ECONOMIC GROWTH CENTER
YALE UNIVERSITY

P.O. Box 208269
New Haven, CT 06520-8269
http://www.econ.yale.edu/~egcenter/

CENTER DISCUSSION PAPER NO. 859

WHAT DRIVES RACIAL SEGREGATION? NEW EVIDENCE USING CENSUS MICRODATA

Patrick Bayer
Yale University

Robert McMillan
University of Toronto

Kim Rueben
Public Policy Institute of California

July 2003

Notes: Center Discussion Papers are preliminary materials circulated to stimulate discussions and critical comments.

The authors would like to thank Fernando Ferreira (University of California, Berkeley) for outstanding research assistance, as well as Pedro Cerdan and Jackie Chou for help assembling the data. We would also like to thank Debbie Reed, Steve Ross, Jon Sonstelie, Chris Udry, Junfu Zhang and seminar participants at Johns Hopkins, UCLA, and Yale for helpful comments. We are grateful to the California Census Research Data Center for providing access to the data and to Ritch Milby in particular. All output included in this paper has been subject to thorough disclosure analysis by Census Bureau officials and meets stringent standards to safeguard confidentiality. Financial support from the Public Policy Institute of California is gratefully acknowledged. Please send correspondence to any of the authors: patrick.bayer@yale.edu, mcmillan@chass.utoronto.ca, or rueben@ppic.org

What Drives Racial Segregation? New Evidence Using Census Microdata

Patrick Bayer, Robert McMillan, and Kim Rueben

Abstract

This paper sheds new light on the forces that drive residential segregation on the basis of race, assessing the extent to which across-race differences in other household characteristics can explain a significant portion of observed racial segregation. The central contribution of the analysis is to provide a transparent new measurement framework for understanding segregation patterns. This framework allows researchers to characterize patterns of segregation, to decompose them in meaningful ways, and to carry out partial equilibrium counterfactuals that illuminate the contributions of a variety of non-race characteristics in driving segregation. We illustrate our approach using restricted micro-Census data from the San Francisco Bay Area that provide a rich joint distribution of household and neighborhood characteristics not previously available to the research community. In contrast to findings in the prior literature, our analysis indicates that individual household characteristics *can* explain a considerable fraction of segregation by race, explaining almost 95% of segregation for Hispanic, over 50% for Asian, and 30% for White and Black households.

JEL Classification: H0, J7, R0, R2

Keywords: Residential Segregation, Racial Segregation, Sorting, Housing Markets

I Introduction

Residential segregation on the basis of race and ethnicity is strikingly evident in cities throughout the United States. In trying to explain observed segregation patterns, it is natural to think that race itself must be a fundamental driving force, working through decentralized household preferences for the race of their neighbors that influence residential choices or through centralized discrimination in the housing market. Yet in his seminal work on the processes underlying segregation, Thomas Schelling (1971) identified a number of alternative mechanisms only indirectly related to race that might drive segregation,[1] noting that a sizeable amount of racial segregation may be explained by sorting on the basis of these other mechanisms, especially if the correlation of race with these other household characteristics is strong.

The goal of this paper is to examine the extent to which across-race differences in household characteristics, including education, income, wealth, language, and immigration status, can explain the observed pattern of racial segregation. This task is facilitated by access to newly available restricted Census data for 1990, which allows us to overcome data limitations that have hampered prior work. These unique data match each household appearing in the long form of the Census with its Census block, an area with approximately 100 residents, not only providing detailed information about each individual, but also in the aggregate a detailed picture of the neighborhood that each individual inhabits, based on an exceptionally rich joint distribution. Our final data set consists of over 240,000 households and 650,000 individuals drawn from 39,000 Census blocks in the San Francisco Bay Area.

With these data in hand, we provide a transparent new measurement framework that allows researchers both to characterize patterns of segregation and to decompose them in meaningful ways using data that provide a joint distribution of race and other individual attributes. Our central methodological contribution is to present an intuitive procedure for carrying out partial equilibrium counterfactuals that illuminate the contributions of a variety of non-race characteristics in driving observed segregation patterns. This procedure is easy to implement, and provides a natural way of exploiting the richness of the type of data set used in our analysis.

[1] For instance, households might sort across residences based on their wealth or income, and information about desirable locations or jobs might flow through social networks that households are part of, leading like households to cluster in similar locations.

The rationale behind the procedure is straightforward: in order to understand the role of an individual characteristic, such as education, in driving racial segregation, we seek to determine how the pattern of racial segregation would change if across-race differences in education were eliminated - in other words, if each race had the empirical distribution of education observed in the population of the Bay Area as a whole. We perform these calculations by examining how the propensity for households to live in a segregated neighborhood varies with education. If the distribution of a household characteristic differs significantly across race *and* this attribute affects the typical racial composition of the neighborhoods in which households of a given race live, the counterfactuals that we develop will lead to the conclusion that this attribute is an important factor driving the segregation of that race.

It is important to emphasize that we are not modeling the underlying sorting process explicitly in terms of underlying tastes, technologies, and endowments: that task is carried out in related work (see Bayer *et al.* (2002)). Consequently, the counterfactual exercises that we carry out are not fully general equilibrium in nature. Rather, our framework enables us to look at conditional racial exposure rates, examining how the typical neighborhood racial composition of households of each race varies with education, income, and other household attributes, thereby providing insights into the relative importance of racial differences in these attributes in driving segregation. Because we observe a rich joint distribution of individual and neighborhood characteristics associated with each household, this simple approach allows us to shed new light on the driving forces behind segregation.

Which forces are most important in shaping observed segregation patterns is an unresolved matter in the prior literature, owing in large part to serious data limitations. Researchers using micro data linking individual households with their neighborhoods have typically had to study sorting over large geographic areas such as counties (Gabriel and Rosenthal (1989)) or PUMAs, Census-defined areas made up of at least 100,000 people (Bajari and Kahn (2001)). In order to use data characterizing the racial composition of smaller geographic areas, such as Census tracts or zip codes, researchers have generally made use of data that are not explicitly linked to individual households. Miller and Quigley (1990) and Harsman and Quigley (1995), for example, compare the degree of racial segregation in a metropolitan area to the degree of stratification on the basis of income and other household characteristics, concluding that sorting on the basis of these other characteristics can explain only a small amount of observed racial segregation.[2]

Of the work examining the forces driving segregation, the study by Borjas (1998) deserves special attention. This paper uses a restricted version of the NLSY, generating neighborhood socio-demographics from the characteristics of other the individuals in the sample who reside in the same zip code. Unlike prior work, it links data on individuals with information about quite narrowly defined local neighborhoods, shedding light on whether individuals of different races are more or less likely to live in neighborhoods with many others of the same race, controlling carefully for potentially relevant individual characteristics. In essence, we seek to extend the underlying analysis conducted by Borjas (1998) to explicitly examine the degree to which differences in individual characteristics across race can explain the observed level of segregation in a large metropolitan area.[3]

To illustrate the value of our framework, the first part of our analysis documents the patterns of racial segregation in the Bay Area, revealing marked differences in the exposure of households of a given race to households of their own and other races. Here, we show that a significant amount of segregation is missed if researchers use aggregate measures of segregation, at the county, PUMA or even tract levels, drawing attention to the value of having detailed disaggregate data. We then examine whether individual non-race characteristics help explain observed segregation patterns. In contrast to findings in the previous literature, our analysis indicates that individual household characteristics *can* explain a considerable fraction of segregation by race. Taken together, the correlation of race with other observable household attributes explains almost 95 percent of segregation for Hispanic households, over 50 percent for

---

[2] In addition to these studies, a number of researchers have attempted to use data characterizing differences in the prices paid for comparable houses by households of different races to distinguish whether segregation arises because of centralized discriminatory practices or decentralized residential location decisions made by individual households. Notable papers in this line of research include King and Mieszkowski (1973), Schnare (1976), Yinger (1978), Schafer (1979), Follain and Malpezzie (1981), Chambers (1992), Kiel and Zabel (1996), and Cutler, Glaeser, and Vigdor (1999). These papers provide mixed evidence concerning whether Black households pay a premium for comparable housing, suggesting that the existence of such a premium may vary over time and location and by how well the researcher controls for unobserved neighborhood quality. The focus of these studies has been on factors directly linked to race.

[3] The purposes of our analysis differ substantially from those of the Borjas study. In particular, because we are interested in the neighborhood sorting process itself and the extent to which differences in non-race household characteristics can explain observed segregation patterns, we view a metropolitan area rather than the nation as a whole as the relevant economic environment for our analysis. In this case, relative to the NLSY, the restricted Census data provide detailed information on the characteristics of a much wider sample of households observed at a lower level of aggregation, thereby providing a richer view of the underlying socio-demographic composition of each neighborhood for a large metropolitan area such as the Bay Area.

Asian households, and approximately 30 percent for White and Black households. Our analysis also indicates that different factors drive the segregation of different races. Language explains a substantial proportion - more than 30 percent - of Asian and Hispanic segregation, education explains a further 20 percent of Hispanic segregation, while income is the most important non-race household characteristic for Black households, explaining around 10 percent of Black segregation.

The paper is organized as follows: the next section describes the unique data set used in the analysis. Section III then sets out the measurement framework and some basic results relating to segregation patterns. Section IV provides the main economic analysis of the paper, exploring the extent to which the correlation of household characteristics and race can explain the observed patterns of racial segregation in the Bay Area. Section V concludes.

II Data

Our analysis is conducted using an extensive new data set built around restricted Census microdata for 1990. These restricted Census data provide the same detailed individual, household, and housing variables found in the public-use version of the Census, but unlike the public-use data they provide information on the location of individual residences and workplaces at a very disaggregated level, down to the Census block. Thus the restricted Census microdata allow us to identify the local neighborhood each individual inhabits, and to determine the characteristics of that neighborhood far more accurately than has been previously possible with such a large-scale data set.

Our study area consists of six contiguous counties in the San Francisco Bay Area: Alameda, Contra Costa, Marin, San Mateo, San Francisco, and Santa Clara. Though the framework we set out below has broad applicability for understanding segregation patterns, we focus on this area for three main reasons. First, it is reasonably self-contained. Examination of Bay Area commuting patterns in 1990 reveals that a very small proportion of commutes originating within these six counties ended up at work locations outside the area, and similarly a relatively small number of commutes to jobs within the six counties originated outside the area. Second, the area contains a racially diverse population, with significant numbers of Asian, Black, and Hispanic households. And third, the area is sizeable along a number of dimensions: the six counties include over 1,100 Census tracts, and almost 39,500 Census blocks, the smallest unit of aggregation in our data.[4] Our final sample consists of about 650,000 people in just under 244,000 households.

The Census provides a wealth of data on the individuals in the sample – their race, age, level of educational attainment, income, occupation (if working), language ability, marital status, and more. Throughout our analysis, we treat the household as the decision-making unit and characterize each household's race as the race of the 'householder' – typically the household's primary earner. We assign households to one of four mutually exclusive categories of race/ethnicity: Hispanic, non-Hispanic Asian, non-Hispanic Black, and non-Hispanic White.[5] To ensure that our sample is representative of the overall Bay Area population, we employ the individual weights given in the Census. Accordingly, 12.3 percent of households are categorized as Asian, 8.8 percent as Black, 11.2 percent as Hispanic, and 67.7 percent of households as White.[6] The Census housing record provides other information on household characteristics, such as household size, family structure, number of children and languages spoken.

Using individual and household data linked to Census blocks, we have constructed a series of variables characterizing the neighborhood in which a household lives. We define a variety of neighborhoods based on conventional Census boundaries – the block, block group, tract, Public Use Microdata Area (PUMA) and county.[7] These provide the basis for our analysis of segregation. The full list of variables used in the analysis, along with means and standard deviations, is given in the Data Appendix.

---

[4] Our sample consists of all households who filled out the long-form of the Census in 1990, approximately 1-in-7 households. In our sample, Census blocks contain an average of 6 households, while Census block groups – the next level of aggregation up - contain 92 households.

[5] The task of characterizing a household's race/ethnicity raises the issue of what to do with mixed race households. We use the characteristics of the household head to define the race/ethnic makeup of the household, and also omit the households that do not fit into one of these four primary racial categories (0.7 percent of all households). The results of our analysis are not sensitive to these decisions. Our final sample consists of the 243,350 households that fit into these four racial categories and live in a Census block group that contains at least one other household in our sample.

[6] The Census sample is highly representative of the Bay Area's population. If we calculate unweighted samples using the numbers of householders, 12.4 percent of households are characterized as Asian, 7.6 percent as Black, 10.9 percent as Hispanic, and 68.6 percent as White (and only 0.7 percent of households characterized as "Other").

[7] In addition, as we know the latitude and longitude of the area center of each Census block, we define a succession of neighborhoods surrounding a given block that include all households in the sample in blocks within certain radii - half a mile, one mile, two miles etc. Using this approach, we can construct racial, education and income distributions based on the households within a given radius surrounding each Census block.

III Patterns of Racial Segregation in the Bay Area

A. *Measurement Framework*

We begin our empirical analysis by characterizing the patterns of racial segregation in the Bay Area. Given the assignment of households to one of the four primary race categories - Asian, Black, Hispanic, and White - we define dummy variables, r^i_j, that take the value one if household i is of race j, and zero otherwise. For a particular neighborhood definition, we calculate the fractions of households in each of the four racial categories that reside in the same neighborhood as a given household; let the upper-case notation R^i_k signify the fraction of households of race k in household i's neighborhood. By averaging these neighborhood measures over all households of a given race, we construct measures of the average neighborhood racial composition for households of that race. Put another way, we construct measures of the *average exposure*, $E(r_j,R_k)$, of households of a race j to households of race k:

$$(1) \qquad E(r_j, R_k) = \frac{\sum_i r_j^i R_k^i}{\sum_i r_j^i}$$

An alternative and convenient way to construct these exposure rates is to run the following set of simple regressions. For each household i, regress R^i_k on the set of dummy variables r^i_j:

$$(2) \qquad R_k^i = \sum_j g_{jk} r_j^i + w_k^j, \quad k \in \{A, B, H, W\},$$

where k ranges over the four race categories. The resulting parameters g_{jk} are identically the average exposure of households of race j to race k, $E(r_j,R_k)$. This approach also provides a convenient way to distinguish the precision of these exposure rate measures, as the regression in equation (2) also provides standard errors for these measures.

A number of segregation measures are available.[8] We choose to work with measures of segregation based on the exposure rates described above because exposure rates are easy to

[8] See Reardon and Firebaugh (2002). The measure most often used in sociology is the dissimilarity index. Dissimilarity indices, which range between zero and one, provide information about the residential concentration of one race relative to others – specifically, the share of one population that would need to move in order for the races in a region to be evenly distributed (see Cutler *et al*. (1999) for a definition). In contrast, the exposure rate measures used here simply return the average rate of contact between people with specified sets of characteristics. Alternative measures of segregation include entropy measures (described in Massey and Denton (1989)), which summarize the degree to which the racial distributions of neighborhoods within a region differ from the region's overall racial distribution, entropy being maximized for the region when the racial distributions at lower levels of aggregation are the

interpret and can be decomposed in a variety of meaningful ways. It is straightforward, for example, to calculate exposure rates for various subsets of households within each broad category (*e.g.* households of the same race but differing in their education levels), rates that must as a matter of necessity aggregate back up to the average exposure rate for the whole group. Unlike many segregation measures, exposure rates also allow us to examine the propensity of households of any pair of races to live together and to consider the factors that affect this propensity separately for different pairs of races. Thus we can see if households are clustering with specific households of other types rather than just examining own-group sorting patterns.[9]

It is possible to define a neighborhood and thus R^i_k in a number of ways. In the results that follow, we use the standard neighborhood measures given in the Census, rather than neighborhoods falling within given radii around each house.[10] These methods yield very similar results.

B. *Segregation Patterns*

Figure 1 provides information about the racial composition of Census block groups for the geographic core of our study area including San Francisco, Oakland, and Berkeley. [11] Although Black households make up only 9 percent of the Bay Area population, the large number of Census block groups with a majority of Black households indicates a high degree of Black segregation. Though Hispanic households account for a higher proportion of the Bay Area population than Black households, there are far fewer Census block groups in which a majority of households are Hispanic.

Table 1 provides the exposure rate measures described above calculated for Census block groups. The table should be read as follows: consider the measured exposure rates of the typical

---

same as that for the region overall. Borjas (1998) makes use of individual data, constructing a measure of segregation that takes the value one if the proportion of the individual's own ethnic group in the neighborhood is more than twice the proportion that would be expected under random assignment of individuals.

[9] Note that under the current approach, including a household as an observation when constructing the neighborhood racial composition for that household can affect the measured exposure rates for our smaller neighborhood measures - for instance, Census blocks rather than tracts. To avoid this problem, we define the racial makeup of a neighborhood to be the racial makeup of all *other* households in the neighborhood and avoid including the individual household's own observation. It is important to point out that once this adjustment is made, any incorrect measurement of the neighborhood racial composition variables arising because of the small number of observations used to construct our smaller neighborhood measures does not bias the exposure rate measures.

[10] We considered both methods of defining neighborhoods, as the first corresponds to the approach most commonly used in the literature and the second might provide a better approximation to a household's neighborhood in certain cases.

[11] Figure 1 is derived from information in the public-use Census data set.

Asian household at the Census block group level shown in the top panel of the table. Reading across the first row, these measures imply that Asian households live in Census block groups that have on average 23 percent Asian households, 8 percent Black, 12 percent Hispanic, and 57 percent White households. Comparing these numbers to the racial distribution of the Bay Area as a whole, given in the row labeled "Overall" - 12 percent Asian, 9 percent Black, 11 percent Hispanic, and 68 percent White - it is apparent that the typical Asian household lives in a Census block group with approximately twice the fraction of Asian households as would be found if they were uniformly distributed across the Bay Area. In this case, the additional fraction of Asian households in Census block groups in which Asian households reside is almost exactly offset by a reduction in the fraction of White households in these neighborhoods,[12] with Black and Hispanic households being found in roughly the same proportions as their overall proportions for the Bay Area.

Examining the exposure measures for each race at the Census block group level, a clear pattern emerges, with households of each race residing with households from the same race in proportions significantly higher than their proportions for the Bay Area as a whole. The most striking example of such 'over-exposure' of households to other households of the same race occurs for Black households. On average, the typical Black household lives in a Census block group that has almost 5 times the fraction of Black households as the whole Bay Area and over 8 times the average fraction of Black households as are found in the neighborhoods inhabited by White households. The pattern for Hispanic households is similar to that for Asian households, and consistent with the previous patterns, White households on average live in block groups with a lower proportion of other races than would be found if all racial groups were evenly spread across block groups.

We present exposure rates at five levels of aggregation - county, PUMA, tract, block group, and block - in Appendix Table 1.[13] Examining these exposure rates, it is clear that the exposure of households to other households of the same race increases as the size of the geographic unit under consideration declines. While this general trend is not surprising, the

[12] It is worth noting that other segregation measures such as dissimilarity indices would miss the fact that the increased exposure of typical Asian, Black, and Hispanic households to other households of the same race is almost completely offset by a decreased exposure to White households.

[13] The exposure rates shown in Appendix table 1 also include standard errors, revealing, as one would expect with nearly a quarter of a million observations, very precise estimates.

extent to which these measures differ for PUMAs, which contain approximately 50,000 households, and smaller Census areas such as block groups (around 500 households) and blocks (around 50 households) is significant. The exposure rate measures in Appendix Table 1 imply, for example, that an analysis of segregation at the PUMA level, which is the smallest geographic unit specified in the public-use Census microdata, would significantly understate the fraction of immediate neighbors who are of the same race. This points to the importance of using the restricted data for the type of household-level analysis conducted in the current paper.

IV Exploring the Mechanisms Underlying Segregation

Having characterized the general patterns of racial segregation in the Bay Area, we now turn to the main analysis of the paper - examining the extent to which the correlation of race with other household attributes can explain the segregation of each race. In previous studies that have attempted to examine this question, researchers have typically known only the marginal distributions of race, education, income, and other household attributes (see Massey and Denton (1993), (1998), and Harsman and Quigley (1995)). In the current analysis, we seek to exploit the richness of the restricted Census data, in particular the fact that these data provide the *joint* distribution of household characteristics at very low levels of geographic aggregation.

In order to conclude that a particular household characteristic explains observed patterns of racial sorting, we require two conditions to hold. *First, the distribution of this household characteristic must differ significantly across race.* If, for example, the distribution of educational attainment were the same for all races, it seems reasonable to conclude that this factor would have no ability to explain the observed pattern of racial segregation across race. *Second, the attribute in question must affect the typical racial composition of the neighborhoods in which households of a given race live.* If, for example, a household characteristic has no impact on the propensity of households of a given race to live in segregated neighborhoods, it seems reasonable to conclude that altering the distribution of this attribute would have little effect on the aggregate segregation of that race.

To determine the household attributes that satisfy the first condition described above, Table 2 summarizes a series of household attributes by race. It is immediately apparent that households of the four racial categories differ along many other dimensions, including education, income and wealth, family structure, language(s) spoken, and citizenship. Thus a number of household attributes have the potential to explain the segregation of households of each race.

Determining how changes in household characteristics affect the propensity of households of each race to live with households of the same and other races (the second condition above), we extend the exposure rate regression framework developed in equation (2) to allow racial exposure rates to vary with individual household attributes. To measure how household characteristics affect the exposure of households of race j to households of race k, we include interactions of household attributes and household race in the exposure rate regressions:

$$(3) \qquad R_k^i = \sum_m \sum_j g_{jkm} r_j^i x_m^i + \boldsymbol{n}_k^i, \quad k \in \{A, B, H, W\}.$$

Here, each variable x_m represents a household attribute and each parameter, g_{jkm}, describes how attribute x_m affects the exposure of households of race j to race k.

Because the four mutually exclusive categories of household race are interacted with each household attribute in the regressions shown in equation (3), it is possible to produce the same parameters by stratifying the sample by race and running separate regressions for each race. The resulting parameter estimates describe how each household attribute affects the propensity of households of that race to live with households of the race that constitutes the dependent variable. In order to keep the results tractable, we report only four of the full sixteen regressions in Table 3 - those that describe how household attributes affect the propensity of households of each race to segregate from or live with households of the same race.

The first rows of Table 3 show the marginal impact of educational attainment on the propensity of households of each race to live with others of the same race.[14] For example, at the margin, Black households with less than a high school degree live in neighborhoods with 12 percentage points more Black households than Black households with an advanced degree. The next set of rows show the impact of household income on racial stratification. As with education, increases in income lead to more segregation on the part of White households and less on the part of households of other races. Likewise, we find that the impact of income is largest for Black households. The source of income, in addition to the magnitude, is also important. Black and

[14] Exposure rates can be recovered from these estimates by adding coefficients for households of a given race and given characteristics to the race-specific constants at the bottom of each column.

Hispanic households with capital income tend to live with fewer households of the same race, while Hispanic and especially Black households with public assistance income are more likely to be segregated. Not surprisingly, we also find that speaking a language other than English increases the level of segregation for Asian and Hispanic households, as does answering that the household head speaks only some English or no English. There is also an increase in the segregation of households of all races who have recently moved to the US and of all races other than Black households that are naturalized or not US citizens, especially Asian households.

A. Counterfactuals – Treating Conditional Exposure Rates as Primitives

Using the results of the regressions reported in Table 3, we now describe a procedure for conducting a counterfactual that treats the conditional exposure rates g_{jkm} as primitives of the sorting process – that is, it assumes the racial exposure rates of households of a given race with a set of characteristics $\{x_m\}$ are fixed. In order to calculate the effect of eliminating across-race differences in household characteristics on the segregation of race j, we simply adjust the underlying distribution of characteristics for households to reflect those of the population as a whole. This counterfactual does not account for the fact that the exposure rates implied by the regressions in Table 3 might themselves adjust as the underlying characteristics of each race change - we consider an alternative assumption that allows these rates to change in the next subsection.

To make this procedure clear, consider first multiplying each of the conditional exposure rate g_{jkm} by the mean of each household attribute for race j, $\overline{x}_{jm}$, and summing over the included attributes. Such a calculation reproduces the average exposure of households of race j to households of race k:

(4) $$E(r_j, R_k) = \sum_m g_{jkm}\overline{x}_{jm} = g_{jk}$$

Substituting instead the mean of each household attribute from the full sample, $\bar{x}_m$, we calculate the average exposure of households of race j to households of race k under the assumption of *fixed conditional exposure rates* and mean attributes $\bar{X}$, labeled $E(r_j, R_k \mid \bar{X})$:

$$(5) \qquad E(r_j, R_k \mid \bar{X}) = \sum_m g_{jkm} \bar{x}_m$$

A comparison of $E(r_j, R_k \mid \bar{X})$ to $E(r_j, R_k)$ reveals the impact of reducing across-race differences in all of the included household attributes X on the exposure of households of race j to households of race k. Having estimated equation (3) with a full set of interactions, we calculate the *marginal* impact of a particular household attribute on the exposure of race j to race k by replacing $\bar{x}_j$ with $\bar{x}$ for only that attribute.

Table 4 presents the results of this first set of counterfactual simulations. The top panel of Table 4 gives, for each race, the percentage of racial segregation that can be explained by non-racial household characteristics. The first set of rows presents information first shown in Table 1 - that is, the overall distribution of each racial group and the over-exposure of the average household of each race to other households of the same race. The next set of rows presents the over-exposure rate that would occur if there were no differences in household characteristics across each racial group: it estimates the percent of households predicted to live in a neighborhood of the same race using the regression estimates and the overall sample means. Rows 5 and 6 then relate the decline in exposure rates due to differences in household characteristics to that originally found. As the last row in this panel indicates, differences in non-racial household attributes together explain approximately 93 percent of segregation for Hispanic households, 53 percent for Asian households, 32 percent for White households, and 30 percent for Black households. Note that although an equal amount of the over-exposure rates for Black and White households occurs, the relative amount of over-exposure was much higher for Black households.

To understand which household attributes drive the segregation of each race, we decompose the overall percentages reported in the lower panel of Table 4. This lower panel shows the *marginal* effects of five different sets of attributes: educational attainment, income, language, citizenship, and household demographics. In each case, we calculate exposure rates when the distribution of a particular set of attributes for each race is replaced by the mean distribution of that set of households in the overall sample using the approach described above and then list the amount of the decline in over-exposure of households related to the given attribute. We discuss the findings for each race in turn.

For Asian households, the primary driver of segregation relates to language, which alone can account for almost 40 percent of the 'over-exposure' of Asian households to other Asian households. Much of this effect derives from whether another language is spoken rather than how well English is spoken in the household. Since 75 percent of Asian households speak an Asian language, the results imply that Asian households that do not know another language resemble the overall population. Factors related to immigration status and citizenship explain another 8.5 percent of Asian segregation. Income, education, and family structure have little to no explanatory power.

Lower levels of income, as well as the higher probability of drawing public assistance and lower probability of having capital income, increase the segregation of Black households, explaining over 14 percent of the 'over-exposure' of Black households to other Black households. Differences in education and factors related to immigration and citizenship explain another 11 percent of Black segregation, but family structure variables explain very little.

For Hispanic households, almost every included set of household characteristics has some ability to explain Hispanic segregation. As in the case of Asian segregation, more than 30 percent of the residential concentration of Hispanic households can be explained by language differences, with much of this difference coming from speaking Spanish in the house. Lower than average levels of education and income explain another 19 and 10 percent of Hispanic segregation respectively and family structure – in particular, larger household sizes – explains another 14 percent. Notably, factors related to citizenship and immigration explain none of the observed segregation of Hispanic households on the margin. Combined with the similar finding for the relationship between language and immigration for Asian households, these results suggest that households who do not speak another language show little taste for living in neighborhoods with a larger concentration of other households of the same races. Alternately, if a non-immigrant family chooses to speak another language in the home, this is an indication that they also prefer to live in neighborhoods with a higher concentration of other families of the same racial or ethnic background, rather than clustering being caused by an inability on non-English speakers or new immigrants being limited to specific neighborhoods. This preference could be driven by stores or other characteristics of these neighborhoods rather than a preference to live with other like families.

The segregation of White households is driven by a variety of factors. The fact that White households have higher than average levels of income and education combined with the fact that White segregation increases with increasing levels of these characteristics means that a portion of the over-exposure of White households to other White households can be explained by these factors – around 12 percent. Language differences can also account for about 15 percent of White segregation, while immigration status, citizenship, and family structure have almost no explanatory power. The language difference information may reflect that someone else in the household is of another racial or ethnic group.

B. Counterfactuals – Treating Conditional Intensities of Exposure as Primitives

The counterfactuals just described treat conditional racial exposure rates as primitives. If, for example, education were the only household attribute under consideration, this procedure would treat the exposure of Hispanic households with an advanced degree to other Hispanic households as fixed. The counterfactual calculations are then based on adjusting the underlying educational attainment distribution of Hispanic households – in this case, moving more Hispanic household to the upper end of the distribution, thereby decreasing the average exposure of Hispanic households to others of the same race.

As this case illustrates, however, it is likely that a significant increase in the educational attainment of a large number of Hispanic households would alter these underlying conditional exposure rates. The conditional exposure of highly educated Hispanic households to other Hispanic households would almost certainly increase with the increased education of the Hispanic population as a whole. Similarly, the conditional exposure of Hispanic households with less than a high school degree to other Hispanic households would likely decrease as fewer Hispanic households would remain in lower educational attainment categories.

To account for the effect of changing the distribution of household characteristics on the underlying conditional racial exposure rates, we consider a second type of counterfactual simulation that adjusts the underlying conditional exposure rates in a systematic way as the distribution of household characteristics in the underlying population of households of each race changes. Because the notation required for this type of counterfactual simulation becomes very involved, we illustrate the underlying logic of these counterfactuals by working through an example.

As a starting point, we consider the exposure of households in each race-education category to households in every other race-education category. In contrast to the fixed exposure rate assumption used above, we treat as primitive the propensity to live with households in each race-education category relative to the fraction of households in that category in the full sample. We label this relative exposure measure the *conditional intensity of exposure* to households in each race-education category. Thus the exposure of highly educated White households to Hispanic households, for example, is allowed to increase with an upward shift in the Hispanic education distribution, provided highly educated White households have a greater intensity of exposure to highly educated versus poorly educated Hispanic households. Having calculated the new exposure rates implied by the shifts in the education distribution, we repeat the analysis from above using these adjusted exposure rates.

Table 5 shows the results for the own-race exposure of Hispanic households to illustrate the procedure. The upper panel in Table 5 shows the average fraction of Hispanic households in each education category that reside in the neighborhood in which Hispanic households with the education level listed in the row heading reside. For example, the first row provides the average exposure of Hispanic households without a High School diploma to Hispanic households in each education category. As the table shows, an average of 17 percent of the neighbors of Hispanic households without a High School diploma are also Hispanic households without a High School Diploma while an average of only half of one percent are Hispanic households with a post-graduate degree. The next four rows show the same kind of distributional information for Hispanic households with higher education levels, while the final row in this upper panel shows, for comparison, the fraction of the Bay Area's population accounted for by Hispanic households in each education category. The right-most columns of the upper panel of Table 5 calculate the results when conditional exposure rates are treated as a primitive for the sake of comparison.

The middle panel in Table 5 then calculates the intensity of exposure for Hispanic households with a given level of education to Hispanic households in each education category. The intensity of exposure for a given education pair is just the ratio of the average fraction of Hispanic households of a given education level in the neighborhood to the overall fraction of Hispanic households with that education level in the Bay Area. Thus, Hispanic households headed by householders without a High School Diploma are typically exposed to almost four times as many households of the same type than would be expected in the overall sample (16.5 percent vs. 4.4 percent). The fact that almost all of the figures in this middle panel are greater than one implies that Hispanic households are exposed to a greater fraction of Hispanic households in almost every education category than the fraction of Hispanic households in that education category in the Bay Area as a whole. Moreover, the greatest intensities of exposure in the table describe the propensity of Hispanic households with low levels of education to live together.

The bottom panel in Table 5 uses the intensity of exposure measures from the middle panel to calculate new exposure rates under the counterfactual that Hispanic households had the education distribution of the Bay Area as a whole; and recall that the intensity of exposure measures are taken as the primitives of the sorting process in this counterfactual. In this case, a typical Hispanic household with less than a High School Diploma is predicted to live in a neighborhood in which 6.8 percent of households are Hispanic households with less than a High School Diploma. This number is calculated by taking the adjusted fraction of Hispanic households in the Bay Area with less than a High School Diploma – 1.8 percent – and scaling it up by the fixed intensity of exposure rate of 3.8 for that education pair.

The sixth column of this bottom panel shows how the overall own-race exposure of Hispanic households in each education category changes as a result of treating the intensity of exposure measures as primitives. As the figures in this column illustrate, treating the intensity of exposure measures as primitives greatly reduces the exposure of Hispanic households in the lowest education categories to other Hispanic households. Put another way, because Hispanic households with low levels of education have such strong intensities of exposure to other poorly educated Hispanic households, the upward shift in the education distribution dramatically reduces the overall own-race exposure of these households. At the same time, because Hispanic households with a bachelor's degree, for example, tend to be exposed in roughly the same intensity to Hispanic households in all education categories, the overall own-race exposure of these households changes very little.

The rightmost columns of the bottom panel of Table 5 calculate the average exposure of Hispanic households to other Hispanic households using the new exposure rates and new weights based on the education distribution of the full population of the Bay Area. The predicted reduction in the 'over-exposure' of Hispanic households to one another using the intensity of exposure measures as primitives is 55.7 percent compared with 36.9 percent when the conditional exposure rates themselves are treated used as primitives.

As this example makes clear, this type of counterfactual requires exposure rate measures for each distinct category of race and household characteristics interacted with every other distinct category. As the number and type of categories increases, this approach quickly exceeds the capacity of our data, despite the fact that we have almost a quarter of a million observations. Creating separate cells for all of the interactions included in the regressions of Table 4, for example, would require almost one billion distinct cells. In conducting the counterfactuals that treat the intensity of exposure measures as primitives, therefore, we focus on the effects of variables that are likely to have the greatest influence and consider a number of different groupings of household characteristic categories such that the total number of distinct cells is limited to 4096 (64 distinct race-household characteristic categories).

Table 6 presents the results from this exercise. For each distinct grouping, we also report analogous results based on counterfactuals that treat exposure rates as primitives, reported in the 'Fixed Exposure Rates' rows. The first panel of Table 6 sets out the reduction in exposure rates resulting from changing the education distribution. The second panel creates twelve distinct categories of household characteristics (2 education categories x 3 income categories x 2 language categories). The counterfactuals that treat intensity of exposure as a primitive 'explain' a greater percentage of the segregation of each race (measured again here as the percentage reduction in own-race 'over-exposure' relative to the sample mean) than the counterfactuals that treat exposure rates as primitives. (This general finding holds consistently in every alternative grouping that we have tried.) The remaining panels of Table 6 consider additional household characteristics such as immigrant status and public assistance income in the formation of distinct categories of household characteristics. In all cases, the two counterfactuals produce a similar pattern of results with the fixed intensity of exposure counterfactuals increasing the explanatory power by an average of about 60 percent. In the light of these results, we conclude that the counterfactuals described in Table 4 that treat exposure rates as primitives and use the full set of characteristics reported almost certainly *underestimate* the amount of sorting explained by these household characteristics. At the same time, the analysis of Table 6 (especially the final panel) confirms our general findings in the first set of counterfactuals, namely that these other household characteristics explain the vast majority of Hispanic and to a lesser extent Asian segregation, while leaving much of the segregation of Black and White households unexplained.

This points to a direct trade-off between the two types of counterfactuals described in our analysis. While the calculations that use exposure rates as primitives almost certainly understate the ability of household characteristics to explain racial segregation, this approach allows us to simultaneously control for a wide range of household characteristics in the analysis. And while the calculations that use *conditional intensity of exposure* measures as primitives are likely more appropriate counterfactuals, the data requirements quickly grow too large. In light of these limitations, we focus attention primarily on the former set of results, noting that the explanatory power of the included household variables is likely to be significantly but not overwhelmingly greater.

C. Identifying Assumptions and Alternative Explanations

While the inclusion of additional household attributes could further reduce the unexplained portion of racial segregation, we believe that the analysis presented in the previous sub-section includes the household attributes observed in the Census that are most relevant. A number of potential explanations arise for the portion of segregation that cannot be explained by household characteristics and it is important to emphasize that our analysis provides no indication as to the root cause of this portion of segregation. For Black households, for example, this could arise because of the preferences of Black households to live together, the preferences of Asian, Hispanic, or White households to live with others of the same race, the preferences of Asian, Hispanic, or White households to avoid Black households, or systematic differences in demand for housing and other neighborhood amenities across race, among other explanations. Our analysis provides no evidence that can distinguish these and other alternative explanations for the unexplained portion of racial segregation.

In addition to the assumptions concerning the primitives upon which each of the two types of counterfactual procedures are based, two main additional assumptions underpin the general approach and are worth a careful discussion. First, we implicitly assume that individuals are mobile across neighborhoods to the extent that their income and wealth allow. Examining the mobility of households of different races in the Bay Area Census sample lends support to this view: there is quite significant mobility across households of all races and at all points in the income and education distributions.[15] The second is more controversial - that individual household characteristics such as income and educational attainment can be taken as exogenous with respect to the degree of racial segregation, and as such, can be used as explanatory variables.

Here, we contend that, while the degree of racial segregation (and neighborhood effects more generally) may have some effect on individual outcomes, it is likely to be dwarfed by family and individual characteristics as determinants of an individual's income or education. And further, it is not clear that neighborhood effects have a strong impact on individual residential choice. We note that the literature on the strength of causation from neighborhoods to individual outcomes is somewhat mixed, but the view that individual characteristics can be taken as exogenous is not unreasonable as a first step.[16]

Given the emphasis we place on the explanatory power of individual characteristics, it is worth considering the possibility that our 'explanations' are spurious – that our individual

---

[15] For example, in the period 1989-90, between 17 and 29 percent of Black households moved into their current residence – 29 percent for Black households with incomes less that $12,000 per annum and 17 percent for Black households at the top of the income distribution. In the period 1985-88, between 26 and 32 percent of Black households moved into their current residence, depending on income level. The pattern is similar for other races, and similar based on educational attainment rather than income. In the period 1989-90, for instance, between 17 and 28 percent of Black households moved into their current residence, depending on educational attainment, and between 26 and 35 percent moved in between 1985 and 1988.

[16] Carefully-conceived recent research on the strength of neighborhood effects – research that is very careful to deal with non-random sorting of individuals – lends some support to this position. In particular, Katz *et al.* (2001) present evidence from the Moving to Opportunity Program indicating that the effects of neighborhoods on individuals are weak in the short-term. Oreopoulos (2003) finds a similar lack of effect using longer term income as an outcome, based on quasi-random data from Toronto's sizeable public housing program. Other evidence indicates that ethnic network effects may play a role in influencing welfare participation – see, for instance, Bertrand *et al.* (2000), building on the work of Borjas (1992) and Evans *et al.* (1992). It is not clear whether there are strong effects on characteristics such as household income.

characteristics merely proxy for more fundamental unobservables that drive the residential location decision. Two comments are due in response: first, we condition on a set of variables, such as income and education, that are very likely to be fundamental in and of themselves, either because they affect opportunities or preferences for location. The claim that these variables are merely incidental to the sorting process seems hard to sustain. Second, it is possible that a variable such as not speaking English in the home may pick up some more fundamental and unobserved preference for segregation, to the extent that providing intensive English education would not remove an important segregating force. Here, we emphasize that our analysis is still important and informative, providing a metric by which the relative contributions of a vector of potentially relevant observables can be compared with each other. Some of our findings regarding the relative contributions of different observables across race are striking and point to potentially fruitful areas of further research.

V Conclusion

The central contribution of this paper is to provide a transparent framework for studying residential segregation using rich micro data. This framework allows researchers to characterize segregation in an intuitive way, to decompose observed segregation patterns in order to explore the role of individual characteristics, and most importantly, to carry out informative counterfactuals that help shed new light on the forces driving segregation.

Our analysis has taken seriously Schelling's idea that racial segregation may be driven by forces that are only incidental to race. Using exceptionally rich new data drawn from the restricted-access version of the 1990 Census, we have addressed the following question: To what extent can across-race differences in household characteristics, including education, income, wealth, language, and immigration status, explain the observed pattern of racial segregation?

In line with the previous literature, our results indicate that segregation patterns vary markedly by race, though there is a tendency for households of a given race to cluster disproportionately with households of the same race. The extent of this clustering depends to a considerable degree on the definition of neighborhood used and we find that a substantial amount of segregation is missed when segregation patterns are studied at the county, PUMA, or even tract level. In direct contrast to the previous literature, however, our findings indicate that household attributes, including education, income, language, and immigration status, can collectively explain almost 95 percent of the segregation for Hispanic households, over 50 percent for Asian households, and approximately 30 percent for White and Black households. For Hispanic households, racial segregation appears to be primarily a by-product of the sorting that occurs in any metropolitan area on the basis of education, income, language and other household attributes. In contrast, the results suggest that race itself directly contributes to the segregation of Black and White households. The results also provide a great deal of information about how a wide set of household characteristics affect the segregation patterns of households of each race, with a different set of household characteristics serving as the primary driver of the segregation of households of each race.

Though our analysis focused on the San Francisco Bay Area, the method has broader applicability, providing a clean way of both describing and decomposing patterns of neighborhood segregation, and of exploring relevant counterfactuals. Future work could extend this analysis to a more nationally representative sample of metropolitan areas, and focusing on segregation along other dimensions.

References

Bajari, Patrick, and Matthew Kahn (2001), "Why Do Blacks Live in Cities and Whites Live in Suburbs?" unpublished manuscript, Stanford University.

Bayer, Patrick, Robert McMillan, and Kim Rueben, (2002), "The Causes and Consequences of Residential Segregation: An Equilibrium Analysis of Neighborhood Sorting," mimeo, Yale University.

Bertrand, Marianne, Erzo F. P. Luttmer, and Sendhil Mullainathan, (2000), "Network Effects and Welfare Cultures," *Quarterly Journal of Economics*, August: 1019-1055.

Borjas, George J., (1992), "Ethnic Capital and Intergenerational Mobility," *Quarterly Journal of Economics*, CVII: 123-150.

Borjas, George J., (1995), "Ethnicity, Neighborhoods, and Human-Capital Externalities," *American Economic Review*, 85(3): 365-390.

Borjas, George J., (1998), "To Ghetto or Not to Ghetto: Ethnicity and Residential Segregation," *Journal of Urban Economics*, 44: 228-253

Chambers, Daniel N, (1992), "The Racial Housing Price Differential and Racially Transitional Neighborhood," *Journal of Urban Economics*, 32, 214-232.

Cutler, David M., Edward L. Glaeser, and Jacob L. Vigdor (1999), "The Rise and Decline of the American Ghetto," *Journal of Political Economy*, 107(3): 455-505.

Evans, William N., Wallace E. Oates, and Robert M. Schwab, (1992), "Measuring Peer Group Effects: A Study of Teenage Behavior," *Journal of Political Economy*: 966-991.

Foillain, James R. and Stephen Malpezzi, (1981), "Another Look at Racial Difference in Housing Prices," *Urban Studies*, 18: 195-203.

Gabriel, Stuart and Stuart Rosenthal, (1989), "Household Location and Race: Estimates of a Multinomial Logit Model," *Review of Economics and Statistics*, 71: 240-249.

Harsman, Bjorn and John M. Quigley, (1995) "The Spatial Segregation of Ethnic and Demographic Groups: Comparative Evidence from Stockholm and San Francisco," *Journal of Urban Economics*, 37: 1-16.

Katz, Lawrence F., Jeffrey R. Kling and Jeffrey B. Liebman, (2001), "Moving to Opportunity in Boston: Early results of a randomized mobility experiment," *Quarterly Journal of Economics*, CXVI, Issue 2, 607-654.

Kiel, Katherine and Jeffrey Zabel, (1996), "Housing Price Differentials in U.S. Cities: Household and Neighborhood Race Effects," *Journal of Housing Economics*, 5: 143-165.

King, Thomas and Peter Mieszkowski, (1973), "Racial Discrimination, Segregation, and the Price of Housing," *Journal of Political Economy*, 81: 590-606.

Massey, Douglas S., and Nancy A. Denton, (1989), "Hypersegregation in United States Metropolitan Areas – Black and Hispanic Segregation along Five Dimensions," *Demography*, 26: 373-91.

Miller, V. and John M. Quigley, (1990), "Segregation by Racial and Demographic Group: Evidence from the San Francisco Bay Area," *Urban Studies*, 27: 3-21.

Oreopoulos, Philip, (2003), "The Long-Run Consequences of Living in a Poor Neighborhood," working paper, University of Toronto.

Reardon, Sean and Glenn Firebaugh (forthcoming), "Measures of Multigroup Segregation," *Sociological Methodology.*

Schafer, Robert, (1979), "Racial Discrimination in the Boston Housing Market," *Journal of Urban Economics*, 6: 1176-1196.

Schelling, Thomas C., (1971), "Dynamic Models of Segregation," *Journal of Mathematical Sociology*, 1: 143-186 .

Schnare, Ann, (1976), "Racial and Ethnic Price Differentials in an Urban Housing Market," *Urban Studies*, 13: 107-120.

Yinger, John, (1978), "The Black-White Price Differential in Housing: Some Further Evidence," *Land Economics*, 54:197-206.

Figure 1: Segregation Patterns in the Bay Area

Note: This figure provides a geographical depiction of segregation patterns for only the central portion of the full study area used in the analysis. San Francisco is the peninsula shown on the lower left of the figure; Oakland is located to the east of San Francisco directly across the Bay; Berkeley and Richmond are located north of Oakland in the upper right portion of the figure; and the upper left part of the figure shows a portion of Marin County.

Table 1: Racial Segregation in the San Francisco Bay Area

	Average Racial Composition of Census Block Group			
	Percent Asian	Percent Black	Percent Hispanic	Percent White
Household - Asian	22.5%	8.3%	11.7%	57.4%
Household - Black	11.6%	40.1%	11.4%	36.9%
Household - Hispanic Origin	12.9%	9.1%	21.8%	56.2%
Household - White	10.4%	4.8%	9.3%	75.5%
Overall Composition of Bay Area	12.3%	8.8%	11.2%	67.7%

	Asian	Black	Hispanic	White
Over-Exposure to Own Race	10.2%	31.3%	10.6%	7.8%

Note: Each of the first four rows shows the average racial composition of the block groups in which households of the race shown in the row heading reside. For comparison, the fifth row shows the overall racial composition of the Bay Area. The 'Over-Exposure to Own Race' measure is defined for each race as the difference between the fraction of same-race neighbors (in same Census block group) and the overall fraction of households of the same race in the Bay Area.

Table 2: Mean Values of Selected Household Characteristics for Households of Each Race

Variable	Asian	Black	Hispanic	White	Overall
Household head is high school dropout	0.19	0.23	0.39	0.10	0.16
Household head graduated from high school	0.14	0.22	0.22	0.18	0.18
Household head has some college	0.18	0.28	0.19	0.23	0.23
Household head has bachelor's degree	0.33	0.20	0.16	0.32	0.29
Household head has advanced degree	0.16	0.06	0.05	0.17	0.14
Household income less than $12K	0.13	0.26	0.14	0.10	0.12
Household income $12-20K	0.09	0.14	0.12	0.08	0.09
Household income $20-35K	0.18	0.23	0.24	0.19	0.20
Household income $35-50K	0.18	0.16	0.21	0.18	0.18
Household income $50-75K	0.23	0.14	0.19	0.22	0.21
Household income $75-100K	0.12	0.05	0.07	0.11	0.10
Household income more than $100K	0.08	0.03	0.04	0.12	0.10
Household receives public assistance income	0.13	0.21	0.11	0.05	0.08
Household has capital gains or dividend income	0.48	0.17	0.25	0.56	0.48
Household head over 65	0.13	0.17	0.11	0.21	0.18
Household head divorced	0.07	0.20	0.14	0.16	0.15
Number of adults in the household	2.48	1.85	2.40	1.86	2.00
Number of pre-kindergarten children in household	0.31	0.27	0.40	0.17	0.22
Number of children grades K-8 in household	0.46	0.41	0.54	0.22	0.30
Number of children grades 9-12 in household	0.14	0.11	0.14	0.06	0.08
Spanish spoken in household	0.01	0.04	0.68	0.03	0.10
Asian language spoken in household	0.76	0.01	0.02	0.01	0.11
Household head born in US	0.24	0.97	0.54	0.90	0.78
Household head not a US citizen	0.35	0.02	0.31	0.04	0.11
Household head a naturalized citizen	0.41	0.01	0.15	0.06	0.11
Household head entered the US in 1980s	0.33	0.02	0.15	0.02	0.07
Household head entered the US in 1970s	0.26	0.01	0.14	0.02	0.06
Number of Observations	30271	18501	26675	167897	243344

Table 3: Explaining Exposure to Households of the Same Race

Dependent Variable:	% Asian	% Black	% Hispanic	% White
Sub-Sample:	Asian Hhlds	Black Hhlds	Hispanic Hhlds	White Hhlds
Observations	30,271	18,501	26,675	167,897
Adjusted R-squared	0.127	0.156	0.205	0.090
HH Education Level:				
No HS Diploma	0.054	0.118	0.099	-0.077
	(0.010)	(0.013)	(0.006)	(0.004)
HS Diploma	0.018	0.094	0.064	-0.038
	(0.005)	(0.012)	(0.005)	(0.003)
Some College	0.016	0.049	0.036	-0.021
	(0.004)	(0.011)	(0.005)	(0.002)
BA diploma	0.017	0.027	0.024	-0.010
	(0.004)	(0.010)	(0.004)	(0.002)
Household Income Information:				
< $12K	0.055	0.210	0.078	-0.105
	(0.012)	(0.017)	(0.008)	(0.004)
$12K-20K	0.028	0.189	0.066	-0.089
	(0.008)	(0.016)	(0.007)	(0.004)
$20-35K	0.015	0.147	0.063	-0.074
	(0.007)	(0.015)	(0.006)	(0.003)
$35-50K	0.017	0.116	0.046	-0.062
	(0.006)	(0.015)	(0.006)	(0.003)
$50-75K	0.027	0.085	0.033	-0.048
	(0.005)	(0.014)	(0.005)	(0.003)
$75-100K	0.014	0.038	0.014	-0.030
	(0.004)	(0.015)	(0.006)	(0.002)
Receives Public Assistance	0.002	0.053	0.019	-0.045
	(0.006)	(0.008)	(0.005)	(0.003)
Capital Gains or Dividend Income	0.005	-0.017	-0.018	0.008
	(0.002)	(0.007)	(0.003)	(0.001)
Language Spoken in Household:				
Spanish	-0.001	-0.036	0.051	-0.034
	(0.011)	(0.013)	(0.004)	(0.003)
Other European Language	0.011	-0.033	-0.001	-0.010
	(0.006)	(0.014)	(0.008)	(0.002)
Asian Language	0.048	-0.065	0.005	-0.075
	(0.004)	(0.020)	(0.008)	(0.005)
Other Language	0.005	0.005	-0.012	-0.033
	(0.022)	(0.019)	(0.017)	(0.006)
HH English Ability:				
Speaks English Well	0.004	0.006	0.010	-0.014
	(0.003)	(0.027)	(0.004)	(0.004)
Speaks Some English	0.025	0.038	0.034	-0.047
	(0.008)	(0.031)	(0.005)	(0.007)
Speaks No English	0.158	-0.138	0.055	-0.082
	(0.031)	(0.085)	(0.011)	(0.020)
HH Citizenship Status:				
Not Citizen	0.024	-0.059	0.016	0.012
	(0.006)	(0.037)	(0.008)	(0.006)
Naturalized Citizen	0.033	-0.031	0.003	0.003
	(0.006)	(0.035)	(0.007)	(0.005)
Entered Country in 1980's	-0.024	-0.067	-0.026	-0.024
	(0.008)	(0.036)	(0.009)	(0.007)
Entered Country in 1970's	-0.002	-0.060	-0.012	-0.021
	(0.008)	(0.036)	(0.008)	(0.006)
Entered Country pre-1970	-0.005	-0.089	-0.021	-0.005
	(0.007)	(0.031)	(0.007)	(0.005)
Constant	0.093	-0.066	0.023	0.798
	(0.020)	(0.036)	(0.020)	(0.012)

Notes: Each column shows the results of regressing the fraction of households of the race shown in the column heading on the set of household characteristics shown in the rows using only the sub-sample of households of the same race. The regressions also control for marital status and age of householder, number of adults and children in household, military service history of household and ten broad occupation categories for householder. Omitted categorical variables for each set of regressors are: more than a BA for education, income over $100K, speaks only English, speaks English very well, and native born.

Table 4: Reduction of Racial Segregation Explained By Non-Racial Household Characteristics

	Operation	Asian	Black	Hispanic	White
Baseline:					
(1) Representation of Race in SF Bay Area	(1)	12.3%	8.8%	11.2%	67.7%
(2) Exposure to Households of Same Race	(2)	22.5%	40.0%	21.8%	75.5%
(3) "Over-Exposure" to Households of Same Race	(3) = (2) - (1)	10.2%	31.2%	10.6%	7.8%
Controlling for Full Set of Household Characteristics:					
(4) "Conditional Exposure" to Households of Same Race	(4)	17.1%	30.5%	12.0%	73.0%
(5) Percentage Point Decline in Exposure Rate	(5) = (2) - (4)	5.5%	9.4%	9.8%	2.5%
(6) Amount Explained by Household Characteristics	(6) = (5)/(3)	**53.2%**	**30.3%**	**92.5%**	**32.4%**

Household Characteristics	Percentage Reduction in Exposure to Households of Same Race			
	Asian	Black	Hispanic	White
Educational Attainment	**0.8%**	**4.3%**	**19.3%**	**4.9%**
Total Effect of Income	**0.7%**	**14.2%**	**10.2%**	**6.6%**
Income Level	0.7%	10.2%	5.6%	3.9%
Household on Public Assistance Income	0.1%	2.3%	0.5%	1.8%
Has Non-Salary Wealth	0.0%	1.7%	4.0%	0.8%
Household Language Effects	**38.7%**	**3.0%**	**32.3%**	**15.2%**
Non-English Language Spoken	30.3%	3.1%	27.4%	11.7%
English Ability	8.3%	-0.1%	4.8%	3.5%
Total Citizenship Effect	**8.5%**	**6.9%**	**-1.7%**	**1.7%**
Citizen Status	15.2%	2.5%	3.2%	-1.2%
Years in US	-6.8%	4.4%	-4.9%	3.0%
Household Demographics	**1.3%**	**0.3%**	**13.9%**	**1.7%**
Military Service	0.8%	-0.3%	0.8%	-0.3%
Occupation	1.1%	1.6%	4.0%	0.8%
Total	**53.2%**	**30.3%**	**92.5%**	**32.4%**

Notes: Rows (1) - (3) correspond to the exposure rate measures described in Table 1. Row (4) presents the fraction of households of the same race in the neighborhood predicted using the regression coefficients in Table 3 for each race and the overall population means for the full set of household characteristics included on the right-hand side of these regressions. Rows (5) and (6) present the corresponding predicted decline in own-race 'over-exposure'.

The lower panel decomposes the calculated decline in own-race 'over-expsoure' associated with the particular set of household characteristics listed in the row heading. These values are based on predicted exposure rates obtained using the regression coefficients for each race in Table 3, replacing each race's own mean for the set of household characteristics listed in the row heading with the overall mean for the Bay Area population.

Table 5: Reduction in Percentage of Hispanic Segregation Related to Educational Attainment Holding Intensity of Exposure Constant

Panel A

	Average Exposure to Hispanic Hhlds in Educ Category						Fixed Conditional Exposure Rates				
	Hispanic <HS	Hispanic HS Deg	Hispanic Some Col	Hispanic BA	Hispanic > BA	Hispanic Total (1)	Hispanic Educ Distrib (2)	Overall Educ Distrib (3)	Calculating Average Exposure Measure (1)*(2)	(1)*(3)	
Hispanic Households											
No HS Diploma	0.165	0.057	0.038	0.024	0.005	0.289	0.390	0.160	0.113	0.046	
HS Diploma	0.104	0.051	0.034	0.022	0.005	0.216	0.220	0.180	0.048	0.039	
Some College	0.075	0.037	0.032	0.022	0.006	0.172	0.190	0.230	0.033	0.040	
BA Degree	0.057	0.030	0.026	0.024	0.006	0.143	0.160	0.290	0.023	0.041	**Reduction in 'Over-Exposure' to Hispanic Households**
More than BA	0.034	0.022	0.021	0.018	0.010	0.105	0.050	0.140	0.005	0.015	***Fixed Exposure Rates***
Fraction of Total Bay Area Population:	0.044	0.025	0.021	0.018	0.006	0.112			**0.221**	**0.181**	**36.9%**

Panel B

	Intensity of Exposure Measures					
	Hispanic <HS	Hispanic HS Deg	Hispanic Some Col	Hispanic BA	Hispanic > BA	Hispanic Total
Hispanic Households						
No HS Diploma	3.776	2.317	1.784	1.341	0.893	2.580
HS Diploma	2.380	2.073	1.596	1.229	0.893	1.929
Some College	1.716	1.504	1.502	1.229	1.071	1.536
BA Degree	1.304	1.220	1.221	1.341	1.071	1.277
More than BA	0.778	0.894	0.986	1.006	1.786	0.938

Panel C

	Counterfactual: New Exposure Rates Adjusting Education Distribution								
	Hispanic <HS	Hispanic HS Deg	Hispanic Some Col	Hispanic BA	Hispanic > BA	Hispanic Total (1)	Overall Educ Distrib (2)	Calculating Exposure (1)*(2)	
Hispanic Households									
No HS Diploma	0.068	0.047	0.046	0.044	0.014	0.218	0.160	0.035	
HS Diploma	0.043	0.042	0.041	0.040	0.014	0.180	0.180	0.032	
Some College	0.031	0.030	0.039	0.040	0.017	0.157	0.230	0.036	
BA Degree	0.023	0.025	0.031	0.044	0.017	0.140	0.290	0.041	**Reduction in 'Over-Exposure' to Hispanic Households**
More than BA	0.014	0.018	0.025	0.033	0.028	0.118	0.140	0.017	***Fixed Intensity of Exposure***
Fraction of Total Bay Area Population: Adjusting Education Distribution	0.018	0.020	0.026	0.033	0.016	0.112		**0.160**	**55.7%**

Notes: The rows in Panel A describe the average fraction of Hispanic households with the level of education listed in the column heading that reside in the Census block groups of the Hispanic households with the level of education listed in the row heading. For example, an average of 5.7% of the neighbors of a Hispanic household with a BA degree are Hispanic households with a less than a HS degree (first column, fourth row). The right-hand side of Panel A holds conditional exposure rates fixed.

The rows in Panel B rewrite the exposure rates of Panel A as a percentage of the overall fraction of Hispanic households with the education shown in the column heading in the Bay Area. For example, Hispanic households with a HS degree live on average with twice as many Hispanic households with a HS degree as are represented in the Bay Area as a whole (second row, second column).

Using the intensity of exposure measures of Panel B, Panel C recalculates the exposure rate measures of Panel A under the counterfactual that the distribution of education for Hispanic households matched that of the full population of the Bay Area. The distribution of Hispanic households by education category that corresponds to this counterfactual is shown in the last row of Panel C. The right-hand side of Panel C calculates the overall own-race exposure of Hispanic households using the new exposure measures calculated on the left-hand side of Panel C.

Table 6: Exploring Reductions in Residential Segregation Under Different Exposure Scenarios

	Number of Distinct Race-Hhld Characteristic Categories	Percentage Reduction in Exposure Households of the Same Race			
		Asian	Black	Hispanic	White
Educational Attainment Only					
Fixed Exposure Rates	20	1.4%	8.1%	36.9%	6.9%
Fixed Intensity of Exposure	20	2.1%	16.2%	55.7%	10.7%
Education (2) x Income (3) x Language (2)					
Fixed Exposure Rates	48	44.6%	20.1%	61.9%	9.8%
Fixed Intensity of Exposure	48	62.1%	33.2%	78.5%	30.3%
Education (2) x Income (3) x Immigrant Status (2)					
Fixed Exposure Rates	48	19.6%	22.5%	39.6%	21.6%
Fixed Intensity of Exposure	48	21.2%	36.6%	59.1%	22.7%
Education (2) x Income (3) x Public Assistance Income (2)					
Fixed Exposure Rates	48	0.3%	16.9%	30.6%	9.0%
Fixed Intensity of Exposure	48	0.7%	25.8%	50.1%	12.7%
Education (2) x Public Assistance Income (2) x Language (2) x Immigration Status (2)					
Fixed Exposure Rates	64	46.3%	19.6%	57.4%	14.3%
Fixed Intensity of Exposure	64	67.8%	31.2%	76.2%	24.8%

Notes: This table presents the results from several conterfactuals. The rows labeled 'Fixed Exposure Rates' report the results from counterfactuals that treat exposure rates as primitives, while the rows labeled 'Fixed Intensity of Exposure ' report the results from counterfactuals that treat intensity of exposure measures as primitives. Each panel uses interactions of race with the distinct categories of household characteristics shown in each row heading. The education categories distinguish households that have received at least a bachelor's degree; the income categories distinguish: less than $35k, $35-75k, and $75k+; the language categories distinguish those that speak a foreign language; the immigration status categories distinguish native-born US citizens; and the public assistance categories distinguish those receiving any form of public assistance income.

Appendix Table 1:
Racial Exposure Rates at Different Levels Of Aggregation

	Census Block Racial Composition			
	Percent Asian	Pecent Black	Percent Hispanic	Percent White
Asian Household	26.1%	7.7%	11.2%	55.0%
	(0.009)	(0.003)	(0.003)	(0.008)
Black Household	11.2%	42.8%	11.2%	34.9%
	(0.004)	(0.016)	(0.004)	(0.013)
Hispanic Household	12.5%	8.8%	24.9%	53.8%
	(0.003)	(0.003)	(0.008)	(0.008)
White Household	10.0%	4.4%	8.8%	76.8%
	(0.002)	(0.002)	(0.002)	(0.004)
Observations	242218	242218	242218	242218
Adjusted R-squared	0.081	0.288	0.074	0.21

	Census Block Group Racial Composition			
	Percent Asian	Pecent Black	Percent Hispanic	Percent White
Asian Household	22.5%	8.3%	11.7%	57.4%
	(0.008)	(0.003)	(0.004)	(0.009)
Black Household	11.6%	40.1%	11.4%	36.9%
	(0.004)	(0.015)	(0.004)	(0.013)
Hispanic Household	12.9%	9.1%	21.8%	56.2%
	(0.004)	(0.003)	(0.009)	(0.010)
White Household	10.4%	4.8%	9.3%	75.5%
	(0.003)	(0.002)	(0.002)	(0.005)
Observations	243419	243419	243419	243419
Adjusted R-squared	0.108	0.337	0.112	0.273

	Census Tract Racial Composition			
	Percent Asian	Pecent Black	Percent Hispanic	Percent White
Asian Household	21.4%	8.5%	11.9%	58.4%
	(0.008)	(0.003)	(0.004)	(0.008)
Black Household	11.8%	38.3%	11.7%	38.2%
	(0.005)	(0.020)	(0.004)	(0.016)
Hispanic Household	13.1%	9.3%	20.8%	57.0%
	(0.003)	(0.004)	(0.008)	(0.009)
White Household	10.6%	5.0%	9.5%	75.0%
	(0.003)	(0.002)	(0.002)	(0.005)
Observations	243422	243422	243422	243422
Adjusted R-squared	0.105	0.329	0.111	0.27

	PUMA Racial Composition			
	Percent Asian	Pecent Black	Percent Hispanic	Percent White
Asian Household	16.2%	9.1%	12.1%	62.6%
	(0.009)	(0.013)	(0.010)	(0.021)
Black Household	12.9%	25.6%	12.1%	49.4%
	(0.012)	(0.062)	(0.009)	(0.062)
Hispanic Household	13.4%	9.4%	15.7%	61.5%
	(0.010)	(0.014)	(0.016)	(0.026)
White Household	11.5%	6.4%	10.2%	71.9%
	(0.010)	(0.011)	(0.009)	(0.020)
Observations	243425	243425	243425	243425
Adjusted R-squared	0.053	0.194	0.062	0.168

	County Racial Composition			
	Percent Asian	Pecent Black	Percent Hispanic	Percent White
Asian Household	13.9%	8.9%	11.6%	65.6%
	(0.012)	(0.005)	(0.006)	(0.018)
Black Household	12.5%	12.4%	10.7%	64.4%
	(0.008)	(0.013)	(0.008)	(0.016)
Hispanic Household	12.7%	8.4%	11.9%	67.0%
	(0.006)	(0.006)	(0.004)	(0.014)
White Household	11.9%	8.4%	11.1%	68.6%
	(0.018)	(0.027)	(0.013)	(0.033)
Observations	243425	243425	243425	243425
Adjusted R-squared	0.022	0.04	0.013	0.034

Notes:
Standard errors in parentheses.

Data Appendix

This data appendix gives descriptions of and summary statistics on all the variables used in the analysis.
The following summary statistics are based on a sample of 243,350 households drawn from the 6 Bay Area counties.
Person weights drawn from the Census are used when calculating the household and neighborhood level numbers.

Variable Description	Mean	Std. Dev.
household head is high school dropout	0.16	0.36
household head graduated from high school	0.18	0.39
household head has some college	0.23	0.42
household head has bachelor's degree	0.29	0.45
household income less than $12K	0.12	0.32
household income $12-20K	0.09	0.29
household income $20-35K	0.20	0.40
household income $35-50K	0.18	0.39
household income $50-75K	0.21	0.41
household income $75-100K	0.10	0.30
household receives public assistance income	0.08	0.27
household has dividend income	0.48	0.50
sex of household head	1.34	0.47
age of household head	46.98	16.63
household head over 65	0.18	0.39
household head widowed	0.10	0.30
household head divorced	0.15	0.35
household head separated	0.03	0.17
household head never married	0.21	0.41
number of adults in the household	2.00	0.98
number of pre-kindergarten children in household	0.22	0.56
number of children grades K-8 in household	0.30	0.70
number of children grades 9-12 in household	0.08	0.31
Spanish spoken in household	0.10	0.30
Asian language spoken in household	0.11	0.31
other European language spoken in household	0.07	0.26
other language spoken in household	0.01	0.09
household head speaks English well	0.06	0.24
household head speaks some English	0.04	0.19
household head speaks no English	0.01	0.09
household head not a US citizen	0.11	0.31
household head a naturalized citizen	0.11	0.31
household head entered the US in 1980s	0.07	0.26
household head entered the US in 1970s	0.06	0.24
household head entered US pre-1970	0.09	0.29
household head active in military	0.01	0.07
household head previously in military	0.22	0.41
household head in reserves	0.02	0.15