Characteristics of the Joint Distribution of Race and Income Among Neighborhoods

Sean F. Reardon
Joseph Townsend
Lindsay Fox
Stanford University

Paper prepared for submission to

The Russell Sage Foundation Journal: The Spatial Foundations of Inequality

DRAFT: June, 2015

Abstract

We develop and illustrate a general method for describing in detail the joint distribution of race and income among neighborhoods. The approach we describe provides estimates of the average income distribution and racial composition of the neighborhoods of households of a given racial category and specific income level. We illustrate the method using 2007-2011 tract-level data from the American Community Survey. We show, for example, that blacks and Hispanics of any given income level typically live in neighborhoods that are substantially poorer than do whites and Asians of the same income level. Our approach provides a very general method for fully characterizing the joint patterns of racial and socioeconomic segregation, and so may prove useful for understanding the spatial foundations and correlates of racial and socioeconomic inequality.

Introduction

Although racial and socioeconomic segregation are persistent features of the residential landscape, both have changed in the last four decades in the United States. Racial segregation has declined moderately, particularly segregation between white and black households, but remains very high in many places (Logan & Stults, 2011; Logan, Stults, & Farley, 2004). Segregation by income has risen sharply since 1970; most of that increase occurred in the in the 1980s and 2000s (Bischoff & Reardon, 2014; Jargowsky, 1996, 2003; Reardon & Bischoff, 2011; Watson, 2009). Less clear, however, are the trends and patterns of the joint distribution of neighborhoods' racial and socioeconomic characteristics. That is, we do not have a clear description of how much neighborhoods differ (in terms of racial and economic composition) among households with the same income but that differ by race, or among those of the same race but differing by income. Without such a description, it is unclear whether and how changes in racial and economic segregation have altered disparities in neighborhood conditions.

In this paper, we demonstrate a general approach to describing the joint distribution of race and income among neighborhoods. We are not the first to suggest methods of describing features of this joint distribution (see, for example, Logan, 2002), but our approach is much more general and versatile than existing techniques. We show that estimating a set of multidimensional exposure functions is sufficient to generate a wide range of useful statistics regarding the joint distribution of racial and economic composition of neighborhoods, including many of the measures proposed and used in a more *ad hoc* fashion in much of the literature.

Our interest in developing these methods derives from the theoretical and empirical literature describing the ways in which neighborhoods affect their residents' educational, socioeconomic, and health outcomes. In much of the neighborhood effects research, neighborhood poverty (or socioeconomic conditions more generally) is hypothesized to be a key distal driver of neighborhood effects (Acevedo-Garcia & Lochner, 2003; Brooks-Gunn, Duncan, & Aber, 1997; Leventhal & Brooks-Gunn,

2000; Sampson, 2008; Sampson, Raudenbush, & Earls, 1997; Wodtke, Harding, & Elwert, 2011). A neighborhood's income distribution is hypothesized to directly or indirectly affect housing conditions, school and child-care quality, access to healthy food, green spaces, safe playgrounds, social networks, the prevalence of adult role models, and a range of other institutional and collective resources that are beneficial for child development.

Although some research has suggested that neighborhoods have no significant effect on many aspects of children's development, educational success, and social, behavioral, or economic outcomes (Kling, Liebman, & Katz, 2007; Ludwig et al., 2013; Sanbonmatsu, Kling, Duncan, & Brooks-Gunn, 2006), recent rigorous research suggests that neighborhood socioeconomic conditions can have substantial effects on such outcomes, particularly as a result of sustained exposure during childhood (Burdick-Will et al., 2011; Chetty, Hendren, & Katz, 2015; Harding, 2003; Santiago et al., 2014; Wodtke et al., 2011). Indeed, Chetty et al.'s (2015) analysis of the Moving to Opportunity experiment shows that children in families who used a (randomly assigned) housing voucher to move to a low-poverty neighborhood have substantially higher college attendance rates and 31% higher earnings by their mid-twenties than those in a control group not assigned a voucher, a finding that suggests that neighborhood poverty (and/or its correlates and sequelae) is harmful to young children's development.

Given these findings, and the theoretical importance of neighborhood composition, we develop and demonstrate in this paper a highly general approach to measuring the joint patterns of racial and economic neighborhood composition. In effect, this approach relies on estimating the average race-specific income distribution in the neighborhoods of individuals of any specific income and race/ethnicity. The functions describing these distributions can be used for a wide range of types of descriptive analyses, and provide a detailed account of the joint distribution of race/ethnicity and income across neighborhoods. While we provide some example findings that result from our estimated functions, a full description and explanation of the joint distribution is beyond the scope of this paper. Our goal is to

elucidate and illustrate this new approach so that others may use it in settings where a richer description of the interaction between race/ethnicity and income across neighborhoods is fruitful (see, for example, Reardon, Fox, & Townsend, forthcoming).

Measuring Segregation

Hundreds of articles have been devoted to developing and describing ways of measuring racial and economic segregation; hundreds more are devoted to describing their trends (e.g., James & Taeuber, 1985; Massey & Denton, 1993; Massey & Denton, 1988; Reardon & Bischoff, 2011; Reardon & Firebaugh, 2002; Reardon & Owens, 2014). The welter of methodological approaches to measuring segregation is partly due to academics' penchant for methodological hair-splitting. But there are also important theoretical and conceptual distinctions about the features of segregation that are important to measure and understand, and these distinctions lead to different measurement approaches.

Despite the abundance of ways of measuring segregation, most approaches are limited to measuring segregation along a single population dimension at a time. We know, for example, how to measure segregation among two or more racial groups (James & Taeuber, 1985; Massey & Denton, 1988; Reardon & Firebaugh, 2002); among ordered occupational or educational groups (Reardon, 2009); and by income or any other single continuous dimension (Jargowsky, 1996; Reardon, 2011; Reardon & Bischoff, 2011). Methods of measuring multidimensional patterns of segregation, such as the joint distribution of race and income among neighborhoods, however, are less well-developed.

Three approaches have been used to describe features of the joint distribution of racial and economic segregation patterns. One approach measures racial segregation among households of similar income (Adelman, 2004; Darden & Kamel, 2000; Denton & Massey, 1988; J. E. Farley, 1995; Iceland, Sharpe, & Steinmetz, 2005; Iceland & Wilkes, 2006; Massey & Fischer, 1999), or income segregation among households of the same race (Bischoff & Reardon, 2014; R. Farley, 1991; Jargowsky, 1996; Massey

& Fischer, 2003; Reardon & Bischoff, 2011). Typically, this approach relies on evenness measures of segregation, such as the dissimilarity index or similar measures (James & Taeuber, 1985). Some of these studies allow the comparison of overall racial segregation levels with within-income category segregation levels as a means of testing hypotheses about the role of income in shaping racial segregation levels, but they do not provide a clear description of the joint distribution of race and income across neighborhoods.

A second approach looks at the distribution of neighborhoods along a variety of typologies and the interaction of those typologies. For example, Turner and Fenderson (2006) categorize neighborhoods according to how mixed they are on measures of race, ethnicity, nativity, and income. Cross-tabulating these categorizations shows the patterns of interaction between neighborhood racial and income composition. This approach shows the extent to which tracts with very low proportions of low-income residents are predominantly white or predominantly minority. Goetz, Damiano, and Hicks (2015) take a similar approach in this volume, using it to identify what they call "racially concentrated areas of wealth" (RCAWs). They define RCAWs as tracts in which at least 90% of residents are white and over half of residents exceed an income threshold of four times the cost of living adjusted poverty threshold. Such approaches are useful for their specificity, but provide only partial descriptions of the joint distribution of race and income and are dependent on how racial and income distributions are dichotomized.

A third approach relies on so-called "exposure" measures of segregation to describe the average exposure of households of a given race-by-income category to those of another such category. Most commonly, these studies compute different racial groups' exposure to poverty: the average proportion of poor residents in the neighborhoods of members of different racial groups (Logan, 2002, 2011; Massey & Fischer, 2003; Timberlake, 2002, 2007; Timberlake & Iceland, 2007). These measures provide a much more interpretable description of differences in average neighborhood socioeconomic conditions than the evenness measures. Logan (2011), for example, categorizes households by race and three income categories (poor, middle, and affluent) and measures the exposure of various race-by-income groups to

other groups. This approach results in descriptive statements like “Affluent blacks are currently less [exposed to black neighbors] than poor blacks (36.3 vs. 42.9), and also somewhat more exposed to whites (42.9 vs. 39.8)” (Logan, 2011, p. 3). A related approach compares groups’ exposure to some measure of neighborhood quality. Friedman et al (2014) use data from the American Housing Survey to compare neighborhood conditions among middle- and upper-class households of different races/ethnicities; they find that affluent blacks and Hispanics experience inferior neighborhood circumstances relative to affluent whites. Like the neighborhood typology measures, such approaches are dependent on the specific definition of income categorizations that are used.

The more general drawback of all of these approaches is that, unless fine-grained income information is available, they are limited to comparisons based on the income categories reported in the data, which may be relatively crude. Moreover, these income categories may change over time (e.g., the Census or American Community Survey (ACS) have often changed the number and definition of the income categories reported in published tables). Even if they don’t change over time, their location in the income distribution will vary across time and place, because of differences in income distributions. Finally, even within a given place and time, the categories are not necessarily exactly comparable across groups. Suppose we define poor as having an income below $20,000. By this definition, the average “poor” black household will generally have a lower income than the average “poor” white household, simply because the black income distribution is lower than the white distribution. So a comparison of the neighborhoods of poor whites and blacks may be misleading when we base the comparison on broad income categories rather than exact income.

We adopt an approach similar to the third one here, but develop methods of estimating the joint distribution of racial and economic neighborhood composition in ways that are not sensitive to the definitions of income categories provided in the Census or ACS. Our approach allows one to describe 1) the average racial and income distributions in neighborhoods of households of different incomes, 2) the

average racial and income distributions in the neighborhoods of households of different income levels and race/ethnicity, and 3) the average race -specific income distributions in the neighborhoods of households of different income levels and race. These measures are similar to the more standard exposure measures used by others (e.g., the exposure of poor blacks to middle-class neighbors), but differ in that they are fully continuous, rather than categorical, measures of exposure. In effect, they describe the average joint distribution of race and income in the neighborhoods of individuals of any specific income and race/ethnicity. As a result, they can be used for a wide range of types of descriptive analyses, and provide a detailed account of the joint distribution of race/ethnicity and income across neighborhoods.

Estimating Average Neighborhood Income Distributions, by Race and Income

Notation

In this paper, we use g and h to denote racial groups (or other categorical groups); we use p and q to denote income levels, expressed as percentiles (scaled from 0 to 1, for convenience) of the population income distribution; and we use i to index neighborhoods. The function $\rho_x(p)$ describes the income density function in some population x, where x may refer to a specific group and/or neighborhood; correspondingly, the function $R_x(p) = \int_0^p \rho_x(r)dr$ denotes the cumulative income distribution function in population x. Finally we use T_x to denote the count of households in population x, and π_{gi} to indicate the proportion of households in neighborhood i that are in group g.

Primary Estimand and Estimation Approach

Our goal, in general, is to estimate the function $f_g^h(p,q)$ that describes the average cumulative income distribution function of group h in the neighborhoods of members of group g with income p. That is

$$f_g^h(p,q) = \sum_i \left[\frac{T_i \cdot \pi_{gi} \cdot \rho_{gi}(p)}{T \cdot \pi_g \cdot \rho_g(p)} \cdot \frac{T_i \cdot \pi_{hi} \cdot R_{hi}(q)}{T_i} \right].$$

(1)

Note that $f_g^h(p,q)$, defined this way, is interpreted as the weighted average proportion of the households in a neighborhood that are members of group h with incomes less than or equal to q, where the weights are the number of households of group g with income p in each neighborhood. In the segregation literature, such measures are called "exposure" measures, because they describe the average extent to which members of one group (in this case members of group g with income p) are exposed to members of another group (members of group h with incomes less than or equal to q) in their local context (neighborhood in this case) (Lieberson, 1981; Massey & Denton, 1988). In Appendix A, we describe how to estimate the functions $f_g^h(p,q)$ by assuming they can be approximated by a set of multidimensional polynomials of p and q.

Other Quantities of Interest

If we know $f_g^h(p,q)$ for all groups g and h, we can derive a number of additional useful quantities:

Functions describing exposure to overall (not race-specific) neighborhood income distributions, conditional on race and income. The average cumulative income distribution in the neighborhoods of members of group g with incomes p is simply the sum of the corresponding group-specific functions:

$$f_g^t(p,q) = \sum_h f_g^h(p,q).$$

(2)

Functions describing exposure to race-specific neighborhood income distributions, conditional on income. The typical household with income p lives in a neighborhood where members of group h have an income distribution given by

$$f_t^h(p,q) = \sum_g \left[\pi_g \cdot \rho_g(p) \cdot f_g^h(p,q)\right].$$

(3)

The function describing exposure to overall neighborhood income distributions, conditional on income. Combining (2) and (3) above, we can derive the function $f_t^t(p,q)$ which describes the average cumulative income distribution function in the neighborhoods of households with income p:

$$f_t^t(p,q) = \sum_h f_t^h(p,q) = \sum_g \pi_g \rho_g(p) f_g^t(p,q) = \sum_g \pi_g \rho_g(p) \sum_h f_g^h(p,q).$$

(4)

Average neighborhood racial composition, conditional on race and household income. The average racial composition of the neighborhoods of members of group g with income p (the exposure of members of group g and income p to members of group h) is simply $f_g^h(p,1)$. The average racial composition in the neighborhoods of households with income p is likewise given by the functions $f_t^h(p,1)$. Note that $f_g^t(p,1) = f_t^t(p,1) = 1$ by definition (because all households in a neighborhood have incomes less than or equal to 1 by definition).

Average neighborhood race-specific income density functions, conditional on race and income. Because $f_g^h(p,q)$ is a cumulative distribution function, we can obtain the corresponding density function, denoted $\rho_g^h(p,q)$, by taking the derivative of f_g^h with respect to q:

$$\rho_g^h(p,q) = \frac{d}{dq} f_g^h(p,q).$$

(5)

The formula holds when g and/or h is replaced by t as well.

Percentiles of average neighborhood race-specific income distributions, conditional on race and income. First define

$$f_g^{*h}(p,q) = \frac{f_g^h(p,q)}{f_g^h(p,1)}.$$

(6)

Now $f_g^{*h}(p,q)$ describes, for members of group g with incomes p, the weighted average proportion of the households in a neighborhood that are members of group h with incomes less than or equal to q relative to the weighted average proportion of the households in a neighborhood that are members of group h, where the weights are defined as above. If we wanted to know the median income among the group h neighbors of a member of group g with income p, we would find q such that $f_g^{*h}(p,q) = 0.50$. More generally, the $100 \cdot c^{th}$ percentile of the income distribution of members of group h in the neighborhoods of members of group g with income p is $f_g^{*h^{-1}}(p,c)$, where $f_g^{*h^{-1}}(p,c)$ returns the value q such that $f_g^{*h}(p,q) = c$.

Note that $f_g^t(p,1) = f_t^t(p,1) = 1$ by definition, but $f_t^h(p,1) \neq 1$ in general, so

$$f_g^{*t}(p,q) = f_g^t(p,q)$$

$$f_t^{*t}(p,q) = f_t^t(p,q)$$

$$f_t^{*h}(p,q) = \frac{f_t^h(p,q)}{f_t^h(p,1)}.$$

(7)

Thus, estimating $f_g^h(p,q)$ is sufficient to obtain a number of useful functions describing the joint neighborhood distribution of race and income. A number of other standard exposure measures, as well as measures of between group differences in income, can be readily computed from the $f_g^h(p,q)$ functions, as we describe in Appendix C.

Data

In this paper, we use data from the 2007-2011 American Community Survey (ACS) to illustrate

the types of descriptive patterns that can be obtained from our approach. We use census tracts as our definition of neighborhoods. The ACS provides partial cross-tabulations of household counts by income and racial/ethnic categories. In the 2007-11 ACS data, there are 16 categories of income, 7 race categories, and 1 indicator for whether the household is of Hispanic origin. We focus here on five mutually exclusive and exhaustive race/ethnic groups: non-Hispanic Asian, non-Hispanic black, Hispanic, non-Hispanic white, and non-Hispanic other. We use an iterative proportional fitting (IPF) process to estimate the full cross-tabulations of these five race/ethnic categories by income within each census tract, using Public Use Microdata Samples (PUMS) data to seed the IPF tables (for complete details on the construction of the cross tabulations and a discussion of the accuracy of the IPF process, see Reardon et al., forthcoming). The result is a dataset containing estimated counts of the number of households in each race/ethnic group that are in each income category for each tract in the U.S. in 2007-11.

Illustrative Application of the Approach

The purpose of this paper is to illustrate a new way of describing the joint distribution of race and income across neighborhoods. To do so, we use the 2007-2011 ACS data and the estimation methods described in the appendix to compute 80 observed values of $\rho_g(p)$ and 6,400 values of $f_g^h(p,q)$ for each of the values of g, h, p, and q observed in the ACS data. We fit multidimensional polynomials to these data to estimate the continuous functions $\rho_g(p)$ and $f_g^h(p,q)$. Using these, we derive the other functions described above and construct a set of illustrative figures to demonstrate a number of ways that the estimates can be used to describe the joint neighborhood distributions of race and income. All of our calculations use income percentiles scaled from 0-1 as noted above, but the illustrative figures below show income on a percentile scale from 0-100 for ease of interpretation.

Figure 1. Race-Specific Income Distributions, All Households in U.S., 2007-2011

A first step in estimating the exposure measures is calculating race-specific household income densities, described by the functions $\rho_g(p)$. These density functions are presented (stacked) in Figure 1, which shows the proportion of households of a given race/ethnicity at each percentile of the national household income distribution. The horizontal axis measures household income in percentiles (with corresponding dollar amounts noted) of the national household income distribution. The vertical axis is population proportion. Reading the figure vertically, then, describes the proportion of households of each race among all households at a given income percentile. The shaded area for each group describes the group's income distribution.

Figure 1 illustrates the unequal income distributions among white, black, Hispanic, and Asian households in the United States. Black households are disproportionately concentrated at the lower end of the income distribution; Hispanics are disproportionately in the bottom half of the distribution, while White and Asian households are disproportionately above the national median income. Nonetheless, a majority of low-income households are white in the United States, by virtue of their much larger population share. Although the patterns in Figure 1 have been demonstrated in previous research, our estimation approach facilitates the presentation of these patterns in terms of percentile ranks of the national income distribution.

Figure 2. Percentiles of Average Neighborhood Household Income Distributions, by Own Household Income, 2007-2011

Figure 2 presents the 25th, 50th, and 75th percentiles of the average neighborhood income distributions of households at each point in the income distribution. These income distributions are described by the function $f_t^t(p,q)$. Rather than plot the full surface described by $f_t^t(p,q)$, however, Figure 2 plots selected percentiles of neighborhood income distributions. To compute these values, we construct the function $f_t^{t^{-1}}(p,c)$ (by numerically inverting $f_t^t(p,q)$) for the values $c \in \{.25,.50,.75\}$

and for $p \in \{.01, .02, \ldots, .99\}$. For example, to identify the 50th percentile income in the typical neighborhood of a household with 5th percentile income, we set $f_t^t(.05, q)$ equal to 0.50 and solve for q via numerical interpolation. The horizontal axis represents a household's own income; the vertical axis represents neighborhood household income.[1]

As an example of how to read this figure, consider households with income of $20,000, which is approximately at the 18th percentile of the national income distribution. Such households, on average, live in neighborhoods where 25 percent of households have incomes at or below the 20th percentile of the *national* household income distribution (about $22,000); where the median of the average neighborhood household income distribution is roughly equal to the 42nd percentile of the national household income distribution (about $44,000); and where the 75th percentile of the average neighborhood household income distribution is at about the 68th percentile of the national household income distribution (roughly $77,000). Although Figure 2 shows only the 25th, 50th, and 75th percentiles of the average neighborhood income distributions, recall one could similarly construct these lines for any desired percentile.

Figure 2 makes clear that households with higher incomes live, on average, in neighborhoods with higher household income distributions. The steepness of the lines in Figure 2 describes the association between a household's own income and the 25th, 50th, and 75th percentile household income in the neighborhood. Indeed, the steepness of these lines provide an intuitive measure of income segregation (Reardon et al., forthcoming). Consider the 50th percentile line. A flat line would indicate no or little income segregation: households at any income level live, on average, in neighborhoods with the same median income. Steep lines would indicate a strong association between one's own income and

[1] While the exposure measures themselves are calculated in percentiles, the axes need not be presented in percentiles. The axes can be re-scaled and shown in dollars, or even log dollars. The dollar figures here, as well as in all following figures, are 2008 dollars and correspond directly to the thresholds of the 16 income categories in the ACS data. For convenience, here and elsewhere, the axes are labeled in terms of both income percentiles and dollars.

that of one's neighbors. As both axes are presented in percentile terms, the maximum value of the slope, averaged over the range of percentiles, is one. The lines are steeper in the right side of Figure 2, indicating that segregation among upper-income households is moderately larger than among lower-income households, consistent with other research on income segregation (Bischoff & Reardon, 2014; Reardon & Bischoff, 2011; Reardon et al., forthcoming). The lines here have average slopes of roughly 0.25-0.35, suggesting segregation is roughly one-quarter to one-third as high as its theoretical maximum (which would only occur if all households lived in neighborhoods where they and their neighbors had identical incomes.

Figure 3. Average Cumulative Neighborhood Income Distributions, by Own Household Income, 2007-2011

Figure 3 presents similar information as Figure 2, but in a different way. Figure 3 presents $f_t^t(p,q)$ for values of $p \in \{.05, .25, .50, .75, .95\}$ and for value of $q \in [0,1]$. In Figure 3, the estimated income exposure function (the exposure of households with incomes of p to those with incomes less than or equal to q) is drawn as a function of q for various values of p. To see the connection between the two figures, note that the 50th percentile (red) line in figure 2 corresponds to where each of the lines in figure 3 crosses the value 50 on the vertical axis. The gray line in figure 3, representing households with income at the 5th percentile, crosses this line around where q equals 38. This means that, on average, half of the neighbors of households with 5th percentile incomes have incomes below the 38th percentile, and half have incomes above the 38th percentile, making the 38th percentile the median neighborhood income for households at the 5th percentile. Figure 2 shows this as well: on the red line, when p (scaled here from 0-100) equals 5, median neighborhood income is at the 38th percentile of the national income distribution. Drawing the functions as in Figure 3 makes clear again that segregation between the affluent and the middle class is greater than between the middle class and the poor: the horizontal spaces between the $p = 50$, $p = 75$, and $p = 95$ lines are greater than between the $p = 5$, $p = 25$, and $p = 50$ lines,

indicating larger discrepancies in neighborhood income distributions as household income increases.

Figure 4. Average Neighborhood Racial Composition, by Household Income, All Households in U.S., 2007-2011

Figure 4 shows the average neighborhood racial composition for households of different incomes. These are given by the functions $f_t^h(p, 1)$. In contrast to figure 1, Figure 4 shows the average racial composition of households' neighborhoods, not the population racial composition. On the vertical axis, the typical racial composition of the neighborhood sums to 100%, and the figure shows how the racial composition of the average neighborhood changes as a function of own household income. Higher-income households, on average, have more white and Asian neighbors and fewer black and Hispanic neighbors than lower-income households.

Note that figure 4 looks relatively similar to figure 1. If neighborhoods were sorted perfectly by income, then these two figures would be identical, because every household would have only neighbors with their same income, who would, by definition, have on average the same racial composition as the population at that income level. That said, there are other patterns that may make Figure 4 similar to Figure 1. For example, if neighborhoods were sorted perfectly by race, but not at all by income within racial groups, Figures 1 and 4 would again be identical. Thus, the similarity of Figures 1 and 4 is not particularly informative about the relative extent of racial and income segregation that underlie them.

Figure 5. Average Neighborhood Racial Composition, by Household Income and Race, 2007-2011

What isn't clear from Figure 4 is whether households of different races but the same income typically live in racially similar neighborhoods. Figure 5 is similar to Figure 4, but presents average neighborhood racial composition as a function of both householders' race and income, as described by the functions $f_g^h(p, 1)$. For example, the top left panel shows the average neighborhood racial

composition for Asian households, conditional on their household income percentile. This panel shows, for example, that Asian households at the 50th percentile of the income distribution live in neighborhoods where, on average, roughly 50% of households are white, 10% are black, 20% are Hispanic, and 20% are Asian. Note that here, and throughout the paper, income percentiles are always measured in terms of the overall national income distribution, not group-specific income distributions. Of course, the axes could be scaled to reflect race group-specific household income distributions, in percentiles or dollars, if that were the goal of the description.

One striking feature of figure 5 is the high proportion of same-race households in the neighborhoods of each race group, regardless of income. For example, the average neighborhood racial composition for Asian households shows that, across the income distribution, nearly 20% of households in the neighborhood are Asian households, despite the fact that Asian households make up only roughly 5% of the population. Likewise, even high-income black households typically live in neighborhoods that are over 40% black and less than 50% white. Similar patterns are evident for each race group, but are most extreme for whites. White households live in neighborhoods that are around 80% white, and this racial isolation is consistent across the income distribution. In part this pattern results from between-region racial composition patterns. Many low-income white households are in rural areas and parts of the country with few non-white residents; as a result, most poor whites live in predominantly white neighborhoods. Nonetheless, the general patterns in Figure 5, particularly when compared to Figure 4, indicate high levels of racial segregation, even conditional on income.

Figure 6. Neighborhood Median Income, by Household Income and Race, All Households in U.S., 2007-2011

Figure 6 describes the average neighborhood income distributions for households of different races, by household income. Each line describes the median income in of the average neighborhood income distribution for households of a given race group, as a function of their income. The lines come

from the $f_g^{t^{-1}}(p, .50)$ functions. The figure is similar to Figure 2 (which shows the $f_t^{t^{-1}}(p, .50)$ function), but shows only the median of the average neighborhood income distribution (not the 25th and 75th percentiles), and presents a separate line for each race group.

The most striking feature of Figure 6 is that Asian and white households live in neighborhoods where the median income of their neighbors is much higher than in the neighborhoods of similar income Hispanic and black households. The vertical distance between the lines yields a comparison of neighborhood conditions between households of different races. For example, poor black and Hispanic households live in neighborhoods where the median income is roughly two-thirds that of white and Asian households that are equally poor. As in figure 2, the steepness of the lines is an indication of the degree of income segregation within each race group. The figure indicates that income segregation is higher for all groups in the upper half of the income distribution relative to the lower half of the income distribution.

Another way to compare neighborhood conditions of the different groups is to look at the horizontal differences between the lines. From this perspective, the figure shows that black and Hispanic households typically live in similar neighborhoods (in terms of their median income) as white households with much lower incomes. Black households with incomes of roughly $60,000, for example, live in neighborhoods with median incomes similar to those of white households earning roughly $12,000. This means that black households, on average, need to earn about five times that of poor white households to live in a similar neighborhood. Hispanic households must earn 3.7 times that of whites (see Reardon et al., forthcoming).

Figure 7. Race-specific Neighborhood Median Income, by Household Income, All Households in U.S., 2007-2011

Figure 6 above illustrates the income distribution in the typical neighborhood of households of different races. Figure 7, in contrast, illustrates the opposite: the income distributions of each race in the

average neighborhood of households of a given income. Specifically, Figure 7 plots the functions $f_t^{*h^{-1}}(p, .50)$. Each line represents the median of the income distribution of a specific group for typical households of specific incomes. For example, a typical household at the 25th percentile of the income distribution lives in a neighborhood where the median black income is at roughly the 30th percentile of the national income distribution and the median white income is at roughly the 47th percentile. The figure shows that, on average, Asian and white income distributions are higher than those of blacks and Hispanics for all values of household income, suggesting that, for most households in the United States their black and Hispanic neighbors are poorer than their Asian and white neighbors. Indeed, across the income distribution, the typical household's black neighbors have median incomes roughly $20,000 less than its white neighbors—a substantial difference. These patterns have important implications for perceptions of racial differences. If households were sorted only by income, the average households would experience no racial differences in income among their neighbors. The patterns here (as well as in Figures 5 and 6) indicate that households are sorted not only on income, but on race as well. The average person looking at his/her neighbors experiences black and Hispanics as poorer than their white and Asian neighbors.

Figure 8. Race-specific Neighborhood Median Income, by Household Income and Race, 2007-2011

Figure 8 is similar to Figure 7, but shows race-specific median incomes in the neighborhoods of households of each race. Specifically, each line in Figure 8 is one of the $f_g^{*h^{-1}}(p, .50)$ functions. For example, the green line in the top left panel indicates that the typical Asian household at the median of the national income distribution lives in a neighborhood where the median income among Hispanic households is around the 46th percentile of the national income distribution. As in the other figures showing the f^{-1} functions, one could choose other percentiles of these race-specific distributions to display as well.

In Figure 8, the steepness of the same-race lines indicates the degree of within-race segregation. The top left panel suggests, for example, that Asian households are highly segregated by income—low income Asians live in neighborhoods where their Asian neighborhoods are poor, on average, while high income Asians have much higher-income Asian neighbors. Within-group income segregation is also high for black households, but is somewhat lower among white and Hispanic households. Similar to Figure 7, these findings also imply that perceptions of racial differences in income may differ across races and by income level. For example, very poor Asians typically have white and Hispanic neighbors who earn substantially more than themselves, while affluent Asians generally live in neighborhoods where they are the highest-income group. The average Hispanic, black or white resident of any income typically experiences Asians as wealthier than all other race groups.

Figure 9. Metropolitan Variation in Neighborhood Median Income, by Household Income, Ten Largest Metropolitan Areas by Population, 2007-2011

Each of the previous figures describes patterns for the United States as a whole. The methods we have described here can be applied to smaller geographic regions as well. Figure 9 provides an example of this. It shows median neighborhood income, as a function of household income, for each of the 10 largest metropolitan areas in the United States.[2] The lines come from metropolitan area-specific functions ${f_{mt}^{t}}^{-1}(R_m(p), .50)$, where $R_m(p)$ is a function that converts national income percentiles to local income percentiles of metropolitan area m (that is, it is the cumulative income distribution function for metropolitan area m). The figure indicates, for example, that households at the 50th percentile of the national income distribution in the Washington, D.C. metropolitan area live in neighborhoods where the median income is above the 65th percentile of the national income distribution.

A notable feature of this figure is that both axes are shown in the national income distribution to

[2] We define metropolitan areas using the 2003 Office of Management and Budget metropolitan division codes, and we rank these areas based on their total population in 2010.

allow for comparisons across metropolitan areas. Although the graph could be constructed using local income distributions, that would obscure comparisons among households of the same income in different metropolitan areas, because the 50th percentile of Chicago's income distribution is not the same as the 50th percentile of New York's income distribution. Using a common scale for income (percentiles of the national income distribution) makes evident that households in some metropolitan areas live, on average, in very different neighborhoods than similar income households in other areas. For example, Washington, DC households earning $60,000 live in much higher income neighborhoods than do similar income households in Los Angeles; in fact, Washington, DC households earning $60,000 typically live in neighborhoods similar to those of Los Angeles households earning $150,000.

Also evident in Figure 9 is the fact that income segregation varies across metropolitan areas. As above, the steepness of the lines in Figure 9 provide an intuitive measure of income segregation. In the Minneapolis metropolitan area, for example, segregation is relatively low in comparison to the Dallas or Houston metropolitan areas. In Dallas and Houston, for example, high- and low-income families live, on average, in neighborhoods that are very different in median income levels; the difference in neighborhood conditions in Minneapolis, particularly between high- and middle-income families, is much less pronounced.

Discussion

The approach we have outlined here provides a variety of ways of characterizing the joint patterns of racial and socioeconomic segregation. A full characterization of these patterns is provided by the group-specific income distributions (the ρ_g functions, in our notation) and the set of exposure functions that describe the average neighborhood income distributions conditional on race and income (the f_g^h functions), but simply reporting the parameters of these functions is neither feasible nor particularly informative (in our illustration here, these functions are together characterized by a total of

480 parameters). Instead, we have chosen to illustrate key features of these functions in a series of figures, each of which highlights a different aspect of the joint distribution.

One could, of course, derive additional statistics from these functions. The slopes of the lines in Figures 2 and 8, for example, may be useful as measures of segregation. The vertical or horizontal distances between the lines in Figure 6, likewise, might be thought of as measures of racial inequality in neighborhood conditions net of differences due to between-race differences in household income. Measures of between-group differences in racial composition of neighborhoods (evident in Figure 5) may be useful for measuring and understanding racial segregation. Statistics of these types can be derived from the estimated ρ_g and f_g^h functions and then may be usefully compared across time or metropolitan areas to assess changes or variation in patterns of racial/economic segregation.

Our goal here was to describe a general approach to measuring joint patterns of racial and socioeconomic segregation. Given that, a discussion of the substantive implications of the patterns illustrated in our figures here is beyond the scope of this paper, but a few features of the figures are particularly striking. First, the figures clearly show that there are large racial differences in neighborhood racial and economic composition, even conditional on income. That is, equally poor white, black, Hispanic, and Asian households are located in very different neighborhoods than one another; for example, black households typically live in similar neighborhoods as white households making $40,000-50,000 less. This is very consistent with prior research showing that economic disparities are insufficient to explain racial segregation and that middle-class blacks live in poorer neighborhoods than most whites (Adelman, 2004; Friedman, Gibbons, & Galvan, 2014; Logan, 2002, 2011; Pattillo, 1999; Reardon et al., forthcoming; Timberlake, 2002, 2007; Timberlake & Iceland, 2007) (See also Lareau & Goyette, 2014; Pattillo, 2005 for a useful review of this literature). If racial segregation were simply the result of racial differences in income, we would expect racial differences in neighborhood composition to disappear once we condition on household income. The figures here clearly show that they do not.

Second, the figures reveal something about the income levels of households that different racial and income groups might encounter in their neighborhood. Figures 7 and 8 show that the typical household, regardless of income level or race, lives in a neighborhood where black and Hispanic neighbors have lower incomes than white and Asian neighbors. Indeed, the black and Hispanic neighbors of high-income households have lower median incomes, on average, than the white and Asian neighbors of low-income households (see Figure 7). This pattern may play a role in shaping racial stereotypes.

Third, Figure 9 shows there is substantial variation among metropolitan areas in the patterns of exposure to high- and low-income neighbors, conditional on income. Not shown here, but straightforward to compute from the methods described here, are metropolitan patterns of racial differences in neighborhood economic conditions. A full description of variation across metropolitan areas in the joint neighborhood distribution of race and income would likely reveal considerable variability.

Recent scholarship demonstrates that neighborhood economic conditions affect child development and opportunities for educational and economic success. For example, Chetty et al. (2015) demonstrate that moving to a lower-poverty area has a substantial positive effects on the life course trajectory of young children. Given this evidence, it is likely that variation in economic neighborhood conditions across racial groups, households of different incomes, and metropolitan areas may lead to disparities in developmental, educational, and economic outcomes. In other words, segregation matters for children's outcomes.

In general, children growing up in poor families face a double disadvantage. Their families have fewer private resources than richer families, and they tend to live in poorer neighborhoods, meaning they have access to fewer contextual resources as well. Even more troubling, Figure 6 illustrates that low-income black and Hispanic children face a triple disadvantage relative to middle-class white children: not only do their families have fewer private resources and live in poorer neighborhoods than middle-class

children, but they also live in much poorer neighborhoods than equally poor white children. Given that neighborhood conditions matter for children's development, the joint patterns of racial and economic segregation described here suggest that children of different races and incomes face drastically different life opportunities.

The methods described here provide a consistent way of quantifying disparities in neighborhood conditions that is largely independent of the specific income thresholds used in tabulating income in the ACS data. This makes possible much more precise comparisons of racial and economic neighborhood conditions across place, time, and population groups than has been used in prior work. We expect that these methods will enable researchers to more carefully investigate the patterns, causes, and consequences of racial and economic segregation.

References

Acevedo-Garcia, D., & Lochner, K. A. (2003). Residential Segregation and Health. In I. Kawachi & L. F. Berkman (Eds.), *Neighborhoods and Health* (pp. 265-287). New York, NY: Oxford University Press.

Adelman, R. M. (2004). Neighborhood Opportunities, Race, and Class: The Black Middle Class and Residential Segregation. *City & Community, 3*, 43-63.

Bischoff, K., & Reardon, S. F. (2014). Residential Segregation by Income, 1970-2009. In J. Logan (Ed.), *Diversity and Disparities: America Enters a New Century*. New York: The Russell Sage Foundation.

Brooks-Gunn, J., Duncan, G. J., & Aber, J. L. (Eds.). (1997). *Neighborhood poverty: Context and consequences for children* (Vol. 1). New York: Russell Sage Foundation.

Burdick-Will, J., Ludwig, J., Raudenbush, S. W., Sampson, R. J., Sanbonmatsu, L., & Sharkey, P. (2011). Converging Evidence for Neighborhood Effects on Children's Test Scores: An Experimental, Quasi-experimental, and Observational Comparison. In G. J. Duncan & R. J. Murnane (Eds.), *Whither Opportunity? Rising Inequality and the Uncertain Life Chances of Low-Income Children* (pp. 255-276). New York: Russell Sage Foundation.

Chetty, R., Hendren, N., & Katz, L. F. (2015). *The effects of exposure to better neighborhoods on children: New evidence from the Moving to Opportunity experiment*. Harvard University.

Darden, J. T., & Kamel, S. M. (2000). Black residential segregation in the city and suburbs of Detroit: Does socioeconomic status matter? *Journal of Urban Affairs, 22*(1), 1-13.

Denton, N. A., & Massey, D. S. (1988). Residential segregation of Blacks, Hispanics, and Asians by socioeconomic status and generation. *Social Science Quarterly, 69*(4), 797-817.

Farley, J. E. (1995). Race still matters: The minimal role of income and housing costs as causes of housing segregation in St. Louis, 1990. *Urban Affairs Review, 31*, 244-254.

Farley, R. (1991). Residential segregation of social and economic groups among blacks, 1970-80. In C. Jencks & P. E. Peterson (Eds.), *The urban underclass* (pp. 274-298). Washington, DC: Brookings Institution.

Friedman, S., Gibbons, J., & Galvan, C. (2014). Declining segregation through the lens of neighborhood quality: does middle-class and affluent status bring equality? *Social Science Research, 46*, 155-168. doi: 10.1016/j.ssresearch.2014.03.003

Harding, D. J. (2003). Counterfactual models of neighborhood effects: The effect of neighborhood poverty on dropping out and teenage pregnancy. *American Journal of Sociology, 109*(3), 676-719.

Ho, A. D., & Haertel, E. H. (2006). Metric-Free Measures of Test Score Trends and Gaps with Policy-Relevant Examples. Los Angeles, CA: Center for the Study of Evaluation, National Center for Research on Evaluation, Standards, and Student Testing, Graduate School of Education & Information Studies.

Ho, A. D., & Reardon, S. F. (2012). Estimating Achievement Gaps From Test Scores Reported in Ordinal 'Proficiency' Categories. *Journal of Educational and Behavioral Statistics, 37*(4), 489-517.

Holland, P. (2002). Two measures of change in the gaps between the CDFs of test score distributions. *Journal of Educational and Behavioral Statistics, 27*, 3-17.

Iceland, J., Sharpe, C., & Steinmetz, E. (2005). Class differences in African American residential patterns in US metropolitan areas: 1990–2000. *Social Science Research, 34*(1), 252-266. doi: 10.1016/j.ssresearch.2004.02.001

Iceland, J., & Wilkes, R. (2006). Does Socioeconomic Status Matter? Race, Class, and Residential Segregation. *Social Problems, 53*(2), 248-273.

James, D. R., & Taeuber, K. E. (1985). Measures of segregation. *Sociological Methodology, 14*, 1-32.

Jargowsky, P. A. (1996). Take the money and run: Economic segregation in U.S. metropolitan areas. *American Sociological Review, 61*(6), 984-998.

Jargowsky, P. A. (2003). Stunning progress, hidden problems: The dramatic decline of concentrated poverty in the 1990s *Living Cities Census Series* (pp. 24). Washington, DC: Brookings Institution.
Kling, J. R., Liebman, J. B., & Katz, L. F. (2007). Experimental Analysis of Neighborhood Effects. *Econometrica, 75*(1), 83-119.
Lareau, A., & Goyette, K. (2014). *Choosing Homes, Choosing Schools*. New York: Russell Sage Foundation.
Leventhal, T., & Brooks-Gunn, J. (2000). The neighborhoods they live in: The effects of neighborhood residence on child and adolescent outcomes. *Psychological Bulletin, 126*(2), 309-337.
Lieberson, S. (1981). An asymmetrical approach to segregation. In P. C., R. V. & S. S. (Eds.), *Ethnic segregation in cities* (pp. 61-82). London: Croon Helm.
Logan, J. R. (2002). Separate and Unequal: The Neighborhood Gap for Blacks and Hispanics in Metropolitan America (pp. 20). Albany, NY: Lewis Mumford Center for Comparative Urban and Regional Research, State University of New York, Albany.
Logan, J. R. (2011). Separate and Unequal: The Neighborhood Gap for Blacks, Hispanics and Asians in Metropolitan America. In J. R. Logan (Ed.), *US2010*: US2010 Project, Brown University.
Logan, J. R., & Stults, B. (2011). The Persistence of Segregation in the Metropolis: New Findings from the 2010 Census. In J. R. Logan (Ed.): US2010 Project, Brown University.
Logan, J. R., Stults, B. J., & Farley, R. (2004). Segregation of minorities in the metropolis: Two decades of change. *Demography, 41*(1), 1-22.
Ludwig, J., Duncan, G. J., Gennetian, L. A., Katz, L. F., Kessler, R. C., Kling, J. R., & Sanbonmatsu, L. (2013). Long-Term Neighborhood Effects on Low-Income Families: Evidence From Moving To Opportunity. *NBER working paper, 18772*.
Massey, D. S., & Denton, N. (1993). *American Apartheid: Segregation and the Making of the Underclass*. Cambridge, MA: Harvard University Press.
Massey, D. S., & Denton, N. A. (1988). The dimensions of residential segregation. *Social Forces, 67*(2), 281-315.
Massey, D. S., & Fischer, M. J. (1999). Does Rising Income Bring Integration? New Results for Blacks, Hispanics, and Asians in 1990. *Social Science Research, 28*, 316-326.
Massey, D. S., & Fischer, M. J. (2003). The Geography of Inequality in the United States, 1950-2000. *Brookings-Wharton Papers on Urban Affairs*, 1-40.
Pattillo, M. (1999). *Black Picket Fences: Privilege and Peril Among the Black Middle Class*. Chicago: University of Chicago Press.
Pattillo, M. (2005). Black Middle-Class Neighborhoods. *Annual Review of Sociology, 31*(1), 305-329.
Reardon, S. F. (2009). Measures of Ordinal Segregation. In Y. Flückiger, S. F. Reardon & J. Silber (Eds.), *Occupational and Residential Segregation* (Vol. 17, pp. 129-155): Emerald.
Reardon, S. F. (2011). *Measures of Income Segregation*. CEPA Working Papers. Working Paper. Stanford Center for Education Policy Analysis. Stanford, CA. Retrieved from http://cepa.stanford.edu/content/measures-income-segregation
Reardon, S. F., & Bischoff, K. (2011). Income Inequality and Income Segregation. *American Journal of Sociology, 116*(4), 1092-1153.
Reardon, S. F., & Firebaugh, G. (2002). Measures of multi-group segregation. *Sociological Methodology, 32*, 33-67.
Reardon, S. F., Fox, L., & Townsend, J. (forthcoming). Neighborhood Income Composition by Race and Income, 1990-2009. *The Annals of the American Academy of Political and Social Science*.
Reardon, S. F., & Owens, A. (2014). 60 Years After Brown: Trends and Consequences of School Segregation. *Annual Review of Sociology, 40*(1).
Sampson, R. J. (2008). Moving to Inequality: Neighborhood Effects and Experiments Meet Social Structure. *American Journal of Sociology, 114*(1), 189-231.

Sampson, R. J., Raudenbush, S. W., & Earls, F. (1997). Neighborhoods and violent crime: A multilevel study of collective efficacy. *Science, 277*, 918-924.

Sanbonmatsu, L., Kling, J. R., Duncan, G. J., & Brooks-Gunn, J. (2006). Neighborhoods and Academic Achievement: Results from the Moving to Opportunity Experiment. *Journal of Human Resources, 41*(4), 649-681.

Santiago, A. M., Galster, G. C., Lucero, J. L., Ishler, K. J., Lee, E. L., Kypriotakis, G., & Stack, L. (2014). Opportunity neighborhoods for Latino and African American children: Final Report. Washington, DC: U.S. Department of Housing and Urban Development, Office of Policy Development and Research.

Timberlake, J. M. (2002). Separate, but how unequal? Ethnic residential stratification, 1980 to 1990. *City & Community, 1*(3), 251-266.

Timberlake, J. M. (2007). Racial and ethnic inequality in the duration of children's exposure to neighborhood poverty and affluence.

Timberlake, J. M., & Iceland, J. (2007). Change in racial and ethnic residential inequality in American cities, 1970–2000. *City & Community, 6*(4), 335-365.

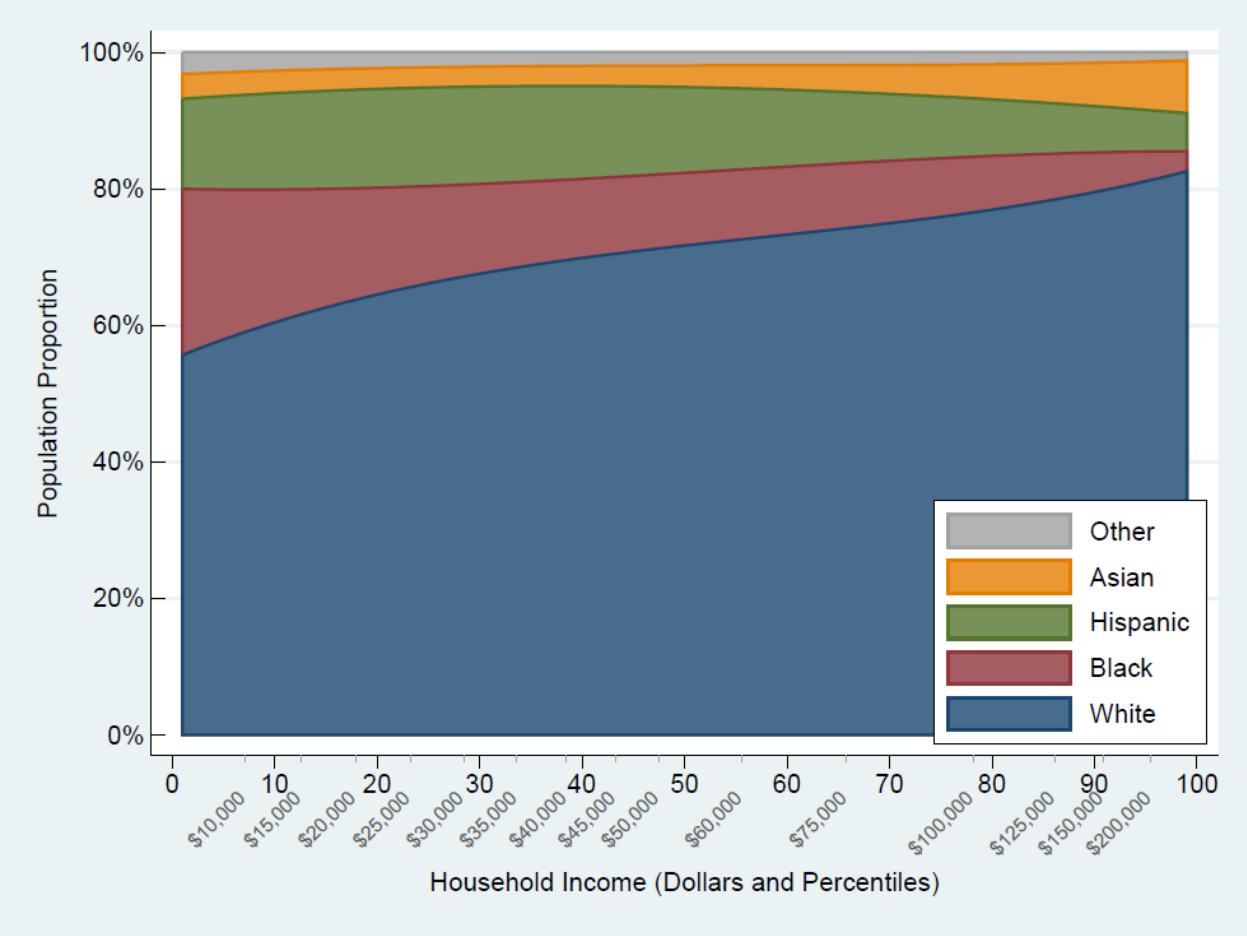
Turner, M. A., & Fenderson, J. (2006). Understanding diverse neighborhoods in an era of demographic change. Washington, DC: Urban Institute, Metropolitan Housing and Policy Center.

Watson, T. (2009). Inequality and the Measurement of Residential Segregation by Income. *Review of Income and Wealth, 55*(3), 820-844.

Wodtke, G. T., Harding, D. J., & Elwert, F. (2011). Neighborhood Effects in Temporal Perspective: The Impact of Long-Term Exposure to Concentrated Disadvantage on High School Graduation. *American Sociological Review, 76*(5), 713-736.

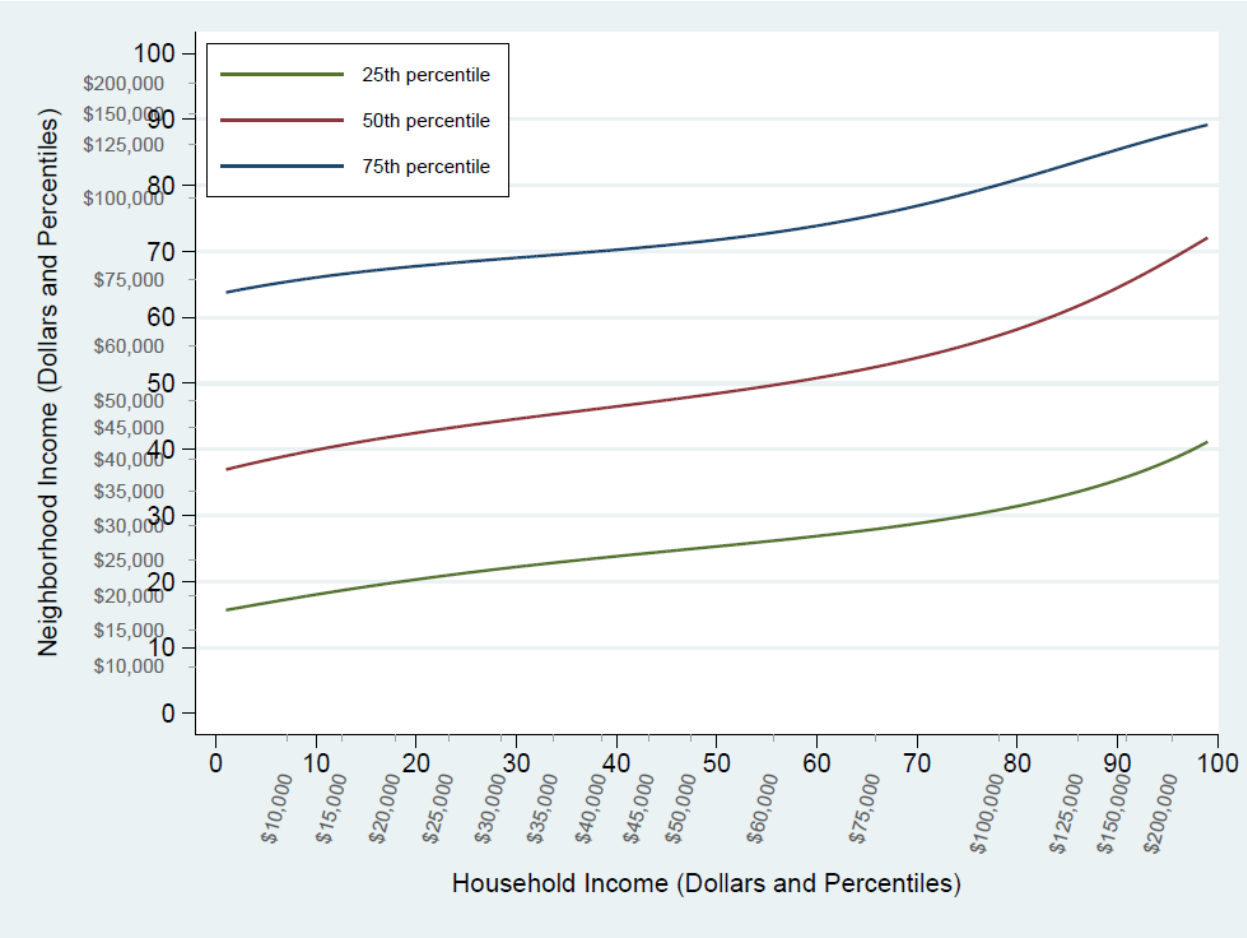
Figures

Figure 1. Race-Specific Income Distributions, All Households in U.S., 2007-2011



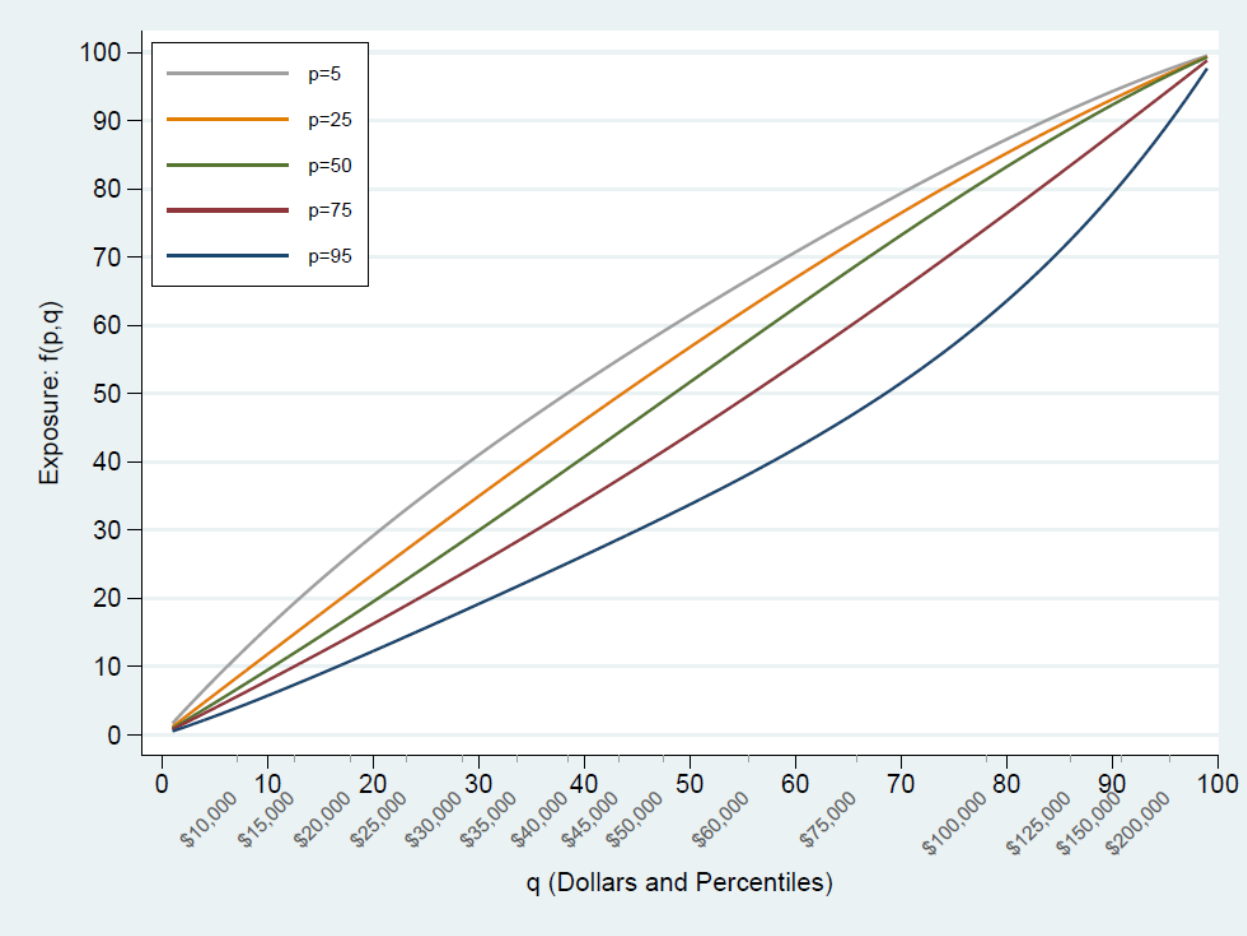
Note. Figure 1 presents the proportion of households in each race group at a given household income percentile for all households in the U.S. for the years 2007-2011. The *x*-axis is household income, on the national scale, in percentiles and dollar figures; the *y*-axis is proportion. The figure shows that, for example, there is a higher proportion of white households than all other groups combined across the distribution of household income, and that there are a higher proportion of white households in the upper end of the income distribution than the lower end.

Figure 2. Percentiles of Average Neighborhood Household Income Distributions, by Own Household Income, 2007-2011



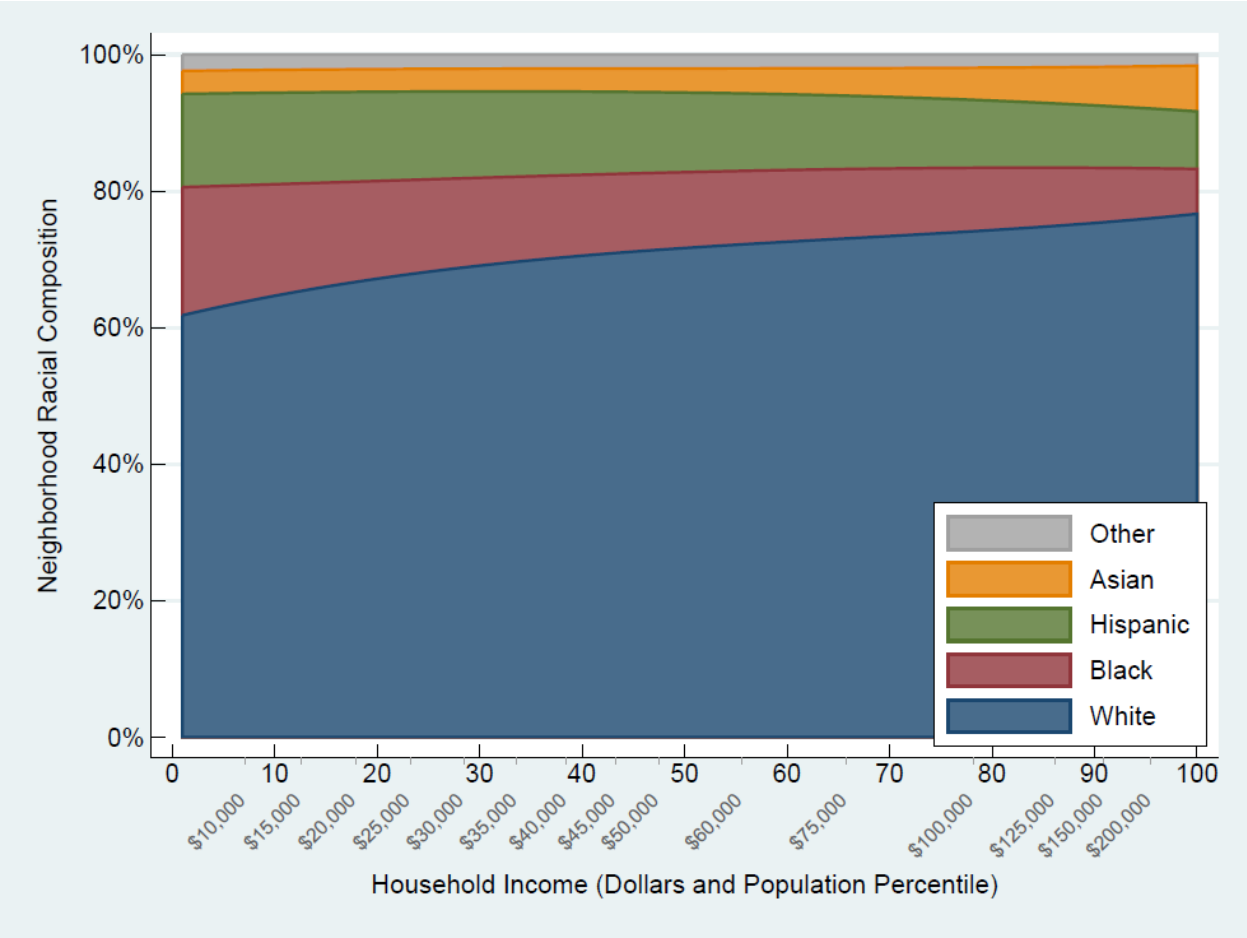
Note. Figure 2 presents neighborhood 25th, 50th, and 75th percentile household income, conditional on own household income, for all households in the U.S. for the years 2007-2011. The *x*-axis indicates household income; the *y*-axis indicates median household income in the neighborhood of a typical household of a given income. For both axes, the percentiles and dollar figures are taken from the national household income distribution. As an example of how to read the figure, consider households earning $60,000/year (roughly the 56th percentile of the household income distribution). Such households live, on average, in neighborhoods where the 25th percentile household income is $27,000 (roughly the 27th percentile of the national household income distribution), the 50th percentile household income is about $53,000 (roughly the 50th percentile of the national household income distribution), and the 75th percentile household income is about $90,000 (roughly the 72nd percentile of the national household income distribution).

Figure 3. Average Cumulative Neighborhood Income Distributions, by Own Household Income, 2007-2011



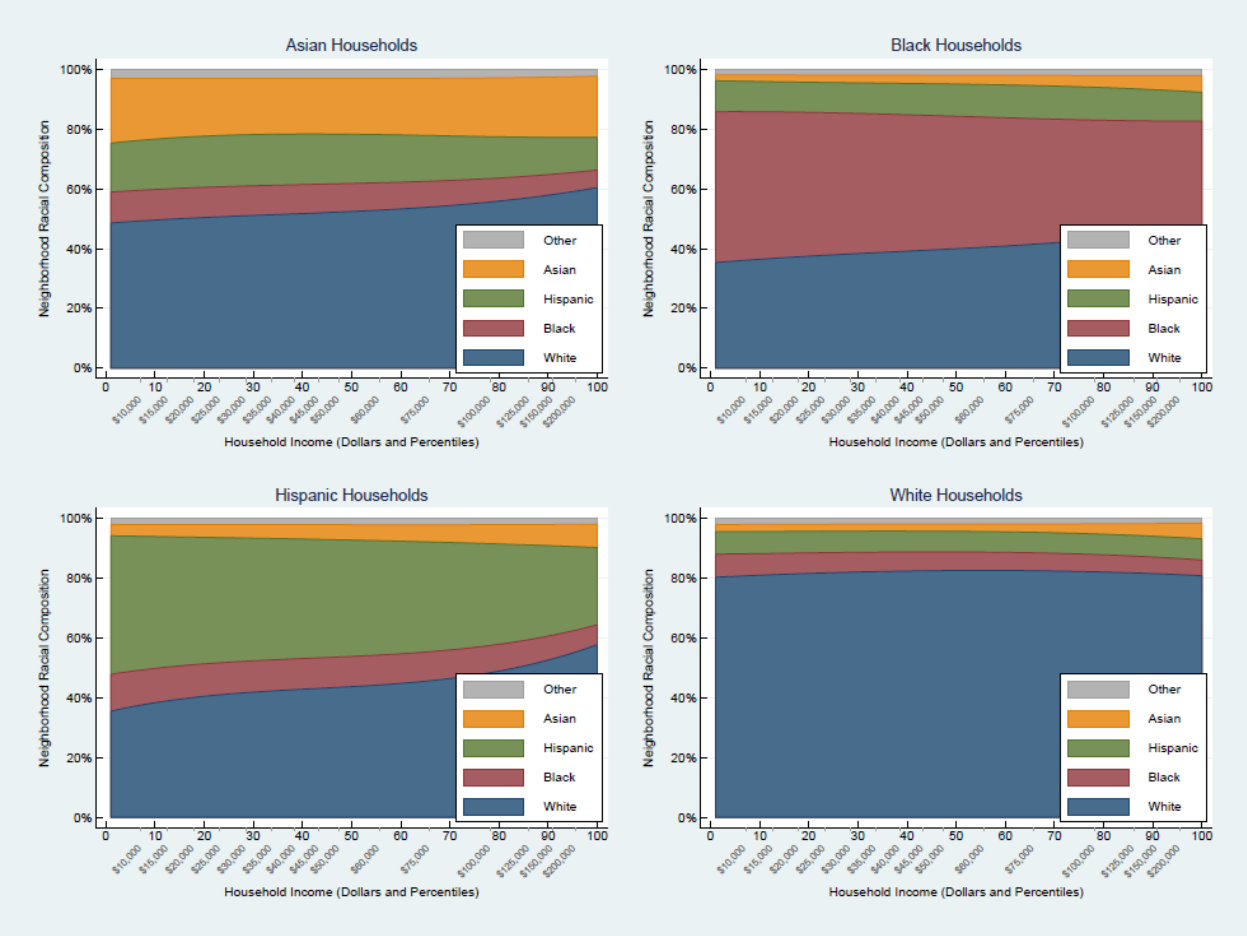
Note. Figure 3 presents the exposure of households with various incomes to the distribution of cumulative household income for all households in the U.S. for the years 2007-2011. The *y*-axis represents the exposure measure, which ranges from 0 (no exposure) to 100 (complete exposure). The x-axis represents the national cumulative household income distribution in both percentiles and selected dollar values. The lines, then, trace the exposure of those with incomes at the 5th, 25th, 50th, 75th, and 95th percentile household incomes to those with incomes equal to or less than the values presented on the *x*-axis. As an example of how to read this figure, consider the red line, and a *y*-axis value of 50. This value of exposure allows us to identify the median income in a given neighborhood – half of all households have incomes below this value, and half have incomes above this value. The red lines suggests that for households that are themselves at the 75th percentile on the household income distribution, the median income in their typical neighborhood is at approximately the 57th percentile of the national household income distribution.

Figure 4. Average Neighborhood Racial Composition, by Household Income, All Households in U.S., 2007-2011



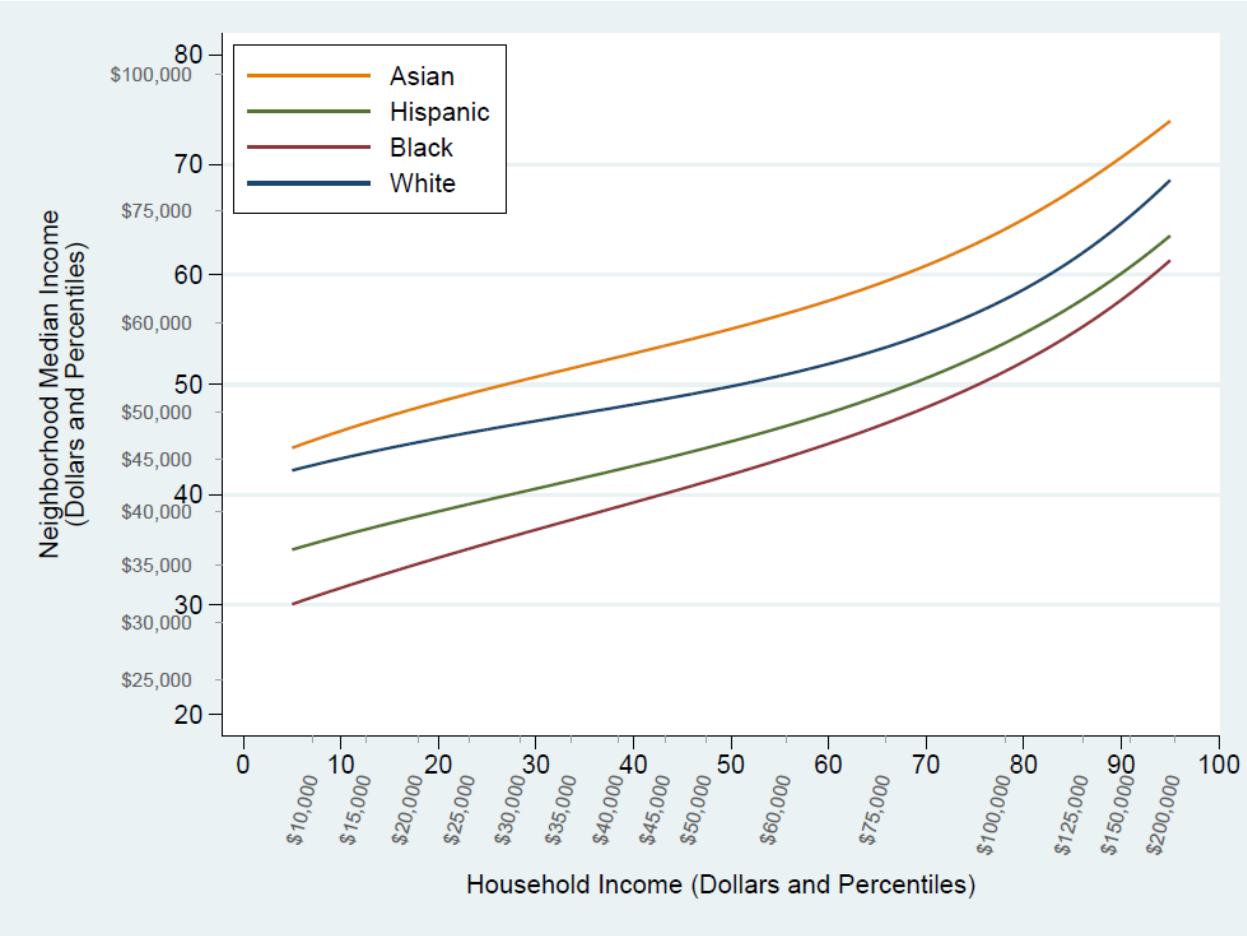
Note. Figure 4 presents the average neighborhood racial composition by household income, for all households in the U.S. for the years 2007-2011. The *x*-axis is household income, on the national scale, in percentiles and dollars. The *y*-axis is proportion. As an example of how to read the figures, consider households at the 40th percentile of the household income distribution. The typical neighborhood household racial composition for households at this income level is close to 64% white, 15% black, 15% Hispanic, 5% Asian, and 1% other.

Figure 5. Average Neighborhood Racial Composition, by Household Income and Race, 2007-2011



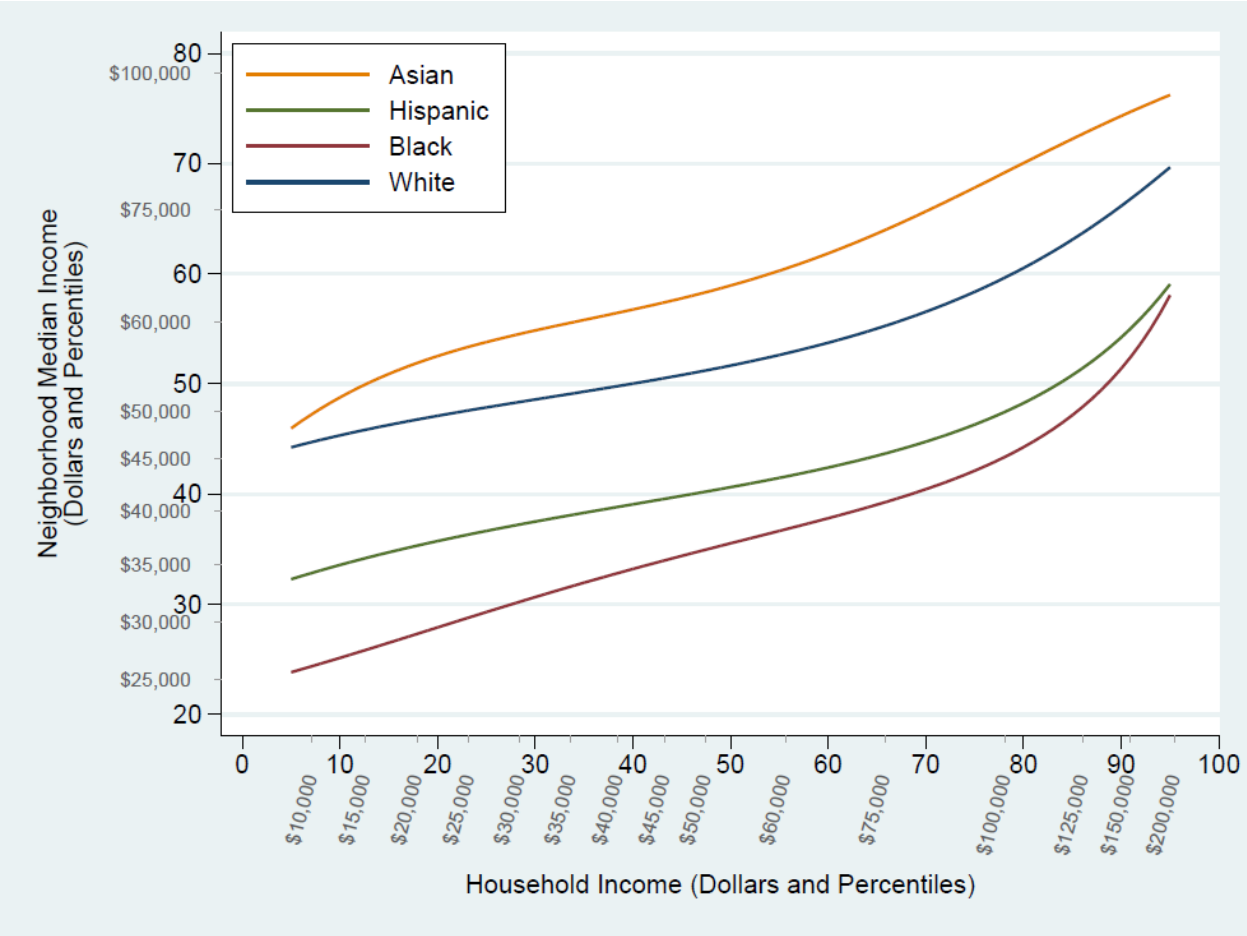
Note. Figure 5 presents four panels, each of which shows the average neighborhood racial composition by household income, for households of a specific racial/ethnic group for the years 2007-2011. The *x*-axis is household income, on the national scale, in percentiles and dollars. The *y*-axis is proportion. Panel 1 presents neighborhood composition for Asian households, panel 2 for black households, panel 3 Hispanic households, and panel 4 white households. As an example of how to read the figure, consider households at the 40th percentile of the household income distribution in panel 1. The typical racial composition in the neighborhoods of Asian households at this income level is close to 50% white, 10% black, 18% Hispanic, 20% Asian, and 2% other.

Figure 6. Neighborhood Median Income, by Household Income and Race, All Households in U.S., 2007-2011



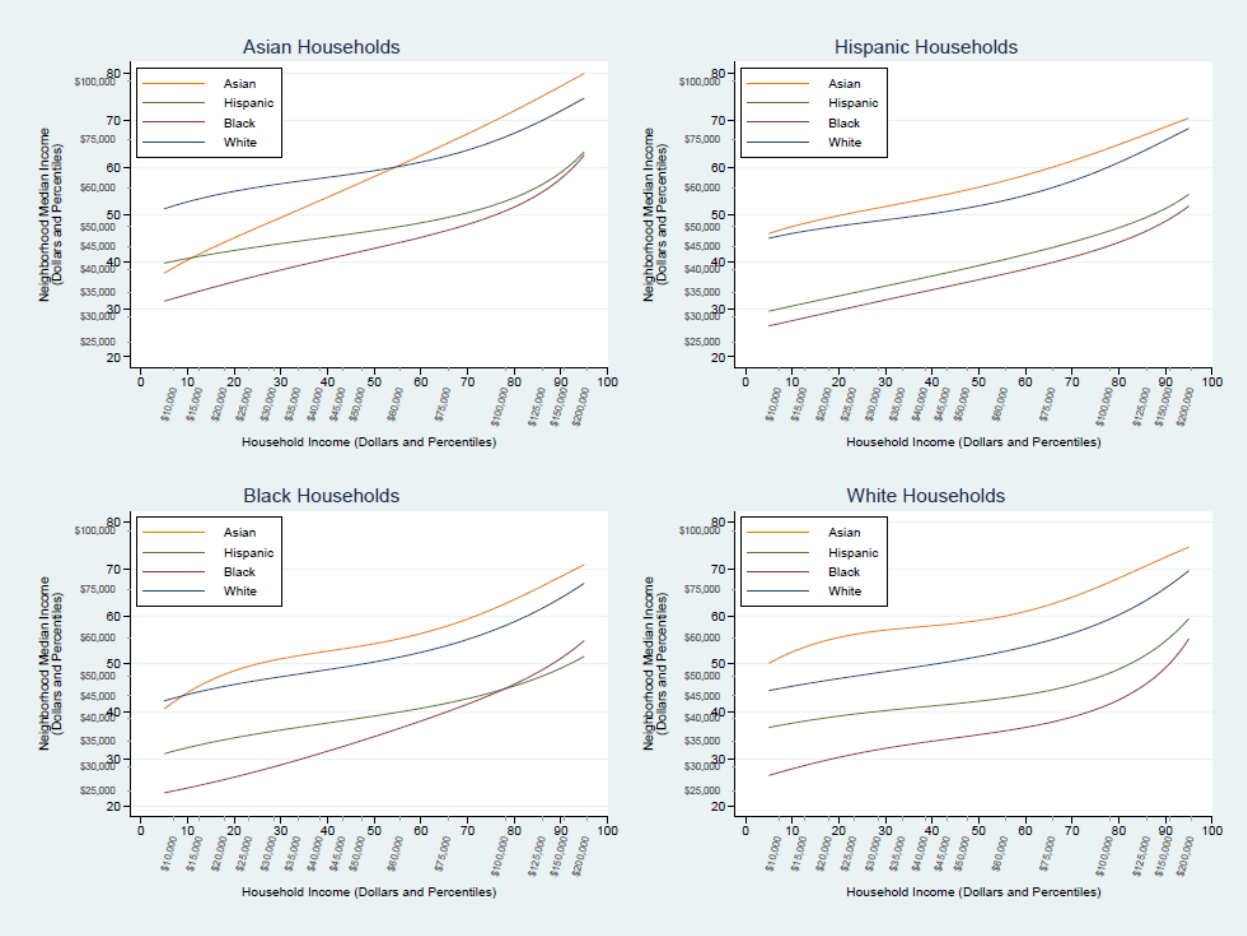
Note. Figure 6 shows neighborhood median household income, conditional on own household income and race/ethnicity, for all households in the U.S. for the years 2007-2011. The *x*-axis is own household income; the *y*-axis is neighborhood median household income. For both axes, the percentiles and dollar figures are taken from the national household income distribution. As an example of how to read the figure, consider White households at the 50th percentile of the national household income distribution. The *y*-axis indicates that such families live, on average, in neighborhoods where the median income is close to $55,000, very close to the median of the national distribution.

Figure 7. Race-specific Neighborhood Median Income, by Household Income, All Households in U.S., 2007-2011



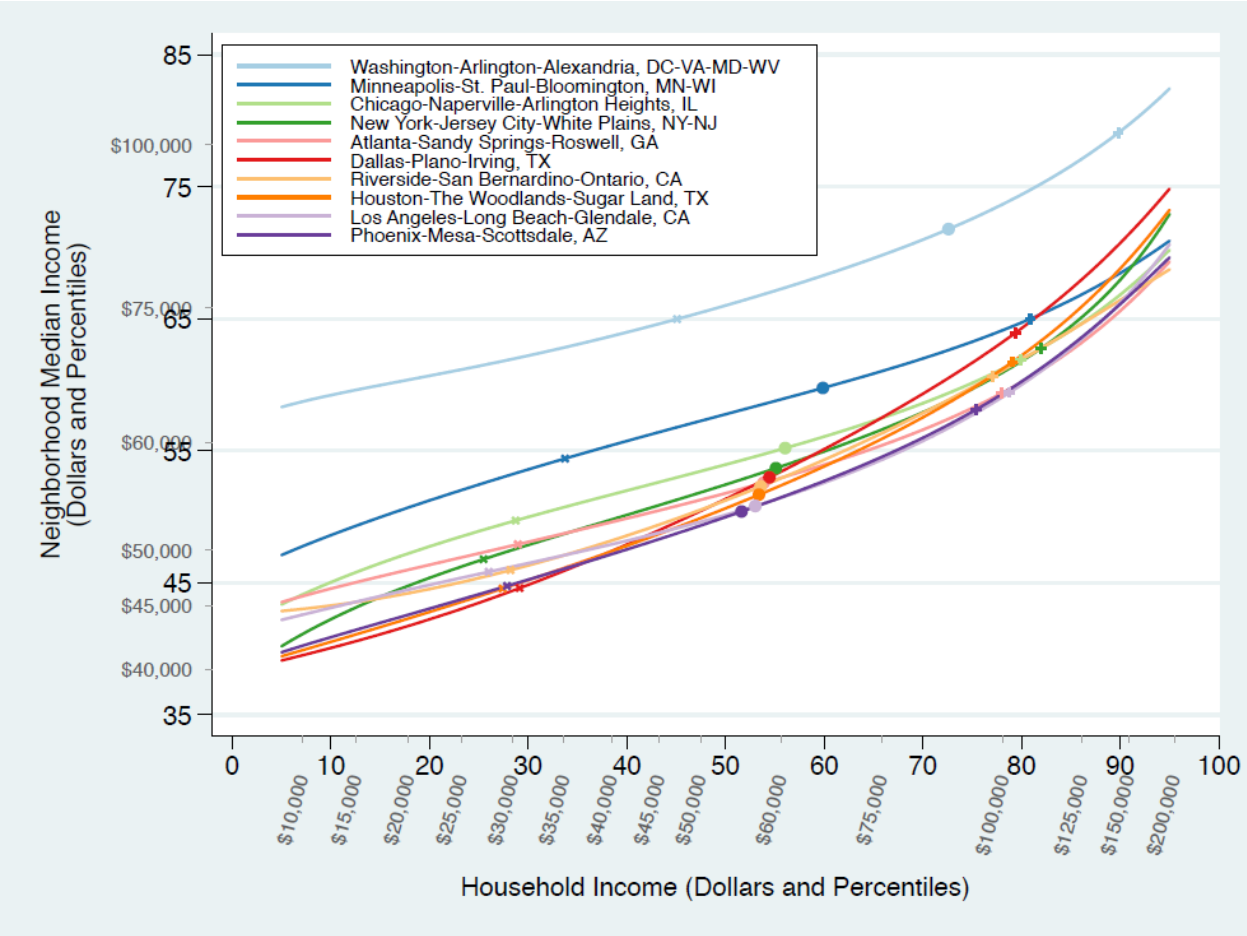
Note. Figure 7 shows race-specific neighborhood median household income, conditional on own household income, for all households in the U.S. for the years 2007-2011. The *x*-axis is own household income; the *y*-axis is neighborhood median household income. For both axes, the percentiles and dollar figures are taken from the national household income distribution. As an example of how to read the figure, consider households at the 50th percentile of the national income distribution (the value of the *x*-axis). The *y*-axis indicates that such families live, on average, in neighborhoods where the median income of blacks is around the 36th percentile of the national income distribution and the median income of Asians is around the 58th percentile.

Figure 8. Race-specific Neighborhood Median Income, by Household Income and Race, 2007-2011



Note. Figure 8 shows race-specific neighborhood median household income, conditional on own household income and race/ethnicity, for the years 2007-2011. The *x*-axis is own household income; the *y*-axis is neighborhood median household income. For both axes, the percentiles and dollar figures are taken from the national household income distribution. As an example of how to read the figure, consider White households (bottom right) at the 50th percentile of the national household income distribution (the value of the *x*-axis). The *y*-axis indicates that such families live, on average, in neighborhoods where the median income of blacks is around the 35th percentile of the national income distribution and the median income of whites is around the 50th percentile.

Figure 9. Metropolitan Variation in Neighborhood Median Income, by Household Income, Ten Largest Metropolitan Areas by Population, 2007-2011



Note. Figure 9 presents neighborhood median household income, conditional on own household income, by metropolitan area for the years 2007-2011. The *x*-axis indicates household income; the *y*-axis indicates median household income in the neighborhood of a typical household of a given income. For both axes, the percentiles and dollar figures are taken from the national household income distribution (not from each metropolitan area). The markers on the lines indicate the 25th, 50th, and 75th percentiles of each metropolitan area's own household income distribution. As an example of how to read the figure, consider households in Minneapolis-St. Paul Bloomington, MN-WI at the 60th percentile of the national income distribution (roughly $66,000). These households typically live in neighborhoods of the Minneapolis-St Paul metropolitan area with median incomes of roughly $64,000, about the 59th percentile of the national income distribution.

Appendix A: Estimating Income Density and Exposure Functions

The estimation of the income density and exposure functions proceeds in three steps. First, we estimate the group-specific income density functions $\rho_g(p)$ and, from them, the group-specific cumulative income distribution functions $R_g(p)$. Second, we estimate the functions $f_t^h(p,q)$ and, from them, the function $f_t^t(p,q)$. Third, we estimate the functions $f_g^h(p,q)$ and, from them, the functions $f_g^t(p,q)$.[3] We do this because the parameter estimates from each step of the model are used to inform the estimation of each subsequent step. Once we have estimated each of these functions, we use them to compute the various exposure functions of interest, as described in the text.

A1. Notation

We use g and h to denote G racial groups; we use p and q to denote income levels, expressed as percentiles (scaled from 0 to 1, for convenience) of the population income distribution; and we use i to index neighborhoods. We use $j, k = 1, \ldots, K$ to index the ordered income categories in which income is reported in the ACS. In the ACS data we use in this paper, there are five mutually exclusive racial/ethnic groups and 16 income categories, so $G = 5$ and $K = 16$ here. Finally we use T_x to denote the count of households in population x, and π_{gi} to indicate the proportion of households in neighborhood i that are in group g.

A2. Data

The data consist of tract-level counts of households of race g with income in category k in census tract i. These counts are denoted T_{igk}. Let $T_{ig\cdot} = \sum_{j=1}^{K} T_{igj}$ denote the total number of households of group g in tract i, let $T_{i\cdot\cdot} = \sum_g T_{ig\cdot}$ denote the total number of households in tract i; and let $T_{\cdot g\cdot} = \sum_i T_{ig\cdot}$ denote the total number of households of group g in the population. The proportion of households of

[3] One approach to estimating $f_g^h(p,q)$ is to estimate $\rho_{gi}(p)$ in each neighborhood i and to estimate $\rho_g(p)$. Then $R_{gi}(p)$ can be estimated as $\hat{R}_{gi}(p) = \int_0^p \hat{\rho}_{gi}(r)dr$. We then estimate $f_g^h(p,q)$ by substituting $\hat{\rho}_{gi}(p)$, $\hat{R}_{hi}(p)$, and the observed group counts and proportions into Equation (1). The potential drawback of this approach is that it requires us to estimate ρ_{gi} from small samples in each neighborhood i and group g. Instead, we adopt an alternative approach, described in detail in the remainder of this appendix (Appendix A).

group g in tract i with incomes in income category k is

$$r_{igk} = \frac{T_{igk}}{T_{ig\cdot}}. \tag{A1}$$

The proportion of households of group g in the population with incomes in income category k is

$$r_{\cdot gk} = \frac{\sum_i T_{igk}}{T_{\cdot g\cdot}}. \tag{A2}$$

We denote the corresponding proportion of households with incomes in category k or below as $r_{ig\leq k} = \sum_{j=1}^{k} r_{igj}$ and $r_{\cdot g\leq k} = \sum_{j=1}^{k} r_{\cdot gj}$, respectively. Finally, the proportion of the total population of households that have income in income category k is denoted p_k or q_k:

$$p_k = q_k = \frac{\sum_i \sum_{g=1}^{G} T_{igj}}{\sum_i \sum_{g=1}^{G} \sum_{j=1}^{K} T_{igj}}. \tag{A3}$$

A3. Computing Exposure Measures

From the ACS data, we compute $G^2K^2 = 6{,}400$ values of f_{gj}^{hk} (for each combination of g, h, j, k, where g and h index five racial groups and j and k index 16 income categories). Each f_{gj}^{hk} is an exposure index of members of group g with income in category j to members of group h in income category k or below. These are computed from the tract-by-group-by-income category counts as

$$f_{gj}^{hk} = \sum_i \left[\frac{T_{ig\cdot} \cdot r_{igj}}{T_{\cdot g\cdot} \cdot r_{\cdot gj}} \cdot \frac{T_{ih\cdot} r_{ih\leq k}}{T_{i\cdot\cdot}} \right]. \tag{A4}$$

We also compute $GK^2 = 1{,}280$ values of f_{tj}^{hk} (for each combination of h, j, k). Each f_{tj}^{hk} is an exposure index of members of the population with income in category j to members of group h in income category k or below. These are computed from the tract-by-group-by-income category counts as

$$f_{tj}^{hk} = \sum_i \left[\frac{\sum_g \left(T_{ig\cdot} \cdot r_{igj} \right)}{\sum_g \left(T_{\cdot g\cdot} \cdot r_{\cdot gj} \right)} \cdot \frac{T_{ih\cdot} r_{ih\leq k}}{T_{i\cdot\cdot}} \right]. \tag{A5}$$

A4. Estimating $R_g(p)$ and $\rho_g(p)$

We assume the group-specific income density functions (the $\rho_g(p)$'s) can be well-approximated as polynomials of order C:

$$\rho_g(p) \approx \sum_{c=0}^{C} \alpha_c^g p^c. \tag{A6}$$

While this may be unrealistic if income were measured in dollars (because the long right tail of the cumulative income distribution function cannot be modeled well as a polynomial), it is much less problematic when p measures income in percentiles of the population income distribution. In this case, the $\rho_g(p)$ functions must satisfy a few conditions. Specifically,

$$\int_0^1 \rho_g(r)dr = 1;$$
$$\sum_{g=1}^{G} \pi_g \rho_g(p) = \rho(p) = 1. \tag{A7}$$

The first condition in (A7) is simply the property that the density function have integral 1. The second follows from the fact that the population income percentile distribution is, by definition, uniform. These two conditions imply, respectively, that

$$\sum_{c=0}^{C} \frac{1}{c+1} \alpha_c^g = 1 \quad \forall\, g,$$

and

$$\sum_{g=1}^{G} \pi_g \alpha_c^g = \begin{cases} 1 \; if \; c = 0 \\ 0 \; if \; c \neq 0 \end{cases}. \tag{A8}$$

Therefore, we can estimate the $\rho_g(p)$ functions by fitting the following regression model to the $GK = 80$ points $(\rho_{\cdot gk}, m_k)$:

$$\rho_{\cdot gk} = \sum_{g=1}^{G} \sum_{c=1}^{C} \alpha_c^g \left(D_g \cdot m_k^c\right) + e_{gk},$$

(A9)

where D_g is an indicator variable taking the value 1 if an observation pertains to group g and 0 otherwise. Let $m_k = \frac{1}{2}(p_{k-1} + p_k)$ be the percentile that falls in the middle of category k.[4] For the analyses reported in this paper, we set $C = 3$. Inspection of the fitted income density functions indicates very good fit with $C = 3$.

In fitting the model, we enforce the following $G + C + 1$ linear constraints:

$$\sum_{c=0}^{C} \frac{1}{c+1} \alpha_c^g = 1 \quad \forall\, g,$$

and

$$\sum_{g=1}^{G} \pi_g \alpha_c^g = \begin{cases} 1 \; if \; c = 0 \\ 0 \; if \; c \in \{1, \dots, C\} \end{cases}.$$

(A10)

This set of constraints is sufficient to satisfy the conditions in (A7). One of the constraints is redundant, so a total of $G + C$ constraints are used to estimate the density functions.

For each group g, the estimated income density function is then a Cth-order polynomial:

$$\hat{\rho}_g(p) = \sum_{c=0}^{C} \hat{\alpha}_c^g p^c$$

(A11)

From $\hat{\rho}_g(p)$ it is straightforward to estimate the group-specific cumulative income distribution function $R_g(p)$ by taking the integral of $\hat{\rho}_g(p)$ on the interval $[0,1]$:

$$\hat{R}_g(p) = \int_0^1 \hat{\rho}_g(r) dr = \sum_{c=0}^{C} \frac{1}{c+1} \hat{\alpha}_c^g p^{c+1}.$$

(A12)

[4] In practice, when we fit (A9), we replace the midpoint m_j^c with $m_j^{c*} = m_j^c + z_{cj}$, where z_{cj} is defined as in Appendix B. In addition, we fit (A9) using weighted least squares regression, with weights as detailed in Appendix B.

A5. Estimating $f_t^h(p,q)$ and $f_t^t(p,q)$

Before estimating the $f_t^h(p,q)$ and $f_t^t(p,q)$ functions, note that it follows from the definition of $f_t^h(p,q)$ (Equation 3 in the paper) that

$$f_t^t(p,q) = \sum_{h=1}^{G} f_t^h(p,q). \tag{A13}$$

Next, note that the $f_t^h(p,q)$ and $f_t^t(p,q)$ functions must satisfy four conditions, by definition:

$$f_t^h(p,0) = 0$$

$$\int_0^1 f_t^h(r,q)dr = \pi_h R_h(q);$$

$$\frac{d}{dq} f_t^t(p,q) = f_t'^t(p,q) = f_t'^t(q,p) = \frac{d}{dp} f_t^t(q,p)$$

$$f_t^t(p,1) = \sum_{h=1}^{G} f_t^h(p,1) = 1. \tag{A14}$$

The first of these simply states that the exposure of any subset of households to the members of another group h with incomes less than or equal to 0, is by definition 0. The second says that, on average, neighborhoods have the same income distribution of each group h as the total population. The third is a symmetry condition that follows from the definition of $f_t^t(p,q)$.[5] And the fourth says that the exposure of any subset of households to households with incomes less than or equal to 1, is by definition 1.

We assume the $f_t^h(p,q)$ functions are well-approximated as polynomial surfaces of order A in p and B in q:

$$f_t^h(p,q) = \sum_{a=0}^{A} \sum_{b=0}^{B} \gamma_{ab}^{th} p^a q^b.$$

[5] To see this, note that $\frac{d}{dq} f_t^t(p,q)$ is the exposure of individuals with income of exactly p to those with income of exactly q: $\frac{d}{dq} f_t^t(p,q) = \sum_i \left[\frac{T_i \rho_i(p)}{T\rho(p)} \cdot \frac{T_i \rho_i(q)}{T_i}\right] = \sum_i \left[\frac{T_i \rho_i(q)}{T\rho(p)} \cdot \frac{T_i \rho_i(p)}{T_i}\right] = \frac{d}{dp} f_t^t(q,p)$.

(A15)

We estimate these functions by fitting the following regression model to the GK^2 points (f_{tj}^{hk}, m_j, q_k):

$$f_{tj}^{hk} = \sum_{h} \sum_{a=0}^{A} \sum_{b=0}^{B} \gamma_{ab}^{th} \left[D_{th} \cdot m_j^a \cdot q_k^b \right] + u_{thpq},$$

(A16)

where D_{th} is an indicator that an observation pertains to exposure to group h.[6]

To ensure that the estimated functions satisfy the conditions in (A14), we impose a set of constraints on the model. Specifically, we impose the following constraints:

$$\gamma_{a0}^{th} = 0 \;\; \forall a \in \{1, \dots, A\}, h \in \{1, \dots, G\}$$

$$A + 1 = B = C + 1;$$

$$\sum_{a=0}^{A} \frac{1}{a+1} \gamma_{ab}^{th} = \frac{1}{b} \pi_h \alpha_{b-1}^h \;\; \forall b \in \{1, \dots, B\}, h \in \{1, \dots, G\};$$

$$\sum_{h=1}^{G} b \gamma_{ab}^{th} = \sum_{h=1}^{G} (a+1) \gamma_{(b-1)(a+1)}^{th} \;\; \forall\, a \in \{1, \dots, A\}, b \in \{1, \dots, a\}.$$

(A17)

Together, these constraints ensure that the conditions in (A14) are met. The first constraint in (A17) implies the first condition of (A14) is met:

$$f_t^h(p, 0) = \sum_{a=0}^{A} \sum_{b=0}^{B} \gamma_{ab}^{th} p^a 0^b = \sum_{a=0}^{A} \gamma_{a0}^{th} p^a = \sum_{a=0}^{A} 0 \cdot p^a = 0.$$

(A18)

Note that the constraint that $\gamma_{a0}^{th} = 0$ for all h and a implies that the γ_{a0}^{th} terms can be omitted from (A16) going forward.

Second, the constraints in (A17), along with (A12) imply the second condition in (A14):

[6] As above, in fitting this model, we replace the midpoints m_j^c with $m_j^c + z_j^c$ and use a weighted least square regression as detailed in Appendix B.

$$\int_0^1 f_t^h(r,q)dr = \sum_{b=1}^{B}\left[\sum_{a=0}^{A}\frac{\gamma_{ab}^{th}}{a+1}\right]q^b = \sum_{b=1}^{B}\left[\frac{\pi_h}{b}\alpha_{b-1}^h\right]q^b = \pi_h\sum_{c=0}^{C}\frac{1}{c+1}\alpha_c^h q^{c+1} = \pi_h R_h(q) \tag{A19}$$

Third, the constraints in (A17) imply the third condition in (A14):

$$\begin{aligned}
\frac{d}{dq}f_t^t(p,q) &= \sum_{a=0}^{A}\sum_{b=1}^{B} b\gamma_{ab}^{tt}p^a q^{b-1} \\
&= \sum_{a=0}^{A}\sum_{b=1}^{B} b\left[\sum_{h=1}^{G}\gamma_{ab}^{th}\right]p^a q^{b-1} \\
&= \sum_{a=0}^{A}\sum_{b=1}^{B}(a+1)\left[\sum_{h=1}^{G}\gamma_{(b-1)(a+1)}^{th}\right]p^a q^{b-1} \\
&= \sum_{a=0}^{A}\sum_{b=0}^{A}(a+1)\left[\sum_{h=1}^{G}\gamma_{b(a+1)}^{th}\right]p^a q^{b} \\
&= \sum_{a=0}^{A}\sum_{b=0}^{A}(b+1)\left[\sum_{h=1}^{G}\gamma_{a(b+1)}^{th}\right]p^b q^{a} \\
&= \sum_{a=0}^{A}\sum_{b=1}^{B} b\left[\sum_{h=1}^{G}\gamma_{ab}^{th}\right]p^{b-1} q^{a} \\
&= \sum_{a=0}^{A}\sum_{b=1}^{B} b\gamma_{ab}^{tt}q^a p^{b-1} \\
&= \frac{d}{dp}f_t^t(q,p)
\end{aligned} \tag{A20}$$

Finally, the constraints in (A17), in conjunction with (A8), ensure that the fourth condition in (A14) is met:

$$\begin{aligned}
f_t^t(p,1) &= \sum_{h=1}^{G} f_t^h(p,1) \\
&= \sum_{h=1}^{G}\sum_{a=0}^{A}\sum_{b=1}^{A+1}\gamma_{ab}^{th}p^a 1^b \\
&= \sum_{a=0}^{A}\sum_{b=1}^{A+1}\left[\sum_{h=1}^{G}\gamma_{ab}^{th}\right]p^a
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{a=0}^{A}\sum_{b=1}^{A+1}\left[\frac{a+1}{b}\sum_{h=1}^{G}\gamma_{(b-1)(a+1)}^{th}\right]p^a \\
&= \sum_{a=0}^{A}\sum_{b=0}^{A}\left[\frac{a+1}{b+1}\sum_{h=1}^{G}\gamma_{b(a+1)}^{th}\right]p^a \\
&= \sum_{a=0}^{A}\sum_{h=1}^{G}\left[\sum_{b=0}^{A}\frac{a+1}{b+1}\gamma_{b(a+1)}^{th}\right]p^a \\
&= \sum_{a=0}^{A}\sum_{h=1}^{G}[\pi_h\alpha_a^h]p^a \\
&= \sum_{a=0}^{A}\left[\sum_{h=1}^{G}\pi_h\alpha_a^h\right]p^a \\
&= \sum_{a=0}^{0}p^a \\
&= 1
\end{aligned}
\tag{A21}
$$

Thus, the constraints in (A17) and (A10) are sufficient to ensure the conditions in (A14) are satisfied. We satisfy the first set of constraints in (A17) by setting $\gamma_{a0}^{th} = 0$ for all a and g. After we set $A + 1 = B = C + 1$, the last two condition in (A17) contain an additional $GB + AB/2$ constraints that are required to estimate $f_t^h(p,q)$. Because $f_t^h(p,q)$ has GB^2 total parameters, there are a total of $AB\left[\frac{2G-1}{2}\right]$ free parameters in the model. With $G = 5$ and $A + 1 = B = 4$, $f_t^h(p,q)$ has 80 parameters, subject to 26 constraints, for a total of 54 freely estimated parameters.

Once we have estimated $f_t^h(p,q)$, it is straightforward to estimate

$$
f_t^t(p,q) = \sum_h f_t^h(p,q). \tag{A22}
$$

A6. Estimating $f_g^h(p,q)$ and $f_g^t(p,q)$

Estimating the functions $f_g^h(p,q)$ and $f_g^t(p,q)$ follows the same logic as above, with some modifications. Recall that above we require that $A + 1 = B = C + 1$. As above, we assume the functions $f_g^h(p,q)$ can be well-approximated as polynomial surfaces of order A in p and $B = A + 1$ in q:

$$f_g^h(p,q) = \sum_{a=0}^{A} \sum_{b=0}^{A+1} \gamma_{ab}^{gh} p^a q^b. \tag{A23}$$

The $f_g^h(p,q)$ and $f_g^h(p,q)$ functions must satisfy several conditions:

$$f_g^t(p,0) = f_g^h(p,0) = 0;$$

$$f_g^t(p,1) = \sum_{h=1}^{G} f_g^h(p,1) = 1;$$

$$f_t^h(p,q) = \sum_{g=1}^{G} \pi_g \rho_g(p) f_g^h(p,q)\,;$$

$$\int_0^1 f_t^h(r,q)dr = \sum_{g=1}^{G} \pi_g \int_0^1 \rho_g(r) f_g^h(r,q)dr = \pi_h R_h(q). \tag{A24}$$

These are satisfied with the following constraints:

$$\gamma_{a0}^{gh} = 0 \;\; \forall a, g, h;$$

$$\sum_{h=1}^{G} \sum_{b=1}^{A+1} \gamma_{ab}^{gh} = \begin{cases} 1 \; if \; a = 0 \\ 0 \; if \; a > 0 \end{cases} \quad \forall \; g;$$

and

$$\gamma_{ab}^{th} = \sum_{g} \pi_g \sum_{c=\max(0,a-A)}^{\min(A,a)} \alpha_c^g \gamma_{(a-c)b}^{gh} \;\; \forall b \in \{1,\dots,B\}, h \in \{1,\dots,G\}, a \in \{0,\dots,\dots,2A\} \tag{A25}$$

where $\gamma_{ab}^{th} = 0$ for all $a \in \{A+1,\dots,2A\}$. The second and third lines of (A25) contain a total of $G(A+1)(2A+2) = 2GB^2$ constraints (160 constraints in our example with $G = 5$ and $B = 4$).

The first condition in (A24) follows from the constraint that $\gamma_{a0}^{gh} = 0$ for all a, g, and h:

$$f_g^h(p,0) = \sum_{a=0}^{A} \sum_{b=0}^{B} \gamma_{ab}^{gh} p^a 0^b = \sum_{a=0}^{A} \gamma_{a0}^{gh} p^a = \sum_{a=0}^{A} (0 \cdot p^a) = 0.$$

(A26)

As above, this implies we can omit the γ_{a0}^{gh} terms from the model. The second condition follows from second constraint:

$$\sum_{h=1}^{G} f_g^h(p,1) = \sum_{h=1}^{G}\sum_{a=0}^{A}\sum_{b=1}^{A+1} \gamma_{ab}^{gh} p^a 1^b = \sum_{a=0}^{A}\left[\sum_{h=1}^{G}\sum_{b=1}^{A+1} \gamma_{ab}^{gh}\right] p^a = p^0 = 1$$

(A27)

The third condition follows from the third constraint:

$$\begin{aligned}
\sum_{g=1}^{G} \pi_g \rho_g(p) f_g^h(p,q) &= \sum_{g=1}^{G} \pi_g \left[\sum_{c=0}^{A} \alpha_c^g p^c\right] \sum_{a=0}^{A}\sum_{b=1}^{A+1} \gamma_{ab}^{gh} p^a q^b \\
&= \sum_{a=0}^{A}\sum_{b=1}^{A+1}\sum_{c=0}^{A}\sum_{g=1}^{G} \pi_g \alpha_c^g \gamma_{ab}^{gh} p^{a+c} q^b \\
&= \sum_{a=0}^{2A}\sum_{b=1}^{A+1}\left[\sum_{g=1}^{G} \pi_g \sum_{c=\max(0,a-A)}^{\min(A,a)} \alpha_c^g \gamma_{(a-c)b}^{gh}\right] p^a q^b \\
&= \sum_{a=0}^{A}\sum_{b=1}^{A+1} \gamma_{ab}^{th} p^a q^b \\
&= f_t^h(p,q)
\end{aligned}$$

(A28)

The fourth condition follows from (A28) and (A19):

$$\sum_{g=1}^{G} \pi_g \int_0^1 \rho_g(r) f_g^h(r,q)dr = \int_0^1 f_t^h(r,q)dr = \pi_h R_h(q).$$

(A29)

While the constraints in (A25) are sufficient to satisfy the conditions in (A24), in practice, we use a subset of constraints that are implied by those in (A25) for computational ease. Specifically, we use first and second sets of constraints from (A25) and an additional set of GB constraints that are implied by

those in the third line of (A25) and the third line of (A17):

$$\sum_{a=0}^{A} \frac{1}{a+1} \sum_{g} \pi_g \sum_{c=0}^{a} \alpha_c^g \gamma_{(a-c)b}^{gh} = \frac{1}{b} \pi_h \alpha_{b-1}^h \quad \forall b, h. \tag{A30}$$

Once we have constrained $\gamma_{a0}^{gh} = 0$ for all a, g, h, the second line of (A25) implies GB constraints. One of the GB constraints in (A30) is redundant, so we invoke a set of $2GB - 1$ total constraints in fitting the $f_g^h(p, q)$ functions.

We estimate the $f_g^h(p, q)$ functions by simultaneously fitting a set of G^2 separate polynomial surfaces of order A in p and order $A + 1$ in b through the G^2K^2 points (f_{gj}^{hk}, m_j, q_k), subject to the constraints described above:

$$f_{gj}^{hk} = \sum_{g} \sum_{h} \sum_{a=0}^{A} \sum_{b=1}^{B} \gamma_{ab}^{gh} \left[D_{gh} \cdot m_j^a \cdot q_k^b \right] + u_{ghpq}, \tag{A31}$$

where D_{gh} is an indicator variable taking the value 1 if an observation pertains to the exposure of group g to group h, and 0 otherwise.[7] Given $G = 5$ and $A + 1 = B = 4$, the functions have a total of 400 parameters, which are subject to 39 linear constraints (159 if we use the full set of constraints in (A25)).

Once we have estimated $f_g^h(p, q)$, it is straightforward to estimate

$$f_g^t(p, q) = \sum_{h} f_g^h(p, q) \tag{A32}$$

which implies that

$$\gamma_{ab}^{gt} = \sum_{h} \gamma_{ab}^{gh}. \tag{A33}$$

[7] As above, in fitting this model, we replace the midpoints m_j^c with $m_j^c + z_j^c$ and use a weighted least square regression as detailed in Appendix B.

Appendix B: Estimating a nonlinear association when the regressor is measured ordinally

This appendix describes one solution to the following general problem: we want to estimate a nonlinear polynomial function describing the conditional mean (given X) of a variable Y when X is measured in a set of ordered categories rather than continuously.

Some notation

As above, suppose income, a continuous variable denoted by X, is categorized into K categories, defined by $K-1$ ordered thresholds $c_1, c_2, \dots, c_{K-1}$. Instead of observing X, we instead observe $c \in \{1, \dots, K\}$ where $c = j \;\; iff \;\; c_{j-1} < X \leq c_j$, where $c_0 = -\infty$ and $c_K = +\infty$. In addition, let p denote income in percentile ranks, scaled from 0 to 1 (so that $p = CDF(x)$ and $p_j = CDF(c_j)$ for $j \in \{0, \dots, K\}$, where $CDF(x)$ is the cumulative income distribution function in the population of interest). Let $m_j = \frac{1}{2}(p_{j-1} + p_j)$ be the percentile that falls in the middle of category j. Let $w_j = \frac{1}{2}(p_j - p_{j-1})$ be half the width of income category j. Note that since p measures income percentile ranks in the population of interest, p is uniformly distributed on the interval $[0,1]$ and its density function is $\rho(p) = 1$, by definition.

Let Y measure some characteristic of an individual, where Y may be binary or continuous. Our goal is to estimate the function $f(p) = E[Y|p]$ describing the conditional expectation of Y given p, despite the fact that we only observe c and Y. Our approach is the following: 1) estimate the mean value of Y (and its sampling variance) among individuals in each income category j; denote these $\hat{\bar{Y}}_j$ and $\hat{v}_j$, respectively; 2) assign income category j a value of p equal to $m_j = \frac{1}{2}(p_{j-1} + p_j)$, the midpoint of the interval (p_{j-1}, p_j); 3) regress $\hat{\bar{Y}}_j$ on a polynomial function of m_j using weighted least squares regression, weighting the observations by $1/\hat{v}_j$.

One complication that arises is that, if the function $f(p)$ is nonlinear, then $E[Y|c = j] \neq E[Y|p = m_j]$. That is, the mean value of Y within an income category will not necessarily equal the mean value of Y among those with incomes at the exact midpoint of the income category. In the curvature of

$f(p)$ is substantial and/or the income categories are wide, simply regressing $\hat{\bar{Y}}$ on m may lead to bias in the estimated β_a's. To remedy this potential bias, we make an adjustment to the m_j^a's.

Suppose that $f(p)$ is well-approximated by a polynomial of order A:

$$f(p) \approx \sum_{a=0}^{A} \beta_a p^a. \tag{B1}$$

We can express the average value of Y in category j as

$$\bar{Y}_j = \frac{\int_{p_{j-1}}^{p_j} \rho(r) f(r) dr}{\int_{p_{j-1}}^{p_j} \rho(r) dr}. \tag{B2}$$

Because $\rho(p) = 1$, this is

$$\begin{aligned}
\bar{Y}_j &= \frac{\int_{p_{j-1}}^{p_j} \sum_{a=0}^{A} \beta_a r^a \, dr}{p_j - p_{j-1}} \\
&= \sum_{a=0}^{A} \beta_a \frac{\left(p_j^{a+1} - p_{j-1}^{a+1}\right)}{(a+1)\left(p_j - p_{j-1}\right)} \\
&= \sum_{a=0}^{A} \frac{\beta_a}{(a+1)} \sum_{b=0}^{a} p_j^{a-b} p_{j-1}^{b} \\
&= \sum_{a=0}^{A} \frac{\beta_a}{(a+1)} \sum_{b=0}^{a} \left(m_j + w_j\right)^{a-b} \left(m_j - w_j\right)^{b} \\
&= \sum_{a=0}^{A} \beta_a \left[m_j^a + \frac{1}{a+1} \sum_{b=0}^{a} \left(m_j + w_j\right)^{a-b} \left(m_j - w_j\right)^{b} - m_j^a \right] \\
&= \sum_{a=0}^{A} \beta_a \left[m_j^{a*} \right]
\end{aligned} \tag{B3}$$

where $m_j^{a*} = m_j^a + z_{aj}$ and

$$z_{aj} = \frac{1}{a+1}\sum_{b=0}^{a}\left[\left(p_j\right)^{a-b}\left(p_{j-1}\right)^{b} - m_j^a\right] = \frac{1}{a+1}\sum_{b=0}^{a}\left[\left(m_j + w_j\right)^{a-b}\left(m_j - w_j\right)^{b} - m_j^a\right].$$

(B4)

Note that the z_{aj}'s in (B4) can be simplified. For example, for $a \in (0,1,2,3,4)$, we get

$$z_{0j} = 0$$

$$z_{1j} = 0$$

$$z_{2j} = \frac{w_j^2}{3}$$

$$z_{3j} = m_j w_j^2$$

$$z_{4j} = 2m_j^2 w_j^2 + \frac{1}{5}w_j^4$$

(B5)

(B3) implies that $\bar{Y}_j = E[Y|c = j]$ is a simple polynomial function of m_j unless $A = 0$ or $A = 1$ (that is, unless $f(p)$ is a linear function). If $f(p)$ is non-linear, $\bar{Y}_j$ is a linear combination of $m_j^{0*}, m_j^{1*}, \dots, m_j^{A*}$. As a result, we can estimate $f(p)$ by regressing $\hat{\bar{Y}}_j$ on the m_j^{a*}'s rather than on the m_j^{a}'s:

$$\hat{\bar{Y}}_j = \sum_{a=0}^{A}\beta_a m_j^{a*} + u_j, \;\; u_j \sim N\left[0, \hat{v}_j\right].$$

(B6)

In (A9) above, $\hat{\bar{Y}}_j = \pi_{gj}$, the proportion of households in income category j who are members of group g. In (A16) and (A33) $\hat{\bar{Y}}_j = f_{tj}^{hk}$, the average proportion one's neighbors who are members of group h and who have incomes at or below some category k. The sampling variance of $\hat{\bar{Y}}_j$ in either case will be proportional to the width of the income category (because this is proportional to the number of households in that category in the population) and $\hat{\bar{Y}}_j\left(1 - \hat{\bar{Y}}_j\right)$ (the variance of a proportion). Because the estimates of WLS are invariant under a linear scaling of the weights, we set $\hat{v}_{gj} = w_j\hat{\bar{Y}}_{gj}\left(1 - \hat{\bar{Y}}_{gj}\right)$.

Appendix C: Other Quantities of Interest

Given $f_g^h(p,q)$ for all groups g and h, we can derive a number of additional useful quantities. Several of these are described in the text. Here, we describe two additional quantities of interest.

Standard exposure measures. We can obtain additional exposure measures, such as the exposure of members of group g with incomes between p_{min} and p_{max} to members of group h in with incomes between some q_{min} and q_{max}, by computing

$$\frac{\int_{p_{min}}^{p_{max}} \rho_g(r)\left[f_g^h(r,q_{max}) - f_g^h(r,q_{min})\right]dr}{\int_{p_{min}}^{p_{max}} \rho_g(r)dr}. \tag{C1}$$

A useful special case of this is the exposure of those in group g with income less than or equal to p to those in group h with income less than or equal to q. Denoted $F_g^h(p,q)$, this is

$$F_g^h(p,q\,) = \frac{\int_0^p \rho_g(r) f_g^h(r,q)dr}{\int_0^p \rho_g(r)dr} \tag{C2}$$

For example, the exposure of group g to poor neighbors would be $F_g^t(1,q_{poverty})$, where $q_{poverty}$ is the income value that corresponds to the poverty line. Thus, measures of "exposure to poverty" used in much of the segregation literature (Logan, 2011; Timberlake, 2002) are special cases of the measurement approach we describe here. Note that in the special case where $p = q = 1$, $F_g^h(1,1)$ is a standard exposure measure of racial segregation, the exposure of group g to group h (usually denoted ${}_gP_h^*$). In our notation, this standard exposure measure can be written

$${}_gP_h^* = F_g^h(1,1\,) = \int_0^1 \rho_g(r) f_g^h(r,1)dr. \tag{C3}$$

Standardized measures of between-group differences neighborhood income distributions,

conditional on household income. We might want to measure the difference between the average neighborhood income density functions for two groups g_1 and g_2, conditional on p; that is, for any given value of p, we want to measure the difference between the distributions $\rho_{g_1}^t(p,q)$ and $\rho_{g_2}^t(p,q)$. We could do this by measuring, for example, the difference in their medians (i.e., by comparing ${f_{g_1}^t}^{-1}(p,.50)$ and ${f_{g_2}^t}^{-1}(p,.50)$), but this this would not provide a summary measure of the overall difference in the distributions. A useful summary measure of the degree of overlap of two distributions is the probability that a randomly chosen value from one distribution is larger than a randomly chosen value from the other. In our case here, this is the probability that a randomly chosen member of the neighborhood of the typical group g_1 household with income p has an income higher than that of a randomly chosen member of the neighborhood of the typical group g_2 household with income p. This probability is equal to

$$Pr_{g_1>g_2}(p) = \int_0^1 f_{g_1}^t\left(p, {f_{g_2}^t}^{-1}(p,c)\right) dc. \tag{C4}$$

This probability can be converted to the V statistic, a non-parametric measure of the difference between two distributions:

$$V_{g_1 g_2}(p) = \sqrt{2}\Phi^{-1}\left(Pr_{g_1>g_2}(p)\right), \tag{C5}$$

where $\Phi^{-1}(\cdot)$ is the probit function. Here $V_{g_1 g_2}(p)$ is a function of p that describes the extent of overlap between the typical neighborhood income distributions. V can be interpreted as the standardized difference between the means of two normal distributions with the same degree of overlap as the distributions of interest, so it is interpretable as a "pseudo effect size" (Ho & Haertel, 2006; Ho & Reardon, 2012; Holland, 2002).