Counterfactual Dissimilarity: Can Changes in Demographics and Income Explain Increased Racial Integration in U.S. Cities?*

Paul E. Carrillo†
George Washington University

Jonathan L. Rothbaum‡
U.S. Census Bureau

November 2018
PRELIMINARY DRAFT

Abstract

Urban areas in the U.S. have experienced important changes in racial/ethnic distributions over the last two decades. In the average urban area today black-white racial integration has increased by about 10 percent between 1990 and 2010. Changes in racial and ethnic distributions and gentrification are often associated with changes in residents' demographic characteristics, such as income, education and age. This paper applies a non-parametric spatial decomposition technique using complete (restricted-use) microdata files from the 1990 Decennial Long Form Census and 2008-2012 American Community Surveys to assess what portion of the changes in racial distributions can be attributed to changes in individual characteristics. We find that that, on average, a little over a third of the observed increase in integration can be accounted for by changes in observed individual characteristics.

Keywords: Counterfactual Distribution, Decomposition, Spatial Econometrics
JEL Codes: C14, R23, R30

*We would like to acknowledge financial support from the GWU Institute for International Economic Policy (IIEP). This work is released to inform interested parties of ongoing research and to encourage discussion of work in progress. The views expressed on (statistical, methodological, technical, or operational) issues are those of the authors and not necessarily those of the U.S. Census Bureau. This paper has been cleared by the Census Bureau's Disclosure Review Board release authorization number CBDRB-FY18-060.

†Email: pcarrill@gwu.edu.

‡Email: jonathan.l.rothbaum@census.gov.

1 Introduction

Urban areas in the U.S. have experienced important changes in their racial and ethnic distributions over the last two decades. Both the location of minorities residing in cities as well as racial integration within the city has substantially changed across time. With a few exceptions, racial integration has improved in almost all urban areas. Meanwhile, other demographic characteristics of U.S. city residents such as their income, education and age have also experienced important changes. What portion of the changes in the racial composition and racial integration of a city can be accounted for by changes in the demographic characteristics of their residents? To assess this question, our paper employs a novel extension of a non-parametric spatial decomposition technique and the complete microdata files from the 1990 Decennial Long Form Census and 2008-2012 American Community Survey. We find that about one third of the increase in racial integration can be accounted for by changes in households' demographic characteristics. To perform the decomposition our paper employs simple and intuitive methods. We first estimate the *counterfactual* spatial distribution of white and minority residents. That is, we simulate the spatial distribution of city residents one would observe in the current period, if individuals had the same demographic characteristics as those prevalent in another base period. The computation of counterfactual distributions allow us then to calculate counterfactual dissimilarity indices which can be compared with the observed indices to perform the decomposition.[1]

Dissimilarity indices are calculated using restricted-use data from the U.S. Census. We use data spanning 20 years (1990 and 2010) because changes in spatial distributions occur

[1] Counterfactual distributions (or counterfactual distributional statistics) are at the core of decomposition methods in economics. In the classical Oaxaca-Blinder wage decomposition researchers often simulate the counterfactual mean: how the mean wage of a demographic group would look like if they experience returns of a counterfactual group. More recent papers estimate the counterfactual at every point of the distribution. For example, Albrecht et al. (2003) use quantile decomposition techniques to simulate the counterfactual distribution of female wages: the wage distribution of females if they had the same demographic characteristics (endowments) as males. In a recent application in urban economics, McMillen (2008) simulates the distribution of home prices in Chicago in 2005 assuming that home characteristics remain constant as in 1995 (among other counterfactual simulations). For an excellent overview of decomposition methods in economics, see Fortin et al. (2011).

slowly. We have access to the complete microdata file from the 1990 Decennial Long Form Census and to the internal 2012 5-year microdata file (2008-2012) from the American Community Survey. Since dissimilarity indices depend on the level of geographic aggregation we make an effort to calculate changes over time across consistent geographic boundaries. This guarantees that any changes in the indices are due to changes in where people live, not changes in how locations are divided into census tracts. Overall, the average dissimilarity index across the 99 largest U.S. urban areas declined by 6.1, which is 10.7 percent of the 1990 average. Results also suggest that, in both periods, there is substantial variation of racial integration across urban areas. Our dissimilarity indices are highly correlated with but not equal to other indices in the literature computed with publicly available data (Frey, 2010). At the very least, our efforts should provide new dissimilarity indices that can be safely used to track patterns of racial integration over time.

Using aggregate city-level data, we also show that dissimilarity indices are highly correlated with the urban area's demographic characteristics. And more importantly, these urban areas have experienced substantial changes between 1990 and 2010. For example, the average city in 2010 features more educated and more single individuals than in 1990. Changes in the aggregate characteristics of urban areas mask important changes in the demographic composition *within* cities. Presumably, as a result of redevelopment, filtering, and changes to the urban infrastructure, the spatial distribution of racial groups within cities could have changed over time. And these changes mechanically determine levels of racial integration and corresponding dissimilarity indices.

How do changes in demographic characteristics, such as income, education and age, in *each* city account for changes in racial integration? The method we propose takes within-city variation into account to estimate spatial counterfactual distributions in each city and for each racial group. That is, we simulate the spatial distribution of whites and blacks in each urban area in 2010 assuming these individuals had the same demographic characteristics as individuals in 1990. The counterfactual distributions allows us to compute dissimilarity indices and to perform a simple decomposition. Our results suggest that, on average, changes in demographic characteristics can account for about 37 percent of the changes in racial integration between this period. The decomposition can be visualized in a series of intuitive maps that show the contribution of each neighborhood to the dissimilarity index. For space constraints, in the paper we only show results for the Washington, DC, and Los Angeles urban areas, but present an online appendix with figures for the 99 largest urban areas in the U.S.

Our paper adds to a long literature seeking to measure and understand how racial integration has changed over time. Empirical estimates agree that integration declined from the late 19th to mid 20th century (around 1970), but has increased since (Timberlake and Iceland, 2007; Rugh and Massey, 2014; Intrator et al., 2016; Firebaugh and Farrell, 2016; Logan and Parman, 2017). Since 1970, other characteristics of neighborhoods blacks and whites live in has converged as well. For example, Firebaugh and Farrell (2016) show that from 1980-2010, blacks and whites live in more similar neighborhoods as measured by their poverty rates and household median income; this convergence is due to improvements in black neighborhood characteristics and declines for white neighborhoods and has been greater for poverty and median income than for integration by race. Our paper is also related to a long literature focused on the residential location of minorities. The vast majority of papers study the determinants of urban segregation between blacks and whites (Reardon and O'Sullivan, 2004; Cutler et al., 1999; Galster and Cutsinger, 2007, for example),[2] and many have suggested that minority segregation is strongly associated with demographic characteristics (Bayer et al., 2004; Iceland and Wilkes, 2006; Iceland and Scopilliti, 2008; Lichter et al., 2010). While it is clear that changes in racial integration are closely associated with changes in demographic characteristics, our study is the first to simulate counterfactual dissimilarity indices, perform a formal decomposition, and evaluate what portion of the changes in racial integration can be accounted for by changes in individuals' demographic characteristics. Assessing the de-

[2] A growing number of studies also analyze the segregation patterns between Hispanics and other groups (Bayer et al., 2004; Johnston et al., 2007; Iceland, 2004, for example).

terminants of racial integration is important, since it has been shown that integration can affect individual outcomes, including health (Jackson et al., 2000), child education and income (Andrews et al., 2017; Chetty et al., 2016), and poverty and inequality (Ananat, 2011), among others.

Our paper also brings a methodological contribution to the literature. To compute counterfactual spatial distributions we follow the methods proposed by Carrillo and Rothbaum (2016) who add *space* to the seminal framework of DiNardo et al. (1996). Carrillo and Rothbaum (2016) show how to simulate a counterfactual *spatial* distribution: the distribution of a variable across space assuming that other observable characteristics are those from a counterfactual group. In order to compute the counterfactual dissimilarity indices, we modify and extend Carrillo and Rothbaum (2016)'s approach and show how to construct counterfactual spatial distributions of population subgroups, which allow to compute counterfactual statistics. While the method has been tailored to address our specific research question, the general approach can be applied in other settings, and we hope it is useful in other applications.

The rest of the paper is structured as follow. Section 2 provides description of our data and calculation of dissimilarity indices. Section 3 discusses the determinants of racial integration. In section 4 we describe our methods to perform the decomposition and present results. Finally, the last section concludes.

2 Racial Integration in U.S. Urban Areas

To compute the imbalance of racial distributions in U.S. cities we use a standard approach. In each urban area we compute the black-white racial dissimilarity index in 1990 and 2010 to evaluate both the level of racial imbalance and changes over time. The dissimilarity index D in urban area j measures the proportion of the population of one of the two groups that would need to move to achieve full racial balance across space. Mathematically, it is computed as follows:

$$D_j = \frac{1}{2} \sum_{i=1}^{N} \left| \frac{b_{ij}}{B_j} - \frac{w_{ij}}{W_j} \right|. \tag{1}$$

Here b_{ij} is the black population in the i^{th} sub-area (census tract), B_j is the total black population of urban area j, w_{ij} is the white population of the i^{th} census tract, and W_j is the total white population in urban area j for which the dissimilarity index is being calculated.

The index is calculated using data from the U.S. Census. Since changes in spatial distributions occur slowly, we use data spanning 20 years. To estimate dissimilarity indices in 1990, we use the complete microdata file from the 1990 Decennial Long Form Census. The 2010 dissimilarity indices are estimated using the internal 2012 5-year microdata file (2008-2012) from the American Community Survey. The 1990 Long Form includes about one sixth of all households in the United States and the ACS 5-year file covers about one eight of all households.

As changes in the level of aggregation can affect the calculated level of racial imbalance, we would like to calculate changes over time across consistent geographic boundaries.[3] For the city boundaries, we use the July 2015 delineation of Core Based Statistical Areas (CBSA). CBSAs are collections of counties that correspond to metropolitan and micropolitan areas in the United States. However within each CBSA, the block and tract boundaries change over time. In order to have constant geographies, we use the block relationship files provided by the U.S. Census Bureau.[4] For blocks with changed boundaries, we group the smallest set of blocks in each period so that there are no overlaps in the boundaries between any set of standardized blocks across the two periods. An example of this process is shown in Figure 1.[5] We then assign each set of standardized blocks into the 2010 census tract in

[3]This guarantees that any changes are due to changes in where people live, not changes in how locations are divided into census tracts. However, it also means that our estimates will differ slightly from others in the literature. For example, Frey and Myers (2005) uses block groups as defined in each census year to calculate the dissimilarity index.

[4]Available at https://www.census.gov/geo/maps-data/data/relationship.html. The 2000 relationship files relate the blocks from 1990 to 2000 and the 2010 relationship files relate blocks from 2000 to 2010. We link these files to create the relationships from 1990 to 2010.

[5]The standardization procedure is discussed in greater detail in the Appendix.

which the majority residents in the 2010 standardized blocks live. As an example, for the Washington, DC CBSA, in 98 percent of cases for both years, all blocks in the standardized group are in the same tract.

To show the dissimilarity calculation, we focus on Washington, DC. First, we calculate the share of blacks and whites in each standardized census tract, shown in Figure 2, Panels A and B. These maps confirm that the majority of blacks live in the South-East areas of the city, while a large fraction of whites live in the West. These simple maps allow to visually inspect the extent of racial integration in an urban area. It is also useful to estimate the contribution to the dissimilarity index of each location i (census tract). To achieve this purpose we first identify locations (1) with a larger share of whites than in the CBSA ($\frac{w_{ij}}{W_j} > \frac{b_{ij}}{B_j}$), and (2) with a larger share of blacks than in the CBSA ($\frac{w_{ij}}{W_j} < \frac{b_{ij}}{B_j}$). In Panel C of Figure 2, areas in (1) are shaded blue whereas areas in (2) are shaded red. As expected, we confirm that the locations in the far East and far West of the city center greatly contribute to its racial imbalance. Note that this Figure provides a very intuitive and simple way to visualize racial integration across space and to compute the dissimilarity index.[6]

Finally, to assess how racial integration has changed over time, we start by computing the dissimilarity indices in both 1990 and 2010. Overall, the Washington DC dissimilarity index decreased from 66.5 in 1990 to 61.6 in 2010. This change, however, masks interesting changes in the spatial distributions of the black and white population that can be fully appreciated by examining Figure 3. We plot the contributions to the 1990 and 2010 dissimilarity indices in Panels A and B of Figure 3, respectively, with their corresponding change in Panel C. These figures provide interesting insights. It is clear that the share of white population near the city center has increased over time, leading to lower contributions to the dissimilarity index. The share of blacks, on the other hand, has substantially increased in areas at the far East portions of this MSA, leading to greater contributions to the index. While some areas have become more integrated, others display less integration; the former effect dominates

[6] By adding one half of (the absolute value of) each area's contribution we can estimate the index shown in equation 1.

the latter leading to a decline in the overall index. In sum, to understand changes in racial integration, it is important to look beyond the aggregate dissimilarity indices and analyze the contribution of each geographic unit to the index.

Besides Washington DC, we also present similar results for Los Angeles CBSA in Figure 4. Racial integration experienced a small overall increase between 1990 and 2010 (dissimilarity index declined form 71.6 to 68.1 during this period). Most of the decline seems to be due to improvements in racial integration in the central areas of LA.

We evaluate changes in racial distributions in each of the 99 largest metro areas in the United States, as evaluated by Frey (2010).[7] Due to space constraints we cannot present detailed maps and results in the paper but these are available in an online appendix. In each of these locations, we calculate the dissimilarity index in 1990 and 2010 given the constant geographies. The results are shown in Table 1. In 91 percent of these cities, racial imbalance declined between 1990 and 2010. Overall, the dissimilarity index declined by an average of 6.1, which is 10.7 percent of the 1990 average.[8]

3 Determinants of Racial Integration

A growing number of papers suggest that racial and minority integration are strongly associated with the population's demographic characteristics (Bayer et al., 2004; Iceland and Wilkes, 2006; Iceland and Scopilliti, 2008; Lichter et al., 2010). Racial integration has been found to be determined in part by income, education, marital status, among other variables. In this section we show that, as suggested in the literature, racial integration is strongly correlated with demographic characteristics in our sample of U.S. cities. We also document significant changes in the demographic characteristics of urban areas between 1990 and 2010.

To show that dissimilarily indices are correlated with the aggregate features of a city, we estimate a conventional mutivariate regression model and present results in Table 2.

[7]Frey (2010) evaluates dissimilarity indices in the 100 largest cities. The final number of CBSAs in our study is 99 as Poughkeepsie, NY was combined into the New York City CBSA in the most recent delineation.

[8]5.5 weighted by the 2010 population, which is 8.3 percent of the initial weighted average.

The dependent variable is the dissimilarity index. Independent covariates include (log of) percapita income, the proportion of the population who are married and who have a college degree, as well as a set of variables that describe the population's age composition. The first and second column use the 1990 and 2010 samples, respectively. The third column displays results when observations for both years are pooled and includes city fixed effects. Despite the relatively low degrees of freedom, it is clear that racial integration is strongly correlated with the demographic characteristics of urban areas. For example, areas that have experienced an increase in percapita income of, say 10%, on average also experienced an increase of the dissimilarity index of about 0.03. Urban areas that have a larger share of "young professionals" (age 26 to 35) and more educated individuals also have more racial integration. In sum, even at the aggregate CBSA level, dissimilarily indices are correlated with many of the aggregate features of a city.

The demographic characteristics of urban areas have substantially changed between 1990 and 2010. Table 3 shows that cities in 2010 have more educated and more single individuals.[9] In addition, both the proportion of people born in the U.S. as well as the urban population's age distribution exhibit substantial changes during this time period. Changes in the *average* demographic characteristics of urban areas could mask important changes *within* cities. And the demographic composition of neighborhoods within cities may precisely be the key determinant of the overall city's racial integration. To illustrate how much the composition of neighborhoods within cities has changed, we present in the first and second panels of Figure 5 average percapita income (as a fraction of the median income in the CBSA) in Washington DC in 1990 and 2010, respectively. We zoom in and focus in the District (rather than in the full CBSA) to facilitate visualizing these changes. One can clearly see that there have been important changes in the distribution of income across space over this period: some areas that exhibit lower than average income in 1990, are populated with people earning higher than average income in 2010 (particularly at the center of the city). We arrive to

[9] Between 1990 and 2010, the proportion of the population with a college degree (legal spouse) has increased (decreased) by about 6 (4) percentage points.

a similar conclusion when we study the spatial distribution of college educated individuals (Figure 6). In sum, both the average characteristics of individuals living in cities as well as the composition of neighborhoods within cities in the U.S. has substantially changed over the past two decades.

To what extent do changes in demographic characteristics account for changes in racial integration? To address this question we need to first assess how dissimilarity indices would have looked like in *each* urban area in 2010, had the spatial composition of their neighborhoods remained the same as in 1990. A conventional Oaxaca-Blinder approach using Tables 2 and 3 is not appropriate because they cannot account for variation of the composition of the neighborhoods *within* cities. The next section develops a method that takes spatial within-city variation into account and that may appropriately decompose changes of the dissimilarity index in each city as the sum of two components: the part than can be "explained" by changes in observed demographic characteristics and a part that remains "unexplained."

4 Decomposition

In this section, we first develop and present the methods required to simulate counterfactual dissimilarity indices. We then perform the decomposition and present results.

4.1 Methods

The method described in this section is an extension of Carrillo and Rothbaum (2016) who apply the insights of DiNardo et al. (1996) to analyze spatial distributions. To keep our exposition self-contained, we carefully review the decomposition method and highlight the methodological contribution in our specific application.

Before providing details it is useful to define some basic notation. Let random vector $[y_1, y_2, X]$ denote the residence location coordinates of individuals in an urban area and X

denote a random vector of relevant covariates (such as income and education, for example.) Let $Y = [y_1, y_2]$ be our variable of interest (the unconditional residence choice of individuals) and $T = t_0$ and $T = t_1$ refer to the two mutually exclusive periods (1990 and 2010) in each of the areas we analyze. Unlike Carrillo and Rothbaum (2016), we also focus on two mutually exclusive groups in the population $G = g_b$ and $G = g_w$; in our empirical application these two mutually exclusive groups correspond to blacks ($G = g_b$) and whites ($G = g_w$).

Let $f(y_1, y_2, X \mid T = t_j, G = g_k)$ denote the joint probability density function in each period, where $j = \{1, 0\}$, $k = \{b, w\}$ and notice that the unconditional density functions $f(y_1, y_2 \mid T = t_j, G = g_k)$ measure the spatial distribution of individuals of each group at each period and can be easily identified from the data. The spatial distribution of the population in period t_1 is equivalent to the marginal joint density of random vector $[y_1, y_2 \mid T = t_1, G = g_k]$,

$$f(y_1, y_2|T = t_1, G) = \int f(y_1, y_2|x, T = t_1, G)h(x|T = t_1, G)dx \tag{2}$$

where T is a random variable describing the period from which an observation is drawn and x is a particular draw of observed attributes of individual characteristics from a random vector of characteristics X. $f(y_1, y_2|x, T = t_1, G)$ is the (conditional) density of vector Y given that a particular set of attributes x have been picked, and $h(x|T = t_0, G)$ is the probability density of individual attributes evaluated at x.

Note that we are conditioning on each racial group G as well. This is the main methodological difference between the approach in this paper and the methods proposed by Carrillo and Rothbaum (2016). While it may look straightforward, this is a powerful generalization of the method that facilitates the construction of functions of counterfactual densities including dissimilarity indices and any other related statistic. This generalization of the method could also be applied in many other applications.

The density of Y in period t_0 is defined similarly.

$$f(y_1, y_2|T = t_0, G) = \int f(y_1, y_2|x, T = t_0, G)h(x|T = t_0, G)dx \quad (3)$$

We highlight again that for each group G and time period T, y_1 and y_2 are observed in the data; hence the spatial distributions can be estimated using any parametric on non-parametric method.

For a particular group G, we would like to assess how the density of Y in period t_1 would look if the individual attributes x (age, education and income, for example) were the same as in period t_0. We denote this counterfactual distribution as as $f_{x_1 \to x_0}$ and express it symbolically as[10]

$$f_{x_1 \to x_0}(y_1, y_2 \mid G) = \int f(y_1, y_2|x, T = t_1, G)h(x|T = t_0, G)dx \quad (4)$$

DiNardo et al. (1996) proposed a semi-parametric approach to estimate such counterfactual. Using Bayes' rule they recognized that

$$\frac{h(x|T = t_0, G)}{h(x|T = t_1, G)} = \frac{\frac{\Pr(T=t_0|X=x,G)}{\Pr(T=t_0|G)}}{\frac{\Pr(T=t_1|X=x,G)}{\Pr(T=t_1|G)}} = \frac{\frac{\Pr(T=t_0|X=x,G)}{1-\Pr(T=t_0|X=x,G)}}{\frac{\Pr(T=t_0,G)}{1-\Pr(T=t_0,G)}} = \tau_{x_1 \to x_0}(x|G) \quad (5)$$

One may use Equation 5 to substitute $h(x|T = t_0, G)$ in Equation 4 and thereby obtain Equation 6.

$$f_{x_1 \to x_0}(y_1, y_2|G) = \int f(y_1, y_2|x, T = t_1, G)h(x|T = t_1, G)\tau_{x_1 \to x_0}(x|G)dx. \quad (6)$$

Notice that this expression differs from Equation 2 only by $\tau_{t_1 \to t_0}(x|G)$. Typically $\tau_{t_1 \to t_0}(x|G)$ is referred to as "weights" that should be applied when computing the counterfactual spatial distribution of our variable of interest. The true weights are unknown, but they can be estimated.

[10] The subscript $x_1 \to x_0$ indicates that the attribute data from period t_0 will be "replaced" by data from period t_1.

The estimation algorithm is simple and can be summarized in a set of steps assuming that a random sample of observations for periods t_0 and t_1 and groups g_b and g_w is available. Denote N_{jk} the sample sizes of each of these four mutually exclusive samples and let i index each observation. Then perform the following steps for *each* subgroup G:

- Step 1: Estimate $P(T = t_0|G)$ using the share of observations where $T_i = t_0|G$; that is, compute: $\hat{\Pr}(T_i = t_0|G) = N_{t_0,g_k}/(N_{t_0,g_k} + N_{t_1,g_k})$.
- Step 2: Estimate $P(T = t_0|X = x, G)$, by estimating a logit model using the pooled data. The dependent variable equals one if $T_i = t_0$ and explanatory variables include the vector of individual attributes x_i.
- Step 3: For the subsample of observations where $T_i = t_1$, estimate the predicted values from the logit $\hat{\Pr}(T_i = t_0|X = x_i, G) = \exp(x_i\hat{\beta})/\left(1 + \exp(x_i\hat{\beta})\right)$, where $\hat{\beta}$ is the parameter vector from the logit regression. Then, compute the estimated weights $\hat{\tau}_{x_1\to x_0}(x|G)$.
- Step 4: For the subsample of observations where $T_i = t_1$, compute the joint density of coordinates $[y_1, y_2|T = t_1, G]$ applying the estimated sample weights given by $\hat{\tau}_{x_1\to x_0}(x|G)$.

We would like to assess the differences between the spatial distributions of interest. In particular, for each group G, we seek to assess how much of the changes in the joint distribution of $[y_1, y_2 \mid G]$ between period t_0 and t_1 can be accounted for by changes in individual attributes X at a given point (y_1, y_2). This decomposition can be performed as follows

$$f(y_1, y_2|t_1, G) - f(y_1, y_2|t_0, G) = \{f(y_1, y_2|t_1) - f_{x_1\to x_0}(y_1, y_2|G)\} - \{f(y_1, y_2|t_0) - f_{x_1\to x_0}(y_1, y_2|G)\} \tag{7}$$

The first term in brackets on the right hand side of the equation above measures the portion of the changes that can be accounted for by differences in the distribution of covariates. The second term measures the "unexplained" part of the changes in the distribution of vector Y.

We are now ready to describe how counterfactual dissimilarity indices can be computed. First notice that, in each period T, we have used the spatial distributions $f(y_1, y_2|T, G = g_b)$ and $f(y_1, y_2|T, G = g_w)$ to estimate the dissimilarity indices described in equation 1. To be precise, if the spatial distribution of the population of blacks and whites is estimated using a kernel within each subarea a_i (census tract), these estimates can be used to compute the dissimilarity index at time $T = t_0$ as follows:

$$D_0 = g\left(f(y_1, y_2|T = t_0, G = g_b), f(y_1, y_2|T = t_0, G = g_w)\right),$$

where the function $g(.)$ applies the correct transformation to the densities to compute equation 1. Formally, given a set of locations $a_i \in A$ in period t, the dissimilarity formula in equation 1 is equal to:

$$D_t = \sum_{a_i \in A} \left| \iint_{a_i} f(y_1, y_2|T = t, G = g_w)\, dy_1 dy_2 - \iint_{a_i} f(y_1, y_2|T = t, G = g_b)\, dy_1 dy_2 \right|,$$

where the geographic unit of aggregation a_i can be tracts, block groups, blocks, etc. To be consistent with other studies we let a_i denote a census tract.

Similarly, one can compute the dissimilarity index at period $T = t_1$

$$D_1 = g\left(f(y_1, y_2|T = t_1, G = g_b), f(y_1, y_2|T = t_1, G = g_w)\right).$$

We define the counterfactual integration rate D_c as the dissimilarity index that one would find in period t_1 if the demographic characteristics of individuals of both groups (black and white) remain constant as in period t_0. This statistic can be computed by using the counterfactual spatial densities described above

$$D_c = g\left(f_{x_1 \to x_0}(y_1, y_2|T = t_1, G = g_b), f_{x_1 \to x_0}(y_1, y_2|T = t_1, G = g_w)\right).$$

The decomposition is straightforward:

$$D_1 - D_0 = (D_1 - D_c) + (D_c - D_0).$$

The first term in the right hand side measures the portion of the changes of the dissimilarity index that can be accounted for by changes in characteristics. The second term shows the portion of the gap that remains unexplained.

Given the dissimilarity contribution of each tract, we can also decompose the dissimilarity in each location (CBSA in this paper) by additional geographies, such as CMSA, MSA, or county to give the aggregate contribution of these larger areas to the CBSA total.

4.2 Results

We apply our spatial counterfactual technique to determine the counterfactual dissimilarity index in 2010 given the 1990 individual and household characteristics for blacks and whites. In order to apply this method we need to observe the exact location of households as well as their demographic characteristics in both periods. The restricted-use data from the U.S. Census (1990 Decennial Long Form Census and 2008-2012 American Community Survey) provides this information.

The counterfactual dissimilarity index was calculated separately for each CBSA. We first estimate the spatial distribution of each racial group in both periods, which allows us to compute unconditional dissimilarity indices. Then, for each race group, we estimate a logit model where the dependent variable equals to 1 if the observation belongs to the 1990 sample, and zero otherwise; covariates include individual's demographic characteristics such as income, age and education. Results from the logit model are used to compute the sample weights in equation 5. The sample weights are then used to estimate a counterfactual distribution of the population (of each race) in 2010 assuming that their demographic characteristics remain the same as in 1990. Finally, one can use the counterfactual spatial distribution to estimate a counterfactual dissimilarity index and perform the decomposition at each CBSA. In sum, for each race group in each urban area, we reweight the observations before calculating the dissimilarity index so that we have a separate counterfactual spatial distribution for whites and blacks from which to calculate the dissimilarity.

As an example, we walk through the process for Washington, DC. In Figure 3, we show the dissimilarity index in Washington, DC in 1990 and 2010, as well as the difference between the two at the standardized tract level ($D_{1990}, D_{2010}, and\, [D_{2010} - D_{1990}]$ respectively). In order to calculate the counterfactual ($D_{2010 \to 1990}$), we regress a dummy for year observed (1 if in 1990 longform, 0 if in 2012 5-year ACS) on the observable individual and household characteristics in the pooled data set from the 1990 longform and 2012 ACS 5-year microdata, with separate regressions for white and black individuals. Results are shown in Table 4 and deserve some discussion. Note that for both racial groups higher levels of education are associated with being in the 2010 ACS sample. Put differently, the average education level in Washington DC has increased between 1990 and 2010. Also note that marital status other than never married (the excluded dummy) is associated with being in the 1990 longform sample, suggesting that the share of married individuals has decreased between 1990 and 2010.

Estimates from the logit models in Table 4 allow us to compute sampling weights which are applied to estimate the counterfactual distribution of the population in 2010. That is we estimate how the spatial distribution of blacks and whites in 2010 would look if they had the same demographic characteristics as in 1990. We use these counterfactual distributions to compute the counterfactual dissimilarity index; results are are shown in Figure 7. In Panel A we show the counterfactual contribution to the dissimilarity index in 2010 given the black and white demographic characteristics from 1990 ($D_{2010 \to 1990}$). In Panels B and C, we decompose the change in dissimilarity shown in Figure 3, into the portion explained by changes in demographic characteristics ($D_{2010} - D_{2010 \to 1990}$), and the portion unexplained by demographic changes ($D_{2010 \to 1990} - D_{1990}$). Washington, DC experienced a decline of 4.9 in dissimilarity from 66.5 to 61.6. Of that decline, 1.4 (28 percent) can be accounted for by changes in demographics.

In Figure 4, we show the dissimilarity index for Los Angeles, CA in 1990 and 2010 (Panels A and B) as well as the difference (Panel C). We apply the spatial counterfactual and show the resulting dissimilarity in Figure 8 Panel A, along with the explained, and unexplained changes (Panels B and C). Los Angeles experienced a dissimilarity decline of 3.4, of which 1.6 (46 percent) can be accounted for by changes in demographics.

We summarize the counterfactual dissimilarity across the 99 CBSAs in our analysis in the last two columns of Table 1. The average share of the change in racial integration accounted for by changes in observable characteristics is 37 percent. A scatter plot of the share explained against the percentile in the distribution of share explained is shown in Figure 9. Most CBSAs are clustered 20 to 50 percent of the change in integration explained by changes in demographic characteristics.[11]

5 Conclusions

Racial integration (as measured by the dissimilarity index) declined on average by about 10 percent between 1990 and 2010 in U.S. urban areas. In order to assess how much of the change in racial integration can be accounted for by changes in individuals' demographic characteristics, we apply a novel spatial non-parametric decomposition technique. Our approach allows us to decompose the change of the dissimilarity index in *each* urban area into an "explained" and "unexplained" component. Results suggest that, in the average city, about one third of the dissimilarity index decline is due to changes in residents' demographic characteristics, such as education, income and age.

Results have important academic and policy implications. From an academic perspective, this paper shows how to perform decomposition of spatial distribution's statistics. The dissimilarity index is the specific focus of our application, but the technique works for any other

[11] At the 25th, 50th, and 75th percentiles, changes in observable characteristics account for 15, 39, and 58 percent respectively of the change in integration between 1990 and 2010.

function/statistic of a spatial distribution. We hope our technique fosters future research in this area.

For policy purposes, our results highlight the extent to which education, income and other demographic factors can determine racial integration. Importantly, in certain urban areas they play a more important role than in others. It is important to note, however, that our decomposition should be interpreted as an *accounting* exercise. Further research is needed to specify and estimate counterfactuals that have a causal interpretation.

References

Albrecht, J., A. Björklund, and S. Vroman (2003): "Is There a Glass Ceiling in Sweden?" *Journal of Labor Economics*, 21, 145–177.

Ananat, E. O. (2011): "The wrong side(s) of the tracks: The causal effects of racial segregation on urban poverty and inequality," *American Economic Journal: Applied Economics*.

Andrews, R., M. Casey, B. L. Hardy, and T. D. Logan (2017): "Location matters: Historical racial segregation and intergenerational mobility," *Economics Letters*.

Bayer, P., R. McMillan, and K. S. Rueben (2004): "What drives racial segregation? New evidence using Census microdata," *Journal of Urban Economics*, 56, 514 – 535.

Carrillo, P. E. and J. L. Rothbaum (2016): "COUNTERFACTUAL SPATIAL DISTRIBUTIONS," *Journal of Regional Science*, 56, 868–894.

Chetty, R., N. Hendren, F. Lin, J. Majerovitz, and B. Scuderi (2016): "Childhood environment and gender gaps in adulthood," in *American Economic Review*.

Cutler, D. M., E. L. Glaeser, and J. L. Vigdor (1999): "The Rise and Decline of the American Ghetto," *Journal of Political Economy*, 107, 455–506.

DiNardo, J., N. M. Fortin, and T. Lemieux (1996): "Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach," *Econometrica*, 64, 1001–1044.

Firebaugh, G. and C. R. Farrell (2016): "Still Large, but Narrowing: The Sizable Decline in Racial Neighborhood Inequality in Metropolitan America, 1980–2010," *Demography*.

Fortin, N., T. Lemieux, and S. Firpo (2011): "Chapter 1 - Decomposition Methods in Economics," Elsevier, vol. 4 of *Handbook of Labor Economics*, 1 – 102.

Frey, W. and D. Myers (2005): "Racial segregation in US metropolitan areas and cities, 1990-2000: patterns, trends, and explanations," *Population Studies Center Research Report*.

Frey, W. H. (2010): "New Racial Segregation Measures for Large Metropolitan Areas: Analysis of the 1990-2010 Decennial Censuses," .

Galster, G. and J. Cutsinger (2007): "Racial Settlement and Metropolitan Land-Use Patterns: Does Sprawl Abet Black-White segregation?" *Urban Geography*, 28, 516–553.

Iceland, J. (2004): "Beyond Black and White: Metropolitan residential segregation in multi-ethnic America," *Social Science Research*, 33, 248 – 271.

Iceland, J. and M. Scopilliti (2008): "Immigrant residential segregation in U.S. metropolitan areas, 1990–2000," *Demography*, 45, 79–94.

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

Intrator, J., J. Tannen, and D. S. Massey (2016): "Segregation by race and income in the United States 1970–2010," *Social Science Research*.

Jackson, S. A., R. T. Anderson, N. J. Johnson, and P. D. Sorlie (2000): "The Relation of Residential Segregation to All-Cause Mortality: A Study in Black and White," *American Journal of Public Health*, 61590, 615–617.

Johnston, R., M. Poulsen, and J. Forrest (2007): "Ethnic and Racial Segregation in U.S. Metropolitan Areas, 1980-2000," *Urban Affairs Review*, 42, 479–504.

Lichter, D. T., D. Parisi, M. C. Taquino, and S. M. Grice (2010): "Residential segregation in new Hispanic destinations: Cities, suburbs, and rural communities compared," *Social Science Research*, 39, 215 – 230.

Logan, T. D. and J. M. Parman (2017): "The National Rise in Residential Segregation," .

McMillen, D. P. (2008): "Changes in the distribution of house prices over time: Structural characteristics, neighborhood, or coefficients?" *Journal of Urban Economics*, 64, 573 – 589.

Reardon, S. F. and D. O'Sullivan (2004): "Measures of Spatial Segregation," *Sociological Methodology*, 34, 121–162.

Rugh, J. S. and D. S. Massey (2014): "SEGREGATION IN POST-CIVIL RIGHTS AMERICA: Stalled Integration or End of the Segregated Century?" *Du Bois Review: Social Science Research on Race*, 11, 205–232.

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

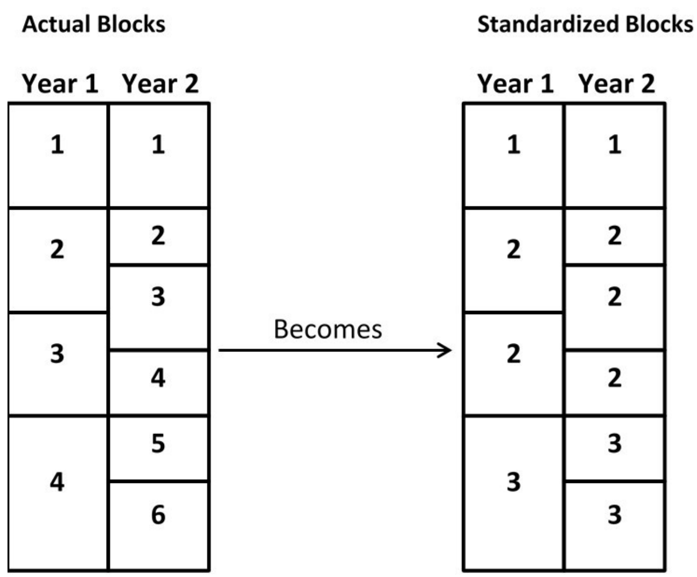
Appendix: Detailed Creation of Constant Geographies across Census/ACS Years

In this section, we discuss how we construct standardized census tracts that are constant between the 1990 longform census and the 2012 5-year ACS. We use mapping and boundary shapefiles provided by the National Historic Geographic Information System (NHGIS). First, we take the the census block files from the 1990, 2000, and 2010 census. For each CBSA, we selected only those blocks with their centroid within the 2015 CBSA boundaries. In the DC CBSA, there are 47,367 blocks in the 1990 census, 52,157 blocks in the 2000 census, and 91,418 blocks in the 2010 census. Using the block relationship files from the U.S. Census Bureau, we constructed standardized blocks (S-blocks) relating blocks from: 1) 1990 to 2000 and 2) 2000 to 2010 separately. To create the S-blocks, for a given block in period 1, we did the following:

1. Find all the blocks in period 2 that matched the initial period 1 block.
2. Find any additional period 1 blocks that matched to the set of period 2 blocks from step 1.
3. For the additional period 1 blocks repeat steps 1 and 2 until no new period 1 and 2 blocks are added to the set of matched blocks.
4. Assign all matched period 1 and 2 blocks to a set of grouped blocks to create a single S-block.
5. Repeat this for all period 1 blocks until all blocks in both periods are assigned to an S-block. We then linked the 1990-2000 and 2000-2010 S-blocks as above to create ones that spanned the 1990 to 2010 geographies. Therefore, any block boundaries that changed between 1990-2000 and 2000-2010 were aggregated into the smallest set of grouped blocks that had the same boundaries in 1990 and 2010, for a total of 29,740 S-blocks.

In order to assign each S-block into a 2010 census tract, we calculated the share of the 2012 5-year ACS population in each S-block in each census tract and assigned the S-block in both periods to the tract with the highest share. As an example, in Washington, DC for both periods, over 98 percent of S-blocks were in the same 2010 tract. An example of this procedure is shown in Figure 1.

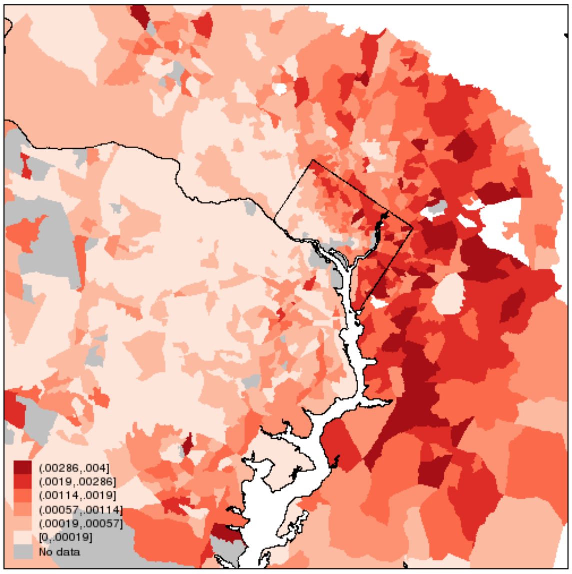
Figure 1:
Block Relationship Mapping Example



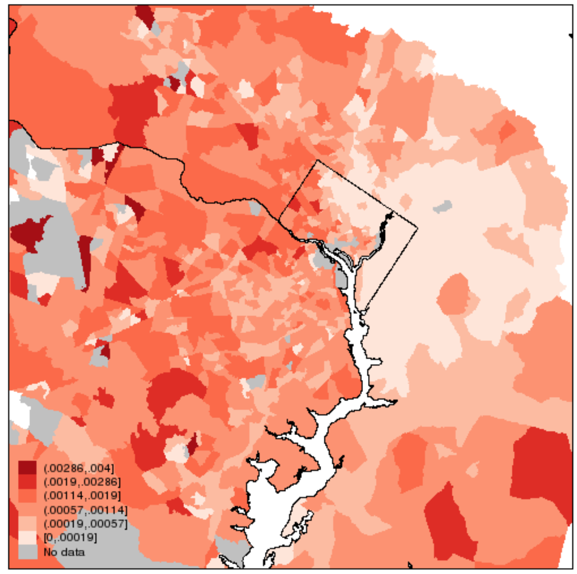
Notes: This figure shows a simple example of the block relationship mapping used to construct consistent geographies in the 2010 DC CBSA. It contains examples of the various types of blocks that are aggregated into larger groups of blocks that have the same boundaries in both periods.

Figure 2:
Racial Integration in Washington, DC, 2010

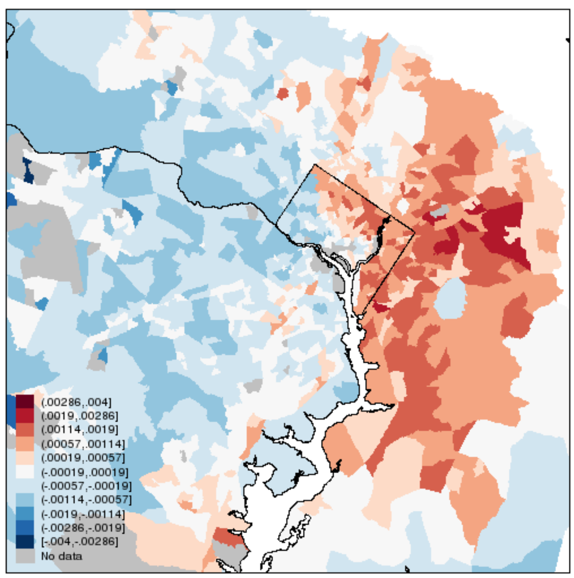
A. Blacks



B. Whites

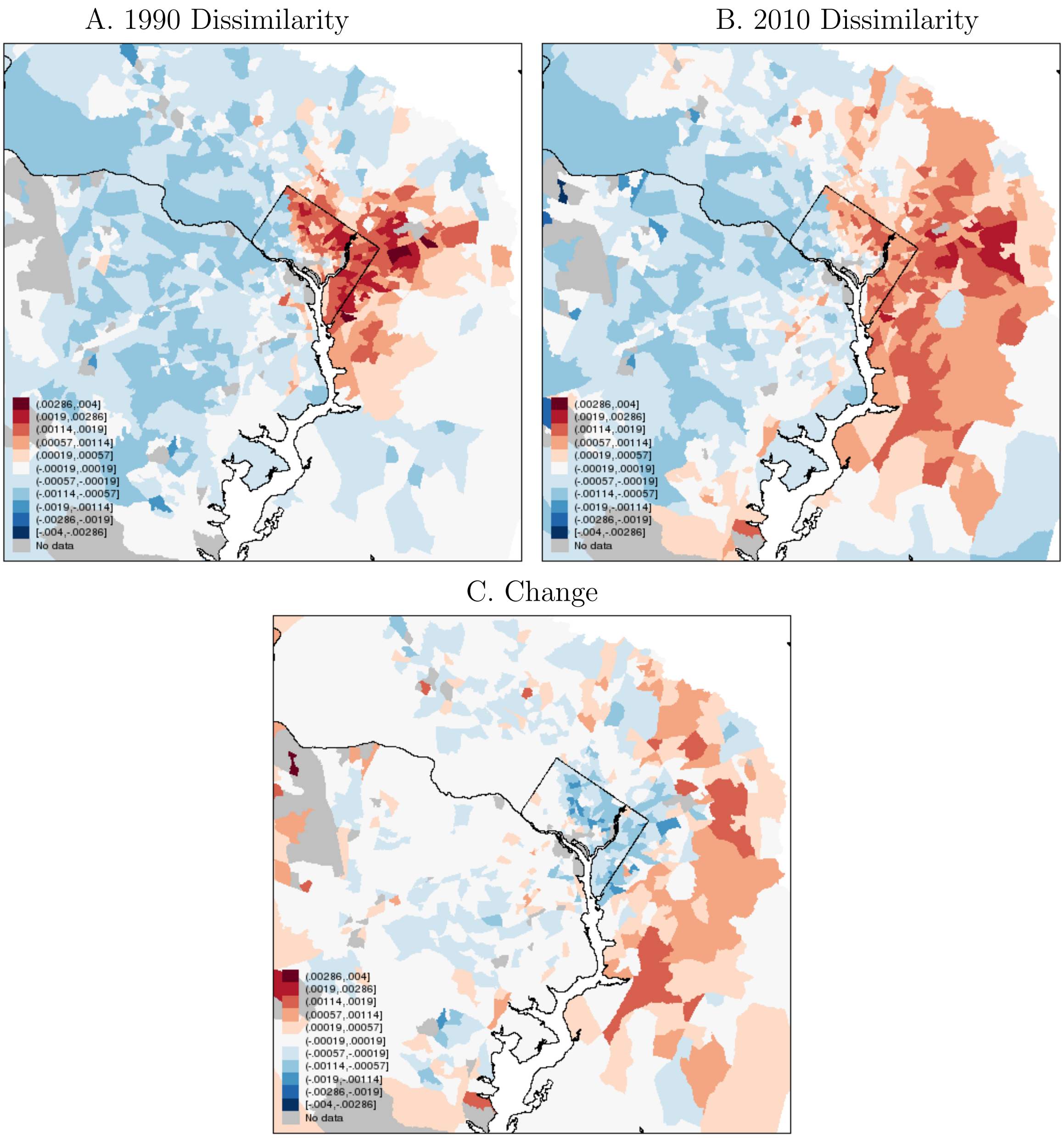


C. Dissimilarity



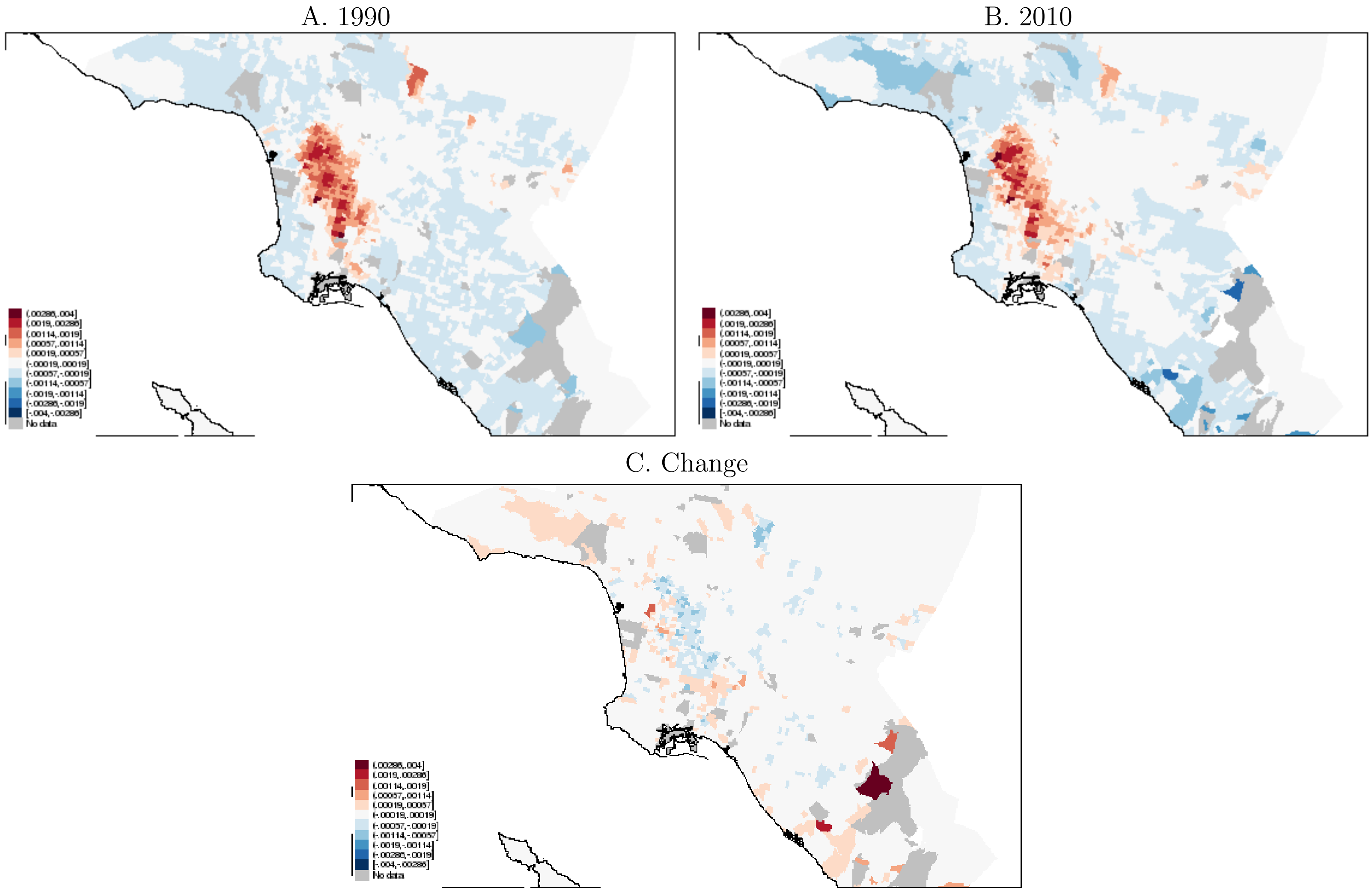
Notes: These maps show the distribution of blacks (Panel A) and whites (Panel B) in the Washington, DC CBSA. Darker areas in the first two panels have a higher share of the population for that group. Panel C shows the contribution of each tract to the dissimilarity index. As the dissimilarity is $D_j = \frac{1}{2}\sum_{i=1}^{N}\left|\frac{b_{ij}}{B_j} - \frac{w_{ij}}{W_j}\right|$, it can be decomposed into the contribution of each location i where 1) the share of whites in the area is greater than in the CBSA $\frac{w_{ij}}{W_j} > \frac{b_{ij}}{B_j}$ and 2) the share of blacks in the area is greater than in the CBSA $\frac{w_{ij}}{W_j} < \frac{b_{ij}}{B_j}$. Tracts with more whites as in 1) are shaded blue, and tracts with more blacks as in 2) are shaded in red.

Figure 3:
Contributions to Dissimilarity Index in Washington, DC



Notes: These maps show the contribution of each tract to the dissimilarity index (DI). As the dissimilarity is $D_j = \frac{1}{2}\sum_{i=1}^{N}\left|\frac{b_{ij}}{B_j} - \frac{w_{ij}}{W_j}\right|$, it can be decomposed into the contribution of each location i where 1) the share of whites in the area is greater than in the CBSA $\frac{w_{ij}}{W_j} > \frac{b_{ij}}{B_j}$ and 2) the share of blacks in the area is greater than in the CBSA $\frac{w_{ij}}{W_j} < \frac{b_{ij}}{B_j}$. Tracts with a larger fraction of whites are shaded blue, and tracts with a larger fraction of blacks are shaded in red. In Panel C, we display the difference in contributions to the DI by tract. Tracts where the share of whites increased relative to the CBSA average over time are shaded in blue and tracts where the share of blacks increased relative to the CBSA average are shaded in red.

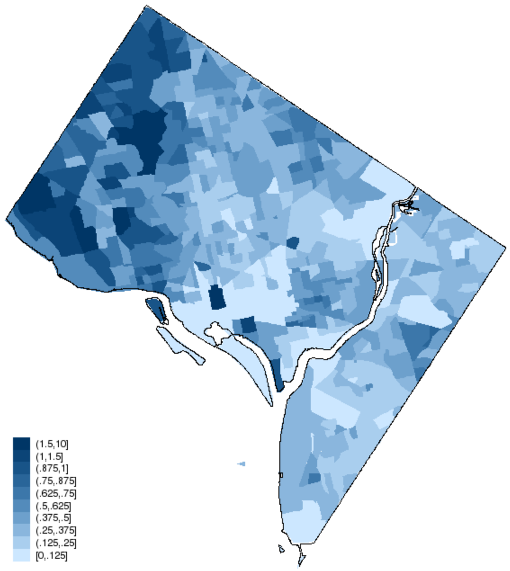
Figure 4:
Contributions to Dissimilarity Index in Los Angeles, CA CBSA



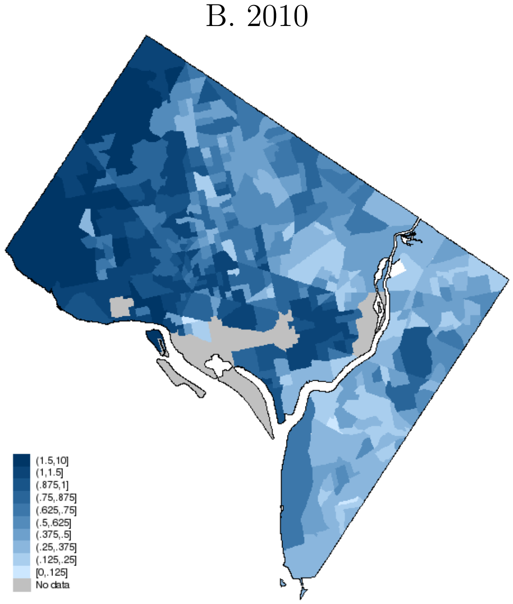
Notes: These maps show the contribution of each standardized tract to the dissimilarity index in the Los Angeles, CA CBSA in 1990 (Panel A) and 2010 (Panel B). As the dissimilarity is $D_j = \frac{1}{2}\sum_{i=1}^{N}\left|\frac{b_{ij}}{B_j} - \frac{w_{ij}}{W_j}\right|$, it can be decomposed into the contribution of each location i where 1) the share of whites in the area is greater than in the CBSA $\frac{w_{ij}}{W_j} > \frac{b_{ij}}{B_j}$ and 2) the share of blacks in the area is greater than in the CBSA $\frac{w_{ij}}{W_j} < \frac{b_{ij}}{B_j}$. Tracts with more whites as in 1) are shaded blue, and tracts with more blacks as in 2) are shaded in red. In Panel C, tracts where the share of whites increased relative to the CBSA average over time are shaded in blue and tracts where the share of blacks increased relative to the CBSA average are shaded in red.

Figure 5:
Household Income (Relative to the Median) in Washington DC

A. 1990



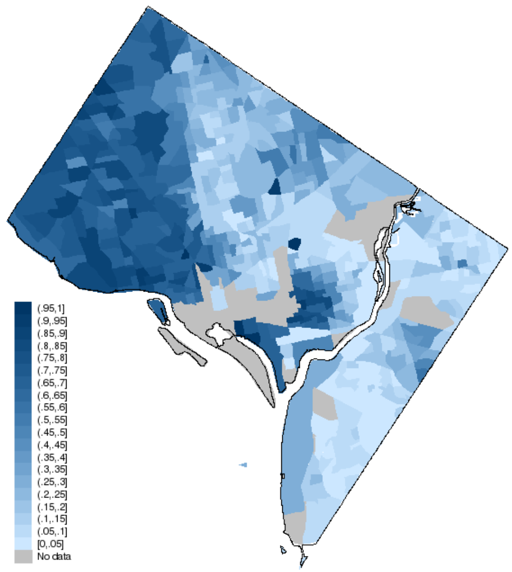
B. 2010



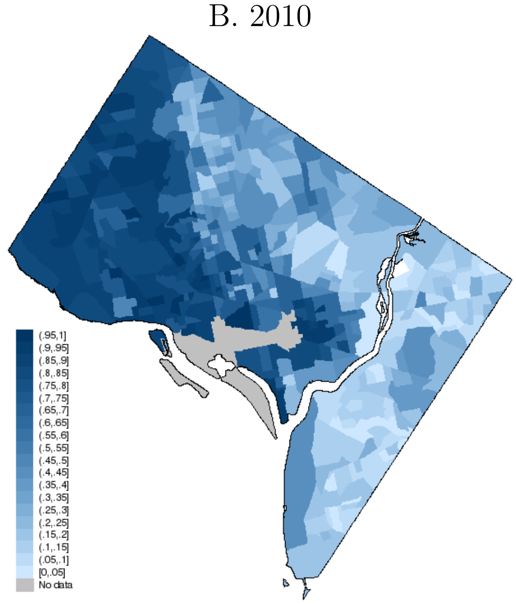
Notes: These maps plot average percapita income relative to the median income (of the CBSA) by census tract. Darker areas show census tracts with percapita income above the median.

Figure 6:
Share of Population with College Degree in Washington DC

A. 1990



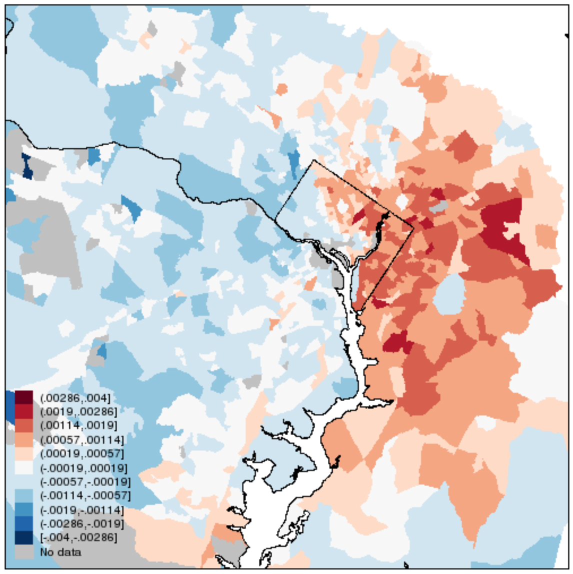
B. 2010



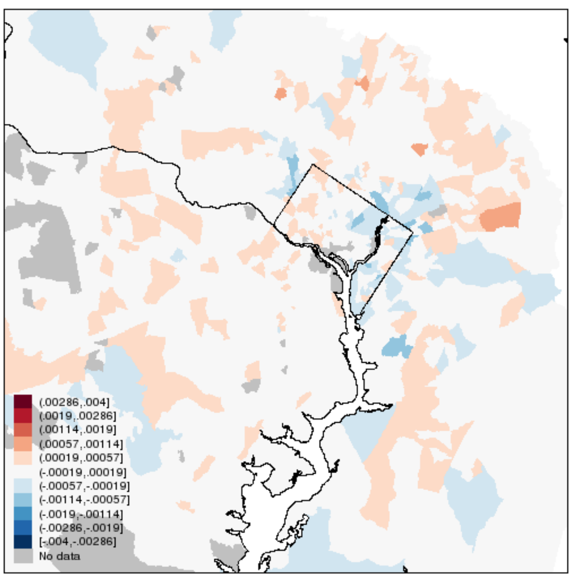
Notes: These maps plot the share of the population with college degree by census tract.

Figure 7:
Counterfactual Dissimilarity in Washington, DC

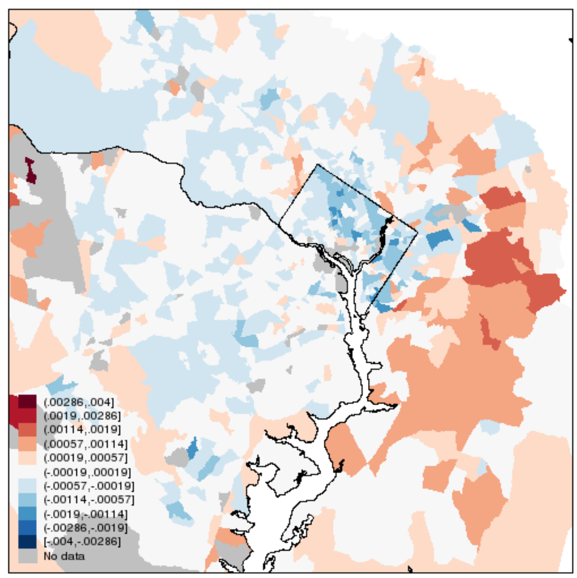
A. Counterfactual Dissimilarity



B. Explained Change ($D_{2010} - D_{2010\to1990}$)



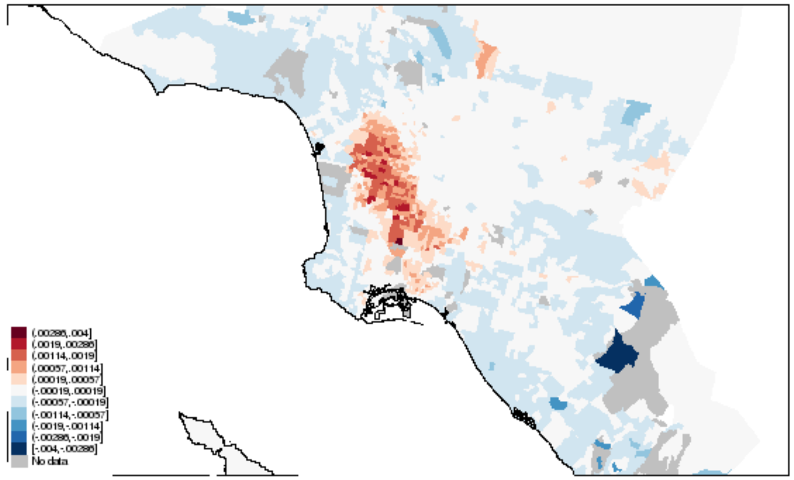
C.Unexplained Change ($D_{2010\to1990} - D_{1990}$)



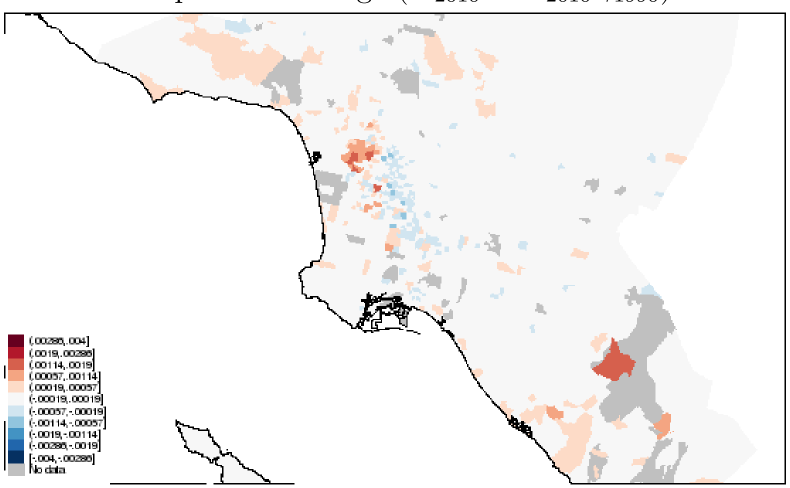
Notes: These maps show the counterfactual dissimilarity index (DI) by standardized tract in Washington, DC calculated by reweighting the 2012 5-year ACS observations to the 1990 longform sample demographic characteristics, shown in Table 4. Panels B and C decompose the observed change in the DI in Washington, DC in to the portions explained by the changes in demographic characteristics ($D_{2010} - D_{2010\to1990}$) and unexplained by changes in demographics ($D_{2010\to1990} - D_{1990}$). In Panel A, tracts with a higher share of whites than the CBSA are shaded blue, and tracts with a higher share of blacks are shaded in red. In Panels B and C, tracts where the share of whites increased relative to the CBSA average over time are shaded in blue and tracts where the share of blacks increased relative to the CBSA average are shaded in red.

Figure 8:
Counterfactual Dissimilarity in Los Angeles, CA CBSA

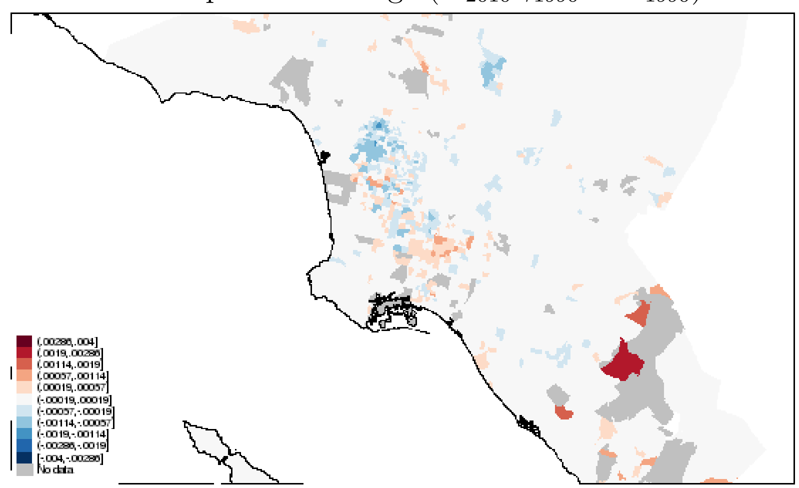
A. Counterfactual Dissimilarity



B. Explained Change ($D_{2010} - D_{2010 \to 1990}$)

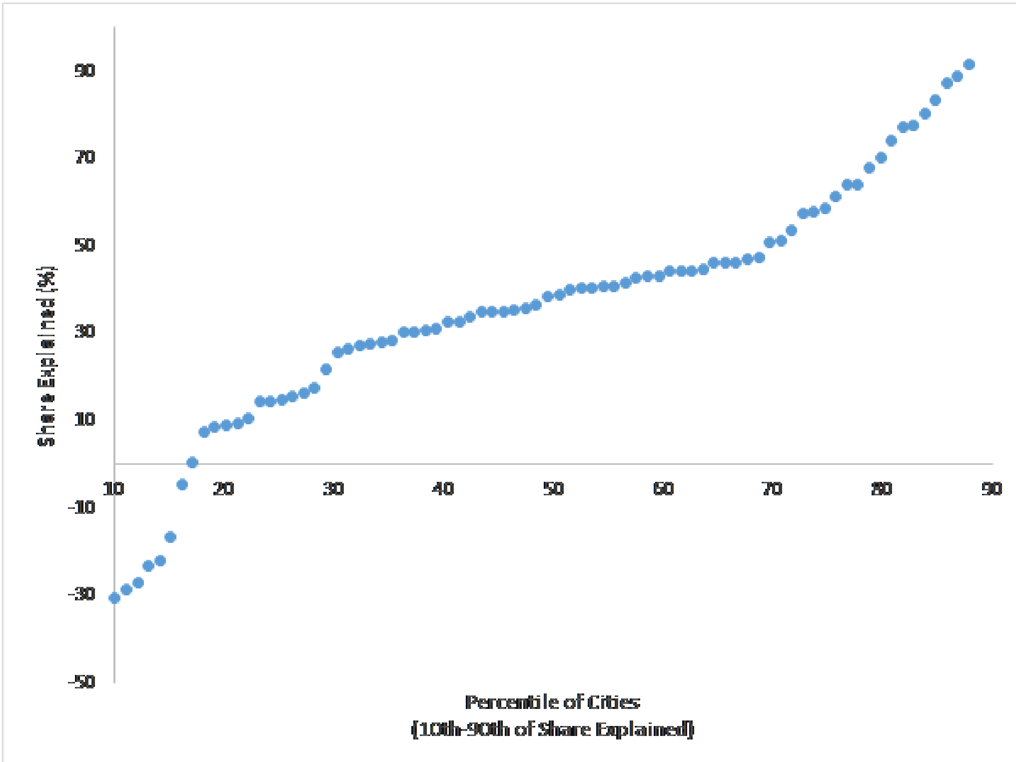


C.Unexplained Change ($D_{2010 \to 1990} - D_{1990}$)



Notes: These maps show the counterfactual dissimilarity index by standardized tract in Los Angeles, CA calculated by reweighting the 2012 5-year ACS observations to the 1990 longform sample demographic characteristics, shown in Table 4. Panels B and C decompose the observed change in racial integration in Washington, DC in to the portions explained by the changes in demographic characteristics ($D_{2010} - D_{2010 \to 1990}$) and unexplained by changes in demographics ($D_{2010 \to 1990} - D_{1990}$). In Panel A, tracts with a higher share of whites than the CBSA are shaded blue, and tracts with a higher share of blacks are shaded in red. In Panels B and C, tracts where the share of whites increased relative to the CBSA average over time are shaded in blue and tracts where the share of blacks increased relative to the CBSA average are shaded in red.

Figure 9:
Distribution of Dissimilarity Index Changes Explained by Changes in Observables



Notes: This figure shows the share of the changes in the dissimilarity index explained by changes in observable characteristics, with cities order from the lowest share explained to highest. For clarity, the figure is truncated at the 10th and 90th percentiles.

Table 1:
Changes in Dissimilarity Index (DI) from 1990 to 2010 in Largest U.S. CBSAs

City	DI: 1990	DI: 2010	Change	Explained	Share Explained
Akron, OH	71.2	61	-10.2	-3.5	33.8
Albany, NY	63.6	64.5	0.9	0.4	47.2
Albuquerque, NM	40.6	39.4	-1.2	-4.1	356
Allentown, PA	55.6	49.3	-6.3	-5.6	89
Atlanta, GA	67.8	59.3	-8.5	-1.5	17.6
Augusta, GA	47.3	45.4	-1.9	-1.5	80.3
Austin, TX	54.9	52.8	-2.1	-1.9	91.6
Bakersfield, CA	53.5	52	-1.6	-3.6	226.9
Baltimore, MD	72	65.8	-6.1	-2.2	35.4
Baton Rouge, LA	60.6	58.2	-2.4	-4.3	177.3
Birmingham, AL	72.2	67.5	-4.7	-1.3	28.3
Boise City, ID	56.3	45.3	-11	-5.6	51.2
Boston, MA	71	65.9	-5.1	-0.5	10.6
Bradenton, FL	76.1	57.8	-18.3	-5	27.4
Bridgeport, CT	69.4	69.6	0.2	3.1	1547
Buffalo, NY	81	73.3	-7.8	-0.7	9.5
Cape Coral, FL	80.5	63.7	-16.9	-7.3	43.2
Charleston, SC	50.9	39.5	-11.3	-3.9	34.8
Charlotte, NC	56.9	54.3	-2.6	-1.2	47.4
Chattanooga, TN	74.2	66.2	-8	-2.6	32.7
Chicago, IL	84.1	76.9	-7.3	-2.2	30.7
Cincinnati, OH	76.6	68.1	-8.5	-3.8	44.8
Cleveland, OH	83.2	74.5	-8.7	-0.1	0.7
Colorado Springs, CO	45.4	41.7	-3.8	1.4	-37.3
Columbia, SC	52.7	49.3	-3.4	-1.6	46.4
Columbus, OH	68.6	62.6	-6	-1.5	25.6
Dallas, TX	61.9	57.8	-4	-3.1	77
Dayton, OH	75.7	63.9	-11.8	-4.1	35.1
Denver, CO	65.6	63.3	-2.3	0.5	-21.8
Des Moines, IA	67.4	55.6	-11.8	-5.1	42.8
Detroit, MI	88	75.3	-12.7	-1.9	14.6
El Paso, TX	48.9	36.4	-12.5	2.1	-16.7
Fresno, CA	53.6	54	0.4	-0.4	-101.5
Grand Rapids, MI	75.2	66.3	-8.8	-3.9	44.3
Greensboro, NC	54.5	55.4	1	-1.1	-110.4
Greenville, SC	50.5	43.2	-7.3	-4.7	64.1
Harrisburg, PA	76.4	67.5	-9	-3.5	38.9
Hartford, CT	70.6	65.3	-5.3	-1.9	35.7

Continued on next page

Table 1 – *Continued from previous page*

City	DI: 1990	DI: 2010	Change	Explained	Share Explained
Honolulu, HI	45.8	43.1	-2.7	-1.6	57.4
Houston, TX	64.7	61.3	-3.4	-3	87.4
Indianapolis, IN	75.9	65.8	-10.1	-4	40.1
Jackson, MS	61.6	56.8	-4.8	-4	83.4
Jacksonville, FL	60.1	54.8	-5.2	-1.7	32.6
Kansas City, MO	73.5	61.2	-12.4	-3.8	30.4
Knoxville, TN	63.1	56.4	-6.7	-5.2	77.7
Lakeland, FL	64	47	-17	-3.7	21.8
Lancaster, PA	66.2	55.3	-10.9	-5	46.3
Las Vegas, NV	49.5	36	-13.5	-5.5	40.7
Little Rock, AR	60.8	58.4	-2.4	-1.4	57.8
Los Angeles, CA	71.6	68.1	-3.4	-1.6	46
Louisville, KY	69.9	58.5	-11.4	-3.5	31.2
Madison, WI	56.2	54.6	-1.6	-4	255.3
McAllen, TX	67.3	66.7	-0.7	-7.5	1126.6
Memphis, TN	67.7	63.2	-4.5	0.2	-4.7
Miami, FL	73.6	65.4	-8.1	-3.5	43.1
Milwaukee, WI	83.6	81.1	-2.6	-0.9	36.6
Minneapolis, MN	63.8	55.4	-8.4	-3.7	44.2
Modesto, CA	41.7	42.8	1	-3.1	-293.2
Nashville, TN	61.7	54.8	-6.9	-3.5	51
New Haven, CT	70.3	64	-6.3	-1.9	30.4
New Orleans, LA	67.6	64.1	-3.6	-2.1	58.6
New York, NY	81.6	78.1	-3.5	-1.4	40.7
Ogden, UT	57	46	-11	-7.7	70.4
Oklahoma City, OK	61.1	54.2	-7	-2.7	38.5
Omaha, NE	71.3	62.4	-8.9	-4.8	53.7
Orlando, FL	61	51.3	-9.8	-7.3	74.3
Oxnard, CA	45.6	50.5	4.9	-4.9	-98.8
Palm Bay, FL	53.7	46	-7.7	-1.3	16.3
Philadelphia, PA	76.3	68.6	-7.6	-0.7	8.7
Phoenix, AZ	50.3	46.8	-3.5	-1	27.8
Pittsburgh, PA	70.7	67.2	-3.5	-0.3	8.9
Portland, OR	65.7	50.2	-15.5	-2.2	14.4
Providence, RI	63.2	55.8	-7.4	1.7	-23.3
Provo, UT	74.5	48	-26.5	-18	67.8
Raleigh, NC	47.1	42.5	-4.6	1.4	-30.7
Richmond, VA	59.2	54.6	-4.6	1.4	-31.1
Riverside, CA	45.6	45.9	0.2	0.4	158.9
Rochester, NY	69.3	67.1	-2.2	-4.5	206.3

Continued on next page

Table 1 – *Continued from previous page*

City	DI: 1990	DI: 2010	Change	Explained	Share Explained
Sacramento, CA	55.9	55.6	-0.3	1.7	-522
Salt Lake City, UT	52.2	51.1	-1.1	-2	183.6
San Antonio, TX	53.7	52.3	-1.4	-2.2	160
San Diego, CA	57.7	53.9	-3.8	-1	26.3
San Francisco, CA	66.7	62.9	-3.8	1.8	-46.5
San Jose, CA	42.1	45	2.9	-0.8	-27.2
Scranton, PA	67.4	59.3	-8.1	2.3	-28.8
Seattle, WA	57.6	51.2	-6.5	2.2	-34.4
Springfield, MA	70	66.3	-3.8	-0.3	7.5
St. Louis, MO	78	72.5	-5.4	-3.3	61.1
Stockton, CA	59.1	51.2	-7.9	-3.5	44.3
Syracuse, NY	76.6	69.8	-6.8	-1	14.6
Tampa, FL	71.2	56.9	-14.3	-5.8	40.4
Toledo, OH	74.8	64.7	-10.1	-1.6	15.8
Tucson, AZ	43.9	41.3	-2.6	-1.1	41.5
Tulsa, OK	64.3	57.2	-7	-2.8	40.3
Virginia Beach, VA	52.1	48	-4	-1.4	34.9
Washington, DC	66.5	61.6	-4.9	-1.4	28.2
Wichita, KS	64.1	61.3	-2.8	-6.2	216.9
Worcester, MA	55.2	55.5	0.4	2.7	735.4
Youngstown, OH	75.3	69.8	-5.5	-3.5	64.2
Average	63.8	57.6	-6.1	-2.4	

Notes: This table shows the dissimilarity index in each of the largest CBSAs in the U.S. For each city, the dissimilarity is calculated from the 1990 longform census and the 2012 5-year ACS file, with the change in dissimilarity over the 20-year period shown in the third column. For each city, we have applied the spatial counterfactual technique to calculate the dissimilarity in 2010 given the demographics from the 1990 longform census ($D_{2010\rightarrow 1990} - D_{1990}$). We then calculate the change in dissimilarity over the 20-year period that can be explained by changes in observable characteristics ($D_{2010} - D_{2010\rightarrow 1990}$) and show results in the fifth column of this table. The last column displays the share of the overall change than can be explained by changes in the observables (demographic characteristics of individuals such as income, age and education).

Table 2:
Determinants of Racial Integration
Dependent Variable: Dissimilarity Index

	1990	2010	Pooled
Log Percapita Income	0.289*	0.008	0.296***
	(0.170)	(0.222)	(0.091)
Share Individuals Age 18 - 25	-0.250	-3.617**	0.861
	(1.109)	(1.380)	(0.725)
Share Individuals Age 26 - 35	-3.223**	-2.627*	-3.001***
	(1.285)	(1.338)	(0.952)
Share Individuals Age 36 - 45	-1.780	-3.568*	-1.079
	(1.656)	(1.896)	(0.907)
Share Individuals Age 45 - 55	1.309	0.426	-0.652
	(2.010)	(1.687)	(0.888)
Share Individuals Older than 55	0.482	-1.036*	-0.543
	(0.472)	(0.540)	(0.543)
Share College Degree	0.107	1.009	-1.281***
	(0.637)	(0.688)	(0.383)
Share Married Individuals	-2.136***	-1.805***	1.047**
	(0.460)	(0.549)	(0.402)
Share Born in US	0.553**	0.119	-0.237
	(0.219)	(0.188)	(0.184)
Year 2010			-0.087*
			(0.048)
Constant	-1.221	2.303	-1.644*
	(1.715)	(2.523)	(0.846)
R^2	0.423	0.309	0.968
N	99	99	198
MSA Fixed Effects	No	No	Yes

Notes: This Table shows results of a linear regression model. The sample includes the 99 largest U.S. MSAs. The dependent variable is the dissimilarity index. Standard errors robust to heteroscedasticity are shown in parenthesis. *, **, and *** denote statistical significance at the 10 , 5 and 1 percent level, respectively.

Table 3:
Average Characteristics of U.S. MSAs

Variable	1990	2010	Difference
Dissimilarity index	0.638	0.576	-0.061
Percapita Income (USD thousands)	24.80	28.94	4.145
Share Individuals Age 18 - 25	0.116	0.105	-0.011
Share Individuals Age 26 - 35	0.180	0.134	-0.045
Share Individuals Age 36 - 45	0.148	0.137	-0.012
Share Individuals Age 46 - 55	0.098	0.145	0.046
Share Individuals Older than 55	0.195	0.230	0.035
Share Married Individuals	0.441	0.401	-0.041
Share College Degree	0.145	0.208	0.063
Share Born in US	0.922	0.869	-0.053

*Notes:*Table displays (unweighted) average characteristics of the 99 largest US MSAs.

Table 4:
Counterfactual Logit Regression Results for Washington, DC

Variable	Black	White	Variable	Black	White
	(1)	**(2)**		**(1)**	**(2)**
Person Variables			**Household/Family Variables**		
Education			Log Income	0.777	0.293
1-8 Grade or Less			(Relative to CBSA Median)	0.026	0.011
			Children in HH	-0.277	-0.442
9-12 (No HS Diploma)	-0.187	-0.075		(0.018)	(0.013)
	(0.025)	(0.019)	Age 65+ in HH	-0.25	-0.189
HS Diploma	-0.845	-0.576		(0.02)	(0.015)
	(0.03)	(0.025)	Share non-citizen in HH	-0.183	-0.123
Some College	-0.842	-0.568		(0.057)	(0.044)
	(0.031)	(0.025)	Married Couple in HH	0.102	0.082
Bachelor's	-0.805	-0.761		(0.018)	(0.015)
	(0.036)	(0.026)	Most Educated HH Member		
Graduate	-0.682	-0.729	1-8 Grade or Less	2.067	1.468
	(0.041)	(0.027)		(0.064)	(0.057)
Age			9-12 (No HS Diploma)	1.877	1.447
< 18				(0.036)	(0.035)
			HS Diploma	1.111	0.935
18-25	1.004	0.703		(0.026)	(0.016)
	(0.035)	(0.027)	Some College	0.66	0.574
26-35	1.275	0.919		(0.023)	(0.012)
	(0.043)	(0.032)	Bachelor's	0.3	0.29
36-45	0.855	0.95		(0.023)	(0.01)
	(0.056)	(0.038)	Graduate		
46-55	0.356	0.416			
	(0.07)	(0.046)	Family Size		
56-65	0.308	0.377	1 Person	0.046	0.121
	(0.085)	(0.054)		(0.023)	(0.015)
66-75	0.726	0.962	2 People		
	(0.102)	(0.064)			
76+	0.458	0.686	3 People	0.159	0.153
	(0.123)	(0.075)		(0.019)	(0.012)
Citizenship			4 People	0.121	0.129
Born in US				(0.021)	(0.014)
			5 People	0.143	0.119
Born US Territory	0.165	-0.035		(0.024)	(0.016)
	(0.139)	(0.133)	6 People	0.207	0.125
Born Abroad	-0.303	-0.054		(0.029)	(0.021)
	(0.064)	(0.025)	7 People	0.552	0.149
Naturalized	-1.44	-0.374		(0.037)	(0.031)
	(0.034)	(0.02)	8 People	0.048	-0.154
Not a Citizen	-0.44	0.018		(0.054)	(0.048)
	(0.047)	(0.038)	9 People	0.74	-0.197
Age	-0.02	-0.024		(0.07)	(0.074)
	(0.002)	(0.001)	10 People	0.842	0.34
Male	-0.039	0.004		(0.096)	(0.104)
	(0.012)	(0.007)			
Marital Status					
Married	0.767	0.653			
	(0.023)	(0.017)			
Widowed	1.235	0.975			
	(0.033)	(0.023)			
Divorced	0.618	0.396			
	(0.025)	(0.017)			
Separated	1.176	0.684			
	(0.034)	(0.032)			
Constant	-1.056	-0.081			
	(0.035)	(0.021)			
Number Obs.	213,000	591,000			

Notes: This table shows the pooled 1990 census longform and 2012 5-year ACS results for blacks and whites in Washington, DC. In (1) only blacks are included in the sample and in (2) only whites are included. In each case the dependent variable is a dummy for whether the observation is from the 1990 sample. The regression results are used to reweight observations in the 2012 5-year ACS sample to match the demographics of the 1990 sample. Standard errors are shown in parenthesis. Number of observations have been rounded to the nearest thousand.