NBER WORKING PAPER SERIES

ON THE MEASUREMENT OF SEGREGATION

Federico Echenique
Roland G. Fryer, Jr.

Working Paper 11258
http://www.nber.org/papers/w11258

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 2005

We are grateful to Kim Border, Toni Calvo-Armengol, David Card, Joan Esteban, Drew Fudenberg, Edward Glaeser, Jerry Green, Oliver Hart, James Heckman, Caroline Hoxby, Matthew Jackson, Kevin Lang, Edward Lazear, Steven Levitt, Erzo Luttmer, Kevin Murphy, Derek Neal, Sarah Reber, and Jesse Shapiro for useful comments and suggestions. We are especially grateful to Lawrence Katz and Glenn Loury for advice, encouragement, and detailed comments on previous drafts. We thank seminar participants at Berkeley, Boston, Brown, Caltech, Carnegie Mellon, Harvard, McGill, NBER, Pompeu Fabra, Torcuato Di Tella, Universitat Autonoma de Barcelona, and Vanderbilt. Patricia Foo, Alex Kaufman, and Roland Pongou provided exceptional research assistance. A portion of this paper was written while Fryer was a visitor at Institut d'Anàlisi Económica at Universitat Autonoma in Barcelona, Spain. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

On the Measurement of Segregation
Federico Echenique and Roland G. Fryer, Jr.
NBER Working Paper No. 11258
March 2005
JEL No. Z13, C0

ABSTRACT

This paper develops a measure of segregation based on two premises: (1) a measure of segregation should disaggregate to the level of individuals, and (2) an individual is more segregated the more segregated are the agents with whom she interacts. Developing three desirable axioms that any segregation measure should satisfy, we prove that one and only one segregation index satisfies our three axioms, and the two aims mentioned above; which we coin the Spectral Segregation Index. We apply the index to two well-studied social phenomena: residential and school segregation. We calculate the extent of residential segregation across major US cities using data from the 2000 US Census. The correlation between the Spectral index and the commonly-used dissimilarity index is .42. Using detailed data on friendship networks, available in the National Longitudinal Study of Adolescent Health, we calculate the prevalence of within-school racial segregation. The results suggests that the percent of minority students within a school, commonly used as a substitute for a measure of in-school segregation, is a poor proxy for social interactions.

Federico Echenique
The Division of the Humanities and Social Sciences
228-77
California Institute of Technology
Pasadena, CA 91125
fede@hss.caltech.edu

Roland G. Fryer, Jr.
Department of Economics
Harvard University
Littauer Center, M-7
Cambridge, MA 02138
and NBER
rfryer@fas.harvard.edu

1 Introduction

Ethnic and racial segregation is an important and well-studied social phenomenon. For over 50 years, social scientists have been concerned with measuring the extent, and estimating the impact of segregation in education, housing, and the labor market. The result of this scholarship has been nearly 20 different indexes of segregation, and a consensus that the spatial separation of many minorities from jobs, role models, health care, and quality local public goods is a leading cause of racial and ethnic differences on many economic, social, and health related outcomes (Almond, Chay, and Greenstone 2003, Borjas 1995, Case and Katz 1991, Kain 1968, Cutler and Glaeser 1997, Massey and Denton 1993, Collins and Williams 1999). Fundamental to understanding the potential impact of segregation is measuring it.

We propose a new approach to measuring segregation based on two premises: (1) a measure of segregation should disaggregate to the level of individuals, and (2) an individual is more segregated the more segregated are the agents with whom she interacts. Having a measure of segregation with the flexibility to disaggregate to the level of individuals opens up windows of opportunity for empirical work, and a better understanding of the mechanisms by which segregation affects economic outcomes. We also desire a measure that gives a larger level of segregation for individuals whose contacts are more segregated. Consider Figure 1, which depicts the distribution of blacks across metropolitan Detroit. There is a large oval in the center of the city containing almost exclusively black households. Any measure of segregation should report that the household in the epicenter is more segregated than a household equidistant from the center and the edge, even when each household has all black neighbors. These are two features that are absent in all existing measures of segregation.

We use networks – individuals and their connections – as our mathematical framework. In this framework, we propose three specific axioms that any measure of segregation on a network should satisfy. We prove that one and only one index satisfies these axioms and the two broad principles above; which we coin the "Spectral Segregation Index" (SSI). The axioms require that: (a) [Monotonicity] if all individuals in Network A have more interactions with agents of the same race than in Network B, then Network A is more segregated than B; (b) [Linearity] an individual is more segregated the more segregated are the agents with whom she interacts, and this relationship takes on a linear form; and (c) [Homogeneity] if all individuals in a network have half of their interactions with members of the same race, the index of segregation is one-half. The latter condition normalizes the index.

We defer a formal definition of the SSI to Section 4. Informally, the SSI measures the connectedness of individuals of the same race. Put differently, the SSI captures the growth of race-specific capital in a network. Consider the following thought experiment. Let there be a network of many individuals. Suppose that, in each period of time, individuals possess a degree of own-race capital—how adept they are at reciting 'Tu Pac' or how well they whistle Vivaldi. In each time-period they transmit some of their capital to other individuals of the same race with whom they interact. The SSI will, as time passes, approximate the growth rate of own-race capital, providing a measure of connectedness and own-race social interaction. In highly segregated areas (areas with high SSI) same-race fads will grow quickly because of the frequency of same-race social interactions.

The SSI has important advantages over existing measures of segregation. First, as a gauge of residential segregation, it is invariant to arbitrary partitions of a city, in that it does not depend on the way that local communities choose to draw their census regions. This is important for a variety of reasons – most crucially having to do with the permanent structure of census regions and the relative mobility of populations. Second, it allows us to investigate how segregated multiple minority groups are, allowing one to compare the segregation of Asians, Blacks, Hispanics, Native Americans, and so on, within and across cities.[1] The SSI makes it possible to compare Hispanic segregation across cities, compare the Hispanics of east Los Angeles from the Hispanics in south Los Angeles, or compare them to Blacks in Chicago. Third, our index allows one to analyze the full distribution of segregation, allowing researches to move beyond aggregate statistics, which can be misleading. For instance, while the average Black may be more segregated than the average Hispanic, it is plausible that the most segregated Hispanic is more segregated than any Blacks. Indeed, we find that this is the case for residential segregation in the US. This is potentially an important distinction for empirical work. Fourth, there are inherent multiplicative effects captured by SSI which other indexes omit. That is, an individual's susceptibility to group-transmitted influences depends on how many contacts the individuals have with members of the group, and the susceptibility of her contacts. Fifth, the SSI can be used for calculations of concentration far beyond the measurement of racial segregation. Natural applications of our index include

[1] Another way to analyze multiple groups with existing indexes is to calculate the weighted average of several dichotomous indexes (see Reardon and Firebaugh 2002). It is not clear how to interpret the findings from such an exercise.

the measurement of traffic, power in organizations, concentration of buyer-supplier networks, academic field specialization, segregation of friendship networks, measures of social popularity, and so on. Generally, the SSI provides a mathematical tool adept at measuring the clustering of particular nodes in networks.

The SSI has some disadvantages as well. It depends on the quality of the information one can obtain about social interactions. In the case of residential segregation, for example, the information is restricted to where individuals live and not how they interact within a city. Unlike other indexes, however, as better information on the nature of social interactions is obtained, the SSI becomes a sharpened proxy of those interactions. Second, it is sensitive to the fraction of individuals in a network who have the race/ethnicity under study. We address this issue by calculating a "baseline" SSI, and comparing actual SSI to the baseline. Finally, implementing the SSI can be computationally demanding, though our applications demonstrate that the computational tasks are feasible.[2]

After formally deriving the SSI, we apply the index to two well-known social phenomena: measuring the extent of residential and school segregation. We begin by evaluating the extent of segregation across major cities in the US, using data from the 2000 Census. The results we obtain are interesting, and in some cases quite surprising. The most segregated cities for Asians (including Pacific-Islanders), Blacks, Hispanics and Whites are Honolulu, HI, Detroit, MI, McAllen, TX, and Lowell, MA respectively. On average, Blacks are more segregated than any other racial group, but the most segregated Hispanics are more segregated than the most

[2] To ease this burden, we have posted results from some of the more computationally intense calculations on the following webpage: http://post.economics.harvard.edu/faculty/fryer/projects.html

segregated Blacks. However, the real power of the SSI is the ability to measure segregation at disaggregated levels, allowing one to measure the intensity of same-race clusters or uncover the most segregated city blocks in America. The largest minority ghetto in the US consists of Hispanics in Los Angeles, CA – 17,909 blocks are connected to each other! The second-largest ghetto is comprised of Blacks in Detroit, MI, and the most segregated ghetto consists of Blacks in Jackson, MS. The most segregated city block in America is a Hispanic block in Western San Antonio, TX. It is important to emphasize that the latter results cannot be obtained with any of the existing measures of segregation.

We also apply the SSI to the measurement of within-school segregation patterns, using data on friendship networks available in the National Adolescent Study of Health (Addhealth). Our analysis uncovers many new facts. First, the common practice of using the percentage of black students in a school as a substitute for within-school segregation measures, is a poor proxy for actual social interactions. When black students are relatively scarce in a school, they tend to be integrated. As their share of the student population increases, segregation increases dramatically, hitting a ceiling when blacks comprise roughly twenty percent of the student population. Schools that have twenty percent or more black students exhibit severe within-school racial segregation.

Second, we correlate individual-level segregation with several traits. More segregated Black students are less likely to smoke (a behavior predominant among white teens) and have lower vocabulary test scores. More segregated Asians are less likely to skip school, they have higher vocabulary test scores, put in more effort, and report being happier. Among

Hispanics, more segregation is associated with less smoking, lower vocabulary test scores, lower probability of attending college, and lower grades. Students of all races are less likely to date interracially when schools are more segregated.

The organization of the paper is as follows. Section 2 provides background and a brief discussion of existing indexes of segregation. Section 3 provides an example that previews many of the ideas and intuitions involved in the general modeling. Section 4 provides an axiomatic derivation of our new index of segregation. Section 5 uses the SSI to estimate the prevalence of residential and school segregation across. Section 6 concludes. There are two appendices. Appendix A contains the technical proofs of all formal results and Appendix B provides a theoretical foundation for Baseline SSI.

2 Background and Previous Literature

At an abstract level, segregation is the degree to which two or more groups are separated from each other. However, practical definitions can be quite distinct from one another, conceptually and empirically. Massey and Denton (1988) group existing indexes into five classes: evenness, exposure, concentration, centralization, and clustering, which they take to resemble the totality of what is usually meant by "segregation." Evenness refers to the differential distribution of two groups across areas in a city. Measures of exposure are designed to approximate the amount of potential contact and interaction between members of different groups. Concentration indexes measure the relative amount of physical space occupied by a minority group. Centralization is the extent to which a group is located near the center of an urban area, and clustering measures the degree to which geographic units inhabited by minority members abut one another, or cluster spatially. Of the five dimensions of segregation, only two are used in the vast majority of applied work in the social sciences: evenness and exposure. Economists ultimately care about the degree to which segregation affects social interactions. For this purpose, concentration and centralization are inadequate, and measures of clustering are largely avoided due to their sensitivity to the number and population of census regions.

The most popular measure of segregation is the "dissimilarity" index (developed by Jahn, Schmid, and Schrag 1947), a measure of evenness.[3] Suppose a city is divided into N sections. The dissimilarity index measures the percentage of a group's population that would have to change sections for each section to have the same percentage of that group as the whole city. In symbols:

$$\text{index of dissimilarity} = \frac{1}{2}\sum_{i=1}^{N}\left|\frac{black_i}{black_{total}} - \frac{nonblack_i}{nonblack_{total}}\right|, \tag{1}$$

where $black_i$ is the number of blacks in area i, $black_{total}$ is the total number of blacks in the city as a whole, $nonblack_i$ is the number of non-blacks in area i, and $nonblack_{total}$ is the number of non-blacks in the city. The dissimilarity index has the appealing feature that it is invariant to the size of a minority group.

A second commonly-used measure of segregation is "isolation," a measure of exposure.

[3] Other measures of evenness include the Gini coefficient (the mean absolute difference between minority proportions weighted across all pairs of geographic units, expressed as a proportion of the maximum weighted mean difference), the Atkinson index (similar to Gini coefficient, but allows researchers to decide how to weight geographic units which are over or under the city-wide distribution), and Entropy (the weighted average of each geographic units deviation from the racial entropy of the city as a whole).

As Blau (1977) recognized, Blacks can be evenly distributed among residential areas in a city, but experience little exposure to non-Blacks if they are a relatively large proportion of the city. Isolation measures the extent to which Blacks are exposed only to one other, rather than to non-Blacks. The index is computed as the minority-weighted average of each section's minority population:

$$\text{index of isolation} = \frac{\sum\left(\frac{black_i}{black_{total}} \cdot \frac{black_i}{person_i}\right) - \left(\frac{black_{total}}{person_{total}}\right)}{\min\left(\frac{black_{total}}{person_i}, 1\right) - \left(\frac{black_{total}}{person_{total}}\right)}, \tag{2}$$

where $person_i$ refers to the total population of area i and $person_{total}$ refers to the total population.[4]

Dissimilarity and isolation possess at least two undesirable properties. First, they explicitly depend on the arbitrary ways in which cities are partitioned into sections (e.g. census tracts).[5] That is, fixing the location of minorities and non-minorities in a city and re-drawing the sections can drastically change the measure of segregation. An exaggerated example is depicted in Figure 2. The city depicted in the figure has a dissimilarity index of 0 – perfect integration – when sections are drawn vertically and has a dissimilarity index of 1 – extreme segregation – when sections are drawn horizontally; no household has moved. Similarly, vertical partitions yield an isolation index of 0 whereas horizontal partitions produce an index of .5. This is a highly undesirable property of any segregation index, as it may artificially

[4]Another commonly used measure of exposure is the interaction index, which is the inverse of the isolation index presented above.

[5]We are not the first to draw attention to this flaw in measures of segregation, see Cowgill and Cowgill (1951), Appendix A in Tauber and Tauber (1965), and Massey and Denton (1988). While this property is problematic for measures of residential segregation, it is less likely to effect measures of occupational or school segregation - where there is a natural clustering of individuals.

indicate that a city is more or less segregated as a function of how the tracks are drawn.

Second, existing measures are not defined when trying to measure segregation at the level of individuals. It is difficult to correctly identify the relationship between segregation and outcomes without individual-level variation in segregation. As a descriptive matter, individual segregation may be more useful than city-wide segregation. Rather than correlate individual economic outcomes with city-wide segregation, one can correlate individual outcomes with individual measures of segregation. On the other hand, the right level of aggregation depends on the problem at hand; group-level, neighborhood, or city-level segregation may be the appropriate level of aggregation in many applications. It is an open empirical question, one that cannot be answered without a measure that disaggregates to the individual level.[6]

The literature in economics involving the measurement of segregation is small (Phillipson 1993, Hutchens 2001, Frankel and Volij 2004). Similar to our exercise, their approach is axiomatic – identifying desirable properties that indexes should possess. But, the literature takes an arbitrary partition of a city as given, and uses the partition to identify indexes axiomatically. As such, there is little in common with our approach.

[6] This critique is conceptual – not purely data driven. That is, existing measures are not equipped to measure segregation at the level of individuals, irrespective of the available data.

3 A Motivating Example

Before moving to a full description of the model, we present an example where we calculate our index, and discuss informally some of its properties. Consider City 1, depicted in Figure 3. The nodes in City 1 represent households. Each household can be one of three races: black, white, or gray. In the figure, household $(A, 1)$ is white, $(B, 2)$ is black, $(B, 4)$ is gray, and so on.

There are three steps involved in calculating the SSI for City 1. First, we need to define who is a neighbor of whom. Here, we pick a simple definition; two nodes are neighbors if they are adjacent.[7] In general, however, this is the most important decision in implementing the index. Second, we represent a minority's neighborhood relations in a "social-interactions matrix." The ij entry of the matrix equals $1/4$ if minority-members i and j are neighbors; otherwise it is 0. Why $1/4$? Black and gray households have 4 neighbors each, and for the purposes of this example, we assume that their relations with each neighbor is equally

[7] This restrictive definition is solely for the purposes of the motivating example.

intense. For blacks in City 1 the interaction matrix is:

$$
\begin{array}{c}
\\ (B,2) \\ (C,2) \\ (D,2) \\ (B,3) \\ (C,3) \\ (D,3) \\ (D,4) \\ (D,5)
\end{array}
\begin{array}{c}
\begin{array}{cccccccc} (B,2) & (C,2) & (D,2) & (B,3) & (C,3) & (D,3) & (D,4) & (D,5) \end{array} \\
\begin{bmatrix}
0 & \frac{1}{4} & 0 & \frac{1}{4} & 0 & 0 & 0 & 0 \\
\frac{1}{4} & 0 & \frac{1}{4} & 0 & \frac{1}{4} & 0 & 0 & 0 \\
0 & \frac{1}{4} & 0 & 0 & 0 & \frac{1}{4} & 0 & 0 \\
\frac{1}{4} & 0 & 0 & 0 & \frac{1}{4} & 0 & 0 & 0 \\
0 & \frac{1}{4} & 0 & \frac{1}{4} & 0 & \frac{1}{4} & 0 & 0 \\
0 & 0 & \frac{1}{4} & 0 & \frac{1}{4} & 0 & \frac{1}{4} & 0 \\
0 & 0 & 0 & 0 & 0 & \frac{1}{4} & 0 & \frac{1}{4} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{4}
\end{bmatrix}
\end{array}
$$

Similarly, the "gray graph" gives a 4×4 matrix.

The interactions matrix is always a given in the calculation of the SSI. Any information about the intensity of relations between households can (and should) be incorporated in the matrix. For example, in some applications one can infer the relative importance of particular interactions. In measuring residential segregation, however, all information regarding interactions is gleaned from geographical distance; the matrix thus looks like the one for City 1, except households will have different numbers of neighbors.

The third and final step is to calculate the largest eigenvalue of the matrix, and the associated eigenvector. The eigenvalue is the SSI. The eigenvector, after normalizing such that the minority-level SSI is the average of the individual-level values, gives the individual-level values of SSI. The results are depicted in Table 1. $\widehat{S}^h$ denotes the SSI for blacks/grays while $\widehat{s}_i^h$ denotes the individual-level SSI for each household in City 1.

Table 1: Spectral segregation for Grays in City 1.

i	$(B,4)$	$(B,5)$	$(C,4)$	$(C,5)$	$\widehat{S}^h$
$\widehat{s}_i^h$	.5	.5	.5	.5	.5

Table 2: Spectral segregation for Blacks in City 1.

i	$(B,2)$	$(B,3)$	$(C,2)$	$(C,3)$	$(D,2)$	$(D,3)$	$(D,4)$	$(D,5)$	$\widehat{S}^h$
$\widehat{s}_i^h$	.596	.606	.875	.912	.668	.787	.378	.152	.622

Blacks are more segregated than grays. Households $(B,2)$, $(B,3)$, $(C,2)$ and $(C,3)$ are situated like the grays, thus the blacks must be at least as segregated as the grays. All grays have the same degree of segregation. The most segregated black household is (C,3), the least segregated black is (D,5).

The example provides a snap-shot of how an individual measure of segregation allows the identification of new testable implications. If segregation is strongly correlated with poverty, less schooling, and poor health, we would expect $(D,3)$ to be poorer, less educated, and have lower life expectancy relative to $(D,5)$ – all else equal. Further, the SSI allows one to ask questions that were not possible before. For example, In City 1, who is more segregated, household $(B,2)$ or household $(C,4)$? Each has two black neighbors and two white neighbors. The Spectral index gives a higher level of segregation to $(B,2)$, because $(B,2)$'s black neighbors are more segregated than $(C,4)$'s. If segregation reflects the strength of a household's same race social interactions, as a result of who its neighbors are, the Spectral index seems to capture the right thing.

City 2 in Figure 3 doubles the size of City 1 by adjoining an exact replica. Black and gray segregation in City 2 is the same as in the original city, because neither blacks or grays have new neighbors of the same race. The segregation of whites, however, increases substantially in City 2, as they have become more connected. Notice, if we eliminated columns 6 and 7, black and gray segregation would increase. In essence, the SSI provides a weighted average of all the minority components throughout a network.

4 The Spectral Segregation Index

A. Basic Definitions

Let $G = (V, E)$ denote a graph with vertex-set V and edge-set E, $E \subseteq [V]^2$. [8] The elements of V are vertexes (or nodes) of G, and the elements of E are two-element subsets of V which represent the edges of G. Vertexes represent the individuals in a network, and edges represent relationships. The adjacency matrix of G, with $V = \{1, 2, ...n\}$, is the $n \times n$ matrix $A = (a_{ij})$ whose $(i, j)-$ entry a_{ij} is 1 if there is an edge between i and j, and 0 otherwise. Two vertexes are adjacent (or neighbors) if they are connected by an edge.

B. The Model

[8] We use the most basic notions in Graph Theory. A reader can consult any graph-theory textbook, for example Diestel (1997). Some of the ideas we use are from the field of Spectral Graph Theory; see e.g. Cvetković, D., Rowlinson, P., and Simić, S. (1997) for a comprehensive treatment.

For each $i, j \in V$ let there be a real number r_{ij} that measures how much i interacts with j. Interactions are assumed to be non-negative and bilateral; $r_{ij} \geq 0$ and $r_{ij} > 0$ if and only if $r_{ji} > 0$. Let $R = (r_{ij})$ be the matrix of interaction-measures. R is a non-negative (not necessarily symmetric) $|V| \times |V|$ matrix.

The matrix R is a primitive of our model. One can assume different Rs, depending on the detail with which one can judge how intense interactions are.[9] Social science can play a key role in deriving the matrix R. If, through derivation of an economic model, or detailed participant observation, we know that there exist peculiar behavioral patterns in the manner in which individuals interact (Chicagoans never talk with people South of them, Cambridge residents do not befriend anyone across the Charles, or students rarely interact with other students outside their school district, e.g.) we can incorporate these behaviors into the matrix R. It is important to emphasize that as empirical researchers get better information regarding the structure of social interactions, the index will become a sharpened proxy for those interactions.

Let $H = \{1, 2, ...K\}$ be a finite set of races and ethnic groups, and let $a : V \rightarrow H$ denote an assignment of individuals to a race. Thus, $h = a(v)$ implies that, under assignment a,

[9] There are many ways to think of constructing R. Our leading example is the *neighborhood model*. In this setup, individuals either interact if they are neighbors or do not interact if they are not (this still leaves room for indirect interactions: my interactions, through my neighbors, with my neighbors' neighbors). Thus, $r_{ij} \in \{0, 1\}$. A more relaxed version of the neighborhood model is one in which r_{ij} can take one of a number of finite values, and where larger values reflect longer geographical distances – degrees of "neighborhoodness" say. For example, let r_{ij} be 0 if i and j are further than 10 miles apart, and $r_{ij} = 1$ if they are 1 mile apart, $r_{ij} = 0.9$ if they are 2 miles apart, $r_{ij} = 0.8$ if they are 3 miles apart, and so on. An even different model is where there is an interaction time constraint. Suppose that for each i there is a set $F_i \subseteq V$ of individuals with whom i could feasibly interact. Then $r_{ij} \in \left\{0, \frac{1}{T}, \frac{2}{T}, ..., 1\right\}$ and $\sum_{j \in F_i} r_{ij} \leq 1$. These models are just a few ways one can derive the matrix of social interaction, R. Our theoretical approach is agnostic on this dimension. The structure of R is an important decision for applied researchers.

individual v belongs to race h. Let $A = H^V$ be the set of all such assignments.

Given an assignment $a \in A$ and a matrix of social interactions R, let $\beta = (a, R)$ denote the assignment-interactions pair. An assignment-interactions pair defines a matrix of "same-race interactions," $B(\beta) = (b_{ij})$, where entry b_{ij} of $B(\beta)$ is 0 if $a(i) \neq a(j)$, and r_{ij} if $a(i) = a(j)$.

Let $n^h(\beta)$ be the number of h-race individuals under β, and let $B^h(\beta)$ be the $n^h(\beta) \times n^h(\beta)$ submatrix of $B(\beta)$ where the rows (and columns) i are such that $h = a(i)$. When an assignment-interactions pair $\beta = (a, R)$ is understood, we shall abuse notation in two ways: we denote by h the set of individuals assigned to race h, so h sometimes denotes $a^{-1}(h)$, and we use B^h to denote $B^h(\beta)$. Finally, let $G^h(\beta)$ be the graph with vertex set h, and where ij is an edge if and only if $b_{ij} > 0$.

C. Three Axioms of a Desirable Measure of Segregation

Before describing the axioms, it is useful to introduce a bit more notation. Given $\beta = (a, R)$, for an individual i, let N_i^β be the set of individuals j that interact with i ($r_{ij} > 0$), and are the same race ($a(i) = a(j)$). So N_i^β is the set of neighbors of i in $G^h(\beta)$.

We say that $\beta' = (a', R')$ has more race h-segregation than $\beta = (a, R)$ if, under β', every agent in race h has at least as many same-race interactions. That is, letting $B^h(\beta) = (b_{ij})$ and $B^h(\beta') = (b'_{ij})$, $b_{ij} \leq b'_{ij}$ for all i and j in h. In the simplifying case in which $r_{ij} \in \{0, 1\}$ (which we refer to as the the neighborhood model), this occurs when $(N_i^\beta)_{i \in V}$ and $(N_i^{\beta'})_{i \in V}$ satisfy $N_i^\beta \subseteq N_i^{\beta'}$, for all $i \in V$.

A *segregation index* is a map

$$\beta \mapsto \left(S^h(\beta), \left(s_i^h(\beta)\right)_{i\in h}\right),$$

where $s_i^h(\beta) \geq 0$, and $S^h(a)$ is the average of the $s_i^h(a)$ over $i \in h$.

Our definition of a segregation index reflects our requirement that segregation be measured at the individual level. Individual segregation is measured in the same units as racial segregation. Race-h segregation is the average of the segregation of all individuals of race h.

Axiom 1 states that if two graphs have the same number of race-h nodes, then the graph with more h-segregation is more segregated.

Axiom 1 (Monotonicity) *If β' has more h-segregation than β, and $n^h(\beta) = n^h(\beta')$, then $S^h(\beta) \leq S^h(\beta')$.*

Our second axiom is a normalization of the index. Let $d > 0$ be a scalar. An assignment-interactions pair β is *h-homogeneous of degree d* if, for all i in h, $\sum_{j\in N_i^\beta} b_{ij} = d$. In the neighborhood model, an assignment is h-homogeneous if N_i^β has exactly d households for all i with $h = a(i)$.

Axiom 2 (Homogeneity) *Let β be h-homogeneous of degree d, then $S^h(\beta) = d$.*

Our third axiom is the most controversial. We want the segregation of an individual i to depend on the segregation of the individuals with whom she interacts. Axiom 3 requires this dependence to take a linear form.

Let C_i be the connected component of $G^h(\beta)$ that i belongs to. That is, C_i is the set of individuals j with some connection to i, those with some connections to the individuals with some connection to i, and so on. Let $S^{C_i}(\beta)$ be the average segregation (s_i^h) of individuals in C_i.

Axiom 3 (Linearity) *If $S^{C_i}(\beta) > 0$,*

$$s_i^h(\beta) = \frac{1}{S^{C_i}(\beta)} \sum_{j \in N_i^\beta} r_{ij} s_j^h(\beta)$$

D. Interpreting the Axioms

The above axioms are, for the most part, intuitive. We require the segregation of a household i to depend on how many of i's neighbors are the same race as i. Monotonicity and Homogeneity are expressions of this requirement. Monotonicity requires that if City A's same race neighbors have more interactions than City B's, and they have the same number of that race, City A exhibits a higher degree of segregation than B. Homogeneity says that if a city were regular, so everyone has exactly the same amount of same-race interactions, the index should simply be this fraction. Thus, Homogeneity describes the index's unit of measurement; providing a way to interpret a value of the index. If $S^h(\beta) = .4$, then we know that there is more segregation than if every individual of minority h had a weighted average of one-third of their time with other members of race h, but less than half. Homogeneity also provides a "scale free" like property: If City A has more households than City B, but each household in both cities has the same fraction of same-race neighbors, the index will report the same level of segregation for both cities.

The linearity axiom is our key innovation. We require $s_i^h(\beta)$ to depend on the strength of i's h–specific interactions. As described in the Introduction, if one considers Figure 1, which depicts the distribution of blacks across metropolitan Detroit, it seems evident that individuals in the center of the of the city's black ghetto should be measured as more segregated than those closer to the edge. Linearity is one embodiment of this requirement.

E. Deriving the Spectral Segregation Index

Fix a race h. Let C_k, $k = 1, 2, \dots K$, be the connected components of $G^h(\beta)$. Abusing notation, let C_k also denote the submatrix of B^h with columns (and rows) indexed by the elements of C_k. Let λ_k be the largest eigenvalue of C_k, and x_k be its associated eigenvector, normalized so its entries add to one. Note that λ_k and x_k must exist by the Perron-Froebenius Theorem.

The *Spectral Segregation Index (SSI)* is the index

$$\beta \mapsto \left(\hat{S}^h(\beta), (\hat{s}_i(\beta))_{i\in h}\right),$$

where $\hat{S}^h(\beta) = \sum_{i\in h} \frac{\hat{s}_i(\beta)}{n^h(\beta)}$ and $\hat{s}_i(\beta) = \lambda_k x_{ki} \left|C_k\right|$.

The next theorem presents our main theoretical result; that SSI satisfies monotonicity, homogeneity, and linearity, and no other index does.

Theorem 1 *A segregation index satisfies Monotonicity, Homogeneity and Linearity if and only if it is the Spectral Segregation Index.*

The key idea in the proof of Theorem 1 is quite simple: Linearity requires that, when $S^{C_i}(\beta) > 0$, $S^{C_i}(\beta)s_i^h(\beta) = \sum_{j \in N_i^a} r_{ij} s_j^h(\beta)$, which is equivalent to requiring that $\S^{C_i}(\beta)$ be an eigenvalue of C_k. And, by a corollary to the Perron-Froebenius Theorem for irreducible non-negative matrices, C_k has exactly one positive eigenvalue. Thus, Linearity is almost by itself a definition of the SSI. While not a deep axiomatization of SSI, we think Theorem 1 is still valuable because the three axioms are easy to interpret economically, while the index itself is not. The axioms allow one to evaluate the relatively basic assumptions behind the SSI. [10]

Without assuming linearity, we would be unable to derive a unique numerical index. If, for example, the linearity assumption is replaced with a monotonicity condition – higher segregation among i's same-race neighbors imply higher $s_i^h(\beta)$ – one cannot pin down a specific numerical index. The situation is analogous to that of income distribution measures, where general properties lead to orderings of Lorenz curves, that do not allow one to compare any two distributions. [11]

We state two additional properties of SSI.

[10]Palacios-Huerta and Volij (2004) provide a characterization of eigenvectors of irreducible matrices in terms of more primitive axioms. Their methods are not applicable to our problem because we also need—crucially, in fact—to characterize the eigenvalue. In Palacios-Huerta and Volij's model, the eigenvalue is always fixed.

[11]In our framework a Lorenz-curve-type ordering is readily obtained: let race h be more segregated in β than in β' if the distribution of $(\sum_j r'_{ij})$ dominates that of $(\sum_j r_{ij})$.

Proposition 1 *If $i \in h$ has at least one same-race neighbor, $\hat{s}_i^h(\beta) > 0$. If i has no same-race neighbors, $\hat{s}_i^h(\beta) = 0$.*

Proposition 2 *If C_k, $k = 1 \ldots K$ are the connected components (the irreducible submatrices) of $B^h(\beta)$, then*

$$\hat{S}^h(\beta) = \sum_{k=1}^{K} \left(\frac{|C_k|}{n_a^h} \right) \hat{S}^{C_k}(\beta),$$

and $S^{C_k}(\beta)$ is the largest eigenvalue of C_k. So $\hat{S}^h(\beta)$ is the weighted average of the components' largest eigenvalues.

Proposition 1 says that SSI will identify the isolated individuals in a network, as those individuals have an SSI of 0. Proposition 2 demonstrates that SSI is the average SSI over the connected components of $G^h(\beta)$. These connected components are particularly interesting in some applications. In residential segregation, they can be interpreted as ghettos, and in school segregation as same-race cliques.

F. The Economics of SSI.

The three axioms provide the precise assumptions underlying the SSI. An alternative way to envision the SSI is through a model of group-specific capital transmission. SSI is a measure of how fast same-race capital is disseminated as a result of social contacts.[12]

Here, suppose that $G^h(\beta)$ is a connected graph; without this assumption, the result will hold in each connected component of $G^h(\beta)$.

[12] We thank Erzo Luttmer for suggesting this interpretation.

Let x_i be a measure of how much race-specific capital an individual i has. Suppose that, in each period t, individual i's h-capital grows depending on how much h-specific capital her contacts have, and on how much i interacts with them.

Specifically, suppose that

$$x_{it} = x_{it-1} + \sum_{j \in h} b_{ij} x_{jt-1}, \tag{3}$$

and that x_{i0} is given, for all i. While simple, this model has been used recently to study cultural transmission in networks (Brueckner and Smirnov 2004) and has a direct link to SSI.

Proposition 3 *For all* $(x_{0j})_j$, *and all* i,

$$\lim_{t \to \infty} \frac{x_{it}}{x_{it-1}} = 1 + S^h(\beta)$$

Proposition 3 shows that we can interpret SSI as the rate of growth of same-race capital. It follows from a familiar calculation in Perron-Froebenius theory. In economics the result is reminiscent of the balanced growth result in the theory of Leontief systems (see e.g. Dorfman, Samuelson and Solow (1958)).

Examples of this type of cultural transmission (or any diffusion process) include language (Lazear 1999) and the choice of first names (Fryer and Levitt 2004). In a simple model of culture and language, Lazear (1999) shows that incentives to assimilate by learning to speak the native language are decreasing in the size of an ethnic enclave. Fryer and Levitt (2004) argue that the choice of distinctive first names is a cultural expression, and show that this practice is more common in highly segregated areas. Both of these papers are consistent with the basic model of cultural transmission described above and, *ipso facto*, with our measure of segregation.

Another interesting feature of SSI is that it captures certain multiplier effects in a network; this feature is related to the dynamics of the model described in equation 3. An individual's susceptibility to own-race influences (patterns of speech, names, and other race specific behavior) depends on how many contacts the individual has with their own-race and the susceptibility of those contacts.

Consider the following thought experiment, depicted in Figure 4, which captures the essence of the multiplier effects. Network A in Figure 4 has 3 individuals of the same race and one individual (4) of a different race. To illustrate the multiplier effects captured in SSI, Network B changes the race of Individual 4 to be identical to the others.

	1	2	3	4	$\hat{S}$
Before	0.67	0.67	0.67	0	0.5
After	0.78	0.78	0.91	0.42	0.72

Table 3: SSI before and after the change

Table 3 shows the levels of segregation before and after Individual 4 changes race. Initially, the segregation of 1, 2, and 3 is unchanged. The segregation of 4 adjusts from 0 to 1.3. In a second period, 4's change induces a change in 3's segregation, but in no other individual. In a third period, the change reaches (by linearity) individuals 1 and 2. Now, the changes in 1 and 2 affect 3 again, and ultimately 4, and so on. Figure 5 demonstrates how the levels of segregation ultimately converge to the levels shown in Table 3, which captures the multiplier effects inherent in SSI.

G. Graph-theoretic properties of SSI

We provide two results that help interpret the SSI. The first relates SSI to how many neighbors individuals have. The second result shows how SSI measures the connectivity of the h-race network.

The degree of a vertex v, $d(v)$, is the number of edges at v. Let $d_{\min}(C) = \min\{d(v) | v \in V\}$ denote the minimum degree of G, $d_{\max}(C) = \max\{d(v) | v \in V\}$ represents its maximum degree, and $\overline{d}(C) := \frac{1}{|V|}\sum_{v\in V} d(v)$ the average degree of G. If all vertexes of G have the same degree, then G is *regular*. In this section we assume the neighborhood model ($r_{ij} \in \{0,1\}$).

A natural alternative to SSI is the average degree (or average number of own-race social interactions). We discard the average degree because it fails linearity. The following result provides a relationship between the measure of spectral segregation of a minority group and the degree of the minority graph associated with that group.

Proposition 4 *Let $d_{\min}$, $\overline{d}$ and $d_{\max}$ be the minimum, average, and maximum degrees of B^h, respectively. Then*

$$d_{\min} \leq \overline{d} \leq \hat{S}^h \leq d_{\max}$$

Let d_i be the number of same-race neighbors of household i. Proposition 4 proves that, Homogeneity notwithstanding, $\hat{S}^h(\beta)$ is larger than the average d_i over the individuals with $a(i) = h$. This relationship is strict when a city is irregular. The intuition is related to the feedback mechanism involved in the Linearity axiom. Recall, a household is more segregated the more segregated its neighbors are, and this cascades through the network similar to our thought experiment in the previous section. Proposition 4 indicates that the limit of this process is larger than the average degree of a minority graph.

SSI is a measure of the connectedness of race-h individuals in a network, an interpretation that was implicit in the discussion in Section E. Now we use walks in a graph to bring out the relation between SSI and connectivity. A *walk* of length k is a sequence of (not necessarily different) vertexes v_1, v_2, ..., v_k, v_{k+1} such that for each $i = 1, 2, ..., k$ there is an edge from v_i to v_{i+1}. A walk is closed if $v_{k+1} = v_1$. Let W_i^θ be the number of walks of length θ that individual $i \in h$ can take in B^h, and define $W^\theta = \sum_i W_i^\theta$. Let W_{ij}^θ be the number of walks of length θ between individual $i \in h$ and $j \in h$. A graph is bi-partite if its vertex-set admits a partition into 2 classes such that every edge has its ends in different classes. The graphs one encounters in applications of SSI are never bi-partite.

Proposition 5 *For θ sufficiently large: (1) $\frac{W_i^\theta}{(\hat{S}^h(\beta))^{\theta-1}}$ is approximately proportional to $\hat{s}_i^h(\beta)$, and the constant of proportionality is independent of i; (2) $\sqrt[\theta]{W^\theta / n^h(\beta)}$ approximates $\hat{S}^h(\beta)$; and (3) if B^h is non-bipartite, W_{ij}^θ is approximately proportional to $(\hat{S}^h(\beta))^{\theta-2}\hat{s}_i^h(\beta)\hat{s}_j^h(\beta)$.*

Proposition 5 (1) says that, as θ grows, $W_i^\theta(\hat{S}^h(\beta))^{\theta-1}$ converges. Thus $\hat{S}$ measures the growth in the number of walks that i can take. Further, it converges to something proportional to $\hat{s}_i$, thus individual SSI measures explain the differences, among individuals, in how many walks they can take relative to $\hat{S}$. Statement (2) in Proposition 5 says that $W^\theta \sim n^h(\beta)\left(\hat{S}^h(\beta)\right)^\theta$. The total number of walks will grow at rate $\hat{S}^h(\beta)$ (a statement which is similar, and has a similar proof, to that of Proposition 3). Finally, (3) says that two individuals' measures are related to how many walks there are between the two individuals, relative to the total number of walks (given by $\hat{S}^h(\beta)$, in light of Statement (2)).

5 Two Applications of SSI: Measuring Residential and School Segregation

Here we develop two illustrative applications of SSI: estimating residential segregation, and racial segregation of friendship networks in schools.

5.1 Residential Segregation

We begin by explaining how we take the model in Section 4 to data. The first step involves constructing the graph $G^h(\beta)$ of same-race interactions for each race and city that we analyze. As an example consider Figure 6, which depicts Asian blocks around Downtown Boston. A block is considered "Asian" if Asians represent the largest share of individuals in that block.

We assume that two blocks are neighbors if they are within one kilometer of each other. The second panel of Figure 6 illustrates circles of 1 kilometer radius around the centroid of each Asian block to depict the relevant neighborhood.[13] From this, we can construct the graph of Asians in Downtown Boston. The next step is to calculate the intensities of

[13] We have used one kilometer radii because one kilometer is the median radius of a census tract (1.03), and tracts are the traditional notion of neighborhood in the literature. Our results alter little when we change criterion to 0.5 or 1.5 kilometers.

neighborhood interactions; the r_{ij} of Section 4. We obtain the total number, d_i, of neighbors of block i, i.e. the number of blocks that are within one kilometer of i, and let r_{ij} be $1/d_i$. The justification for using $1/d_i$ is that we think of i as having a budget constraint for social interactions, and we assume that i spends an equal amount of that budget on all its neighbors. Of course, if one had more information about these interactions, one could do something better.

The graph, and the intensities, give us B^h. We then calculate the largest eigenvalue of (each irreducible submatrix of) B^h, and thus obtain the SSI. [14]

A. The Data

The ideal data to estimate residential segregation would contain information on the nature of interactions with other households. This information is, of course, not available. In lieu of this, we use block-level data from the 2000 US Census, restricting our sample to all Metropolitan Statistical Areas (MSAs).[15] Census blocks contain, on average, 300 households, and are approximately 100 meters in radius. We identify a block with the race/ethnicity of the majority of its inhabitants. This assumption is not too problematic, as blocks are strikingly homogeneous: 94.3% of Iowans live in a homogeneous census block and so do 77% of Texans. Save Washington DC, more than 60% of the blocks in all states contain households of only one race (for half the states, 80% or more of the blocks contain only one race).

[14] The Matlab programs to calculate all indexes reported in the paper are available at http://post.economics.harvard.edu/faculty/fryer/fryer.html

[15] We have a proposal being reviewed at the US Census Bureau that would provide household level data for 25 states; those in which it has been collected.

One issue with our application of SSI to residential segregation is that it ignores block density.[16] To correct for this, one could assign all individuals in a census block to the centroid of that block, and run the resulting individual-level estimation. This method, however, is computationally very costly. Measuring the segregation of Chicago amounts to solving the eigenvalues of a matrix whose dimensions are well over 1 million $\times$ 1 million.[17]

C. Baseline Residential Segregation

Since SSI for race h is a measure of the connectivity of the race-h network (or of the growth of race-h social capital) it will tend to be larger in cities with larger fractions of race-h individuals, even if individuals located at random in the city.

We refer to the SSI one would expect to see in a city when individuals locate at random as *Baseline SSI*. We provide estimates of both SSI, and of the SSI in excess of Baseline SSI.

We have obtained measures of Baseline SSI by simulating random assignment of races to large regular (in a graph-theoretic sense) cities with the corresponding fraction of race-h inhabitants. Figure 7 shows the results of our simulations. On the horizontal axis is the fraction of race-h inhabitants, while the vertical axis shows the average SSI. When the share of race-h inhabitants in a city is relatively small, SSI mirrors the percent race-h in a city closely. This is to be expected. When race-h inhabitants are relatively few and assigned to a city graph at random, linearity has little power to alter SSI from percent black.

[16] This likely induces little error in the estimates of segregation, given our definition of neighbor usually encompasses several blocks. In areas such as New York, however, this limitation may be quite restrictive.

[17] Typical cities such as Austin, Texas and Tampa, Florida currently take one day to compute. Imputing all individuals to the centroid of their block in these cities would increase computing time to approximately 1 year.

As the fraction of race-h individuals increases, however, SSI significantly departs from the percentage of race-h in a city. We have used only large cities, as we can prove (See Appendix B) that baseline SSI converges as a city grows. In fact the simulations show the convergence to be quite fast.

D. The Extent of Segregation Across Cities

Table 4 reports the top 10 most segregated cities for Whites, Blacks, Hispanics, and Asians (including pacific-islanders). Detroit is the most segregated city for Blacks; Lowell, MA for whites; McAllen, TX for Hispanics and Honolulu, HI for Asians . The list seems quite intuitive. It also confirms the statement in Section C: SSI is clearly correlated with the size of a minority group. The latter point begs for a distinction between "raw" segregation, as measured by SSI, and "behavioral" segregation: the segregation in excess of baseline segregation. It is unclear which is most closely related to economic outcomes. Behavioral segregation tells us more about preferences, while the original SSI is a better measure of the growth of same-race capital. We are agnostic; they simply measure different things.

Table 5 reports the top 10 most behaviorally segregated cities. The rankings change dramatically. Detroit, the most segregated city for Blacks in Table 1, is not even in the top 10. The most segregated cities after taking out the baseline for Asians, Blacks, Hispanics, and Whites are: Los Angeles, CA; Milwaukee, WI; Flagstaff, AZ; and Pine Bluff, AR, respectively. Approximately 11% of households in Milwaukee are black, implying a SSI of .1145. The actual measure of segregation is a factor of 9 larger. Indeed, to generate the level of SSI in Milwaukee, assuming blocks were assigned a race at random, Blacks need to comprise 80% of the population. It is also interesting to note that behavioral segregation is larger for blacks than for other minorities. That is, the difference between the actual percentage, and the one needed to generate the same SSI at random, is largest for blacks.

We have emphasized how the SSI allows one to consider more disaggregated units than the city. One of the most interesting units is the agglomeration of same-race blocks: racially homogenous ghettos, which SSI identifies endogenously as connected components (see Section 4). This is related to city-wide SSI, but SSI weights the ghetto's SSI against members of the same race in other parts of the city, who are more integrated. For Blacks and Whites, the largest ghetto is Detroit – implying an enormous amount of city-wide segregation. Remarkably, 87% of black blocks in Detroit comprise one large ghetto. The largest connected component is San Francisco for Asians, and Los Angeles for Hispanics. Hispanics in Los Angeles comprise the largest minority ghetto in America; 17,909 Hispanic blocks are connected.

Table 6 provides more facts that are only obtainable with SSI; the top 15 most segregated blocks in America. A Hispanic block in San Antonio is the most segregated block, followed by a Black block in Lafayette, LA and an Asian block in Los Angeles.[18]

Table 7 presents a correlation matrix of popular measures of segregation. These measures include dissimilarity, isolation, Gini coefficient, exposure, entropy, and interaction. Also included in the matrix are SSI, SSI minus the baseline, and the ranking of cities based solely on the their fraction of Blacks. All measures were calculated using data at the census block

[18]The table reports the most segregated block per unique PMSA. Because of Linearity, the (literal) top 15 most segregated Blocks are all in San Antonio. To avoid this redundancy, we report blocks from unique PMSAs.

level for 326 MSAs. The Spectral index has surprisingly little correlation with dissimilarity, gini, entropy, and interaction – averaging less than .5 – and high correlation with isolation and exposure; averaging more than .90. Given the nature of the isolation and exposure indexes, it is not surprising that SSI is more correlated with the measures relative to the others. However, we did not expect there to be such a striking similarity between them. Thus, as a measure of residential segregation, our measure is very similar to existing measures of exposure with the added ability to disaggregate to the level of individuals, and a well-understood theoretical foundation. Taking out the relevant baseline, SSI becomes even less correlated with dissimilarity and isolation. The fraction black in a city is highly correlated with SSI, but linearity assures that this correlation is less than perfect.

E. The Impact of Residential Segregation on Economic Outcomes

The economic literature on the effects of segregation on outcomes is impressive. Case and Katz (1991) show that youths in a central city are affected by the characteristics of their neighbors. Almond, Chay, and Greenstone (2003) show that segregation of hospitals in the Jim Crow era had a significant negative effect on infant mortality. Using evidence from the Moving to Opportunity experiment, Katz, Kling, and Liebman (2001) and Kling, Liebman, and Katz (2005) provide evidence that moving individuals to lower poverty neighborhoods has substantial effects on mental and physical health of parents and children.

Cutler and Glaeser (1997) is one of the most influential papers in economics on the impact of segregation. They use the dissimilarity index as a measure of segregation. We re-estimate the impact of black segregation on economic outcomes with Cutler and Glaeser's specification. Econometrically, we estimate models of the form:

$$outcome_i = X'\beta + \beta_1 segregation_j + \beta_2 segregation_j * black_i + \varepsilon_i, \tag{4}$$

where $outcome_i$ is measured at the individual level and $segregation_j$ is measured at the MSA level, and compare the results obtained with SSI and the dissimilarity index.

Identical to Cutler and Glaeser (1997), we correlate measures of segregation with various economic and social outcomes for young people aged 20-30. We choose to focus on younger individuals for three reasons. First, they are most susceptible to group level influences as a result of social interactions. Second, the problems of mobility across metropolitan areas is more easily avoided. Third, and most importantly, it mirrors the specifications in Cutler and Glaeser (1997). Data from the 5% Census Public Micro Use Sample are used.

Outcome measures are divided into 3 categories: educational attainment, labor market, and social outcomes. Educational attainment is measured as the probability an individual graduates from high school or college. There are two measures of labor market outcomes. The first is whether or not an individual is idle (not working and not employed). The second is earning (sum of wages, salary, and self-employment income). In all specifications, we use the natural logarithm of earnings, conditional on the individual not being in school and reporting positive earnings. The final outcome variable is a social outcome – whether a woman is an unmarried mother. While these may seem like a motley set of outcomes to include, they are identical to those analyzed in Cutler and Glaeser (1997) and serve as proxies for a broad set of social and economic variables.

Tables 8 presents a series of ordinary least squares estimates of the relationship between segregation and outcomes for persons aged 20-24 and 25-30, using three different measures of segregation. All measures of segregation have been normalized such that they have a mean of zero and a standard deviation of one. Along with a mutually exclusive and collectively exhaustive set of racial dummies, we include controls for gender, age dummies in one year increments, logarithm of MSA population, percent black, median household income and percent of the labor force employed in manufacturing. Each of these variables are also interacted with a dummy variable for black. Standard errors are clustered at the MSA level.

The top panel of Table 8 replicates Cutler and Glaeser's (1997) results using the dissimilarity index. The bottom two panels estimate the same specification using SSI and the Isolation index. Results differ sharply between SSI and dissimilarity, which is not surprising given the low correlation between the two indexes reported in Section D. On each outcome, cities with higher dissimilarity indexes have inferior outcomes: less likely to graduate from high school or college, more likely to be idle, earn less money, and more likely to be a single mother. SSI paints a different portrait. On every outcome-measure, the effect of segregation on black outcomes decreases markedly; only idleness and single motherhood are statistically distinguishable from zero. The bottom panel presents results using the isolation index. As one might expect, the estimates using the isolation index lie in between the dissimilarity and SSI extremes, though the results are more in line with the SSI.

It is quite surprising that estimates of the impact of segregation on our set of outcomes differ so substantially when one employs the Spectral index in lieu of the dissimilarity index. Thinking carefully about these differences and related models of social interaction that can inform us about why one type of segregation is worse than another is an important line of inquiry, which we leave for future work.

5.2 School Segregation

There is a growing literature on the effects of school segregation on achievement. Guryan (2004) estimates that half of the decline in black dropout rates between 1970 and 1980 is attributable to desegregation plans. Crain and Strauss (1985) find that students randomly offered the chance to be bussed to a suburban school were more likely to work in professional jobs nearly 20 years after the experiment. Meta-analytic studies, however, typically find no consensus reached about the effect of school integration on black student achievement (Armor 2002, Cook 1984, Crain and Mahard 1981, St. John 1975).

While the literature on the relationship between school segregation and achievement has amassed an impressive array of facts, the vast majority of it has focused on one aspect of the problem – segregation across schools.[19] This is very much in the legacy of the *Brown* Decision; ensuring the placement of black kids in traditionally white school so they can get the benefit of better resources, more qualified teachers, and so on. Traditional measures of segregation calculate the fraction of whites at the typical black student's school (entropy) or the percentage of blacks that would have to change schools to get an equal distribution

[19] Clotfelter, Ladd, and Vigdor (2003), which calculates dissimilarity at the classroom level, is a nice counter-example.

across a district (dissimilarity). Neither of these indexes are equipped to measure the nature of interactions *within* schools.

Estimates of within-school interactions are vitally important for a myriad of educational policies: ranging from the efficacy of bussing programs to the design of optimal class composition. It is one thing to have black kids present in the majority of white schools; it is something quite different to have black students be apart of the social networks of the typical white kid.

A. The Data

The National Longitudinal Study of Adolescent Health (Addhealth) database is a nationally representative sample of 90,118 students entering grades 7 through 12 in the 1994-1995 school year. A stratified random sample of 20,745 students was given an additional in-home interview; 17,700 parents of these children were also interviewed. Thus far, information has been collected on these students at 3 separate points in time: 1995, 1996, and 2002. There are 175 schools in 80 communities included in the sample, with an average of more than 490 students per school, allowing within school analysis. Students who are missing data on race or grade level are dropped from the sample.

A wide range of data are gathered on the students, as described in detail on the Addhealth website (http://www.cpc.unc.edu/projects/addhealth). For our purposes, the most interesting and unique aspect of the Addhealth data is the detailed information regarding friendship associations in schools. All students contained in the in-school survey were asked, "List your closest male/female friends. List your best male/female friend first, then your next best friend, and so on." Students were allowed to list as many as 5 friends from each sex. Each friend can be linked in the data and the full range of covariates in the in-school survey (race, gender, grade point average, etc) can be gleaned from each friend.

B. The Nature of School Segregation

The empirical framework we implement here is similar to that employed in Section 5.1. First, we calculate the spectral index for each school, and for each individual within a school. The theory requires that all interactions be bilateral. Empirically, however, a non-trivial portion of students listed as friends individuals who did not list them. There are two potential ways to handle this: we can include an individual i as a friend if either i or j list them as a friend; or we can opt for a more restrictive definition – including i only if both i and j list them as a friend. We opt for the former – less restrictive – definition.

Figures 8 depicts the relationship between the percentage of minority students in a school and the level of segregation for each minority student in that school, using the Addhealth database. Each observation is a school. Grade levels 7-12 are combined.

Many researchers assume the relationship is linear (see, for example, Orfield 1983). This seems to be true for Whites, Asians, and to a lesser extent Hispanics. For Blacks, however, the relationship between percent own-race in a school and own-race segregation is highly non-linear. As the percentage of black students increases from zero to twenty-five percent, black segregation rises sharply. Above twenty-five percent, Blacks are near complete segregation. This has important implications for the bussing programs and the optimal organization of schools, among other things.

B. The Impact of School Segregation on Academic Achievement

Table 9 presents the first estimates of individual level measures of segregation on individual outcomes. We estimate models of the form:

$$\begin{aligned} outcome_i &= X\beta + \gamma segregation_i + \xi_1 black \cdot segregation_i + \xi_2 asian \cdot segregation_i \\ &\quad + \xi_3 hispanic \cdot segregation_i + \varepsilon_i, \end{aligned}$$

including school-level fixed effects.

We examine eight outcome variables: 4 social and 4 academic. The social variables include smoking, skipping school (without a valid excuse), interracial dating, and whether or not a student is happy at their school. Smoking and skipping school are answers to the question, "During the past 12 months, how often did you..." Answer choices range from never to nearly everyday. Interracial dating is a dichotomous variable equal to 1 if the student reports ever dating interracially and zero otherwise. Happiness measures whether or no students report being happy at their school. The academic variables include: Peabody Vocabulary Test (PVT) scores, whether or not a student plans to attend college, grades in the previous grading period, and a measure of how much effort the student exerts. All responses (including grades) are self-reported.

For blacks, individuals who are more segregated are less likely to smoke (a behavior predominant among white teens) and have lower test scores. Asian students who are more connected to each other are less likely to skip school, have lower test scores, put in more effort, and report being happier. Among Hispanics, more segregation is associated with less smoking, lower test scores, lower probability of attending college, and lower grades. Interestingly, students of all races are less likely to date interracially when schools are more segregated. Similar results are obtained when one excludes school fixed-effects.

6 Conclusion

For decades, social scientists have used measures of evenness and exposure to estimate the prevalence and impact of segregation in housing, firms, and schools. These measures have many limitations, which we have discussed throughout. This paper develops a new measure of segregation based on two key ideas: a measure of segregation should disaggregate to the level of individuals, and an individual is more segregated the more segregated are the agents with whom they interact. Developing three axioms that any segregation measure should satisfy, our main result shows that one and only one segregation index satisfies our three axioms, and the two aims mentioned above—the Spectral Segregation Index. To illustrate the potential of the index, it is applied to two well-known social problems: measuring residential and school segregation and several new insights are gleaned. We hope the Spectral index will be a useful tool for applied researchers interested in the agglomeration of individuals in networks.

References

[1] Almond, Douglas, Kenneth Chay, and Michael Greenstone. 2003. "Civil Rights, the War on Poverty, and Black-White Convergence in Infant Mortality in Mississippi" mimeo. MIT Department of Economics.

[2] Armor, David J. 2002. "Desegregation and Academic Achievement," in School Desegregation in the 21st Century, Christine H. Rossell, David J. Armor, and Herbert J. Walberg, Eds. (Westport, CT) pp. 148-187.

[3] Bell, W. 1954. "A Probability Model for the Measurement of Ecological Segregation." *Social Forces*, 32, 357-364.

[4] Blau, Peter. 1977. *Inequality and Heterogeneity: A Primitive Theory of Social Structure.* New York: Free Press.

[5] Borjas, George. 1995. "Ethnicity, Neighborhoods, and Human-Capital Externalities," *American Economic Review*, pp. 365-390.

[6] Brueckner, J. and Smirnov, O. 2004. "Workings of the Melting Pot: Social Networks and the Evolution of Population Attributes." mimeo. The University of Illinois at Urbana Champaign.

[7] Case, Anne, and Lawrence Katz. 1991. "The Company You Keep: The Effects of Family and Neighborhood on Disadvantaged Youths." *NBER Working Paper No. 3705.*

[8] Clotfelter, Charles T., Helen F. Ladd and Jacob L. Vigdor. 2003. "Racial Segregation in Modern Day Public Schools." mimeo. Duke University.

[9] Collins, Chiquita and David R. Williams. 1999. "Segregation and Mortality: The Deadly Effects of Racism?" *Sociological Forum* 14:495-523.

[10] Cook, Thomas. 1984. "What Have Black Children Gained Academically from School Integration? Examination of Meta-Analytic Evidence," National Institute of Education (Washington, DC).

[11] Cowgill, Donald. and Mary Cowgill. 1951. "An Index of Segregation Based on Block Statistics." *American Sociological Review*. 16, 825-831.

[12] Crain, Robert and Jack Strauss. 1985. "School Desegregation and Black Occupational Attainments: Results from a Long-Term Experiment," Center for Social Organization of Schools, Johns Hopkins University.

[13] Crain, Robert and Rita Mahard. 1981. "Minority Achievement: Policy Implications of Research," in *Effective School Desegregation: Equity, Quality, and Feasibility*, Willis D. Hawley Ed. Sage Publications (California: Beverly Hills) pp. 55-84

[14] Cutler, D. and Glaeser, E. 1997. "Are Ghettos Good or Bad?" *Quarterly Journal of Economics*, 112, 827-872.

[15] Cutler, D., Glaeser, E., and Vigdor, J. 1999. "The Rise and Decline of the American Ghetto." *Journal of Political Economy*, 107 (3), 455-506.

[16] Cvetković, D. and Rowlinson, P. 1990. "The Largest Eigenvalue of a Graph: A Survey." *Linear and Multilinear Algebra*, Volume 28, pp. 3-33.

[17] Cvetković, D., Rowlinson, P., and Simić, S. 1997. *Eigenspaces of Graphs.* Cambridge University Press. Cambridge:UK.

[18] Dorfman, R. Samuelson, P. and Solow, R.. 1958. *Linear Programming and Economic Analysis*, McGraw-Hill.

[19] Diestel, R. 1997. *Graph Theory.* Springer Verlag, Ney York: NY.

[20] Frankel, David and Oscar Volij. 2004. "Measuring Segregation." mimeo. Iowa State University.

[21] Fryer, R. and Levitt, S. 2004. "The Causes and Consequences of Distinctively Black Names," *Quarterly Journal of Economics*, Vol. CXIX, (3) 767-805.

[22] Guryan, J. 2004. Desegregation and Black Drop-Out Rates. *American Economic Review.*

[23] Hutchens, Robert. 2001. "Numerical Measures of Segregation: Desirable Properties and Their Implications," *Mathematical Social Sciences*, 13-29

[24] Jahn, Julius A., Calvin F. Schmid, and Clarence Schrag. 1947. "The Measurement of Ecological Segregation" *American Sociological Review* 103: 293-303.

[25] Kain, John, 1968. "Housing Segregation, Negro Employment, and Metroplitan Decentralization," *Quarterly Journal of Economics*, LXXXII, 175-197.

[26] Katz, Lawrence, Jeffrey R. Kling, and Jeffrey B. Liebman. 2001. "Moving to Opportunity in Boston: Early Results of a Randomized Mobility Experiment." *Quarterly Journal of Economics* 116, 607-54

[27] Kling, Jeffrey R., Jeffrey B. Liebman, and Lawrence Katz. 2005. "Experimental Analysis of Neighborhood Effects." Working Paper. Princeton University.

[28] Lazear, Edward. 1999. "Culture and Language," *Journal of Political Economy*, S6 pp. S95-S129.

[29] Massey, D. and Denton, N. 1988. "The Dimensions of Residential Segregation." *Social Forces*, 67:281-315

[30] Massey, D. and Denton, N. 1993. *American Apartheid: Segregation and the Making of the Underclass.* Cambridge, MA: Harvard University Press.

[31] Palacios-Huerta, I and Volij, O "The Measurement of Intellectual Influence," *Econometrica.* v72(3), pp. 963–977.

[32] Philipson, Tomas 1993. "Social Welfare and Measurement of Segregation." *Journal of Economic Theory*, 60(2), pp. 322-334

[33] Orfield, Gary. 1983. *Public School Desegregation in the United States, 1968-1980.* Washington, DC: Joint Center for Political Studies.

[34] Reardon, S. and Firebaugh, D. "Measures of Multigroup Segregation." *Sociological Methodology* 32:33-67.

[35] St. John, Nancy H. 1975. *School Desegregation Outcomes for Children*, John Wiley and Sons (NY: New York).

[36] Tauber, Karl. and Alma Tauber. 1965. *Negroes in Cities: Residential Segregation and Neighborhood Change*. Aldine.

7 Appendix A: Proofs

Proof of Theorem 1.

The proof of Theorem 1 proceeds by stating and proving 5 lemmas that together establish the theorem.

The first lemma unifies some results about irreducible matrices. Mainly the lemma restates the Perron-Froebenius Theorem.

Lemma 2 *Let C be a real, non-negative, irreducible matrix. Then A has a real, positive, eigenvalue λ with associated eigenvector y. Such that*

1. *y is strictly positive, so $y_i > 0$ for all i, and y is the unique, up to a scalar multiple, strictly positive eigenvector of C;*
2. *λ is larger than $|\sigma|$, for any other eigenvalue σ of C; in particular, λ is larger than any other real eigenvalue.*

Proof. By the Perron-Froebenius Theorem (Theorem 8.4.4 in Horn and Johnson), C has a real, strictly positive, eigenvalue, λ, with associated strictly positive eigenvector y. The multiplicity of λ is one and λ is larger than $|\sigma|$, for any other eigenvalue σ of C (λ is the spectral radius of C).

Let z be any strictly positive eigenvector, by Corollary 8.1.30 in Horn and Johnson, z is associated to eigenvalue λ. The z is a scalar multiple of y, as λ has multiplicity one. ■

Now we verify that the spectral segregation index satisfies our three axioms.

Lemma 3 *The Spectral Segregation Index satisfies Montonicity.*

Proof. Let $\beta' = (a', R')$ have more h-segregation than $\beta = (a, R)$. Let $B^{h\prime}$ denote the submatrix of $B(\beta')$ whose rows i have $h = a'(i)$. Let B^h denote the corresponding submatrix of $B(\beta)$. That β' has more h-segregation than β implies that, if i is a row in B^h, then i is a row in $B^{h\prime}$.

Let $C' = (c'_{ij})$ be an irreducible submatrix of $B^{h\prime}$. Then the set of rows in C' is the union of the rows in some collection $C_1, C_2, \ldots C_L$ of irreducible submatrices of B^h. Let $C = (c_{ij})$ be the block-diagonal matrix with $C_1, C_2, \ldots C_L$ in its diagonal. Let x' be an eigenvector associated to the largest eigenvalue λ' of C'. Then $C'x' = \lambda' x'$, $x_i > 0$ for all i (Lemma 2), and β' having more h-segregation than β impliy that

$$\lambda' = \frac{1}{x'_i} \sum_{j \in C'} c'_{ij} x'_j \geq \frac{1}{x'_i} \sum_{j \in C'} c_{ij} x'_j \tag{5}$$

Let $\lambda = \max \{|\sigma| : \sigma \text{ is an eigenvalue of } C\}$ be the spectral radius of C. Then, by Horn and Johnson's Theorem 8.1.26,

$$\lambda \leq \max_{i \in C} \frac{1}{x'_i} \sum_{j \in C} c_{ij} x'_j. \tag{6}$$

Statements (5) and (6) imply that $\lambda \leq \lambda'$. But λ' is $S^{C'}(\beta')$ (Proposition 2); so $\lambda \leq \hat{S}^{C'}(\beta')$.

Now we prove that $\hat{S}^{C_l}(\beta) \leq \lambda$, for $l = 1 \ldots L$. Let λ_l be the largest real eigenvalue of C_l. Let x_l be an eigenvector of C_l, associated to λ_l; Let $y = (y_i)_{i \in C}$ be the vector obtained from x_l by letting $y_i = x_{li}$ if $i \in C_l$ and 0 otherwise. Then, since C is block-diagonal, λ_l is an eigenvalue of C, with associated eigenvector y. By definition of λ, since λ_l is real, $\lambda_l \leq \lambda$. But Proposition 2 implies that $\lambda_l = \hat{S}^{C_l}(\beta)$, so $\hat{S}^{C_l}(\beta) \leq \lambda$, for $l = 1 \ldots L$.

Let C'_k, $k = 1, \ldots K$ be the irreducible submatrices of $B^{h\prime}$, and let each C'_k be the union of L_k irreducible submatrices of B^h, C'_{kl} with $l = 1, \ldots L_k$. By Proposition 2

$$
\begin{aligned}
\hat{S}^h(\beta) &= \sum_{k=1}^{K} \sum_{l=1}^{L_k} \frac{|C_k|}{n^h(\beta)} \hat{S}^{C_{kl}}(\beta) \\
&\leq \sum_{k=1}^{K} \hat{S}^{C'_k}(\beta') \sum_{l=1}^{L_k} \frac{|C_k|}{n^h(\beta)} \\
&= \sum_{k=1}^{K} \hat{S}^{C'_k}(\beta') \frac{|C_k|}{n^h(\beta')} = \hat{S}^h(\beta')
\end{aligned}
$$

■

Lemma 4 *The Spectral Segregation Index satisfies homogeneity.*

Proof. Let $a \in A$ be h-homogeneous of degree d. Let $y = \mathbf{1}$, then homogeneity says that $Ay = d\mathbf{1}$, so d is an eigenvalue with eigenvector y. By Lemma 2 d must coincide with λ, the largest eigenvalue of B, and the rescaled eigenvector must coincide with x. So $\hat{S}^h(\beta) = d$. ■

Lemma 5 *The Spectral Segregation Index satisfies linearity.*

Proof. By Proposition 2, $\hat{S}^{C_k}(\beta)$ is an eigenvalue with eigenvector (x_i), the eigenvector in the definition of the spectral index. The, for any i, $s_i(\beta) = S^{C_k}(\beta)x_i\,|C_k| = |C_k|\,(C_k \cdot x|_i)$.

So

$$\begin{aligned} s_i(\beta) &= \sum_{j\in C_k} |C_k|\, r_{ij} x_j \\ &= \frac{1}{\lambda_k} \sum_{j\in C_k} |C_k|\, r_{ij} x_j \lambda_k \\ &= \frac{1}{S^{C_k}(\beta)} \sum_{j\in N_i^a} s_j(\beta) \end{aligned}$$

■

Second, we prove that any index that satisfies the three axioms must coincide with the spectral index. Let $\left(S^h(\beta), (s_i(\beta))_{i\in h}\right)$ be a segregation index that satisfies the three axioms.

Lemma 6 *If $\beta = (a, R)$ is such that $b_{ij} = 0$ for all i and j with $h = a(i) = a(j)$, then $s_i(\beta) = \hat{s}_i(\beta)$ for all i with $h = a(i)$.*

Proof. By Homogeneity, $S^h(\beta) = 0$, so we must have and $s_i(\beta) = 0$ for all i with $h = a(i)$, as $s_i(\beta) \geq 0$ and $S^h(\beta)$ is the average $s_i(\beta)$ over i with $h = a(i)$. Thus the index coincides with the Spectral Segregation Index. ■

Lemma 7 *For any assignment-interaction pair β, $s_i(\beta) = \hat{s}_i(\beta)$ for all i.*

Proof. Let β be an assignment-interaction pair. Fix h and let $B^h(\beta) = (b_{ij})$ If β is such that $b_{ij} = 0$ for all i and j, we are done by Lemma 6. Suppose that $b_{ij} > 0$ for at least one i and j.

Let $\gamma = \min\{b_{ij} : b_{ij} > 0\}$. Let $D = (d_{ij})$ be the matrix defined by $d_{ij} = 0$ if $b_{ij} = 0$, and

$$d_{ij} = \frac{\gamma}{|\{j : b_{ij} > 0\}|}$$

if $b_{ij} > 0$. Let $\beta^0 = (a, R^0)$ be the assignment-interaction pair that differs from β in that submatrix for the i with $h = a(i)$ is D. Then, for all i with $h = a(i)$, $\sum_{i \in h} r^1_{ij} = \gamma$. So β^0 is h-homogeneous of degree γ. By homogeneity, $S^h(\beta^0) = \gamma$.

Now, for all i with $h = a(i)$, and j, $r_{ij} \leq b_{ij}$. So β has more h-segregation than β^0. And $n^h(\beta^0) = |a^{-1}(h)| = n^h(\beta)$. Monotonicity implies then $S^h(\beta) \geq S^h(\beta^0) = \gamma$. So $S^h(\beta) > 0$.

Fix a component C_k such that $S^{C_k}(\beta) > 0$; since $S^h(\beta) > 0$ there must exist at least one. For $i \in C_k$, let $x_i = \frac{s^h_i(\beta)}{|C_k| S^h(\beta)}$. Note that, by definition of $S^{C_k}(\beta)x_i$, $\sum_{i \in C_k} x_i = 1$.

Then $S^{C_k}(\beta)x_i = s_i(\beta)/|C_k| = \frac{1}{|C_k|}\sum_{j \in N^a_i} r_{ij}s_j/S^{C_k}(\beta)$, by Linearity. Then $S^{C_k}(\beta)x_i = \sum_{j \in N^a_i} r_{ij}x_j$. So $S^{C_k}(\beta)x = C_k x$; $S^{C_k}(\beta)$ is an eigenvalue of C_k with eigenvector x.

Now, $s_i(\beta) > 0$ for all i. Since $s_i(\beta) = 0$ for some i would imply, by Linearity, that all $j \in N^a_i$ have $s_j(\beta) = 0$. Then, by recursion, $s_j(\beta) = 0$ for all $j \in C_k$, which would contradict that $S^{C_k}(\beta) > 0$. Hence x is a strictly positive eigenvector.

By Proposition 2 and Lemma 2 now $S^{C_k}(\beta) = \hat{S}^{C_k}(\beta)$, and by the rescaling $\sum_{i \in C_k} x_i = 1$, x must coincide with the defining eigenvector in the definition of the spectral segregation index. Then, $s_i(\beta) = \hat{s}_i(\beta)$ for all i.

Finally, take a component with $S^{C_k}(\beta) = 0$. Then Monotonicity and Lemma 6 imply that $b_{ij} = 0$ for all i and j in C_k. ■

Lemmas (3) through (7), taken together, establish the theorem.

Proof of Proposition 1 If $i \in h$ has at least one same-race neighbor, then i is in C_k, for some irreducible submatrix C_k. Let λ_k be the largest eigenvalue of C_k, and x_k be its associated eigenvector. By Lemma 2, x_k is strictly positive, so $x_{ki} > 0$. Since $\lambda_k > 0$ (Lemma 2), the definition of $\hat{s}_i^h(\beta)$ implies that $\hat{s}_i^h(\beta) > 0$. ■

Proof of Proposition 2

We show that $S^{C_k}(\beta)$ is the largest eigenvalue of C_k. $S^{C_k}(\beta) = \sum_{i \in C_k} s_i(\beta) / |C_k| = \lambda_k \sum_{i \in C_k} x_i$. Since x was normalized so that $\sum_{i \in C_k} x_i = 1$, it follows that $S^{C_k}(\beta) = \lambda_k$. That $S^h(\beta)$ is the weighted average of the $S^{C_k}(\beta)$ follows immediately by definition of $S^h(\beta)$ and $S^{C_k}(\beta)$. ■

Proof of Proposition 3 Let I denote the $n^h(\beta) \times n^h(\beta)$ identity matrix. Let $D = I + B^h(\beta)$. Then equation 3 implies that the vector $x_t = (x_{it})_i$ satisfies $x_t = Dx_{t-1}$, for all t. So $x_t = D^t x_0$. By Lemma 8.4.2 in Horn and Johnson, $1 + \hat{S}^h(a)$ is the largest eigenvalue of D. By Lemma 8.2.7 in Horn and Johnson, there is a matrix L such that

$$\lim_{t \to \infty} (1 + \hat{S}^h(a))^{-t} D^t = L$$

Then,

$$\frac{x_{it}}{x_{it-1}} = (1 + \hat{S}^h(a)) \frac{((1 + \hat{S}^h(a))^{-t} D^t x_0)_i}{((1 + \hat{S}^h(a))^{-t+1} D^{t-1} x_0)_i} \to (1 + \hat{S}^h(a))$$

■

Proof of Proposition 4 See Cvetkovic and Rowlinson (1990).■

Proof of Proposition 5 Let $U = (u_i)$ be the eigenvectors of B^h, normalized to form an orthonormal basis, so $U^T U = I$. Let D be the matrix with the eigenvalues of B^h on the diagonal, and 0 everywhere else. So $A = UDU^T$.

If 1 is the vector with 1 in all its entries, the vector of θ-long walks (W_i^θ) is defined by $(W_i^\theta) = A^\theta 1$. So $(W_i^\theta) = UD^\theta U^T 1$. The (u_i) vectors form a basis, so there are scalars (ξ_i) such that $1 = \sum_i \xi_i u_i$.

Then $(W_i^\theta) = \sum_i \xi_i UD^\theta U^T u_i$. But $U^T u_i = e_i$, the vector with i-th entry 1, and 0 elsewhere. So $(W_i^\theta) = \sum_i \xi_i \lambda_i^\theta U e_i = \sum_i \xi_{i\lambda i}^\theta u_i$. Let $\lambda_1 = S^h$; λ_1 has multiplicity 1, as B^h has a unique non-trivial eigenvector (Theorem 2.1.3 in Cvetkovic, Rowlinson and Simic). So $S^h(\beta) > \lambda_i$, $i = 2, 3 \ldots |h|$.

Then

$$\frac{1}{(S^h(\beta))^{\theta-1}}(W_i^\theta) = S^h(\beta) \sum_i \xi_i \frac{\lambda_i^\theta}{\lambda_1^\theta} u_i \tag{7}$$

$$\rightarrow S^h(\beta)\xi_1 u_i, \tag{8}$$

as $\lambda_i^\theta / \lambda_1^\theta \rightarrow 0$ for all $i \neq 1$. Since u_1 is a scalar multiple of the (x_i) vector in the definition of the spectral index, $S^h(\beta)\xi_1 u_1$ is a scalar multiple of s_i^h.

The second statement is a theorem of Cvetkovic, stated in the survey by Cvetkovic and Rowlinson (1990). The third statement is essentially Theorem 2.2.5 in Cvetkovic, Rowlinson and Simic. ■

8 Appendix B: Baseline Segregation

Here we present a theoretical justification for our "baseline" simulations. SSI converges as a city's size grows, so we can estimate SSI for relatively large cities (the size of 6400 is enough in our simulations).

Let $H = \{0, 1\}$ be the set of races. We are interested in only one race here, so working with $H = \{0, 1\}$ is without loss of generality. Let V_n be set of households, such that if $n \leq m$ then $V_n \subseteq V_m$.

Let $\Omega_n = H^{V_n}$ be the set of possible assignments of households to races. Endow the power set of Ω_n with the probability measure p_k obtained by letting each household be race 1 with probability $\pi \in (0, 1)$, independently of the races of other households.

Let

$$E_n S = \sum_{\omega \in \Omega_n} S(\omega) p_n(\omega)$$

be the expected value of the SSI.

Proposition 6 *There is $\overline{S}$ such that $E_n \uparrow \overline{S}$ as $n \to \infty$.*

Proof. We shall prove that, if $n \leq m$, then

$$\sum_{\omega \in \Omega_n} S(\omega) p_n(\omega) \leq \sum_{\omega \in \Omega_m} S(\omega) p_m(\omega).$$

Since the $E_n S$ are bounded above by 1, the result follows.

Let $q_{n,m}$ be the probability distribution on $H^{V_m \setminus V_n}$ induced by letting each household be race 1 with probability $\pi \in (0,1)$, independently of the races of other households. Abusing notation, we shall use $q_{n,m}$ for the probability distribution induced by $q_{n,m}$ on $\left\{\omega \in \Omega_m : \omega|_{V_n} = \{0\}^{V_n}\right\}$. Then,

$$\begin{aligned}
\sum_{\omega\in\Omega_m} S(\omega)p_m(\omega) &= \sum_{\omega'\in\Omega_n} p_n(\omega')\left[\sum_{\{\omega\in\Omega_m:\omega|_{V_n}=\omega'\}} q_{n,m}(\omega-\omega')S(\omega)\right] \\
&\geq \sum_{\omega'\in\Omega_n} p_n(\omega')\left[\sum_{\{\omega\in\Omega_m:\omega|_{V_n}=\omega'\}} q_{n,m}(\omega-\omega')S(\omega')\right] \\
&= \sum_{\omega'\in\Omega_n} p_n(\omega')S(\omega') \sum_{\{\omega\in\Omega_m:\omega|_{V_n}=\omega'\}} q_{n,m}(\omega-\omega') \\
&= \sum_{\omega'\in\Omega_n} p_n(\omega')S(\omega')
\end{aligned}$$

■

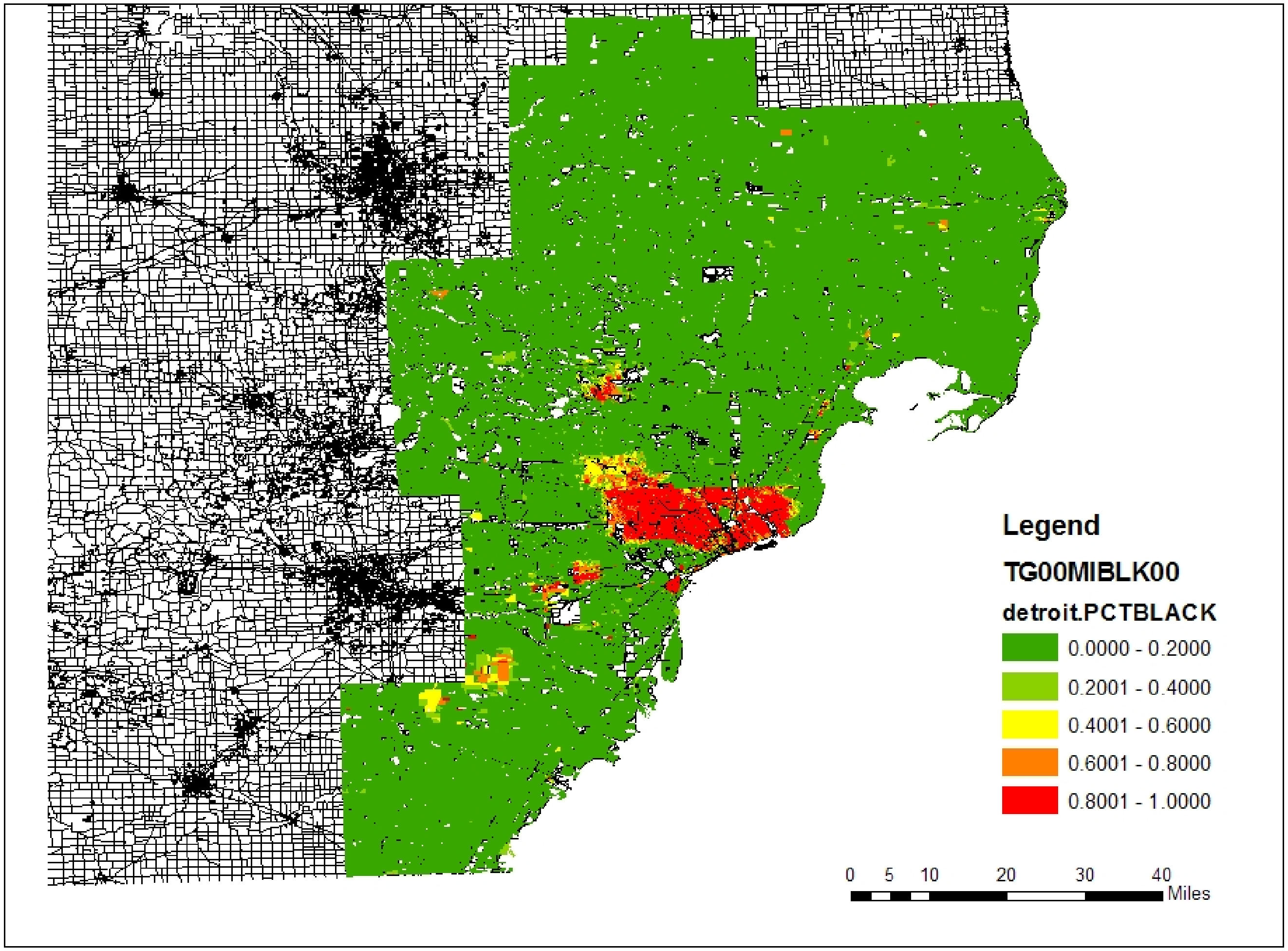


Figure 1: Segregation in Metropolitan Detroit

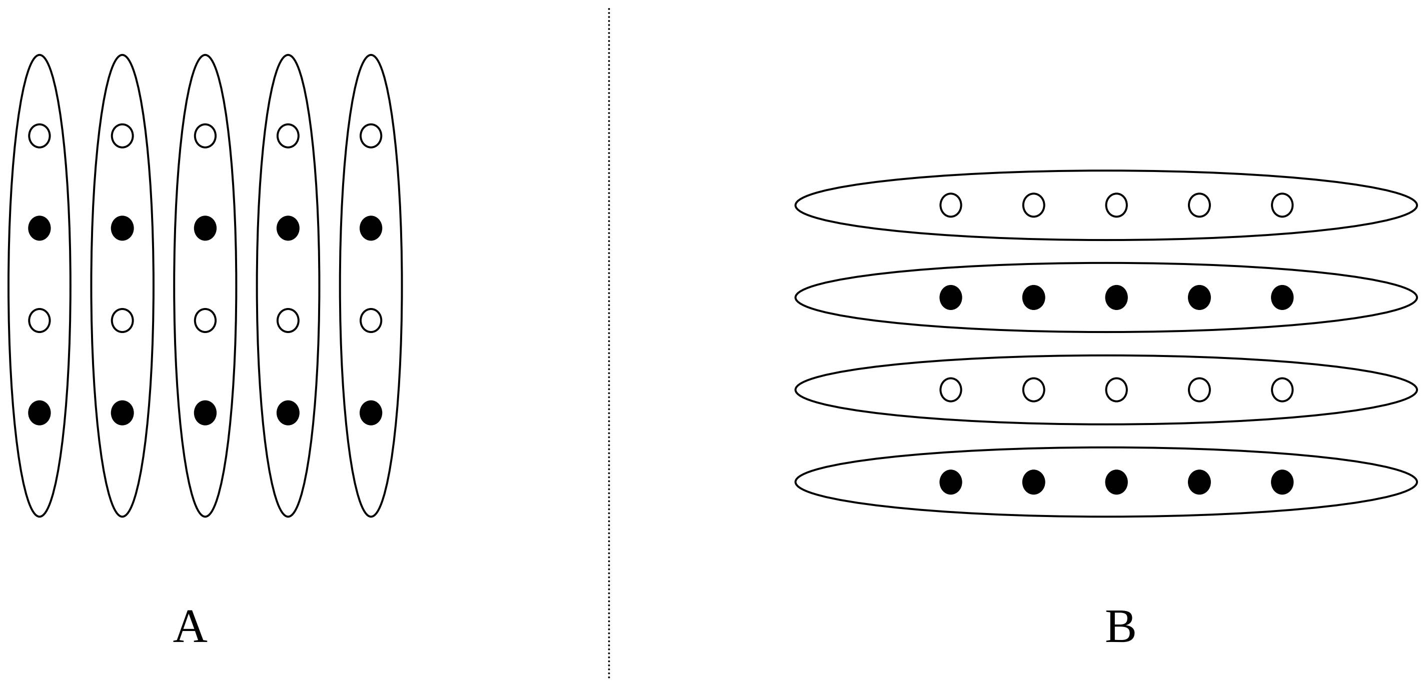
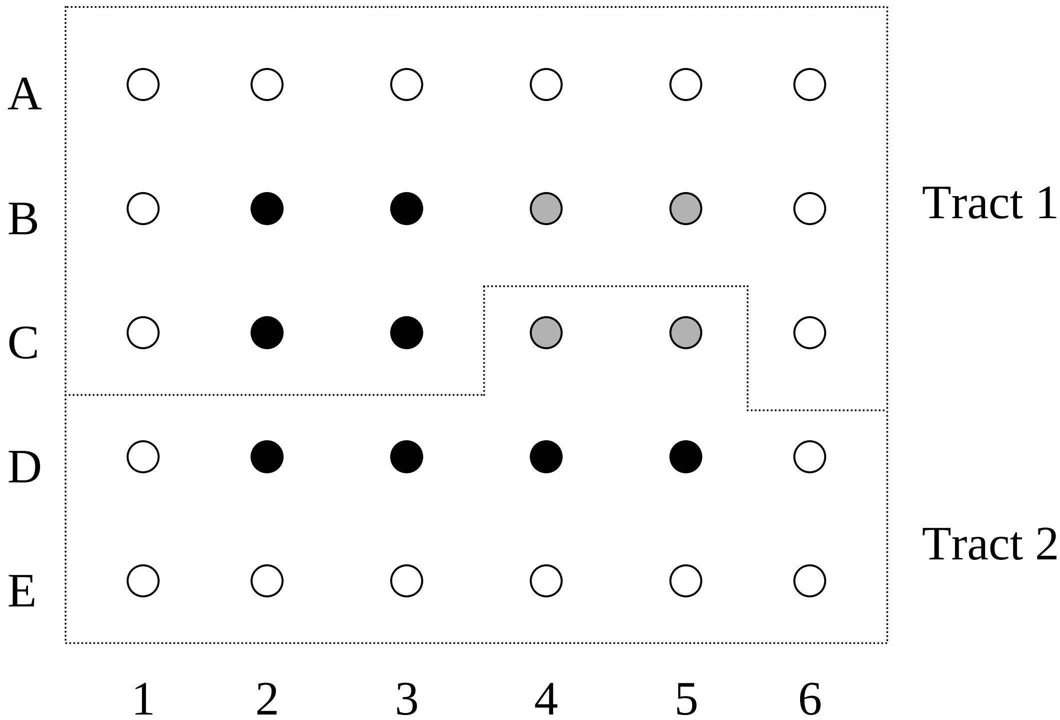
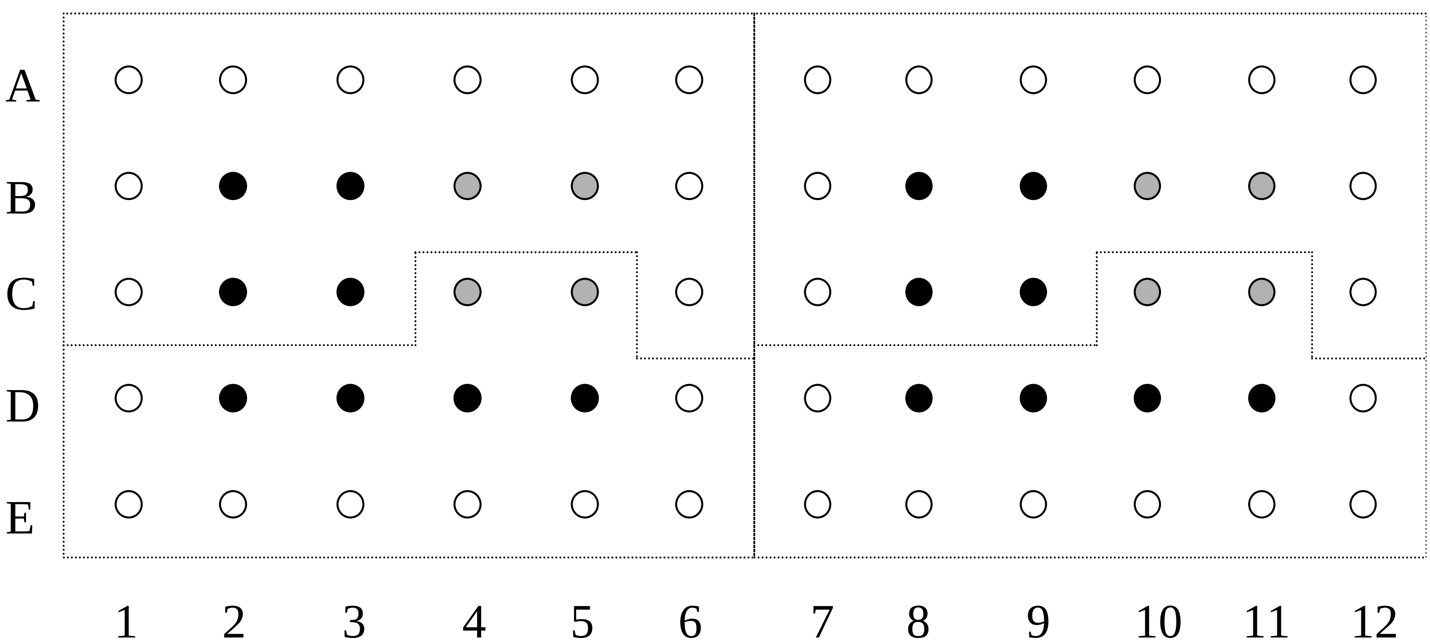


Figure 2: A hypothetical city



CITY 1



CITY 2

Figure 3: Two segregated cities

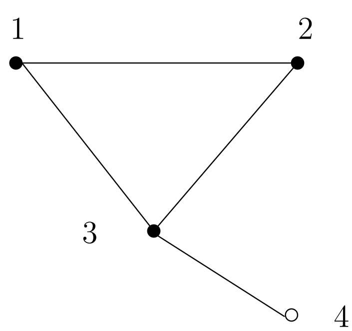


Figure 4: Adding a same-race neighbor.

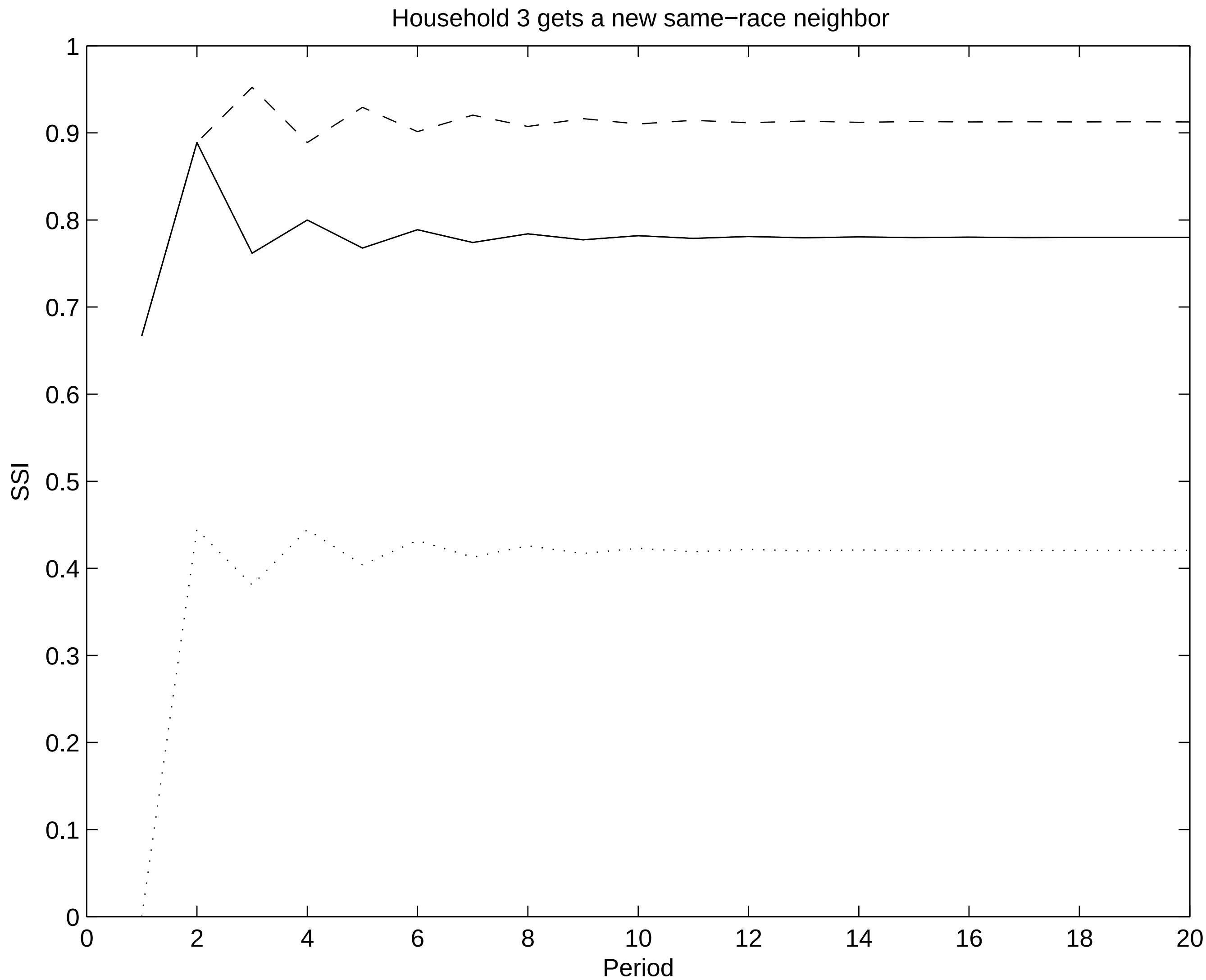


Figure 5: Changes in individual SSI.

Figure 6: Calculating the Spectral Index for Asians near Downtown Boston, MA

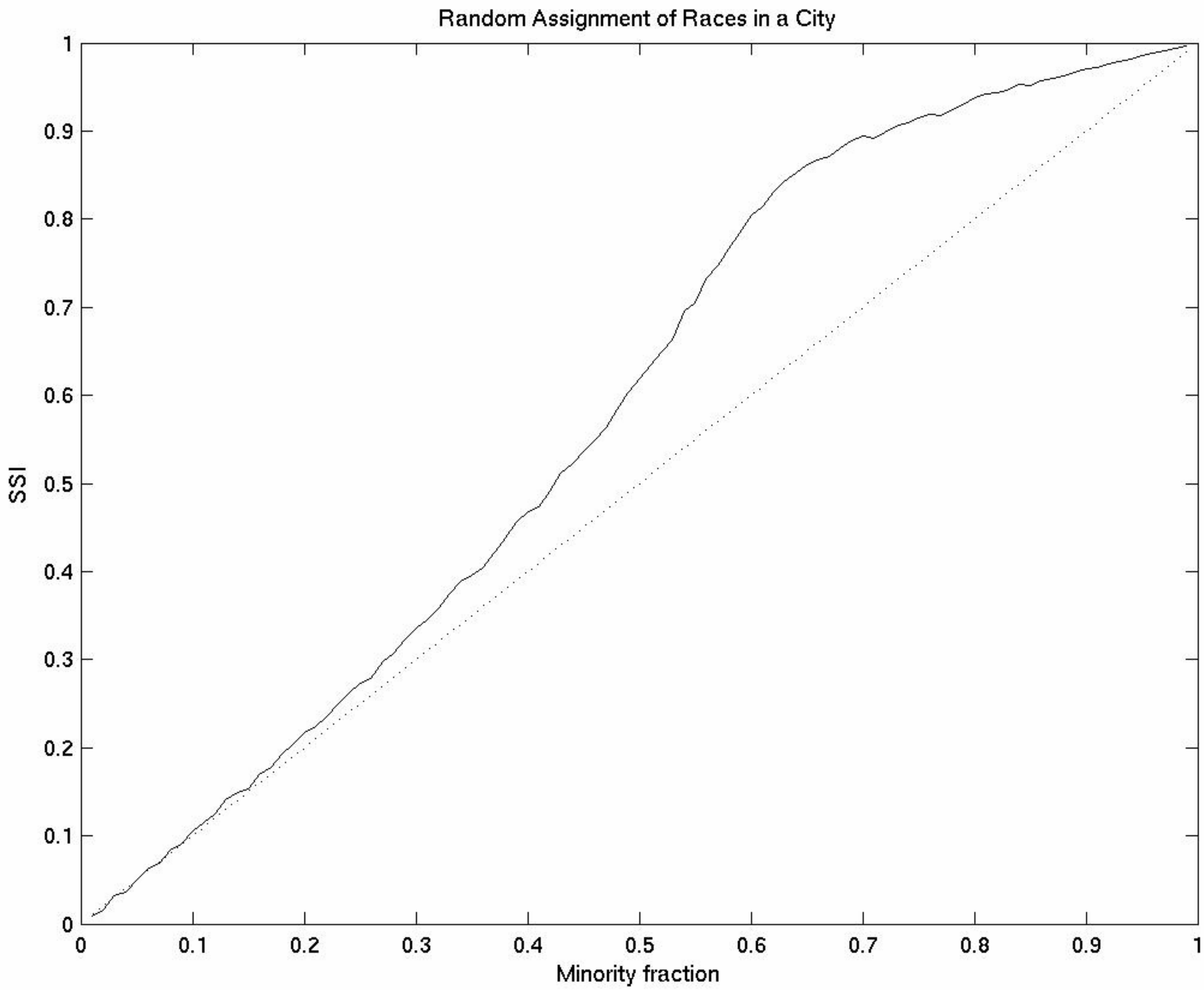


Figure 7: Simulating the Baseline Spectral Segregation Index

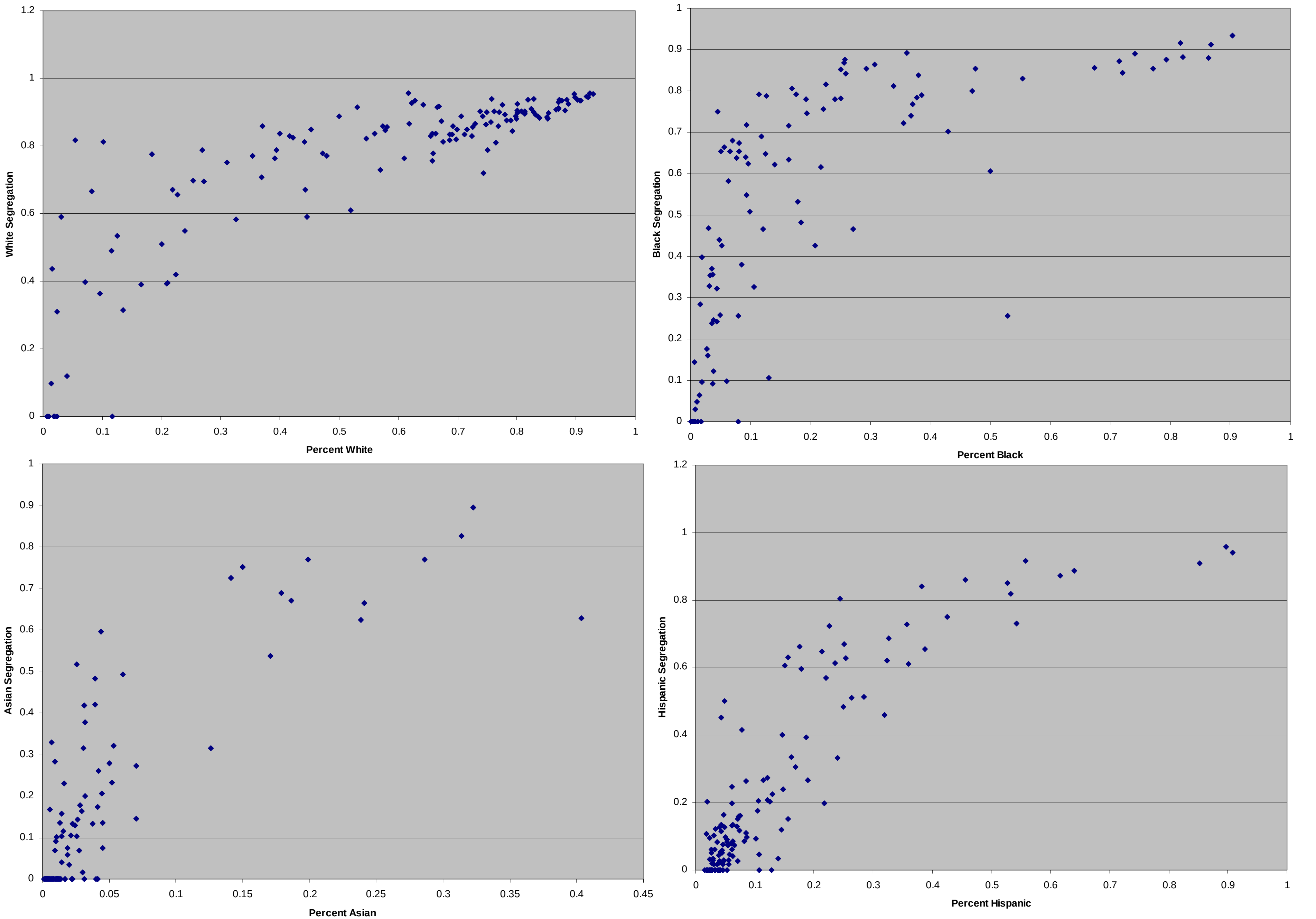


Figure 8: The Relationship Between Group Size and Group Segregation, By Race

Table 4: Top 10 Most Segregated Cities, by Racial Group

Whites	SSI	**Blacks**	SSI	**Asians**	SSI	**Hispanics**	SSI
Lowell, MA	0.99984	Detroit, MI	0.95421	Honolulu, HI	0.93403	McAllen, TX	0.9585
Lawrence, MA	0.99984	Monroe, LA	0.94912	San Francisco, CA	0.8056	Laredo, TX	0.9497
Nashua, NH	0.99966	Milwaukee, WI	0.93605	San Jose, CA	0.71692	Los Angeles, CA	0.939
Sharon, PA	0.99952	Flint, MI	0.93027	Los Angeles, CA	0.65878	El Paso, TX	0.9256
Boston, MA	0.99949	Pine Bluff, AR	0.92744	Vallejo, CA	0.63447	San Antonio, TX	0.9048
York, PA	0.99947	Chicago, IL	0.9206	Oakland, CA	0.56615	Brownsville, TX	0.8769
Barnstable, MA	0.99947	Memphis, TN	0.9166	Anaheim, CA	0.53402	Tuscon, AZ	0.8654
Johnstown, PA	0.99944	Miami, FL	0.91513	Seattle, WA	0.52639	Anaheim, CA	0.8624
Providence, RI	0.99943	Birmingham, AL	0.91449	New York, NY	0.47642	Corpus Christi, TX	0.8322
Springfield, MA	0.99933	Gary, IN	0.91418	San Diego, CA	0.41735	Albuquerque, NM	0.8246

Notes: Calculations performed using block-level data from from all MSAs in the 2000 US Census. The sample includes all census blocks in all MSAs. Racial categories are mutually exclusive. Asians include Pacific Islanders.

Table 5: The Top 'Behaviorally' Segregated Cities, By Racial Group

	PMSA	SSI-Baseline	SSI	% Minority	% Needed to Give SSI
Blacks					
	Milwaukee, WI	0.82155	0.93605	**0.11**	**0.80**
	Saginaw Bay, MI	0.79187	0.89757	**0.1**	**0.72**
	Indianapolis, IN	0.77346	0.87916	**0.1**	**0.68**
	Omaha, NE	0.77144	0.83954	**0.07**	**0.63**
	Chicago, IL	0.7712	0.9206	**0.14**	**0.76**
	Wichita, KS	0.7685	0.8298	**0.06**	**0.62**
	Toledo, OH	0.76485	0.85475	**0.09**	**0.64**
	Fort Wayne, IN	0.75505	0.80525	**0.05**	**0.60**
	Los Angeles, CA	0.7545	0.8444	**0.09**	**0.63**
	Buffalo, NY	0.74787	0.83777	**0.09**	**0.63**
Whites					
	Pine Bluff, AR	0.46306	0.90166	**0.38**	**0.72**
	Honolulu, HI	0.44971	0.66641	**0.2**	**0.53**
	Los Angeles, CA	0.27402	0.97952	**0.55**	**0.93**
	Albany, GA	0.26953	0.90433	**0.51**	**0.73**
	Sumter, SC	0.26029	0.86269	**0.49**	**0.65**
	Memphis, TN	0.24103	0.94653	**0.55**	**0.83**
	Jackson, MS	0.19891	0.90441	**0.55**	**0.73**
	Monroe, LA	0.19399	0.97809	**0.59**	**0.93**
	Oakland, CA	0.18802	0.97212	**0.59**	**0.91**
	Flagstaff, AZ	0.18247	0.94837	**0.58**	**0.83**
Asians					
	Los Angeles, CA	0.54428	0.65878	**0.11**	**0.53**
	Oakland, CA	0.41675	0.56615	**0.14**	**0.47**
	San Francisco, CA	0.6283	0.8056	**0.17**	**0.60**
	Honolulu, HI	0.37023	0.93403	**0.47**	**0.80**
	New York, NY	0.39272	0.47642	**0.08**	**0.41**
	Orange County, CA	0.40882	0.53402	**0.12**	**0.45**
	San Diego, CA	0.35605	0.41735	**0.06**	**0.37**
	San Jose, CA	0.49312	0.71692	**0.21**	**0.55**
	Seattle, WA	0.45829	0.52639	**0.07**	**0.44**
	Vallejo, CA	0.56637	0.63447	**0.07**	**0.51**
Hispanics					
	Flagstaff, AZ	0.71812	0.78622	**0.07**	**0.59**
	Orange County, CA	0.63855	0.86235	**0.21**	**0.65**
	Bridgeport, CT	0.59928	0.68918	**0.09**	**0.54**
	Milwaukee, WI	0.59905	0.63585	**0.04**	**0.51**
	Tucson, AZ	0.58659	0.86539	**0.26**	**0.66**
	Lawrence, MA	0.58459	0.67449	**0.09**	**0.53**
	Fort Worth, TX	0.5657899	0.74309	**0.17**	**0.57**
	Philadelphia, PA	0.56206	0.59886	**0.04**	**0.49**
	San Diego, CA	0.54279	0.77799	**0.22**	**0.59**
	Chicago, IL	0.54143	0.65593	**0.11**	**0.52**

Notes: Calculations performed using block-level data from from all MSAs in the 2000 US Census. The sample includes all census blocks in all MSAs. Racial categories are mutually exclusive. Asians include Pacific Islanders. Baseline SSI calculated from simulations described in Section 5.1.C.

Table 6: Top 15 Most Segregated Blocks

PMSA	Race	CC SSI	Size of CC	Latitude	Longitude	Block SSI
San Antonio, TX	Hispanic	0.995	6166	29.40878	-98.52187	292.260
Lafayette, LA	Black	0.833	406	30.53489	-92.08582	96.407
Los Angeles, CA	Asian	0.871	751	33.9731	-117.9167	94.085
Houston, TX	Hispanic	0.735	226	29.55255	-95.81261	87.263
Brownsville, TX	Hispanic	0.877	559	26.20365	-97.68172	82.148
Fresno, CA	Hispanic	0.919	1714	36.73535	-119.8122	66.056
Atlanta, GA	Black	0.992	6498	33.76681	-84.41813	62.625
Washington, DC	Black	0.993	6465	38.92611	-76.99167	57.627
McAllen, TX	Hispanic	0.988	3541	26.27484	-98.20446	56.180
Tallahassee, FL	Black	0.878	171	30.58215	-84.59776	45.379
Corpus Christi, TX	Hispanic	0.965	1116	27.69691	-97.36656	40.777
Oakland, CA	Black	0.837	1567	37.77278	-122.2013	39.056
San Diego, CA	Hispanic	0.958	2083	32.71384	-117.0652	38.918
Jackson, MS	Black	0.996	1828	32.30819	-90.20749	38.847
Columbus, GA	Black	0.957	1157	32.43829	-84.95276	34.448

Notes: The table reports the most segregated block per unique PMSA. Calculations performed using block-level data from from all MSAs in the 2000 US Census. The sample includes all census blocks in all MSAs. Racial categories are mutually exclusive. Asians include Pacific Islanders. Baseline SSI calculated from simulations described in Section 5.1.C.

Table 7: Correlation Between Existing Measures of Segregation and the Spectral Index

	SSI	Dissimilarity	Isolation	Exposure	Entropy	Gini	% Black	Interaction	SSI-Baseline
SSI	1								
Dissimilarity	**0.419**	1							
Isolation	**0.9283**	0.5594	1						
Exposure	**0.9097**	0.594	0.9538	1					
Entropy	**0.4726**	-0.3811	0.3614	0.3434	1				
Gini	**0.4563**	0.9953	0.6009	0.6266	-0.365	1			
Percent Black	**0.8973**	0.3055	0.9224	0.8432	0.5633	0.3498	1		
Interaction	**0.4684**	-0.3518	0.3743	0.3398	0.9835	-0.3328	0.5678	1	
SSI-Baseline	**0.8913**	0.3882	0.7382	0.8007	0.2	0.4112	0.6149	0.1823	1

Notes: All calculations performed using block-level data from from all MSAs in the 2000 US Census. The sample includes all census blocks in all MSAs. Baseline SSI calculated from simulations described in Section 5.1.C.

Table 8: Ordinary Least Squares Estimates of the Effects of Segregation on Outcomes, by Measure of Segregation

	Age 20-24					**Age 25-30**				
Variable	Education			Income	Social	Education			Income	Social
	High School Graduate	College Graduate	Idle	Log (Earnings)	Single Mother	High School Graduate	College Graduate	Idle	Log (Earnings	Single Mother
Dissimilarity	-0.014	.098	-.034	.058	.043	-.003	.079	-.026	-.082	.079
	(.026)	(.046)	(.031)	(.084)	(.024)	(.019)	(.080)	(.031)	(.088)	(.017)
*Dissimilarity*Black*	-.288	-.145	.432	-.344	.340	-.214	-.311	.331	-.249	.303
	(.046)	(.044)	(.069)	(.179)	(.080)	(.041)	(.072)	(.047)	(.133)	(.077)
SSI	-.001	.009	-.011	.018	.001	-.000	.015	-.011	-.009	.001
	(.004)	(.007)	(.004)	(.009)	(.003)	(.003)	(.011)	(.004)	(.012)	(.002)
*SSI*Black*	-.017	-.007	.025	-.082	.037	-.002	-.021	.013	-.028	.025
	(.009)	(.007)	(.012)	(.031)	(.016)	(.007)	(.012)	(.011)	(.022)	(.012)
Isolation	-.002	.011	-.004	.009	.006	.000	.005	-.001	-.018	.011
	(.004)	(.007)	(.005)	(.013)	(.003)	(.003)	(.013	(.005)	(.014)	(.003)
*Isolation*Black*	-.042	-.020	.061	-.061	.051	-.030	-.040	.041	-.047	.033
	(.007)	(.007)	(.011)	(.028)	(.012)	(.006)	(.011)	(.008)	(.018)	(.012)
Number of Observations	62,797	62,797	62,797	28,092	31,861	81,537	81,537	81, 537	54,141	41,933

Notes: Dependent variables vary by column. All calculations performed using 2000 Census data from the 5% Public Use Micro Sample. Segregation indices are calculated at the census block-level and normalized such that each has a mean of 0 and standard deviation of 1. Controls include gender, age dummies in one year increments, logarithm of MSA population, percent black, median household income and percent of the labor force employed in manufacturing. Each of these variables are also interacted with a dummy variable for black. Standard errors are clustered at the MSA level.

Table 9: The Effect of Individual Level Segregation on Social and Academic Outcomes

	Social				Academic			
	Smoking	Skip School	Interracial Dating	Happiness	PVT Scores	No College	Grades	Effort
Black	-0.143**	-0.010**	0.085**	-0.091**	-0.424**	-0.013*	-0.216**	0.027**
	0.004	0.003	0.017	0.007	0.026	0.006	0.011	0.002
Asian	-0.081**	-0.008*	0.372**	-0.025*	-0.303**	-0.047**	0.258**	0.036**
	0.006	0.004	0.029	0.010	0.042	0.007	0.015	0.003
Hispanic	-0.040**	0.026**	0.460**	-0.016*	-0.426**	0.065**	-0.182**	0.003
	0.005	0.003	0.017	0.007	0.026	0.006	0.010	0.002
Individual SSI	0.000	0.000	-0.007**	0.000	0.001	0.000	0.001	0.000
	0.001	0.000	0.002	0.001	0.003	0.001	0.001	0.000
Black*Individual SSI	-0.003*	-0.001	-0.004	0.003	-0.025*	0.000	0.000	0.000
	0.001	0.001	0.009	0.003	0.013	0.003	0.005	0.001
Asian*Individual SSI	-0.006	-0.008**	-0.067**	0.021**	-0.102**	0.000	0.017	0.004*
	0.005	0.002	0.017	0.006	0.030	0.007	0.014	0.002
Hispanic*Individual SSI	-0.006*	-0.002	-0.015	0.004	-0.047**	0.012**	-0.013**	0.004**
	0.003	0.002	0.014	0.004	0.014	0.003	0.004	0.001
Age	0.029**	0.009**	-0.002	-0.037**	-0.034**	0.021**	-0.024**	-0.011**
	0.001	0.001	0.003	0.001	0.006	0.001	0.002	0.001
Male	-0.002	0.019**	0.004	0.047**	0.124**	0.085**	-0.184**	-0.047**
	0.003	0.002	0.008	0.004	0.014	0.003	0.006	0.001
Mother College Educated	-0.024**	-0.002	0.001	0.031**	0.099**	-0.080**	0.154**	0.006**
	0.004	0.002	0.010	0.005	0.019	0.004	0.008	0.002
Father College Educated	-0.032**	-0.010**	0.014	0.021**	0.078**	-0.075**	0.163**	0.013**
	0.004	0.002	0.012	0.005	0.021	0.004	0.008	0.002
Mother Professional	-0.002	-0.001	0.011	-0.006	0.067**	-0.024**	0.062**	0.003*
	0.004	0.002	0.010	0.005	0.019	0.004	0.007	0.002
Father Professional	-0.008*	0.000	0.018	0.022**	0.127**	-0.048**	0.114**	0.001
	0.004	0.002	0.011	0.005	0.020	0.004	0.008	0.002
Constant	-0.220**	-0.108**	0.117*	1.134**	0.725**	-0.125**	3.216**	0.989**
	0.017	0.011	0.049	0.022	0.088	0.019	0.035	0.008
Observations	78075	77903	9553	73837	14387	69257	72744	79599
R-squared	0.07	0.04	0.37	0.05	0.28	0.1	0.18	0.08

Notes: Dependent variables vary by column. In all cases, dummy variables for missing values and school fixed effects are included. Robust standard errors are beneath the coefficients. * significant at 5%; ** significant at 1%.