social Interactions and Endogenous Association", Working Paper.