Identifying Sources of Racial Homophily in High School Friendship Networks[*]

Sergio Currarini[†] Matthew O. Jackson[‡] Paolo Pin[§]

This Draft: June 11, 2009

Abstract

We analyze homophily in the Adolescent Health data set. We first show that homophily patterns differ by race. Next, using a model of friendship formation that combines choices with random meeting patterns, we use the observed homophily patterns to estimate biases in students' preferences over the races of their friends and biases in the rates at which students of various races meet each other. We find that both biases are significant in these data: students prefer to form friendships with students of their own race, and students of a given race are more likely to meet students of that same race than is reflected by that race's prevalence in the population. Moreover, we find slight but significant differences in preference biases across races with White and Asians exhibiting more preference bias than Blacks and Hispanics. Meeting biases differ substantially significantly, with almost no bias for Whites, some bias for Hispanics, and a very strong bias towards own race for Asians and Blacks. We also find that these biases depend on high school size with larger high schools exhibiting larger meeting biases and more homophily.

Keywords: Networks, Homophily, Segregation, Friendships, Social Networks, Race, Ethnicity, Diversity, Minorities

1 Introduction

Homophily, the term coined by Lazarsfeld and Merton (1954), refers to the pervasive tendency of people to associate with people similar to themselves. The extensive literature on homophily

---

[*] We gratefully acknowledge financial support from the NSF under grant SES–0647867, and the EU-STREP project n. 516446. We thank James Moody for making available the network data from the 'Add Health' data set.

[†] Dipartimento di Scienze Economiche, Università di Venezia and School for Advanced Studies in Venice (SSAV). Email: s.currarini@unive.it

[‡] Department of Economics, Stanford University and the Santa Fe Institute. Email: jacksonm@stanford.edu, http://www.stanford.edu/∼jacksonm/

[§] Dipartimento di Economia Politica, Universitá degli Studi di Siena and Max Weber Programme, European University Institute.

has demonstrated such patterns across many dimensions, including ethnicity, age, gender, religion, profession, as well as things like political opinions and other behaviors.[1] Homophily can arise for many reasons, such as biases in the amount of contact that people have with each other (Blau (1977), Feld (1981), Rytina and Morgan (1982)), preferences for associating with individuals with similar traits, behaviors, or backgrounds (Cohen (1977), Kandel (1978), Knoke (1990), and Currarini, Jackson and Pin (2008) and Bramoullé and Rogers (2009)), competition among groups (Giles and Evans (1986)), social norms and culture (Carley 1991)), and institutional and organizational pressures (Meeker and Weiler (1970), Khmelkov and Hallinan (1999), Kubitschek and Hallinan (1998), Stearns (2004)).

While this bias in favor of similarity is a sound empirical observation, still little is known about the generative processes that are responsible for it. In particular, the distinction between the roles of choice and opportunities in the formation of social ties is crucial to understanding patterns of social interaction and behavior, especially if one wishes to influence them through policy. What makes this task challenging is the fact that choices and opportunities are interconnected, with feedback effects where one affects the other (as discussed in McPherson and Smith–Lovin (1987) and more recently in Franz, Marsili and Pin (2008)). Disentangling the influences on homophily is not only important from a purely scientific perspective, but also in designing and evaluating policies. Homophily has implications in terms of how it affects things like social mobility, decisions to undertake education, and the speed of diffusion and learning in a society (e.g., see Calvo-Armengol and Jackson (2004, 2009), Jackson (2007), and Golub and Jackson (2008)), and school performance (e.g., see Patacchini and Zenou (2006)). If one wishes to influence patterns of friendships, then it is important to understand whether homophily is due to peoples' preferences for the characteristics of their friends, or to biases in the rates at which people meet each other, or to some other force.

In a recent paper, Currarini, Jackson, and Pin (2008) (CJP henceforth), we develop a rational-choice model of friendship formation designed to include roles of both choice and opportunities in network formation. We show that the model generates patterns of friendships that match some observed patterns of homophily in a data set of American High School friendship networks from the Adolescent Health data set.[2] Specifically, the model generates two major stylized facts in students relations: first, students whose ethnic group forms a greater fraction of a school's population tend to form more friendships per capita; and second, a bias in the mix of friendships in favor of one's own race is largest among ethnic groups that form a middle-sized fraction of a school's population. The main contribution of that paper is to trace the first observation to biases in preferences and the

---

[1] The review of McPherson, Smith-Lovin, and Cook (2001) provides a wide overview, and Jackson (2008) provides discussion of some of the recent literature.

[2] Add Health is a program project designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris, and funded by a grant P01-HD31921 from the National Institute of Child Health and Human Development, with cooperative funding from 17 other agencies. 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).

second observation to biases in meeting opportunities; and a central conclusion is then that biases in both preferences and meeting opportunities are needed to generate observed patterns, and also that these biases reinforce each other.

As we show here, there is a fundamental and important issue that was not addressed in CJP: homophily patterns differ significantly across races. We begin this paper by showing that patterns of both the number of friendships formed and the extent to which they are biased towards own race differ significantly across races. This raises the obvious question of what is responsible for these differences. We develop an extension of CJP model that allow for different biases in preferences and meeting rates across races and develop a method for estimating the new model. We then estimate a parameterized version of the model using the networks of friendships in the Add Health data. Our key findings are that biases in favor of own race in preferences and in meeting opportunities both play roles for the different races. Interestingly, we find that preference biases are roughly similar across races, so that a friendship with an individual of a different race provides about 65 to 85 percent of the value of a friendship to a same type individual, on average for each of the races. However, races differ quite significantly in their meeting biases, Whites show almost no bias in meetings, Hispanics exhibit some bias towards meeting their own type, and Asians and Blacks exhibit quite substantial biases towards meeting their own types (a bias parameter almost six times that of Whites). So, although all groups exhibit homophily, estimating this model leads to different explanations for the observed patterns of homophily across races.

We also examine how homophily patterns vary with school size.[3] We find that there is little effect on preference biases, but that larger high schools display evidence of a greater bias towards having students meet other students of their same race.

There are previous papers that shed light on homophily in the Adolescent Health data set such as Moody (2001) and Goodreau, Kitts, and Morris (2009). In particular, Goodreau, Kitts, and Morris (2009) employ new techniques for estimating exponential random graph models to analyze the friendship networks in the Add Health data set. As they include race as an explanatory variable, they obtain some results about differences in friendship formation patterns across races. There are several patterns in our findings that echo their findings. At a very basic level, they observe differences in behavior across races. In this regard, the most interesting point of comparison is their examination of "sociality" across races, where the term sociality refers to an individual's *propensity* to form links. They point to an important distinction between sociality and degree, where sociality is influenced by many other characteristics to eventually influence degree. Differences in degree that remain after other factors are controlled for are attributable to "sociality". In our setting "sociality" is not a primitive, but instead is something that is derived from underlying preferences over racial mixtures and the racial mixing that occurs. Thus, our modeling provides an explanation

---

[3] There is a literature concerning optimal school–size (e.g., Meeker and Weiler (1970), Morgan and Alwin (1980), Coladarci and Cobb (1996), Slate and Jones (2005), and Berry and West (2008)). Weinberg (2007) relates it explicitly to homophily through endogenous associations and extracurricular activities.

for observed differences in outcomes across races. This could explain why Goodreau, Kitts, and Morris (2009) do not find significantly different socialities across races (except blacks having lower sociality), while we find significant differences in the number of friendships formed across all races, once we control for a group's size. Based on our estimation, it is not sociality that accounts for these differences, but differences in racial preferences for racial mixtures coupled with the mixing of races across schools. Another point of complementarity is that Goodreau, Kitts and Morris (2009) include "homophily" variables directly, which account for propensities to form links across races. In this regard, our modeling distinguishes different sources of homophily based on preference biases and meeting biases, and so homophily is derived as an outcome from more basic processes.[4]

Before proceeding, we offer important cautions in interpreting our results. First, this paper should be viewed as a step in a broader research agenda of developing models that will help us to estimate and understand homophily and other patterns of social relationships. While the model provides causal inferences - preference and meeting biases cause certain patterns of homophily to emerge - one should not necessarily conclude that this is the explanation of what is going on in the data. The model is limited to considering only these forces, and it could be that some other forces are responsible for homophily, but since they are un-modeled they are picked up by the forces that we do model. Second, the differences that we find across races and across school sizes could also be due to other unobserved correlates. For example, we find differences in preference biases and in meeting rate biases across races. While these might be racially or culturally influenced, it could also be that some unobserved socio-economic factor that correlates with race in these data impact preferences and meeting biases. Thus, while the findings here do unearth specific relationships and patterns that differ across races, they should be interpreted as calling for the development of richer models and further drilling into these and other data.

2 Measuring Homophily

We begin with some simple definitions that are important in measuring homophily and also in presenting the model.

Individuals have "types," which are for instance their race, gender, age, etc., or a combination of characteristics. Let N_i denote the number of type i individuals in the population, and let $w_i = \frac{N_i}{N}$ be the relative fraction of type i in the population, where $N = \sum_k N_k$.

Let s_i denote the average number of friendships that agents of type i have with agents who are of the same type, and let d_i be the average number of friendships that type i agents form with agents of types different than i. Let $t_i = s_i + d_i$ be the average total number of friendships that type i agents form.

---

[4] Also, our findings of differences in the ways that races meet each other in larger schools versus smaller schools is quite orthogonal to their analysis. The only real role of school size in their analysis is in predicting triadic closure.

The obvious basic homophily index, H_i, measures the fraction of the ties of individuals of type i that are with that same type.

DEFINITION **1** *The homophily index H_i is defined by*

$$H_i = \frac{s_i}{s_i + d_i}.$$

As mentioned above, the index H_i is partly determined by relative populations, and so a benchmark is the relative size of a group, which corresponds to the expectation of the fraction of same-type friendships in a world where friendships are distributed uniformly at random among pairs of agents in the population.

A standard terminology is that a society satisfies baseline homophily if relative friendship rates across types is in proportion to relative frequencies in the population. In particular, the profile $(s, d) = (s_1, d_1, s_2, d_2, \ldots, s_K, d_K)$ satisfies **baseline homophily** if for all i

$$H_i = w_i.$$

The observed tendency of friendships to be biased towards own types beyond the effect of relative population sizes has been referred to as "inbreeding homophily" (see, e.g., Coleman (1958), Marsden (1987), and McPherson Smith-Lovin and Cook (2001)). The profile (s, d) exhibits **inbreeding homophily** for type i if

$$H_i > w_i.$$

Generally, there is a difficulty in simply measuring homophily according to H_i. For example, consider a group that comprises 95 percent of a population. Suppose that its same-type friendships are 96 percent of its friendships. Compare this to a group that comprises 5 percent of a population and has 96 percent of its friendships being same-type. Although both have the same homophily index, they are very different in terms of how homophilous they are relative to how homophilous they could be. Comparing the homophily index, H_i, to the baseline, w_i, provides some information, but even that does not fully capture the idea of how biased a group is compared to how biased it could potentially be. In order to take care of this we use the measure developed by Coleman (1958) that normalizes the homophily index by the potential extent to which a group could be biased.

DEFINITION **2** *The* ***inbreeding homophily*** *index of type i is*

$$IH_i = \frac{H_i - w_i}{1 - w_i}.$$

This index measures the amount of bias with respect to baseline homophily as it relates to the maximum possible bias (the term $1 - w_i$). It can be easily checked that we have inbreeding homophily for type i if and only if $IH_i > 0$, and inbreeding heterophily for type i if and only if

$IH_i < 0$. The index of inbreeding homophily is 0 if there is pure baseline homophily, and 1 if a group completely inbreeds.[5,6]

3 A Preliminary Analysis of the Add Health Data

Before presenting the model and its estimation, we discuss some of the patterns of friendship formation in the Add Health dataset.

3.1 General Patterns of Homophily

We begin by reviewing patterns uncovered by Currarini, Jackson, and Pin (2008).

The first pattern concerns how the average number of friends that students in a given racial group have depends on the fraction of the high school population comprised by that racial group. Each observation refers to the average of a given racial group within a given high school. We have a total of 301 observations (all racial groups that are present in the sampled high schools).

Figure 1 relates the total number of friendships (on average) held by each racial group to the relative size of that group.

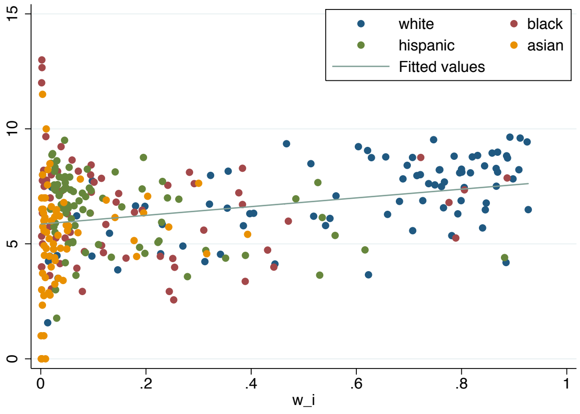

Figure 1: Total number of friends by group size

We observe that the average number of total friends is an increasing function of group size, with a mean of less than 6 friends for groups that are a small fraction of their school increasing up

---

[5] One could also define a heterophily index, which would be $\frac{\frac{s_i}{s_i+d_i}-w_i}{-w_i}$, reflecting the extent to which a group is outgoing. It would be 0 at baseline homophily and 1 if a group only formed different-type friendships.

[6] The measures H_i and IH_i have slight biases in small samples. For example, suppose that there was no bias in the friendship formation process so that we are in a "baseline" society. Then the fraction of other agents that are of type i is $\frac{N_i-1}{N-1}$. Thus, the expected value of $H_i - w_i$ in a baseline society is $-\frac{N-N_i}{N(N-1)}$, which vanishes as N becomes large. The expected value of IH_i is then $-\frac{1}{N-1}$, which is independent of i, vanishing in N, and slightly negative.

to more than 8 friends on average for a racial group that comprises most of a high school. This positive relation between total friendships and group size echoes a pattern noted by Marsden (1988) in a national survey. Regressing the total number of friends t on relative group size w we find that:

$$t = \underset{(0.15)}{5.86} + \underset{(0.37)}{1.86}\, w \tag{1}$$

Both the constant and the coefficient of group size are significant at a 99 percent confidence level (with t-statistics of 4.99 for w and 40.34 for the constant, and $R^2 = 0.076$).[7]

Next, we examine the racial mix of students' friendships. In Figure 2 we plot the homophily index H as a function of relative group size w. The 45 degree line represents the expected outcome if friendships were governed by a purely random process in which case the proportion of same race friends on total friends would follow population shares (i.e., baseline homophily). We instead observe that the homophily index exceeds relative group size for most groups (i.e., inbreeding homophily), with some exceptions for very small and very large groups.

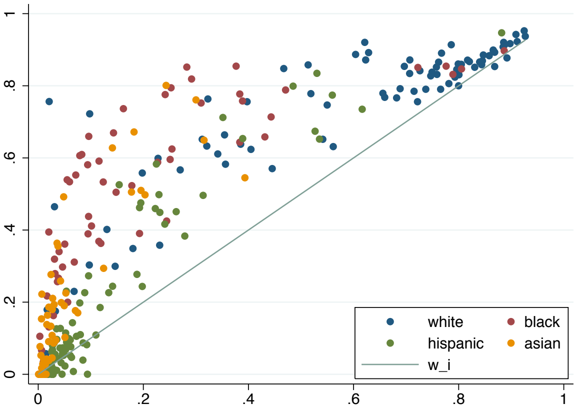

Figure 2: Homophily Index by Group Size

As explained in the previous section, it is informative to normalize groups' inbreeding relative to their inbreeding potential by dividing the difference between the observed index H and the 45 degree line by a factor of one minus the relative group size, to obtain the Inbreeding Homophily Index. As we see in Figure 3, this index varies non-monotonically with relative group size, following a humped shape. Very small and very large groups tend to inbreed very little compared with their inbreeding potential, while, on average, middle sized groups inbreed the most.

Because of the nonlinearity of the relation between IH and w, we regress the index IH on

---

[7] Note that the variance is higher for low levels of w_i. When w_i is very small, there are only a few individuals in a racial group and so a single individual's characteristics matters and there is high variability. For large groups, the numbers are averages over many students and so there is less variation.

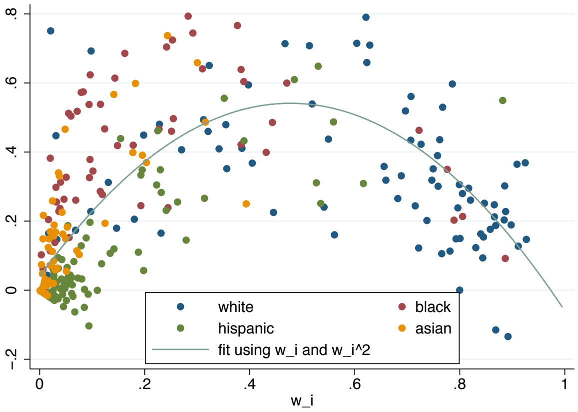

Figure 3: Inbreeding Homophily by Group Size

group size w and on the square of group size w^2 (higher order terms do not significantly improve the fit). We obtain the following relationship:

$$IH = \underset{(0.01)}{.032} + \underset{(0.13)}{2.15}\,w - \underset{(0.15)}{2.35}\,w^2 \tag{2}$$

Both coefficients and the constant term are significant at 99 percent confidence level, with t-statistics of 2.16 for the intercept, of 24.8 for w and of 20.4 for w^2, and with $R^2 = 0.73$.

These significant patterns of the increasing number of friends with group proportion and the non-monotone inbreeding homophily that is maximal with middle proportion groups were the subjects of CJP, where we showed that the increasing number of friends with group size can be explained by a bias in preferences and the nonomontonicity in inbreeding homophily can be explained by a bias in meeting or contact. Roughly, the ideas are as follows. If students are biased in their preferences and prefer to form friendships with their own type, then when group size is relatively larger they have more opportunities to meet their own types and find attractive matches and end up forming more friendships. As we show there, without such a bias in preferences, students' friendship formation patterns are insensitive to group size and so such a preference bias is needed to explain the pattern, at least in the context of our choice-based model where students select their friends. In terms of the inbreeding homophily pattern, simply by adding up friendships across races, if there is no bias in the meeting process then if one group is inbred another group must be outbred. It is then quite easy to see that any model that matches these data will require some bias in the meeting or contact process underlying friendship formation in order to generate inbreeding homophily systematically for all groups.

We now move on to the main topic of this paper, which is how these patterns vary across races and school size.

3.2 A Closer Look at Data: Differences across Schools and Races

A closer inspection of Figures 1 and 3 suggests that the observed relations between relative group size and friendship patterns result from the aggregation of seemingly different patterns for the various races. As we see in figures 5 and 7 below, significant differences are also associated with friendships patterns in schools of different sizes.

A key issue is whether the observed heterogeneity reflects systematic behavioral differences across racial groups and/or across schools. And, if so, whether these differences in behavior are due to varying preference biases across races, to differences in mixing opportunities across races and schools, or to other factors. In this section we take a first step by showing that the observed heterogeneity is significant in the statistical sense; and in Section 4 we propose a formal rational choice model through which we can disentangle the observed heterogeneity in behavior into differences in racial preferences and in meeting opportunities.

3.2.1 The relation between group size and inbreeding behavior

We first look at the relation between relative group size and the inbreeding homophily index by race. We obtain a clear picture of the different trends followed by different races by running separate regressions for each race, and plotting the fitted curves in Figure 4.

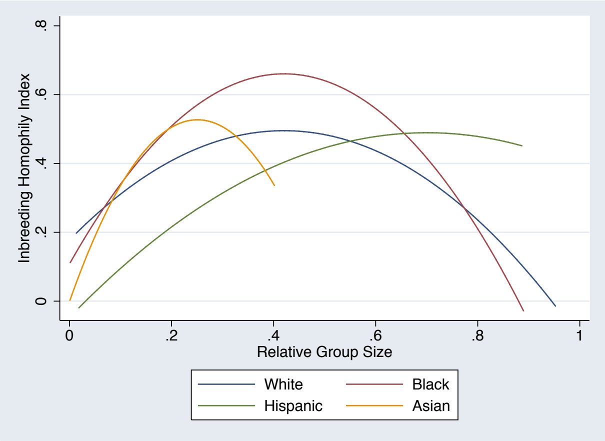

Figure 4: Inbreeding Homophily Index by Race

To get an idea of the effect of school size, we run separate regressions of the inbreeding homophily index against group size and its square for small and large schools. We break the data into two parts, with a threshold that splits the data roughly in half: those schools with more than one thousand students and those with fewer than one thousand students. The separate fits are pictured in Figure 5.

The two quadratic fits suggest a general increase in the inbreeding of all groups in larger schools.

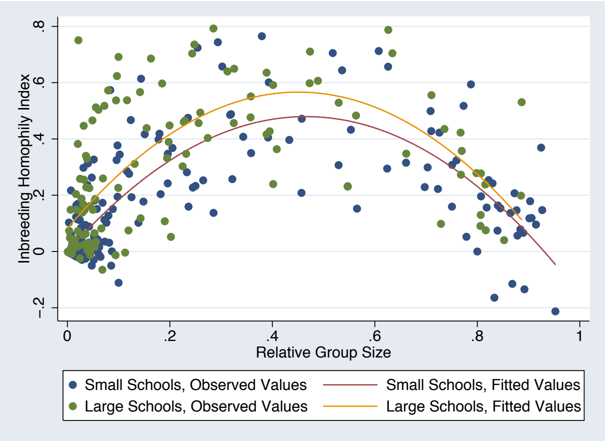

Figure 5: Inbreeding Homophily Index by School Size

Interestingly, this increase seems to be more substantial for smaller groups (and is statistically significant, as found by the significance of the intercept dummy variable for school size in the regression in Table 1).

In order to test the statistical significance of the above differences across races and schools, we run a regression of the Inbreeding Homophily Index against relative group size, controlling for the composition of the sample with respect to race and school size. The results of this regression are are summarized in Table 1.

We control for the effect of race by means of race dummies for Black, Hispanic and Asian groups, and slope dummies for these three races for both w and w^2; we control for school size by means of a dummy variable splitting the sample in large schools (more than 1000 students) and small schools (less than 1000 students). We also control for the interaction effect of school size with both group size and the square of group size. We obtain qualitatively identical results for different thresholds of school size other than the 1000 which splits the data roughly in half.

The parametric tests of Table 1 impose constraints on both the functional form relating variables. In order to obtain further evidence of the significance of the above differences, we run a non parametric Mann-Whitney test on the difference of distributions from which observed data for the various races and schools are drawn. Table 2 summarizes the result of the test on each pair of races. The null hypothesis is here that the observed Inbreeding Homophily Indices for the two races are drawn from the same distribution.

The results of Table 2 are in line with those found in the parametric tests, suggesting a systematic effect of race. There are very highly significant differences between races in all cases with the exception of Black-White and Asian-Hispanic. As this is a much weaker test than the parametric fits, it is not completely clear how to interpret this. Fitting the parametric regressions based on

Inbreeding Homophily Index	*coefficient*	*std. err.*	*t-statistic*	*p-value*
constant	.135	.047	2.83	0.005
Group Size (w)	1.44	.23	6.18	0.000
Group Size Squared (w^2)	-1.66	.23	-7.13	0.000
School Size Dummy(DS)	.059	.024	2.40	0.017
School Size Dummy times w	.27	.22	1.24	0.216
School Size Dummy times w^2	$-.40$	.27	-1.47	0.147
Dummy Black	$-.049$	.052	-0.95	0.344
Dummy Hispanic	$-.19$	.05	-3.73	0.000
Dummy Asian	$-.16$	.051	-3.23	0.001
Slope Dummy Black times w	1.02	.31	3.31	0.001
Slope Dummy Hispanic times w	$-.16$	.34	-0.49	0.627
Slope Dummy Asian times w	2.67	.65	4.13	0.000
Slope Dummy Black times w^2	-1.24	.35	-3.53	0.000
Slope Dummy Hispanic times w^2	.92	.44	2.09	0.038
Slope Dummy Asian times w^2	-6.71	1.89	-3.55	0.000

Table 1: Inbreeding Homophily Index Regressed on School Size and Racial Variables

quadratics picks up a difference between Blacks and Whites and also for Asians and Hispanics (using confidence intervals on the dummies), while the nonparametric rank sum test does not (and is also weaker as it does not correct for school size effects). Fitting the model below will help in further sorting this out, as it provides a different angle altogether.

We can non-parametrically test the effect of school size by means of a Matt-Whitney test on the difference of distributions of the Inbreeding Homophily Index associated with large (> 1000 students) and small (< 1000 students) schools. The null hypothesis is that both samples are drawn from the same distribution. The null hypothesis is rejected at the 99 percent confidence level. Finally going back to the discussion following Figure 5, we test whether the difference in distribution is driven by small racial groups. Interestingly, we obtain that the distributions from large and small schools are not statistically different for majority groups and are statistically different for minorities. In particular, for large groups we obtain a z-statistic of -1.703, with Pr $>$ z equal to 0.089, while for small groups we obtain a z-statistic of -4.057 , with Pr $>$ z equal to 0.0001. We can rephrase this result by saying that the inbreeding behavior of students in groups that comprise small fractions of their school is “more affected” by school size than the behavior of groups that comprise large fractions of their school. A potential explanation for this, that will be consistent with the modeling below, is that small racial groups may find it easier to inbreed in larger rather than in smaller schools, possibly because of the presence of economies of scale in the

Compared Pairs of Races	*z-statistic*	*p-value*
Asian-Black	-4.417	0.000
Asian-Hispanic	1.269	0.2044
Asian-White	-5.041	0.000
Black-White	0.549	0.5828
Black-Hispanic	5.271	0.000
Hispanic-White	-6.036	0.000

Table 2: Mann-Whitney test on Inbreeding Homophily Index

formation of organized *loci* of activity that traditionally favor inbreeding behavior, such as clubs, societies, and other extracurricular activities and organizations (see Weinberg (2007)).

3.2.2 Numbers of Friendships

We now turn to the pattern of how the average numbers of friendships varies as a function of group size. Figure 1 suggests that the relation between total number of friends and relative group size results from quite different patterns across races. Running separate regressions for separate races, we obtain the fitted lines pictured in Figure 6.

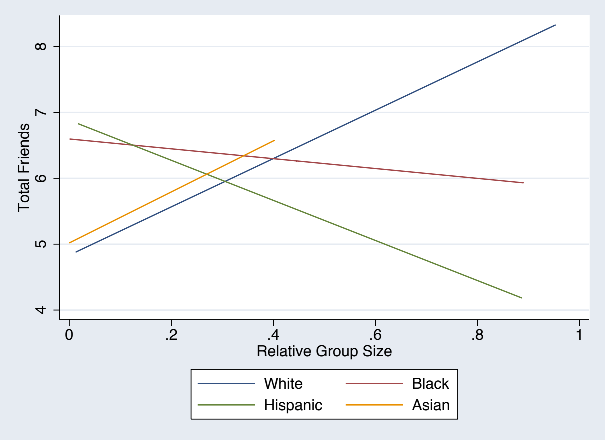

Figure 6: Total Friends by Race

As we see from table 3, the relation between group size and total friends is significantly different from a flat line only for Whites and Hispanics, with positive and negative trends, respectively. The negative pattern of Blacks and the positive pattern for Asians are not statistically significant even at a 95 percent confidence level.

Finally, the effect of school size on total number of friends is illustrated in Figure 7, where total

Number of Friends	*White*	*Black*	*Hispanic*	*Asian*
Constant	4.83	6.60	6.88	5.02
t-statistic	13.79	19.84	32.67	16.22
p-value	0.000	0.000	0.000	0.000
Coefficient	3.67	-.75	-3.04	3.87
t-statistic	6.90	-0.63	-3.11	1.20
p-value	0.000	0.533	0.003	0.236

Table 3: Number of Friends against Group Size: Regressions by Race

friendships slightly decrease in larger schools (as we see from table 4 below, this difference is not statistically significant).

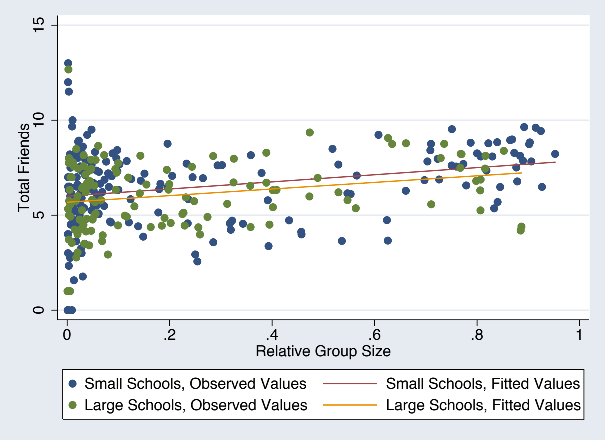

Figure 7: Total Number of Friends by School Size

As we did for the Inbreeding Homophily Index, we run a single regression of the total number of friends against relative group size, controlling for racial composition and school size. Again, we control for the effect of race by means of racial dummies and racial slope dummies for w, and for the effect of school size by means of the dummy DS which splits the sample in schools with more and less than one thousand students. The results are summarized in Table 4.

These results about both the number of friendships and the inbreeding homophily index point out effects of both race and school size on the relation between the relative size of racial groups and their pattern of friendship formation. Race significantly affects both total friendships and their racial mix, and in different ways across races. School size also affects the racial mix of friendships, strengthening the tendency to inbreed; but does not have a significant effect on total numbers of friendships. We again remark that although statistically significant, these results merely indicate

Number of Friends	*coefficient*	*std. err.*	*t-value*	*Pr* $>$ *t-value*
constant	4.95	.49	10.17	0.000
Group Size (w)	3.53	.73	4.80	0.000
School Size Dummy	$-.20$	.28	$-.73$	0.468
School Size Dummy times w	.24	.77	0.31	0.756
Dummy Black	1.73	.54	3.19	0.002
Dummy Hispanic	2.00	.54	3.73	0.000
Dummy Asian	.16	.53	0.30	0.766
Slope Dummy Black times w	-4.35	1.24	-3.50	0.001
Slope Dummy Hispanic times w	-6.60	1.42	-4.659	0.000
Slope Dummy Asian times w	.29	2.77	.31	0.916

Table 4: Number of Friends

an association between race and school size and some patterns of friendship formation, while no casual effect is implied. We now turn to modeling friendship formation in ways that will allow us to estimate different potential forces that might account for these effects.

4 A Preference- and Random Meeting-Based Model of Friendship Formation

As mentioned in the introduction, the model that we use is based on one developed in CJP, and is extended to allow for different preferences and matching rates across races. In the model, friendship formation takes place via a matching pool in which students enter without any friends, are randomly matched to new possible friends and eventually exit the process after having made some friends.[8]

Students are characterized by race and generally we use the term *type*, as the model applies for all sorts of different categorizations, including things like age, gender, or combinations of such attributes. Students have preferences over whom they are friends with, which are potentially sensitive to the racial (or type) mix of these friends. Each racial group i is characterized by its relative size w_i in the school. In line with the notation introduced in Section 2, s_i and d_i denote the number of same and different friends made by a representative student of race i, while $t_i = s_i + d_i$ is the total number of friends.

We consider a steady state of the matching process in which the flow of new students into the matching balances those exiting. Three key elements of the model are: (i) the preferences and

---

[8] These assumptions enrich a class of "search" models that have become standard in the literature on labor markets and also for studying macroeconomic implications, such as the model of Mortensen and Pissarides (1994) where a matching is between a worker and a firm.

resulting choices of the students of how many friendships to form given the matching process; (ii) the random matching process itself, which may be more or less biased in terms of the relative rates at which it matches types; and finally (iii) the steady–state requirements that require that friendships add up across students and that people enter and exit at a similar rate so that the process is in equilibrium. We use the model to calibrate preferences and matching rates across races, evaluating whether there are differences across races or according to school size. This approach complements the purely statistical one in the previous section, since it allows us to infer which forces are affecting homophily and friendship numbers.

4.1 An Agent's Preferences

Each student receives utility from the composition of the set of his or her friends. For the purposes of this model, the agent only distinguishes between "same" and "different"types. This is roughly consistent with empirical evidence. For example Marsden (1988) does not find any significant distinction between friendships and races after accounting for same versus different types. As McPherson, Smith-Lovin and Cook (2001, footnote 7) point out: "The key distinction appears to be same–different, not any more elaborated form of stratification."[9] Thus, at least as a first step, this seems to be the important distinction.

Students of the same type are characterized by their utility function, which may, however, differ across types. The total utility to an agent of type i who has s_i same-type friends and d_i different-type friends is given by:

$$U_i(s_i, d_i) = (s_i + \gamma_i d_i)^{\alpha} \quad , \tag{3}$$

where both γ_i and α are between 0 and 1, so that U_i is increasing in both s_i and d_i.

This simple functional form for preferences has several features worth commenting on. First, the $\alpha(< 1)$ parameter captures the fact that there are diminishing marginal returns to friendships, so that although there are benefits from having more friends, those marginal benefits decrease as more friends are added. Second, the γ_i parameter captures the bias that an individual has in evaluating friendships of same type versus different type. A different-type friend is only worth γ_i as much as a same-type friend. Third, we allow γ_i to vary with type but keep α the same across types. We could extend the model to fuller generality, but at the risk of having too many free parameters and over-fitting the data. The critical difference that we are interested in exploring is racial attitudes towards cross-race friendships and so γ_i is a critical parameter to allow to vary, while the rate at which friendship value diminishes is less pertinent and so we hold that fixed across races.

The race-dependent parameter γ_i quantifies the bias towards own type in preferences: a value of $\gamma_i = 0$ indicates completely biased preferences which attributes no value to friendships with

---

[9] See also the discussion on page 114 of Goodreau, Kitts and Morris (2009).

different types, a value of $\gamma_i = 1$ corresponds instead to preferences which are independent of types (in the economics terminology, this is a case of *perfect substitutes*).

There are costs to meeting people and forming friendships, both in time and energy, and that caps the numbers of friendships that students form. In particular, an agent bears a cost $c > 0$, in terms of opportunity cost of time and resources, for each unit of time spent in the matching process. We will see that the parameter c, although needed to close the model, does not end up playing a significant role in the calibration, and could in principle even be heterogeneous across schools.

4.2 The Matching Process and Decision Problem

The way in which people form friendships is through a matching process. This process is similar to a sort of blind-dating service. Agents are repeatedly matched with other people through a stochastic matching process. Agents can then choose to either form a friendship or not with the other agent with whom they are matched. An agent can then either continue in the process and meet additional people, or exit the process. Given the form of preferences that we have here which are increasing in friendships of both types, agents will accept whatever people they meet as friends, and the main decision is simply when to exit the process. One can enrich the model to have idiosyncratic values to particular matches and to have satiated preferences so that agents reject some friendships, but such embellishments do not really add much (as shown in the supplementary material to CJP (2008)). Note that even though the only decision is when to exit the process, both biases in preferences and biases in the matching will affect that decision. The bias in preferences determines how the agent evaluates what the marginal return from staying in the matching process is relative to its cost, while the bias in the matching process will affect the mix of same versus different people that will be met and thus also the anticipated marginal return from being in the matching process.

Thus, the relevant matching parameter from an agent's decision perspective is the expected rate at which the agent will meet same versus different type friends. Because of the bias and potential heterogeneity in the actions of different types of agents, the relative rate at which agents meet their own type versus different types will not correspond directly to their relative fraction in the population w_i. This would only be true if all agents stayed in the matching process for the same amount of time and the process operated completely uniformly at random and had no bias in matching. Otherwise, we need to keep track of the rate at which a type i agent meets other type i's, which we denote q_i. This parameter will be determined by the decisions and the steady-state.

Given the matching probabilities of same type of agents and different type agents of q_i and $1 - q_i$ per unit of time, respectively, if an agent of type i stays in the matching process for a time t_i, then he or she will end up with $(s_i, d_i) = (t_i q_i, t_i(1 - q_i))$ friends of same and different types, respectively. We solve the model in the case where the actual realizations of the matching are the expected numbers.[10] Thus, an agent of type i solves the following decision problem of how long to stay in the matching process and thus how many friends to have:

$$\max_{t_i} U_i(q_i t_i, (1 - q_i)t_i) - ct_i. \tag{4}$$

Given the utility function described in (3), this is a straightforward maximization problem and it has solution:[11]

$$t_i = \left(\frac{\alpha}{c}\right)^{\frac{1}{1-\alpha}} (\gamma_i + (1 - \gamma_i)q_i)^{\frac{\alpha}{1-\alpha}} \quad . \tag{5}$$

4.3 The Bias in Matching Process and the Steady State

Solving the model requires determining the matching rates. Clearly, there must be some conditions that relate meeting rates across races, since if a person of race i is meeting a person of race j then the converse is also true. Thus, there are cross conditions that constrain the potential configurations of q_i's.

The matching process is described by an $n \times n$ matrix $\mathbf{q} \in [0, 1]^{n \times n}$, where q_{ij} is the fraction of i's meetings per unit of time that are with type j and the matrix is row stochastic so that $\sum_j q_{ij} = 1$.

Let M_i denote the relative stock of agents of type i in the matching process at any time. In particular, $M_i = t_i w_i$. The relative meeting probabilities (q_{ij}'s) depend on the stocks of agents in the society and how they bump into each other, which is captured by a function F, where $\mathbf{q} = F(M_1, \ldots, M_n)$ is the matching that occurs as a function of the relative stocks of agents in the society.

To be well-defined, the matching process needs to balance, so that the number of meetings where an i meets a j is the same as those where a j meets an i. A matching process F is *balanced* at a given M if $\mathbf{q} = F(M_1, \ldots, M_n)$ is such that $M_i q_{ij} = q_{ji} M_j$ for all i and j.

The canonical matching process is one where agents meet each other in proportion to their relative stocks. We call that the *unbiased matching process*, and it is such that $q_{ij} = \frac{M_j}{\sum_k M_k}$. Given that agent's preferences only depend on own an other types, we let $q_i = q_{ii}$ and then $1 - q_i = \sum_{j \neq i} q_{ij}$.

If the matching process were uniformly at random, then it would have to be that $\sum_i q_i = 1$. However, if there is a bias in the meeting process, agents can meet their own types at a rate which is greater than their relative mass in the meeting process. In that case $q_i > \frac{M_i}{\sum_k M_k}$ and so $\sum_i q_i > 1$.

The parametrization of the matching process that we work with is one such that:

$$\sum_i q_i^{\beta_i} = 1 \quad , \tag{6}$$

---

[10] This can be justified by a limit process with an infinite number of agents, as discussed in the appendix of CJP, and renders the analysis tractable.

[11] See Section 4.5 in CJP.

where $\beta_i \geq 1$ is the bias that each type has toward itself in the matching process. The case $\beta_i = 1$ in the unbiased process (uniformly random meetings), and the meeting bias of type i increases with $\beta_i > 1$.

4.4 Calibrating the Model

Using this model, we can estimate the parameters from steady-state conditions.

First, note that from the optimization problem for the agents (5) it follows that for any pair of types i and j,

$$t_i \left(\gamma_j + (1 - \gamma_j) q_j\right)^{\frac{\alpha}{1-\alpha}} = t_j \left(\gamma_i + (1 - \gamma_i) q_i\right)^{\frac{\alpha}{1-\alpha}} \quad . \tag{7}$$

Next, note that t_i, t_j and q_i, q_j are available from the data since t_i is simply the total number of friends, and $q_i = s_i/(s_i + d_i)$ is the relative rate at which type i's meet themselves. Thus, we can estimate the preference parameters (α and the γ_i's) from the data by searching over values which come closest to satisfying (7).

Finally, we can estimate the meeting bias parameters (the β_i's) directly by searching over values which come closest to satisfying (6).

In next section we calibrate the model using the Add Health data.

5 Calibration of the Model

5.1 Differences Across Races

In this section we calibrate the model described in Section 4 to the Add Health data.

The Add Health data actually have six (self-reported) ethnic categories: Asian, Black, Hispanic, White, Mixed, and Missing. For the sake of completeness and to avoid discrepancy with the empirical data, we use a category of *Others*: to include "Mixed" and "Missing" outcomes.

We calibrate the model as follows. Consider a specific choice of values for α, γ_{Asians}, γ_{Blacks}, $\gamma_{Hispanics}$, γ_{Whites} and γ_{Others}. For every pair of races i and j, represented in a given school, we compute the difference between the right hand side and the left hand side of equation (7). For each school we then have an error from summing across each pair of races. The overall error of that choice for α and γ's is then the sum of all of these errors squared across schools. In order to find the "best" choice, we search over a grid of values for the above parameters to find the one that minimizes this sum of squared errors.

Table 5 reports the combination that minimizes the sum of squared errors as described above.

Our calibration of the model with respect to preference bias is consistent with the statistical evidence from Tables 1, 4 and 2.

Next, we calibrate the matching bias parameters based on equation (6). Here we have one equation for each school and the error is the difference between the right hand side and the left

Parameter:	α	γ_{Asians}	γ_{Blacks}	$\gamma_{Hispanics}$	γ_{Whites}	γ_{Others}
Calibrated Value:	.60	.75	.85	.85	.65	.80

Table 5: Best calibration of equation (7) from the Add Health data. The grid is of step .05, from .20 to .85 for α and from .40 to 1 for each of the γ's

hand side of equation (6), and then we sum the squared errors across schools and choose bias parameters to minimize that sum. We first search over a grid of step .5, from 1 to 9. If we hit a corner then we refine the grid to a step of .1 and search again (now from 1 to 3). In this case we hit corners for Whites and Others, as we discuss in more detail below, but not for the remaining races. Table 6 reports the combination that minimizes the sum of squared errors.

Matching bias	β_{Asians}	β_{Blacks}	$\beta_{Hispanics}$	β_{Whites}	β_{Others}
	7	7.5	2.5	1	1

Table 6: Best calibration of equation (6) from the Add Health data. The grid is of step .1, from 1 to 3, for Whites and Others, of step .5, from 1 to 9, for the other three races.

The result $\beta_{Whites} = 1$ identifies a corner solution. One reason for this is that the calculation ends up including some noisy observations which are those corresponding to groups that are very small fractions of their schools. For example, if a group is a few percent of a school, then it can end up just having a few students and their idiosyncratic behavior ends up influencing the error.[12] We are considering averages across all the students of the same type as a good indicator of their matching opportunities, and this assumption relies implicitly on a law of large numbers. For this reason we rerun the calibration excluding all those observed types i whose representative w_i in the school is smaller than a threshold τ. To do this we can define a "threshold" version of equation (6), that accounts for this:

$$\sum_{i:w_i>\tau} q_i^{\beta_i} = \sum_{i:w_i>\tau} w_i \quad . \tag{8}$$

Here only a type i for which $w_i > \tau$ is considered. The sum of the biased opportunities $q_i^{\beta_i}$ of sufficiently represented types should now sum to 1 minus the fraction of those minorities we have excluded (clearly $\sum_{i:w_i>\tau} w_i = 1 - \sum_{i:w_i<\tau} w_i$).

We run a calibration of equation (8) adopting a threshold $\tau = .06$ (which is the minimal value that avoids $\beta_{Whites} = 1$). In this way we consider 237 out of 389 observations (we discard 6 out of 83 Whites, 31 out of 70 Blacks, 41 out of 82 Hispanics, 58 out of 70 Asians and 16 out of 84 Others). Table 7 shows the best fitting β's.

---

[12] We cannot simply re-weight observations, as we need to respect the structural equations from the model.

Matching bias	β_{Asians}	β_{Blacks}	$\beta_{Hispanics}$	β_{Whites}	β_{Others}
	4.5	6	3	1.1	1

Table 7: Best calibration of equation (8), with a threshold $\tau = .06$, from the Add Health data. The grid is of step .1, from 1 to 3, for Whites and Others, of step .5, from 1 to 9, for the other three races.

The obtained result is no longer a corner solution for Whites and is consistent with the qualitative outcomes of Table 6.

5.1.1 Discussion and Significance Testing

The calibrations help illuminate the econometric results discussed in Section 3.

The first thing to note is that all of the preference bias parameters are lower than 1. Thus all races exhibit some bias in preferences towards their own race. When we test whether the preference bias parameters are significantly different from 1, we find that they are significantly different well beyond the 99% confidence level (with an F-statistic of 14.61). The specifics of all of the significance tests appear in the appendix, but are easy to explain: we examine how the errors that are minimized to find the estimates change as the parameters are changed and perform F-tests on the changes in total error.

The second thing to note is that there are some differences between the races. Whites exhibit the strongest preference bias with a γ of .65, so that a friendship with another race provides only 65 percent of the utility of a friendship with same race to a white. Blacks and Hispanics see values of 85 percent for the same parameter and Asians are in between with a value of 75 percent. Those differences are relatively small but some of them turn out to be significantly different from each other. Whites and Asian preference biases are statistically indistinguishable at the 95% level, as are Black and Hispanic preference biases, but we find that White and Asian preferences are significantly more biased than Black and Hispanic preferences (at the 95% significance level).

When we examine the meeting biases, we see much more dramatic differences across races. Whites have nearly none, while that of the other races are quite substantial, with Blacks seeing the largest meeting bias, Asians the second largest, and Hispanics the third. The bias parameter for Blacks is almost six times that of Whites. The differences between each pair of races is highly significant, except between Asians and Blacks (again see the appendix for the details of each test).

Thus, we see that all races exhibit some preference bias, and we see some variation of those biases across races, most notably with Whites and Asians exhibiting slightly more bias than Blacks and Hispanics. However, we see very substantial and significant differences in meeting biases, with almost no bias for Whites, some bias for Hispanics, and very pronounced biases for Asians and Blacks, which would account for the differences in the observed homophily patterns across races (e.g., Figure 4).

The previous interpretations are based on a model such that all students of a given race behave homogeneously. This is an important caveat, since significant differences may arise within groups, and possibly driving average data. Moreover, in deriving conclusions about norms and/or policies it is important to recall that we are not taking into account other socio-economic factors that could be correlated with race and be driving some of the differences in calibration (for instance, a preference for friendships along some other dimension).

5.2 Calibration: School Size and Matching Rates

We now calibrate the model by school size, seeing what differences in biases exist between small and large schools. The threshold determining size is kept at 1000, as in the empirical section. We run separate calibrations for small and large schools. Here we report the results when the bias parameters are set to be the same for each race, but allowing them to vary with school size. We report the full fitting of the parameters for all races by small and large schools in the appendix - the results are essentially the same: there is almost no difference in preference biases by school size and these differences are insignificant (with an F-statistic of 0.032), but there are large changes in meeting biases by school size and these are highly significant (with an F-statistic of 4.714). In particular, when we examine the meeting biases by races and school size, Black meeting biases are 6 for small schools and 9 for large schools,[13] Hispanics vary from 2 for small schools and 6.5 for large schools, where there is no change for Whites (unbiased in each case) and actually a decrease for Asians (from 6.5 for small to 3.0 for large). So again, we see different patterns for Blacks and Hispanics compared to Asians and Whites.

Tables 8 and 9 show the results.[14] For the first fit the $\alpha = .65$ from our previous calibration (again, for the full fitting of all parameters, see the appendix).

Preference Bias	γ_{small}	γ_{big}
	.83	.80

Table 8: Best calibration of equation (7) from the Add Health data, based on school size. The grid is of step .01, from .01 to 1, for both small and big schools.

These calibrations find a larger matching bias in larger compared to smaller schools, but almost no difference in preference bias. The larger bias towards same–race meetings in larger schools than in smaller schools is intuitive. For example, it may depend on the presence of what are often termed

---

[13] This last value is an upper bound on the range used in calibrations. A larger upper bound would further increase the F–statistic, which is already significant at the 99% level.

[14] Here, we have not adopted any threshold for the estimation of β's, because we did not have corner solutions.

Matching Bias	β_{small}	β_{big}
	1.79	2.42

Table 9: Best calibration of equation (6) from the Add Health data, based on school size. The grid is of step .01, from 1 to 9, for both small and big schools.

"loci of activity" such as clubs and sport teams that enlarge the sources of segregation instead of diminishing them (see the discussion at the end of Section 3.2.1).

6 Concluding Remarks

We have used a friendship formation model to study how the homophilous behavior of high school students, and its key determinants, vary across racial groups and according to school size. We have focused on two main forces that drive the generative process of homophily: i) preferences over the racial mix of friendships, and ii) meeting opportunities. We find differences in biases across races and across schools. In particular, both the inbreeding homophily index and the total number of friends as a function of group size differ across races; also, larger schools tend to be associated with more homophilous behavior, especially for small groups. A calibration of the model exhibits both preference and meeting biases, and we find quite significant differences in meeting biases across races and relatively small (insignificant) differences in preference biases across races. Regarding the effect of school size, the evidence suggest that larger inbreeding indexes are to be imputed to larger biases in the meeting process.

While our work is suggestive of different determinants for differences in homophily patterns across races and schools, it is far from conclusive and suggests further research on these issues.

References

[1] Berry, C. R., and M. R. West (2008) "Growing Pains: The School Consolidation Movement and Student Outcomes", *Journal of Law, Economics, and Organization*.

[2] Blau, P.M., (1977) *Inequality and Heterogeneity: A Primitive Theory of Social Structure*, New York: Free Press.

[3] Bramoullé, Y., and B. W. Rogers (2009) "Diversity and Popularity in Social Networks", Mimeo.

[4] Calvo Armengol, A., and M. O. Jackson (2004) "The Effects of Social Networks on Employment and Inequality," *American Economic Review* 94 (3), 426–454.

[5] Calvo Armengol, A., and M. O. Jackson (2009) “Like Father, Like Son: Network Externalities, Parent-Child Correlation in Behavior, and Social Mobility,” *American Economic Journal: Microeconomics* 1 (1), 124–150.

[6] Carley, K. M., (1999) “On the evolution of social organizational networks,” *Res. Sociol. Org.*, 16, 3-30.

[7] Coladarci, T., and C. D. Cobb (1996) “Extracurricular Participation, School Size, and Achievement and Self-Esteem Among High School Students: A National Look,” *Journal ofResearch in Rural Education*, 12 (2), 92–103.

[8] Cohen, J., (1977) ”Sources of peer group homo- geneity”, *Sociol. Educ.* 50:227412

[9] Coleman, J., (1958) “Relational analysis: the study of social organizations with survey methods,” *Human Organization*, 17, 28-36.

[10] Currarini, S., M.O. Jackson, and P. Pin (2008) “An Economic Model of Friendship: Homophily, Minorities and Segregation,” forthcoming in *Econometrica*.

[11] Feld, S.L., (1981) “The Focused Organization of Social Ties”, *The American Journal of Sociology* 86 (5), 1015–1035.

[12] Franz, S., M. Marsili and P. Pin (2008) “Opportunity and Choice in Social Networks”, SSRN Working Papers.

[13] Giles, M.W., and A. Evans (1986) “The power approach to intergroup hostility,” *Journal of Conflict Resolution* 30 (3), 469–486.

[14] Golub, B., and M. Jackson (2009) “How Homophily Affects Learning and Diffusion in Networks”, Mimeo, arXiv:0811.4013v1.

[15] Goodreau, SM, JA. Kitts, and M Morris (2009) “Birds of a Feather, or Friend of a Friend? Using Exponential Random Graph Models to Investigate Adolescent Social Networks, ” *Demography*, 46(1): 103125.

[16] Jackson, M.O., (2007) , “Social Structure, Segregation, and Economic Behavior,” presented as the *Nancy Schwartz Memorial Lecture* at Northwestern University, http://www.stanford.edu/∼jacksonm/schwartzlecture.pdf.

[17] Jackson, M.O., (2008) *Social and Economic Networks*, NJ: Princeton University Press.

[18] Kandel, D. B., (1978) “Homophily, Selection and Socialization in Adolescent Friendships”, *American Journal of Sociology* 84, 427–436

[19] Khmelkov, V. T., and M. T. Hallinan (1999) “Organizational Effects on Race Relations in Schools”, *Journal of Social Issues*, 55 (4), 627-645.

[20] Knoke, D., (1990) “Networks of Political Action: Towards Theory construction”, *Social Forces* 68, 1041–1063

[21] Kubitschek, W. N., and M. T. Hallinan (1998) “Tracking and Students’ Friendships”, *Social Psychology Quarterly*, 61 (1), 1–15.

[22] Lazarsfeld, P.F, and R.K. Merton (1954) “Friendship as a social process: a substantive and methodological analysis,” in *Freedom and Control in Modern Society*, ed. M Berger, 18–66, New York: Van Nostrand.

[23] Marsden, P.V., (1987) “Core Discussion Networks of Americans,” *American Sociological Review* , 52, 122–131.

[24] Marsden, P.V., (1988) “Homogeneity in confiding relations,” *Social Networks*, 10, 57–76.

[25] McPherson, M., L. Smith-Lovin, and J.M. Cook (2001) “Birds of a Feather: Homophily in Social Networks,” *Annu. Rev. Sociol.*, 27, 415–44

[26] McPherson, J. M., and L. Smith-Lovin (1987) “Homophily in voluntary organizations: status distance and the composition of face-to-face groups,” *American Sociological Review*, 52, 370-79.

[27] Meeker, R. J., and D. M. Weiler (1970) *A New School for the Cities.*

[28] Moody, J., (2001) “Race, school integration, and friendship segregation in America,” *American Journal of Sociology*, 107 (3), 679–716.

[29] Morgan, D. L., and D. F. Alwin (1980) “When Less Is More: School Size and Student Social Participation,” *Social Psychology Quarterly*, 43 (2), 241–252.

[30] Mortensen, D.T., and C.A. Pissarides (1994) “Job Creation and Job Destruction in the Theory of Unemployment,” *Review of Economic Studies*, 61 (3), 397–415.

[31] Ornstein, A. C., (1990) “School Size and Effectiveness: Policy Implications”, *The Urban Review*, 22 (3), 239–245.

[32] Page, S.E., (2007) *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies*, Princeton University Press: NJ.

[33] Patacchini, E., and Y. Zenou (2006) “Racial Identity and Education,” SSRN Working Papers.

[34] Rytina, S., and D. L. Morgan (1982) “The Arithmetic of Social Relations: The Interplay of Category and Network,” *The American Journal of Sociology*, 88 (1), 88–113.

[35] Slate, J. R., and C. H. Jones (2005) “Effects of school size: A review of the literature with recommendations”, *Essays in Education*.

[36] Stearns, E., (2004) “Interracial Friendliness and the Social Organization of Schools”, *Youth & Society*, 35 (4), 395–419.

[37] Weinberg, B. A., (2007) “Social interaction with endogenous associations,” NBER Working Paper Series.

7 Appendix - Significance Tests

We consider the following F-statistic:

$$F = \frac{\dfrac{(SSR_{con} - SSR_{uncon})}{p_{uncon} - p_{con}}}{\dfrac{SSR_{uncon}}{n - p_{uncon}}}$$

where SSR stands for ”sum of squared residuals” of the best fit calibration, while p is the number of parameters estimated in the various models, and n is the number of observations: 84. The subscript ”con” stands for the constrained model under the null hypothesis that some of the γ’s are equal to each other and/or take on some values. The subscript ”uncon” stands for unconstrained model, where all parameters are fit as in Section 5.1.

7.1 Significance Tests of the γ Parameters

Let us begin with a general description of how an F-test is performed for a generic fitting exercise.

Consider the following definitions of variables.

- RSS_{con} = Sum of Squared Residuals (Errors) of the constrained model under whatever null hypothesis is being considered. For example, this might involve only searching over grids where some of the γs for different races are equal to each other.
- RSS_{uncon} = Sum of Squared Residuals of the Unconstrained Model, in which all the parameters are allowed to vary along.
- p_{con} = number of parameters fit in the constrained model.
- p_{uncon} = number of parameters fit in the unconstrained model.
- n = number of observations (the number of schools considered)

We then examine the following test statistic which has an F-distribution with $(p_{uncon}-p_{con}, n-p_{uncon})$ degrees of freedom:[15]

$$\frac{(RSS_{con} - RSSS_{uncon})/(p_{uncon} - p_{con})}{(RSS_{uncon})/(n - p_{uncon})}$$

To illustrate this, before presenting the full tables of all of the tests, we first test the null hypothesis that all γ parameters are equal to 1. This is a test of the hypothesis that preferences are not sensitive to race.

We examine a 99% confidence level, and look at the F-statistic with $(5, 78)$ degrees of freedom. The threshold F level for a 99% level is 3.26. We obtain:

$$F = \frac{\frac{15811.0 - 8163.3}{6-1}}{\frac{8163.3}{84-6}} = 14.61^{**} > 3.261.$$

7.2 Significance Tests For All 84 Schools

7.2.1 Preference Biases

	α	γ_w	γ_b	γ_h	γ_a	γ_o	RSS_{con}	RSS_{uncon}	F	95% Threshold	99% Threshold
Unconstrained	0.60	0.65	0.85	0.85	0.75	0.80	-	8163.3	-	-	-
White=Black	0.60	0.85	0.85	0.85	0.75	0.80	8595.3	-	4.13^{*}	3.963	6.971
White=Hispanic	0.60	0.85	0.85	0.85	0.75	0.80	8595.3	-	4.13^{*}	”	”
White=Asian	0.60	0.75	0.85	0.85	0.75	0.80	8244.5	-	0.78	”	”
Black=Hispanic	0.60	0.65	0.85	0.85	0.75	0.80	8163.3	-	0.00	”	”
Black=Asian	0.65	0.65	0.80	0.85	0.80	0.80	8519.1	-	3.40	”	”
Hispanic=Asian	0.80	0.80	0.90	0.90	0.90	0.90	8746.9	-	5.58^{*}	”	”
All = 1	0.20	1.0	1.0	1.0	1.0	1.0	15811.0	-	14.61^{**}	2.332	3.261
All =	0.75	0.90	0.90	0.90	0.90	0.90	9279.1	-	2.67^{*}	2.489	3.570

7.2.2 Meeting Biases

	β_w	β_b	β_h	β_a	β_o	RSS_{con}	RSS_{uncon}	F	95% Threshold	99% Threshold
Unconstrained	1.0	7.5	2.5	7.0	1.0	-	1.7265	-	-	-
White=Black	3.0	3.0	1.0	9.0	1.0	4.4483	-	124.5^{**}	3.962	6.967
White=Hispanic	1.5	7.0	1.5	8.5	1.0	2.2366	-	23.34^{**}	”	”
White=Asian	1.5	6.5	3.5	1.5	1.0	2.7748	-	47.97^{**}	”	”
Black=Hispanic	1.0	5.5	5.5	3.5	1.0	2.1486	-	19.31^{**}	”	”
Black=Asian	1.0	7.5	2.5	7.5	1.0	1.7274	-	0.04	”	”
Hispanic=Asian	1.0	7.5	3.5	3.5	1.0	1.8347	-	4.95^{*}	”	”
All = 1	1.0	1.0	1.0	1.0	1.0	25.8836	-	22107^{**}	2.330	3.258
All =	2.0	2.0	2.0	2.0	2.0	6.2069	-	51.25^{**}	2.487	3.566

---

[15] This presumes independent and normally distributed error terms.

7.3 Significance Tests Allowing for Differences between Large and Small Schools (more or fewer than 1000 Students)

Here we fit different parameters for large and small schools and then get a total error when we allow these parameters to vary. This becomes the unconstrained case for the F-test, and then we compare it to the error when we add the constraint that the parameters not vary with school size.

7.3.1 Preference Biases

	α	γ_w	γ_b	γ_h	γ_a	γ_o	RSS_{con}	RSS_{uncon}	F	95% Threshold	99% Threshold
Ignoring Size	0.60	0.65	0.85	0.85	0.75	0.80	8163.3	-	-	-	-
Small Schools	0.60	0.65	0.85	0.85	0.75	0.80	-	4483.7	-	-	-
Large Schools	0.55	0.60	0.80	0.80	0.70	0.75	-	3657.9	-	-	-
Total Error Small+Large							-	8141.6	.032	2.227	3.063

7.3.2 Meeting Biases

	β_w	β_b	β_h	β_a	β_o	RSS_{con}	RSS_{uncon}	F	95% Threshold	99% Threshold
Ignoring Size	1.0	7.5	2.5	7.0	1.0	1.7265	-	-	-	-
Small Schools	1.0	6.0	2.0	6.5	1.0	-	.9406	-	-	-
Large Schools	1.0	9.0	6.5	3.0	1.0	-	.3688	-	-	-
Total Error Small+Large						-	1.3094	4.714^{**}	2.338	3.275