

# Racial Sorting and the Emergence of Segregation in American Cities<sup>1</sup>

Allison Shertzer (University of Pittsburgh and NBER)

Randall P. Walsh (University of Pittsburgh and NBER)

January 2015

**Abstract:** Residential segregation by race first emerged in the United States as black migrants from the South arrived in northern cities in the early twentieth century. The existing literature emphasizes discriminatory institutions as the driving force behind the particularly rapid rise in segregation over this period. We use newly assembled neighborhood-level data to instead focus on the role of residential sorting by whites. Employing both nonlinear tipping and linear white flight empirical approaches, we show that white departures in response to black arrivals were quantitatively large and accelerated between 1900 and 1930. Our results indicate that sorting by whites can explain between 45 and 60 percent of the observed increase in segregation over this period. Uncoordinated market decisions appear to have been a key mechanism behind the development of racially segregated cities in the United States.

---

<sup>1</sup> Antonio Diaz-Guy, Phil Wetzel, Jeremy Brown, Andrew O'Rourke, Aly Caito, Loleta Lee, and Zach Gozlan provided outstanding research assistance. We gratefully acknowledge the Central Research Development Fund and the Center on Race and Social Problems at the University of Pittsburgh for supporting this work. We are grateful to Elizabeth Cascio, Ethan Lewis, Leah Platt Boustan, Bob Margo, Lowell Taylor, Brian Kovak, Spencer Banzhaf, Tom Mroz, Aimee Chin, Judith Hellerstein, and seminar audiences at the NBER Summer Institute, Carnegie Mellon, and Georgia State for helpful comments. We thank John Logan for assistance with enumeration district mapping and for providing 1940 street files. We also thank David Ash and the California Center for Population Research for providing support for the microdata collection, Carlos Villareal and the Center for Population Economics at the University of Chicago for the 1930 street files, Jean Roth for her assistance with the national Ancestry.com data, and Martin Brennan and Jean-Francois Richard for their support of the project. We are grateful to Ancestry.com for providing access to the digitized census manuscripts. Corresponding author's email: shertzer@pitt.edu (A. Shertzer).

## I. Introduction

Social scientists have argued that residential segregation by race in the United States grew out of collective action by whites and government policies that deliberately disadvantaged black neighborhoods in the early twentieth century. The seminal work on the emergence of segregation (Massey and Denton, 1993) vividly describes coordinated house bombings of recently arrived black families and the formation of neighborhood “improvement” associations that existed solely to maintain the color line with restrictive covenants. Cutler, Glaeser, and Vigdor (1999) echo these findings in their work on U.S. segregation trends, arguing that the premium paid by black families for apartments by 1950 demonstrates that collective action by whites was the driving force behind racial segregation in the first half of the twentieth century. The scholarly consensus can thus be summarized in Denton and Massey’s own words: “racial segregation [in northern cities] was accomplished through violence, collective anti-black action, racially restrictive covenants, and discriminatory real estate practices” (p. 42).

The belief that discriminatory institutions were responsible for segregation is clearly reflected in the policies adopted by the federal government in the second half of the twentieth century to encourage residential integration. Most notably, the Fair Housing Act of 1968 outlawed discrimination by landlords and property sellers on the basis of race. This reduction in allowable discrimination was preceded by Supreme Court decisions that struck down racial zoning ordinances and restrictive covenants in 1916 and 1948, respectively. Yet segregation remained prominent in large cities throughout the remainder of the twentieth century.<sup>2</sup> If these institutions were the driving force behind the creation of the black ghetto, the persistence of high

---

<sup>2</sup> Several papers have argued that black-white segregation declined modestly over the 1980s and 1990s (for instance, see Farley and Frey, 1994). However, other scholarship argued that much of this decline stemmed from an increasing proportion of Hispanics in the neighborhood of the typical black resident rather than from the integration of blacks and non-Hispanic whites (Logan, Stults, and Farley, 2004).

levels of racial residential segregation after the most odious forms of discrimination became illegal remains a puzzle.

In this paper we return to the question of why racial segregation emerged in American cities, revisiting in particular the scholarly consensus that institutional factors were primarily responsible for the emergence of the black ghetto. We focus our empirical analysis on the residential response of white individuals to the initial influx of rural blacks into the industrial cities of the North on the eve of the First World War, asking to what extent white departures in response to black arrivals can account for the rise of segregation in American cities. Our approach can thus be thought of as separately identifying the role of residential sorting from the institutions that have been the focus of previous work.

The early twentieth century decades are of particular importance for understanding the fundamental causes of residential segregation. Figure 1 shows the twentieth century trend in segregation by race for the ten largest northern cities from the dataset used in Cutler, Glaeser, and Vidgor (1999).<sup>3</sup> This figure shows that 97 percent of the twentieth century increase in dissimilarity and 63 percent of the increase in isolation took place by 1930.<sup>4</sup> However, the capacity of scholars to rigorously investigate the mechanisms responsible the emergence of segregation has been limited by the lack of spatial data covering prewar urban neighborhoods in the United States. The first contribution of this paper is a fine-grained, spatially-identified demographic dataset covering ten major cities in 1900, 1910, 1920, and 1930. We digitized maps of census enumeration districts, small administrative units used internally by the census, for each city and census year to develop a dataset with consistent neighborhood borders over

---

<sup>3</sup> The largest according to population in 1880 were: Baltimore, Boston, Chicago, Cincinnati, Cleveland, Detroit, New York (Brooklyn and Manhattan were separate cities at this time), Philadelphia, Pittsburgh, and St. Louis.

<sup>4</sup> Isolation peaked in 1970, with isolation rising from .23 to .66 between 1900 and 1970. However, 63 percent of the overall increase took place by 1930. Dissimilarity likewise peaked in 1950, with 97 percent of the 1900 to 1950 increase (from .64 in 1900 to .81 in 1950) occurring between 1900 and 1930.

time. The expiration of census confidentiality rules allows us to observe age, race, state of birth, and parents' state of birth for each of the approximately 60 million individuals observed living in our sample cities in these census years.

We use this novel dataset to provide the first estimates of residential sorting by race in prewar urban America. The primary difficulty in identifying white "flight" from black newcomers is that black migrants did not exogenously arrive in neighborhoods, and we are particularly concerned that blacks may have located in places already being abandoned by whites. Our empirical strategy to identify the magnitude of white flight from black arrivals takes two forms. We first search for nonlinearities in white population share as a function of black population share that could indicate neighborhood "tipping" in response to black migration from the South. In our second approach we develop an instrumental variable for black population size in neighborhoods in the spirit of the immigration shock literature.

Both strategies provide evidence of white flight from blacks in the early twentieth century; moreover, the flight effect appears to accelerate over the three decades we study. We find robust evidence of nonlinear declines in white population consistent with neighborhood tipping. In particular, the loss of white population is discontinuous between 10 and 15 percent black neighborhood population share in every decade and city, consistent with a Schelling model of tipping (Schelling, 1971). Neighborhoods with black share above the tipping point lost about 30 percent more whites (relative to the city average) than neighborhoods with few blacks in the 1910s. By the 1920s these neighborhoods with black share above the tipping point lost about 45 percent more white population.

In our second empirical approach we relate changes in black populations to changes in white populations, again considering each decade separately. The OLS results suggest one black

arrival in the preceding decade was associated with -.8 and -1.5 white departures in the 1910s and 1920s, respectively, a finding that would indicate little racial animus in location decisions. However, our instrumental variables strategy, which assigns estimated black outflows from southern states to northern cities according to black settlement patterns prior to the Great Migration, indicates that OLS estimates were likely biased against a finding of flight due to black settlement in generally growing neighborhoods. Our IV results indicate that one black arrival was associated with 2.2 white departures in the 1900s and 4.2 white departures by the 1920s. We show that this effect is driven by accelerating flight behavior by both white natives and white immigrants.

In the final portion of our empirical analysis, we use our causal estimates of white flight to estimate how much segregation would have arisen over the 1900 to 1930 period solely as a consequence of sorting in response to black in-migration. Applying our IV estimates of flight to reallocate white population in each decade, we show that between 45 and 60 percent of the observed increase in dissimilarity over these decades can be explained by white flight from neighborhoods with growing black populations. Using a range of assumptions on black settlement behavior, we find that white flight is quantitatively important for segregation levels whether blacks are constrained to live in neighborhoods with existing African American populations or allowed to move into predominantly white neighborhoods.

Our results are not meant to imply that collective action and government policy had no impact on patterns of racial segregation. Discriminatory institutions such as restrictive covenants were no doubt successful in directing black settlement by raising the relative cost of settling in some neighborhoods. However, the results of this paper suggest that segregation would have arisen even without these institutions as a direct consequence of the widespread, decentralized

relocation decisions of white individuals. Furthermore, our findings demonstrate that sorting by whites out of neighborhoods with growing black populations was a quantitatively important phenomenon decades before the postwar opening of the suburbs. Several studies have shown that the willingness of white individuals to depart neighborhoods with rising black populations over the 1940 to 1980 period was an important mechanism through which racial segregation was perpetuated (Card, Mas, and Rothstein, 2008; Boustan, 2010). Our paper adds to this literature by showing that this same mechanism was important for the initial emergence of segregation in the United States. Finally, our findings suggest that even the complete elimination of racial discrimination in housing markets may fail to bring about significant racial integration so long as individual preferences remain unchanged.

The paper proceeds as follows: Section II reviews the evolution of segregation over the twentieth century and gives historical context for the black migration from the South. Section III discusses the construction of the dataset used in this paper. Section IV details both of our empirical approaches for measuring white flight and Section V presents our results and Section VI relates our finding to the observed increase in segregation. Section VII concludes.

## **II. Background on Segregation and Urbanization in the United States**

Residential segregation by race has remained one of the most prominent and enduring features of American cities. Social scientists have linked segregation to a host of adverse minority outcomes, arguing that blacks living in more segregated cities have worse health, human capital accumulation, and labor market outcomes (Wilson, 1996; Cutler and Glaeser, 1997; Card and Rothstein, 2007; Sharkey, 2013). Ananat (2011) provides causal evidence of the impact of segregation on black poverty using historical railroads as an instrument for

contemporary dissimilarity levels. Recent scholarship has also found that segregation is negatively correlated with the intergenerational mobility of the entire American urban population, including whites (Chetty, Hendren, Line, and Saez, 2014). While a broad literature studies the impact of segregation, less empirical work exists to explain the emergence of the phenomenon in the American cities.

In this paper we study the marked rise in racial segregation that occurred between 1900 and 1930 in American cities. We set the stage for our analysis by re-establishing the extant understanding of this rise in segregation levels using our newly constructed spatial data set. We measure segregation using the two most common indices of segregation: isolation and dissimilarity. In constructing isolation indices we follow Cutler, Glaeser, and Vidgor (1997) and compute a modified index which controls for the fact that under the standard approach there is a potential for the index to be sensitive to changes in the overall group share. For each year compute:

$$Isolation\ Index = \frac{\sum_{i=1}^N \left( \frac{blacks_i}{blacks_{total}} \cdot \frac{blacks_i}{population_i} \right) - \left( \frac{blacks_{total}}{population_{total}} \right)}{1 - \left( \frac{blacks_{total}}{population_{total}} \right)} \quad (1)$$

where *population* refers to the population of the enumeration district (*i* subscript) or city (*total* subscript) and *blacks* refers to the racial group’s enumeration district population (*i* subscript) or city population (*total* subscript). The “standard” isolation index simply computes the average percentage of a group member’s neighborhood composed of members of her own group. The modified measure controls for the fact that under random sorting groups with larger overall population shares will, by construction, experience neighborhoods with larger own group shares. The modification addresses this issue by expressing the average exposure share relative to the group’s overall share of the population. This relative measure is then rescaled (hence the

numerator in Equation 1) so that it spans the interval from zero to one. Such an adjustment makes measured isolation less dependent on a group's share of the overall population.

Our second measure is the dissimilarity index (Duncan and Duncan, 1955). For blacks and whites it is defined as:

$$Dissimilarity\ Index = \frac{1}{2} \sum_{i=1}^N \left| \frac{black_i}{black_{total}} - \frac{whites_i}{whites_{total}} \right| \quad (2)$$

where  $black_i$  is the number of blacks in enumeration district  $i$ ,  $black_{total}$  is the number of blacks in the city, and the white variables defined analogously. This index ranges from zero to one with one representing the highest degree of dissimilarity between where whites and blacks in a city reside. Intuitively, the index reveals what share of the black (or white) population would need to relocate in order for both races to be evenly distributed across a city.

The Cutler et al data in Figure 1 was constructed using adjusted ward-level measures for isolation and dissimilarity prior to 1940, the year when census tract data became widely available. We compute the same measures using our enumeration district data discussed below in Section III for comparison and report these results in Figure 2. Segregation indices computed at the enumeration district are much higher than those computed at the ward level, which is to be expected given the smaller scale of these units (the average enumeration district had 1,400 individuals while wards could have as many as 100,000 residents in large cities). However, the trends in ward and enumeration district segregation are nearly parallel, showing a steep increase between 1900 and 1930. Furthermore, the Cutler et al adjusted ward measures are quantitatively similar to the enumeration district measures of both isolation and dissimilarity.

These stylized facts are not new. Scholars have long argued that the groundwork of the black ghetto was laid during the first decades of the twentieth century as black populations in northern cities grew, leading to the sharp increase in racial residential segregation. African

Americans began to migrate to northern cities on the eve of World War I, an event that brought European immigration to a temporary halt while simultaneously increasing demand for industrial production. These wartime developments in the northern labor market coincided with the arrival of the Mexican boll weevil in Mississippi and Alabama (1913 and 1916, respectively), which devastated cotton crops and led to a decline in demand for black tenant farmers (Grossman, 1991). The combination of push and pull factors led to unprecedented out-migration from the South: 525,000 blacks came to the North in the 1910s and 877,000 came in the 1920s (Farley and Allen, 1987).

Of importance to our analysis of white flight is the fact that cities were growing at an unprecedented rate during these initial decades of black migration, particularly from European immigration. In contrast to the postwar era, which saw significant suburbanization and declines in urban population, segregation in the early twentieth century emerged against a backdrop of rapid urbanization. The share of the population residing in central cities grew from 14 to 33 percent between 1880 and 1930, leveling off subsequently.<sup>5</sup> Although some “streetcar suburbs” existed by 1910, white flight in this period can primarily be thought of as departures for neighborhoods outside the urban core but still within city boundaries. The destination neighborhoods for whites fleeing black arrivals were thus similar in their public goods and tax base, unlike the suburban destinations of postwar white flight (Boustan, 2013). In this sense our estimates of white flight in this period may be a better gauge of racial distaste than those from empirical work from later in the twentieth century when whites were leaving cities for different tax and public good combinations that were available in the suburbs.

---

<sup>5</sup> This computation uses the center city status variable from IPUMs samples for 1880 to 1930.

### **III. Enumeration District Data for 1900 to 1930**

The analysis in this paper is based on a novel, fine-scale spatial dataset spanning the years 1900 through 1930 (Shertzer, Walsh, and Logan, 2014). There are two major components of our dataset: census-derived microdata retrieved from Ancestry.com and digitized enumeration district maps. The census-derived microdata cover 100 percent of the population of ten large cities over four census years. For the twentieth century decades (1900, 1910, 1920, and 1930) we collected the universe of census records for Baltimore, Boston, Cincinnati, Chicago, Cleveland, Detroit, New York City (Manhattan and Brooklyn boroughs), Philadelphia, Pittsburgh, and St. Louis from the genealogy website Ancestry.com. To maximize the usefulness of the dataset for our purpose, we selected cities that received substantial inflows of black in-migration. This sample contains the ten largest northern cities in the United States in 1880 and nine out of the ten largest cities in the United States in 1930. The combined population of these cities was 9.3 million in 1900 and over 18 million in 1930, which is about half of the total population in the largest 100 cities in both years.

The microdata we compiled for this paper represent a significant improvement over existing sources of data on early twentieth century urban populations. Ward-level tabulations published by the census are the smallest unit at which 100 percent counts were previously available for the years we study.<sup>6</sup> Wards were large political units used to elect city councilmen, still in use in some cities today, while enumeration districts were small administrative units used internally by the census prior to the shift to mail surveys in 1960. Every Ancestry.com record contains the enumeration district of the individual being surveyed, allowing us to make counts of any population at this level of geography. The other digitized variables from the 1900 to 1930

---

<sup>6</sup> Some census tract tabulations are available beginning in 1910, but we prefer our approach because the enumeration district data is available for all cities and years in our sample and counts of any subpopulation can be created from microdata.

censuses include place of birth, father's place of birth, mother's place of birth, year of birth, marital status, gender, race, year of immigration (for foreign-born individuals), and relation to head of house in addition to place of residence (city, ward, and enumeration district) at the time of the respective census.

To place these individuals in urban space, we created digitized versions of the census enumeration district maps based on information available from the National Archives. We used written descriptions of the enumeration districts that are available on microfilm from the National Archives. The written descriptions from these microfilms has been digitized and made available on the web due to the work of Stephen P. Morse.<sup>7</sup> We also employed a near complete set of physical maps for our census-city pairs located in the maps section of the National Archives. We took digital photographs of these maps as a second source for our digitization effort. Working primarily with geocoded (GIS) historic base street maps that were developed by the Center for Population Economics (CPE) at the University of Chicago, research assistants generated GIS representations of the enumeration district maps that are consistent with the historic street grids.<sup>8</sup> Figure 3 provides an illustration of this process. Here the shaded regions in panel D represent the digitized enumeration districts.

In order to conduct analysis of change over time within neighborhoods, we require neighborhood definitions that are constant across census years. Forming such neighborhoods is challenging for this data because enumeration districts were redrawn for each decadal census and, unlike the case of modern-day census tracts, most changes were more complex than simple combinations or bifurcations. In this paper we employ a hexagon-based imputation strategy. The strategy is illustrated in Figure 4. It involves covering the enumeration district maps (Panel

---

<sup>7</sup> website: <http://stevemorse.org/ed/ed.php>

<sup>8</sup> We used 1940 street maps produced by John Logan for Detroit, Cleveland, and St. Louis.

A) with an evenly spaced temporally invariant grid of 800 meter hexagons (Panel B) and then computing the intersection of these two sets of polygons (Panel C). The count data from the underlying enumeration districts is attached to individual hexagons based on the percentage of the enumeration district's area that lies within the individual hexagon. Panel D presents the allocation weights for a sample hexagon. In the example, 100 percent of four enumeration districts lies completely within the hexagon (136, 139, 140, and 144) while 11 enumeration districts are partially covered by the hexagon. For these partial enumeration districts, only fractions of their counts are attributed to the hexagon, ranging from a minimum of 0.2 percent (155) to 93.6 percent (142).

We form a panel of hexagons with at least 95 percent coverage by enumeration districts from the respective census in each year from 1900 to 1930. We also trim the sample at the 1<sup>st</sup> and 99<sup>th</sup> percentile of both white and black population change for each decade. In Table 1 we provide summary statistics for our hexagon neighborhoods. The neighborhoods have an average population of 3,349 in 1910 and 4,493 in 1930, with the increase in density reflecting the rise in urban population density that occurred over this period. Thus, they are roughly similar in scale to modern-day census tracts. The average white population growth is positive in all years but declining from 683 over the 1900s to 264 over the 1920s, with much of this slowdown due to declining immigration from Europe after World War I and the Immigration Restriction Act of 1921. The average black percent increases from 2.4 to 5.3 percent over the 1910 to 1930 period.

#### **IV. Empirical Strategy**

The objective of the empirical work is to ascertain whether black arrivals had a causal impact on white population dynamics over the 1900 to 1930 period. The primary difficulty in

identifying such an effect is that minorities do not exogenously arrive in neighborhoods, and we are particularly concerned that blacks and immigrants may locate in locations already being abandoned by white natives for other reasons. We use two different strategies to address this concern. First, we search for nonlinearities in white population share as a function of black population share that could indicate “tipping” behavior in the data. Second, we develop an instrumental variable strategy for black population size in our neighborhoods.

### *A. Nonlinear Evidence of White Flight*

The intuition behind our first empirical approach is as follows: we wish to know if the presence of racial and ethnic minorities above some “tipping point” induces white natives to “flee” their neighborhood at an accelerated pace, giving rise to a nonlinear relationship between baseline black population share and the change in white population share. Our methodology is similar in nature to that of Card, Mas and Rothstein (2008) who study neighborhood tipping later in the twentieth century based on the classic Schelling (1971) model. However, we depart from their empirical framework in several ways. First, we use as our neighborhood definition the 800 meter hexagons instead of census tracts, which were not used by the census until 1910. Second, the criteria for identifying specific “tipping” points were less clear in our context.<sup>9</sup> Thus, instead of looking explicitly for tipping points based on specifically defined criteria, we present results of entire nonparametric regressions using local polynomial smoothing.

Specifically, we predict the change in the percentage of whites in a given neighborhood  $i$  located in city  $j$  for the panel of hexagons based on the following non-parametric regression:

$$\Delta WP_{ij}^{t1-t0} = f(BP_{ij}^{t0}) + \epsilon_{ij}. \quad (1)$$

---

<sup>9</sup> We experimented with the two approaches employed by Card, Mas, and Rothstein in their work as well as other candidate methodologies. While the shape of the non-parametric relationships is quite robust, the specific tipping points identified in our data were sensitive to small changes in procedure.

where  $\Delta WP_{ij}^{t1-t0}$  is the de-meanded (by city) percent change in white population over a census decade and  $BP_{ij}^{t0}$  is the percent of the neighborhood composed of African Americans at the start of the decade. We also present results where the dependent variable is the de-meanded (by city) percent change in white native population (defined as third generation or greater) over a census decade. In the results section below, we discuss both the general trend in the share of white and white native populations relative to black population share as well as the presence of non-linear responses that could be evidence of “tipping” behavior. Results are presented by decade to illustrate changes over time in neighborhood population dynamics.

### ***B. Instrumental Variables Estimates of White Flight***

The tipping analysis above bases the identification of “white flight” on the assumption that any observed non-linear relationships arise as a result of white responses to black population shares.<sup>10</sup> While we find the “tipping” results that we present below compelling, there is a concern that our nonlinear approach does not rely on exogenous variation in the black population shares. Our second estimation approach addresses the causality of the white flight effect by instrumenting for black populations across our sample city neighborhoods using heterogeneous outmigration shocks that occurred in southern sending states. Our analysis is in the spirit of the immigration shock literature (Card, 1991; Boustan, Fishback, and Kantor, 2010; Saiz and Wachter, 2011; Cascio and Lewis, 2012).

We begin this analysis by considering a simple OLS model relating the change in black populations to the change in white populations. Because we are interested in understanding the response of native whites to the influx of new black residents, we restrict our attention to neighborhoods that saw increases in the number of black residents. Such a model is:

---

<sup>10</sup> This assumption would hold if all potentially confounding relationships between black share and percent change in white population change were linear or if any confounding non-linear relationships were sufficiently small.

$$\Delta W_{ij}^{t1-t0} = \beta \Delta B_{ij}^{t1-t0} + \eta_j + \epsilon_{ij}. \quad (2)$$

where  $\Delta W_{ij}^{t1-t0}$  ( $\Delta B_{ij}^{t1-t0}$ ) is the change in the number of whites (blacks) in a neighborhood over a decade and  $\eta_j$  is a city fixed effect. The coefficient of interest from this first differences strategy,  $\beta$ , relates the change in share black to the change in share white in a particular neighborhood over the same decade with beyond the city-level average captured by the fixed effect. As with the nonlinear approach, results are presented by decade to illustrate changes over time in neighborhood population dynamics.

In a world where neighborhood choice was unrelated to race and considering housing supply and demand only, ordinary least squares (OLS) estimates of  $\beta$  in equation (2) would be expected to be bounded below by -1. This case corresponds to perfectly inelastic housing supply, where each black arrival would displace exactly one white resident in the absence of racial animus. On the other hand, if housing were elastically supplied, we would expect  $\beta > 0$ . The magnitude of population growth in our fixed-border neighborhoods suggests that housing supply was at least somewhat elastic during this period, so we do not expect negative coefficients to arise purely as a result of supply and demand. However, we must also consider the impact of sorting in response to unobserved neighborhood shocks. Of particular concern are positive neighborhood shocks that encouraged black in-migration and also lead to increased white in-migration. Shared sorting on neighborhood characteristics would bias OLS estimates of  $\beta$  upwards (away from flight). OLS estimates of  $\beta$  will be biased in a negative direction (towards flight) if black arrivals were settling in neighborhoods already being abandoned by whites.

To overcome the bias concerns we leverage arguably exogenous variation in contemporary state-level outmigration rates in combination with pre-1900 patterns of black migration into our sample of northern cities. To do so, we construct an instrument using the

universe of historical census records recently made available by Ancestry.com. Our instrument for  $\Delta B_{ij}^{t1-t0}$  is constructed from two components: estimated black outflows from each state in each decade (1900 to 1930) and settlement patterns established by African American who came to the North before the Great Migration and were thus living in our sample cities 1900.

To estimate the total number of black out-migrants from each state over each census decade, we exploit the 100 percent census microdata samples for 1900 to 1930 and count the number of black individuals who appear outside of their state of birth in each gender, state of birth, and birth cohort cell. For simplicity, we consider only individuals under the age of 60 and aggregate birth cohorts into ten year intervals. To illustrate, for the census year 1900, we count the number of individuals of each gender observed outside each birth state in the 1840-1849, 1850-1859, 1860-1869, 1870-1879, 1880-1889, and 1890-1899 birth cohorts. The total number of out-migrants in each cell is obtained by summing over the number of out-migrants present in each state of residence. To obtain the estimated outflow at the national level by cell over a census decade, we take the difference in the number of out-migrants by the five birth cohort intervals ( $c$ ), two genders ( $g$ ), and 51 states of birth ( $s$ ) appearing in each state:

$$black\_outflow_{cgs}^{t1-t0} = \sum_{k=1}^{51} black\_outmigrants_{icgs}^{t1} - \sum_{k=1}^{51} black\_outmigrants_{icgs}^{t0}$$

where  $k$  indexes the state of residence where the individual was observed (state  $i=51$  is the District of Columbia). Here is the  $j$  subscript for city is suppressed for simplicity.

For the 1900 base year component of the instrument, we count the number of black out-migrants in each birth cohort-gender-state of birth cell present in each neighborhood of our sample in 1900 to obtain  $black\_basepop_{icgs}^{1900}$ . To construct the predicted change in the number

of blacks in a neighborhood  $i$  in decade  $tI$ , we assign the estimated outflows according to the base year population for each cell and sum over each cell:

$$pred\_black_i^{t1-t0} = \sum_{c=1}^5 \sum_{g=1}^2 \sum_{s=1}^{51} \left[ \left( \frac{black\_basepop_{icgs}^{1900}}{black\_outmigrants_{cgs}^{1900}} \right) black\_outflow_{cgs}^{t1-t0} \right]$$

where  $black\_outmigrants_{cgs}^{1900}$  is the national sum of all black out-migrant individuals in the cell in 1900.<sup>11</sup> Our instrument for  $\Delta B_{ij}^{t1-t0}$  is thus  $pred\_black_i^{t1-t0}$ .

Our approach departs from much of the literature on the impact of immigration on local labor markets, where papers allocate the *actual* inflow of immigrants to cities. Because there is no systematic data on internal migration in the United States prior to 1940, we need to allocate *estimated* outflows to cities. However, we are able to observe a rich set of characteristics of black migrants living outside their birth state, in particular year of birth, enabling a close approximation to the true size of outflows in each decade. These two approaches are thus in principal very similar. Following to other papers in this literature, our instrument relies on the fact that blacks departing the South tended to follow a settlement distribution pattern that was similar to that of blacks who had left their state in earlier decades, due to the stability of railway routes and enduring social networks.

For this instrument to have power, two types of variation are needed. First, within a given city the distribution of blacks across neighborhoods must differ by state of origin. To give a sense of the existence of this type of variation, we provide city-level scatter plots in Figure 5 showing by neighborhood the share of black men aged 20 to 29 in 1900 who were born in two pairs of source states. Panel A shows that neighborhoods within Boston, Brooklyn, Chicago,

---

<sup>11</sup> We shift the cohorts for each decade so that individuals of the same age are assigned in the same proportion across time. For instance, outflows of men from Alabama who were born in the 1900-1909 decade and were thus between the ages of 21 and 30 in 1930 were assigned to neighborhoods according to the distribution of men born in Alabama aged 21 to 30 present in 1900.

Cleveland, and Philadelphia all exhibit rich variation in the share of black men from this cohort originating in North Carolina as opposed to Virginia (Panel A). Panel B shows the significant variation across neighborhoods in Chicago, Cincinnati, and St. Louis in the share of the black population originating in Kentucky versus Tennessee.

In addition to differential within city sorting, for our instrument to have predictive power, it should also be the case that variation exists across sending states over time. Figure 6 shows the estimated outflows from the thirteen most important sending states for black men aged 20 to 29 across each of the decades we study in this paper.<sup>12</sup> Texas and Virginia provided relatively more out-migrants during the 1900 to 1910 decade while South Carolina and Georgia were the most significant sending states by the 1920 to 1930 decade. Taken together Figures 5 and 6 suggest the potential predictive power of our instrument. The instrument is further strengthened by the fact that we compute its components separately by birth cohort and gender. Formal F-tests presented below confirm this suggestive evidence regarding the instrument's power.

## **V. Analysis of White Flight in the Early Twentieth Century**

### ***A. Nonlinear Evidence of Tipping***

In this section we present the results from both the nonlinear and causal models of white flight proposed in the previous section. The results from the nonparametric regressions are given in Figure 7 using local polynomial smoothing. We begin with the relationship between black neighborhood share and (de-meaned) change in white population share over the next decade in Panel A. For the 1900 to 1910 decade, a decline in white population relative to the city average is apparent for neighborhoods with more than 10 percent black share, but the effect is somewhat noisy due to the small number of blacks in northern cities in 1900. By the 1910 to 1920 decade,

---

<sup>12</sup> These thirteen states represent between 87 and 92 percent of total black outflows in each year we study.

the relationship is much more precisely estimated, with a discontinuity in white share suggestive of tipping apparent between 10 and 15 percent black share. Neighborhoods with more than 20 percent black population lost about 30 percent more white population than neighborhoods with less than 10 percent black population. By the 1920 to 1930 decade, this affect had accelerated, and neighborhoods above the 20 percent black population mark lost about 45 percent more white population than neighborhoods with small black population shares, with the discontinuity in white population change again readily apparent.

If neighborhoods in these northern cities did exhibit white flight when faced with growing numbers of black residents, a natural question is which whites were doing the fleeing. In Panel B we show the same local regression results for third-generation-or-more natives and first-generation immigrants separately (we show second-generation immigrants as well in the linear regression results in the next section). The results show that the overall pattern of white population loss is driven by natives and immigrants alike. However, the acceleration of the effect over the three decades is more dramatic for the foreign born. While there is virtually no evidence of a nonlinearity over the 1900 to 1910 decade, by the 1920s, neighborhoods above the tipping point lost about 55 percent more white immigrants than neighborhoods with few black residents.

### ***B. OLS and IV Estimates of White Flight***

In our discussion of the nonlinear estimation results in Figure 7, we focused largely on evidence of tipping behavior. However, while the relationships between black share and white population share dynamics were generally nonlinear, they were also monotonic. We thus turn to the second empirical strategy, beginning with OLS estimation of equation (2). These results are

presented in Table 2. Here we follow the literature and consider changes in numbers (controlling for the city average change in white population with city fixed effects) rather than percentages.<sup>13</sup>

The first decade we consider (1900-1910), one black arrival is associated with .47 white *arrivals*, suggesting positive local shocks that attracted both blacks and whites to the same neighborhoods, masking any true flight affect. By the second decade (1910-1920), the sign on this effect has flipped, and one black arrival was associated with the loss of .86 whites. This effect doubles by the final decade (1920-1930), with one black arrival associated with the loss of -1.5 whites. The OLS estimates of white flight thus increase as the black migration out of the South accelerated, finally falling below -1 in the 1920s. We next employ our instrument for black population change to investigate the bias of our OLS coefficients. The IV estimate is -.7 and insignificant in the 1900s but grows to -2.2 in the 1910s before reaching -4.2 in the 1920s. Both of the latter two effects are significant with first stage F-test values over 200, indicating our instrument is strong.

Taken together, the OLS and IV estimates suggest that whites were leaving neighborhoods in response to growing black arrivals, but that this effect is masked by generally positive local shocks that attracted both blacks and whites.<sup>14</sup> As with the nonlinear approach, we decompose the flight effect by white population type, considering first-generation immigrants, second-generation immigrants, and third or more-generation natives using our IV specification. The relative size of each of these groups is roughly similar, facilitating comparison across the specifications presented in Table 3. We see that an increase of one black resident caused a loss of 1.7 natives with little effect on second-generation immigrants over the 1900 to 1910 decade.

---

<sup>13</sup> As discussed in Section III, we drop the 1<sup>st</sup> and 99<sup>th</sup> percentiles of both black and white population changes to ensure that our results are not being driven by outliers in the data.

<sup>14</sup> This finding stands in contrast to what Boustan (2010) finds for the 1940 to 1970 period when measuring flight from central cities to the suburbs. Her OLS coefficients are negative in all years and generally similar in magnitude to the IV results from an estimation strategy similar to ours.

Interestingly, immigrants and blacks seemed to be co-locating in this first decade, with no evidence of flight from blacks (effect is a positive and significant 1.1). Immigrants' residential responses converge on that of natives over time, and by the 1920 to 1930 decade, an increase of one black resident caused a loss of 1.5 natives, 1.2 second-generation immigrants, and 1.4 first-generation immigrants (roughly summing to our overall white flight effect).

In Table 4 we present a series of robustness checks to our main linear regression results. One concern with our instrumental variable is that percent black in 1900 is highly correlated with the urban core and neighborhoods that were set to both attract black population and lose white population. We control for percent black in 1900 in the first set of checks and show our results are slightly larger in magnitude. We also control for the number of blacks in 1900 in the next robustness check, but we cannot do this exercise for the 1900 to 1910 decade because number of blacks in 1900 is used to compute change in black population. The results are reduced in magnitude somewhat but are still sizeable and significant.

We also show our results with the standard inclusion of pre-trends in white population in addition to percent black in 1900. Although the pre-trend may absorb some of the true effect of white flight from black arrivals carrying over from the previous decade, our results for both the 1910 to 1920 and 1920 to 1930 decade are still significant and similar to the baseline. We also present results from an alternate definition of our instrument where only southern states are used to compute black outflows (instead of all fifty states as in our original instrument). Our results are again similar to the baseline, indicating that migration shocks out of the South are driving our instrument.

Finally, one concern with our approach is that black households may be smaller on average than white households, leading to the appearance of "flight" when a white family is

replaced by a black family. Using the relationship to the head of household variable, we create an alternate dataset using only heads of household in the census and rerun our estimation.<sup>15</sup> The results from the 1920s indicate that the arrival of one black household led to the departure of four white households, strongly suggesting that differences in household composition are not driving our findings.

## **VI. How Important was Sorting for the Rise of Segregation in U.S. Cities?**

In this section we use our causal estimates of white flight to ask how much of the observed increase in segregation was due solely to residential sorting. We begin with a simple exercise in Table 5 to demonstrate the population dynamics implied by our empirical results. We focus on the 1920 to 1930 decade, which saw the largest white flight effects. In the first column we reproduce the white flight effect from the instrumented equation (2) with the addition of the 1900 black population control for the full sample.<sup>16</sup> The mean white population in 1920 across the sample is 3883 and the mean black population across the sample is 193. The predicted value from the regression is 281, meaning that the average neighborhood gained 281 whites after flight from blacks is taken into account. The actual average gain in white population is 265. Our model thus closely predicts overall white population dynamics across the sample.

Of course, the nonlinear regression results from Section V.A indicate that the implied impact of white flight should not be constant across the black population share range, and in particular we should expect white population losses to be concentrated in neighborhoods with

---

<sup>15</sup> The head of household dataset contains some significant outliers due to a fraction of a black head of household being assigned to a neighborhood, leading to very large ratios of blacks to black heads of household in areas with very few blacks. Outliers also arise for white household heads due to large institution containing many whites but no household heads. We trim at the 99<sup>th</sup> percentile of the ratio of white to white household heads as well as black to black household heads to remove these outliers. We reran our baseline (full population) specification on this subsample and the results were very similar.

<sup>16</sup> Controlling for 1900 black population improves the precision of the white flight estimates on the subsamples in Table 5.

more than 10 or 15 percent black share. To explore the connection between the two approaches further, we partition the sample by 1920 black share and rerun our specification for neighborhoods with 1 to 5 percent black share, 5 to 10 percent black share, 10 to 20 percent black share, and over 20 percent black share. Although the estimated white flight effect declines as 1920 black share increases, the implied change in white population is only positive (438) for the 0 to 5 percent black neighborhoods. Neighborhoods in the 5 to 10 percent black range are predicted to lose 16 percent of their white population while actual white population loss was 13 percent. For the two largest share black subsamples, our model overpredicts white population loss. In particular, the -2.7 white flight coefficient for the over 20 percent black share subsample implies a loss of 120 percent of a neighborhood's white population while the actual percent change in white population was -41 percent. These results suggest that the true effect of white flight was muted by the positive shocks that attracted both blacks and whites to the same neighborhoods during this period of urbanization.

To quantify the contribution of white sorting on overall segregation trends, we perform a decade-by-decade calculation for the full sample in a spirit similar to Table 5 for decade. We focus our attention on explaining the growth in dissimilarity, which is less correlated with overall black population share than isolation (see Cutler, Glaeser, and Vidgor, 2008).<sup>17</sup> In addition, nearly all of the increase in dissimilarity in large cities occurred by 1930 (see Figure 1), providing further evidence that dissimilarity is not strongly affected by group size. To perform the calculation, we develop hypothetical distributions of both blacks and whites over each decade. We are particularly interested in how sensitive the white flight impact is to different assumptions on the sorting behavior of blacks. For instance, if blacks had been allowed to freely

---

<sup>17</sup> Segregated groups that compose a very small share of the population can exhibit high dissimilarity and low isolation at the same time. Since black share is very small in 1900 in most of our cities, we proceed with dissimilarity only.

settle in neighborhoods across the sample cities, would white flight have been more or less important?

We explore this idea in our calculation, which proceeds as follows:

*Step 1:* Form a consistent panel of neighborhoods that appear in each decade from the primary dataset. This balanced panel has 2,296 neighborhoods.<sup>18</sup>

*Step 2:* Allocate black population growth over the 1900 to 1910 decade to neighborhoods according to one of the three propagation rules we discuss below (Black: By Race, Black: By Population, Black: By Italian).

*Step 3:* Apply the instrumented equation (2) to generate the predicted white population change over the 1900 to 1910 decade in each neighborhood.

*Step 4:* Compute the difference between the total actual white population change in each city and the total predicted white population change in each city and assign the difference to neighborhoods proportionally according to their share of the city's population in 1900.<sup>19</sup> This step ensures that the hypothetical white population in 1910 remains equal to the actual white population in 1910.

*Step 5:* Compute the “with flight” index of dissimilarity for each city using the hypothetical black population distribution under the chosen rule from Step 2 and the adjusted post-flight white population distribution from Steps 3 and 4.

*Step 6:* Compute a no-flight hypothetical white population for 1910 by assigning white population growth over the 1900 to 1910 decade according to the overall population distribution. For instance, a neighborhood that had 5 percent of the city's population in 1900 would be assigned 5 percent of white population growth over the 1900 to 1910 decade.

*Step 7:* Compute the “no flight” index of dissimilarity for each city using the hypothetical black population distribution under the chosen rule from Step 2 and the no-flight white population distribution from Step 6.

*Step 8:* Repeat steps 1-7 for the 1910-1920 period and the 1920-1930 period using 1910 and 1920 as base years, respectively.

---

<sup>18</sup> Dropping the approximately 100 neighborhoods that do not appear in every decade from the full dataset does not significantly affect our estimates.

<sup>19</sup> We also tried assigning the residual white population change evenly across neighborhoods. Both approaches yield similar dissimilarity measures. We note that this step redistributes white population into neighborhoods with black residents, blunting the impact of white flight in the calculation. We also tried ignoring the white population residual (omitting Step 4) and the pattern of our findings was quantitatively similar.

We present an illustration of our findings in Figure 8. The first set of bars shows the actual increase in dissimilarity in the balanced panel of neighborhoods in each decade. The increase in dissimilarity grows each decade from .03 in the 1900s to about .08 in the 1920s for an overall increase of about .165 between 1900 and 1930. For the hypothetical dissimilarity indices, we restrict our attention to the difference in dissimilarity arising solely from the application of a flight effect (Steps 3 and 4) in our hypothetical scenarios. In other words, we consider the difference in the “with flight” and “no flight” dissimilarity indices computed above.

We begin with the black population assignment rule (Black: By Race) that best corresponds to a world where blacks faced significant constraints on where they could live in northern cities. To simulate extreme institutional factors preventing blacks from moving into new neighborhoods, we assign black population growth over the 1900 to 1910 decade according to the distribution of blacks across each city in 1900. Black in-migrants can only settle in existing black neighborhoods under this rule, and the 1910 hypothetical black population is obtained by summing the allocated portion of black inflows over the 1900 to 1910 decade and the 1900 black population for each neighborhood. We then repeat this process using 1910 and 1920 as base years for the latter two decades.

The second set of bars shows the difference in dissimilarity implied by this black allocation rule. In the first decade, dissimilarity actually falls slightly when flight is allowed, consistent with our very small estimate of flight in this decade and pre-1900 patterns of black settlement even more stringent than what we consider. However, by the 1910s, the “with flight” calculation significantly increases dissimilarity relative to the “no-flight” case, with dissimilarity growing by .025 when flight is allowed (relative to the actual increase of .056). By the 1920s, white flight can explain all of the increase in dissimilarity (.081 versus the actual increase of

.078). Considering the cumulative effect of flight in our calculation, white flight can explain 60 percent of the observed increase in segregation as measured by dissimilarity, with the explanatory power concentrated in the 1910s and particularly the 1920s.

We next consider a black allocation rule that assumes no constraints on where black in-migrants could settle. While this rule is not particularly realistic for prewar America, when restrictive covenants could be upheld in court, it is much more appropriate for the current legal and institutional environment. To some extent, this calculation asks if segregation could have emerged even if there had been no discrimination in the housing market. In this rule (Black: By Population), we assign blacks analogously to how we assign whites in Step 6. Specifically, we assign black population growth over the 1900 to 1910 decade according to the distribution of *total population* across each city in 1900. Black in-migrants can settle in any neighborhood in proportion to the neighborhood's relative size under this rule. The 1910 hypothetical black population is obtained by summing the allocated portion of black inflows over the 1900 to 1910 decade and the 1900 black population for each neighborhood. We then repeat this process using 1910 and 1920 as base years for the latter two decades.

The results of this calculation are shown in the next set of bars in Figure 8. Unsurprisingly, dissimilarity drops over the first decade as blacks are allocated to new neighborhoods with only a very weak flight effect on whites. However, white flight can explain about 20 percent of the rise in dissimilarity over the 1910s (an increase of .011 compared with an actual increase of .056) and again nearly 100 percent of the rise in dissimilarity over the 1920s. Interestingly, the white flight from black arrivals would be expected to generate 45 percent of the observed increase in segregation between 1900 and 1930, even without any constraints on where blacks could live. Of course, one objection to this assignment rule is that it ignores black

preferences for living alongside other blacks. In other words, if blacks had been free to choose their neighborhoods freely, what impact would white flight have had on the dissimilarity trend?

To develop an assignment rule that reflects levels of co-ethnic residence that would arise from minority preferences as opposed to discriminatory institutions, we use the distribution of Italian immigrants in the sample in each base year. Italians composed about 1 percent of the sample in 1900 and 2.2 percent in 1920, making them reasonably comparable (albeit smaller) in overall size to blacks. In each base year, we compute the percentiles of black share and the percentiles of Italian share for each city. We then assign blacks to neighborhoods using the Italian percentile distribution. For instance, in Chicago the 80<sup>th</sup> percentile of black population share is .0026 in 1920 while the 80<sup>th</sup> percentile of Italian population share is .0036 (the distribution of Italians is generally less concentrated than the distribution of blacks). Our (Black: By Italian) rule assigns .0036 of Chicago's 1920 to 1930 black inflows to neighborhoods in the 80<sup>th</sup> percentile of black population share in 1920. The final set of bars in Figure 8 shows a very similar pattern in the impact of white flight on dissimilarity as with the case of no constraints. White flight from black arrivals could thus have been expected to generate sizable levels of segregation regardless of the institutional environment faced by blacks. In particular, between 45 and 60 percent of the observed increase in dissimilarity appears to arise as a consequence of white sorting.

## **VII. Conclusion**

Taken in total, our results suggest that the dynamics of white populations likely played a key role in the sharp increase in racial segregation observed over the 1900 to 1930 period. Our nonlinear analysis showed that white population loss in tipping neighborhoods accelerated over

the period. Furthermore, the causal, linear analysis showed that black arrivals caused an increasing number of white departures in each decade: by the 1920s, one black arrival was associated with the loss of more than four white individuals. The robustness of these findings and the way in which they vary across time suggests that evolving changes in white animus was a key factor in rising racial segregation.

White flight was not simply a response to deplorable ghetto conditions developed over decades of black migration to northern cities. Instead, whites appear to have been fleeing black neighbors as soon as the migration from the South got underway, and these market decisions had important impacts on the aggregate level of racial segregation in cities. These findings nuance our understanding of the persistence of segregation in the United States, suggesting that even the complete elimination of racial discrimination in housing markets may fail to bring about significant racial integration so long as the sizeable numbers of white individuals remain willing to move to avoid having black neighbors.

## BIBLIOGRAPHY

- Ananat, E. O. "The wrong side (s) of the tracks: The causal effects of racial segregation on urban poverty and inequality." *American Economic Journal: Applied Economics*, 3(2), 2011, pp. 34-66.
- Ancestry.com. *1900 United States Federal Census* [database on-line]. Provo, UT, USA: Ancestry.com Operations Inc, 2004.
- Ancestry.com. *1910 United States Federal Census* [database on-line]. Provo, UT, USA: Ancestry.com Operations Inc, 2006.
- Ancestry.com. *1920 United States Federal Census* [database on-line]. Provo, UT, USA: Ancestry.com Operations Inc, 2010.
- Ancestry.com. *1930 United States Federal Census* [database on-line]. Provo, UT, USA: Ancestry.com Operations Inc, 2002.
- Beaman, Lori. "Social Networks and the Dynamics of Labor Market Outcomes: Evidence from Refugees Resettled in the U.S." *Review of Economic Studies*, 79(1), 2012, pp.128-161.
- Boustan, Leah Platt. "Was Postwar Suburbanization "White Flight"? Evidence from the Black Migration." *The Quarterly Journal of Economics* 125.1, 2010, pp. 417-443.
- Boustan, Leah Platt. "Local Public Goods and the Demand for High-Income Municipalities," *Journal of Urban Economics*, 76, 2013, pp. 71-82.
- Boustan, Leah Platt, Price V. Fishback, and Shawn Kantor. "The Effect of Internal Migration on Local Labor Markets: American Cities during the Great Depression." *Journal of Labor Economics*, 28(4), 2010, pp. 719-746.
- Card, David, Alexandre Mas, and Jesse Rothstein. "Tipping and the Dynamics of Segregation." *The Quarterly Journal of Economics*, 123(1), 2008, pp. 177-218.
- Cascio, Elizabeth U. and Ethan G. Lewis. "Cracks in the Melting Pot: Immigration, School Choice, and Segregation," *American Economic Journal: Economic Policy*, 4(3), 2012, pp. 91-117.
- Chetty, Raj, Nathan Hendren, Patrick Kline, and Emmanuel Saez. "Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States," *Quarterly Journal of Economics*, 129(3), 2014, pp. 1553-1623.
- Chicago Commission on Race Relations. "The Negro in Chicago: A Study of Race Relations and a Race Riot." 1922.

- Cutler, David M. and Edward L. Glaeser. "Are Ghettos Good or Bad?" *The Quarterly Journal of Economics*, 112(3), 1997, pp. 827-872.
- Cutler, David M. and Grant Miller. "The role of public health improvements in health advances: the twentieth-century United States." *Demography* 42.1 (2005): 1-22.
- Cutler, David M., Edward L. Glaeser and Jacob L. Vigdor. "Is the Melting Pot Still Hot? Explaining the Resurgence of Immigrant Segregation." *The Review of Economics and Statistics*, 90(3), 2008, pp. 478-497.
- Cutler, David M., Edward L. Glaeser and Jacob L. Vigdor. "The Rise and Decline of the American Ghetto." *Journal of Political Economy* 107, 1997, pp. 455-506.
- Edin, Per-Anders, Peter Fredriksson, and Olof Åslund. "Ethnic Enclaves and the Economic Success of Immigrants – Evidence from a Natural Experiment." *The Quarterly Journal of Economics*, 118(1), 2003, pp. 329-57.
- Farley, Reynolds, and Walter R. Allen. *The Color Line and the Quality of Life in America*. Russell Sage Foundation, 1987.
- Ferrie, Joseph P., and Werner Troesken. "Water and Chicago's mortality transition, 1850–1925." *Explorations in Economic History* 45.1 (2008): 1-16.
- Gould, J. D. "European Inter-Continental Emigration. The Road Home: Return Migration from the U.S.A." *Journal of European Economic History* 9 (1), 1980, pp. 41–112.
- Grossman, James R. *Land of Hope: Chicago, Black Southerners and the Great Migration*. University of Chicago Press, 1991.
- Kim, Sukkoo. "Expansion of markets and the geographic distribution of economic activities: the trends in US regional manufacturing structure, 1860–1987." *The Quarterly Journal of Economics* 110.4, 1995, pp. 881-908.
- Kim, Sukkoo. "Changes in the Nature of Urban Spatial Structure in the United States, 1890-2000." *Journal of Regional Science* 47(2), 2007, pp. 273-287.
- Lieberson, Stanley. *A Piece of the Pie: Blacks and White Immigrants since 1880*. Berkeley: University of California Press, 1980.
- Logan, John R., Jason Jindrich, Hyoungjin Shin, and Weiwei Zhang. "Mapping America in 1880: The Urban Transition Historical GIS Project." *Historical Methods* 44(1), 2011, pp. 49-60.
- Massey, Douglas S. and Nancy A. Denton. *American Apartheid: Segregation and the Making of the Underclass*. Cambridge: Harvard University Press, 1993.

Ruggles, Stephen et al. Integrated Public Use Microdata Series: Version 4.0 [Machine-readable database]. Minnesota Population Center, Minneapolis, MD. 2008.

Saiz, Albert and Susan Wachter. "Immigration and the Neighborhood." *American Economic Journal: Economic Policy*, 2011, pp. 169-188.

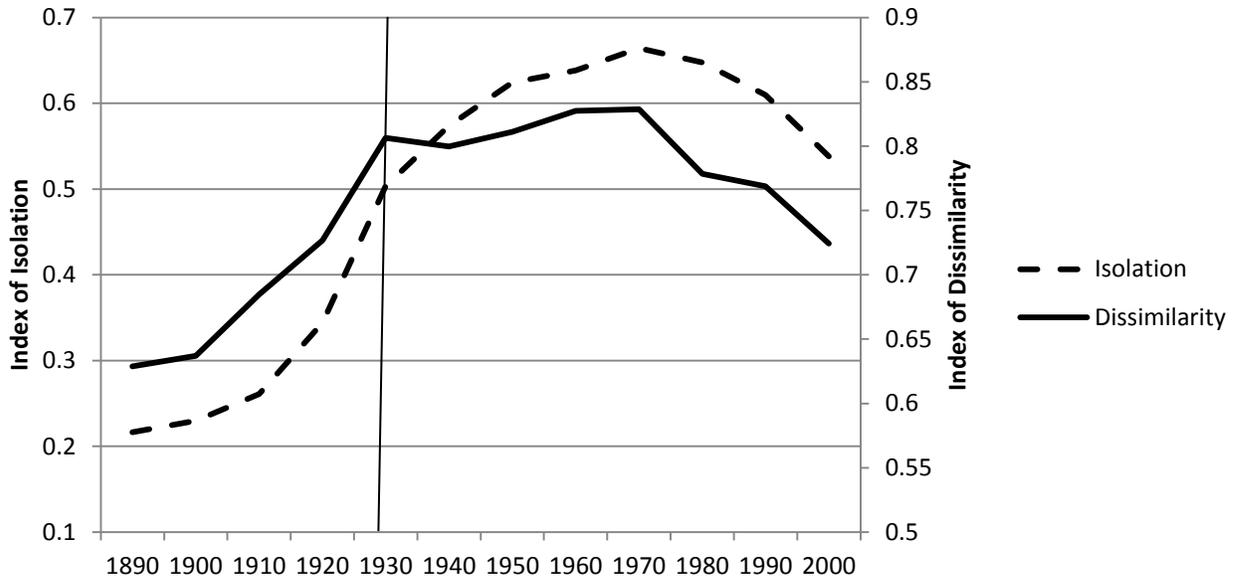
Schelling, Thomas C. "Dynamic Models of Segregation," *Journal of Mathematical Sociology*, 1(2), 1971, pp. 143-186.

Sharkey, Patrick. *Stuck in Place: Urban Neighborhoods and the End of Progress Toward Racial Equality*. Chicago: University of Chicago Press, 2013.

Willcox, Walter F. *Statistics of Migrations, National Tables, United States*. 1929. Retrieved from <http://www.nber.org/chapters/c5134>.

Wilson, William Julius. *When Work Disappears: The World of the New Urban Poor*. New York: Vintage Books, 1996.

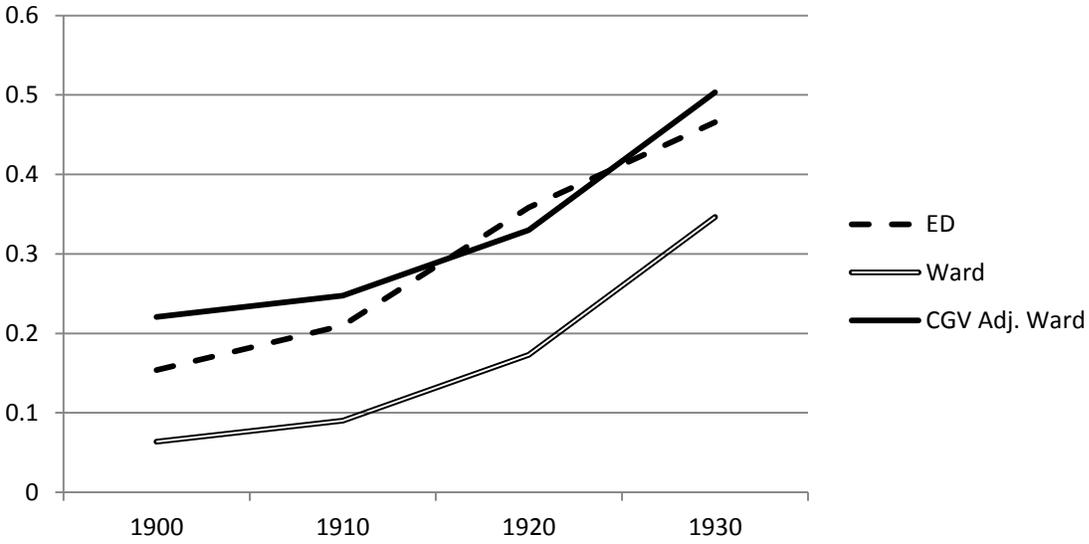
**Figure 1. Segregation Trends in the Largest Ten American Cities, 1890-2000**



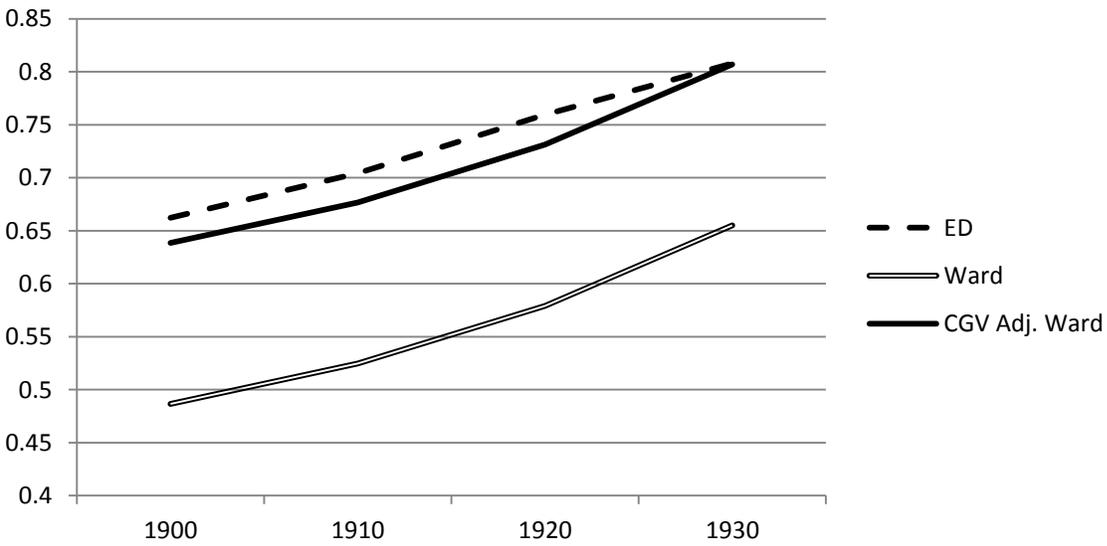
Notes: Data are taken from the dataset used in Cutler, Glaeser, and Vidgor (1999) and show the average segregation indices across Baltimore, Boston, Brooklyn, Chicago, Cincinnati, Cleveland, Detroit, Manhattan, Philadelphia, Pittsburgh, and St. Louis. We employ their adjustment factor to make the ward-level indices from 1930 and before comparable to the 1940 and onward tract-level indices.

**Figure 2. Segregation Trends by Enumeration and Ward, 1900-1930**

**A. Isolation**



**B. Dissimilarity**



Notes: See Figure 1 for notes on the ward and adjusted ward data from Cutler, Glaeser, and Vigdor (1999). The enumeration district segregation averages are computed using the universe of census records from each of the ten sample cities accessed from Ancestry.com.

**Figure 3. Digitizing the Enumeration Districts**



A. Enumeration District Map



B. Digitized Street Map

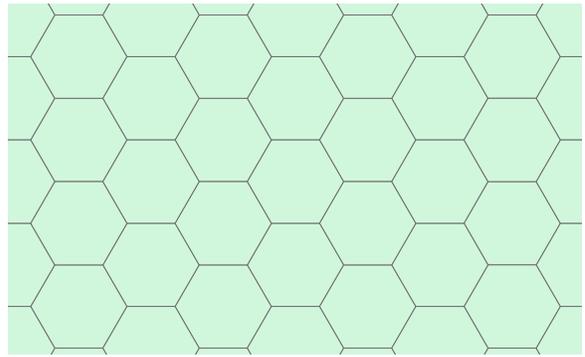
150 Pittsburg City, 12th Ward, Pct 5, bounded by Allegheny River, 31st, Smallman, 28th

C. Enumeration District Descriptions



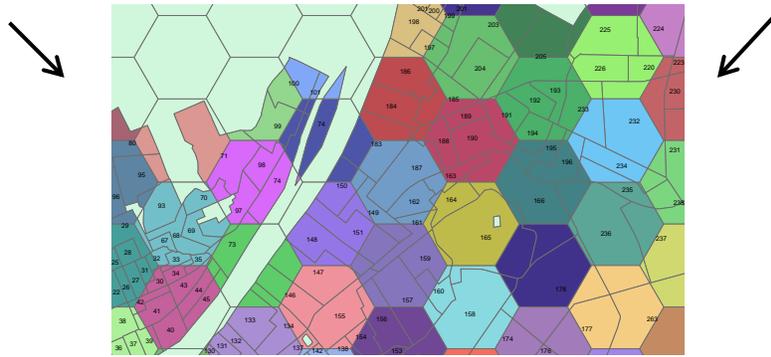
D. Digitized Enumeration District Map (ArcMap)

**Figure 4. Constructing Hexagon Neighborhoods from Enumeration District Maps**

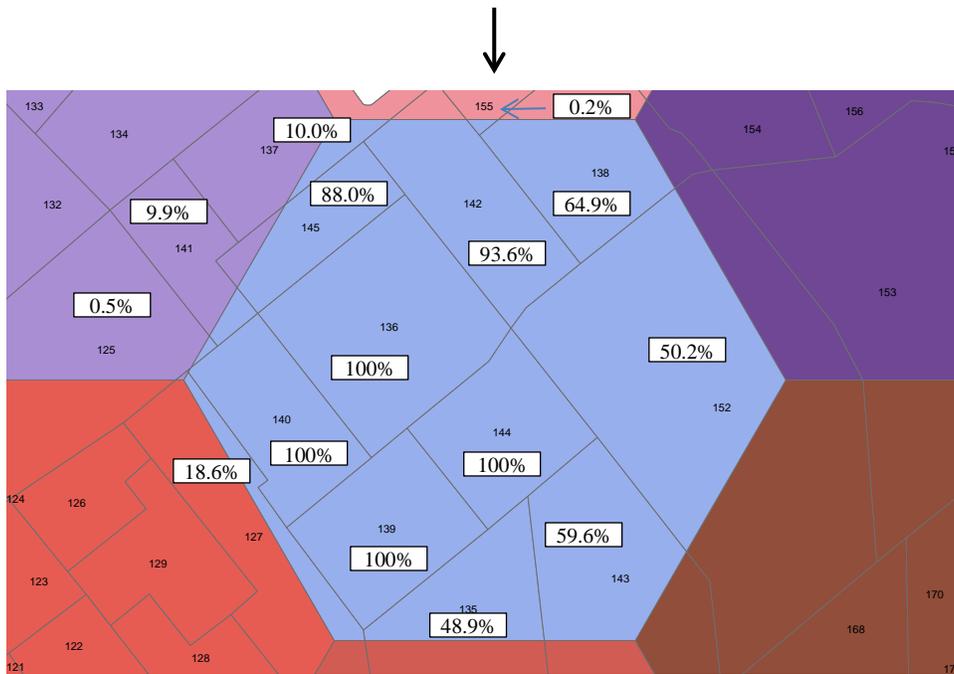


A. Enumeration District Map (1900)

B. Hexagon Grid (Constant across Decades)



C. Intersection between Enumeration Districts and Hexagons

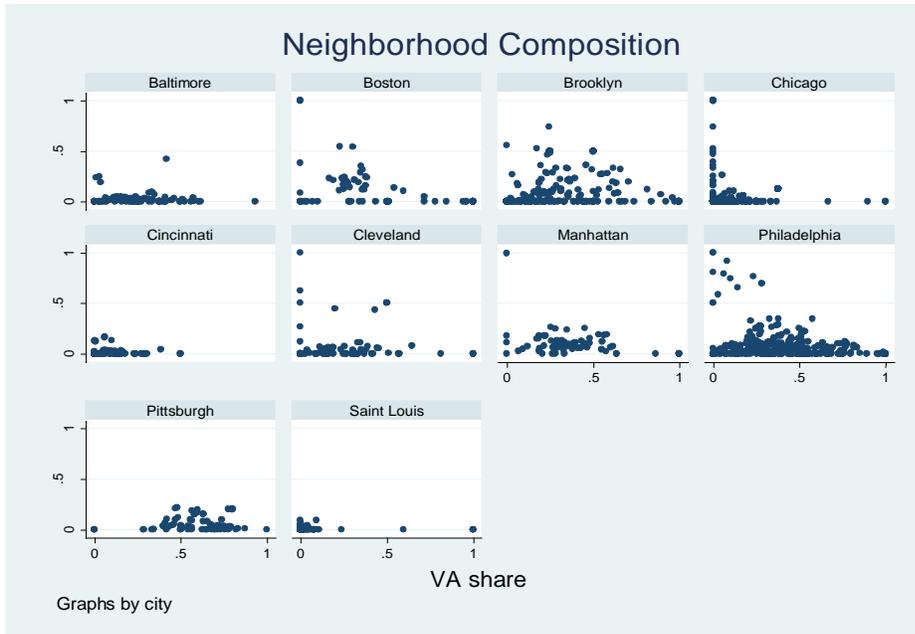


D. Allocating Enumeration District Count Data to Hexagon Neighborhoods

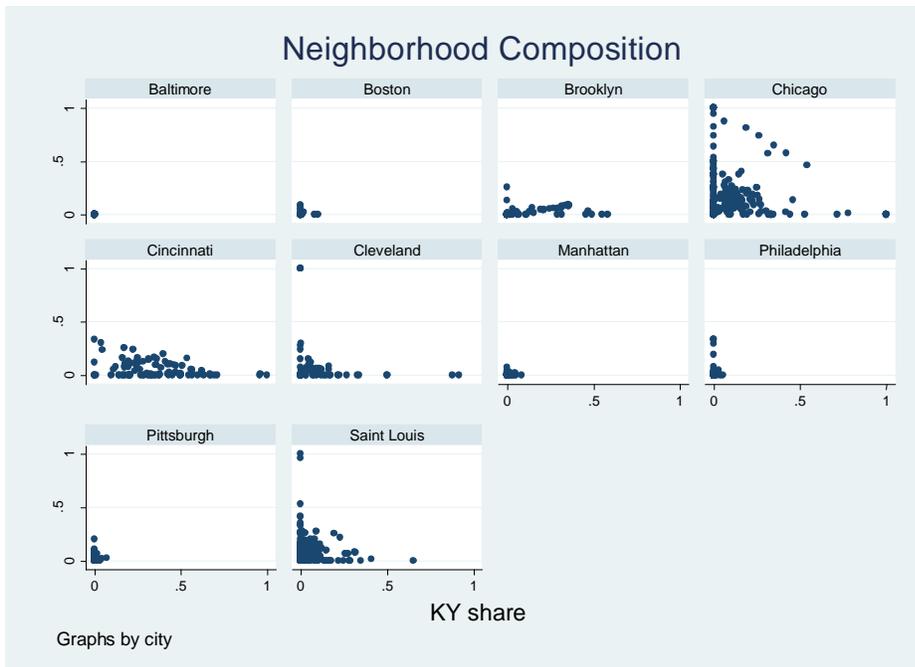
Notes: see Section III for details on the source of the maps and street files used to construct these images.

**Figure 5: Variation in Origin of Black Settlement across Neighborhoods in 1900**

*A. Virginia versus North Carolina*

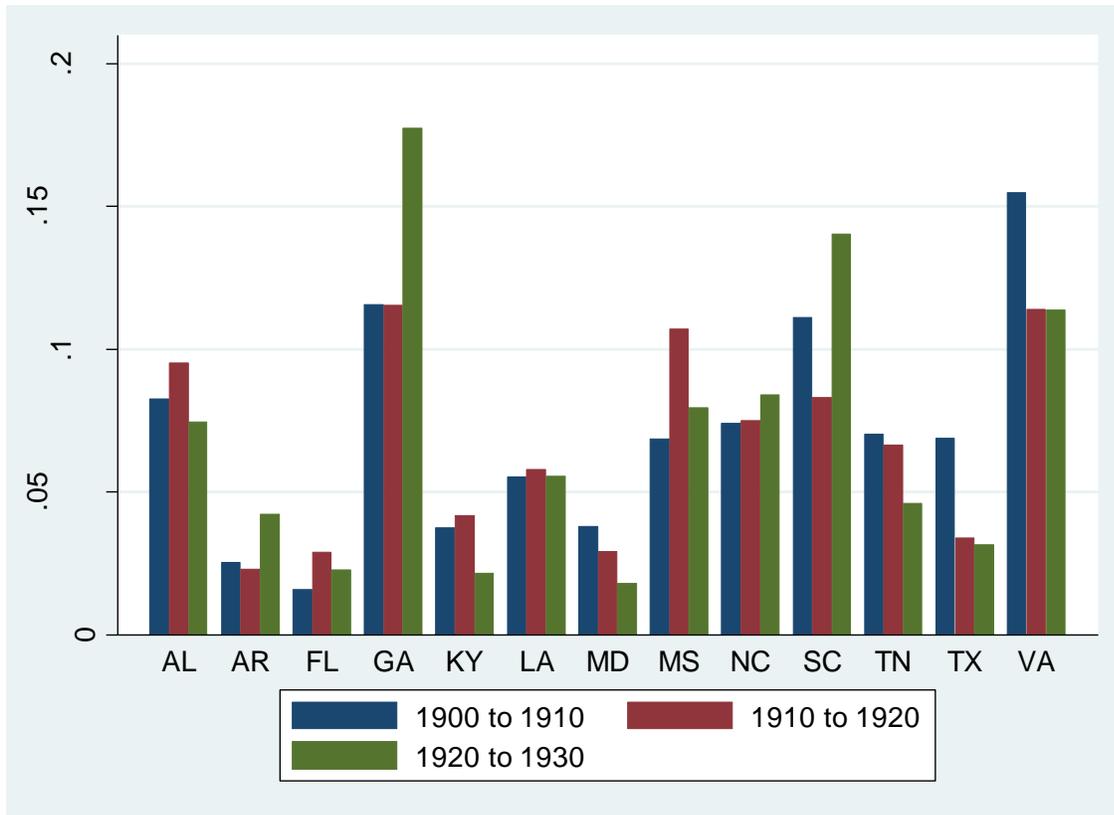


*B. Kentucky vs Tennessee*



Notes: Scatterplots show the share of black men aged 20 to 29 born in each source state out of the total number of black men in the cohort in neighborhood. The shares are computed using the universe of census records with enumeration district identifiers from each city and the hexagon imputation strategy discussed in Section III.

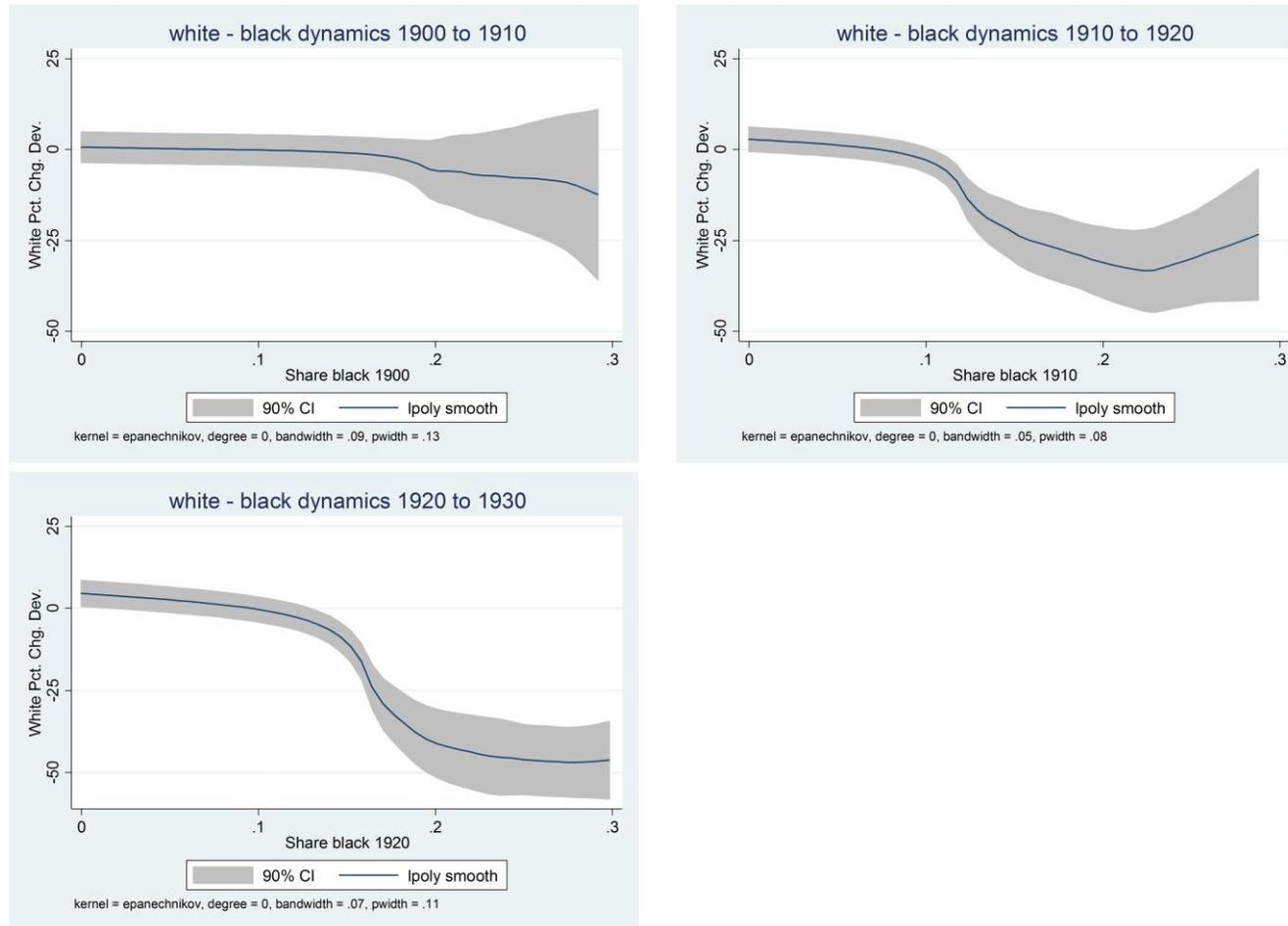
**Figure 6. Variation in Estimated Black Outflows from Southern States by Decade**



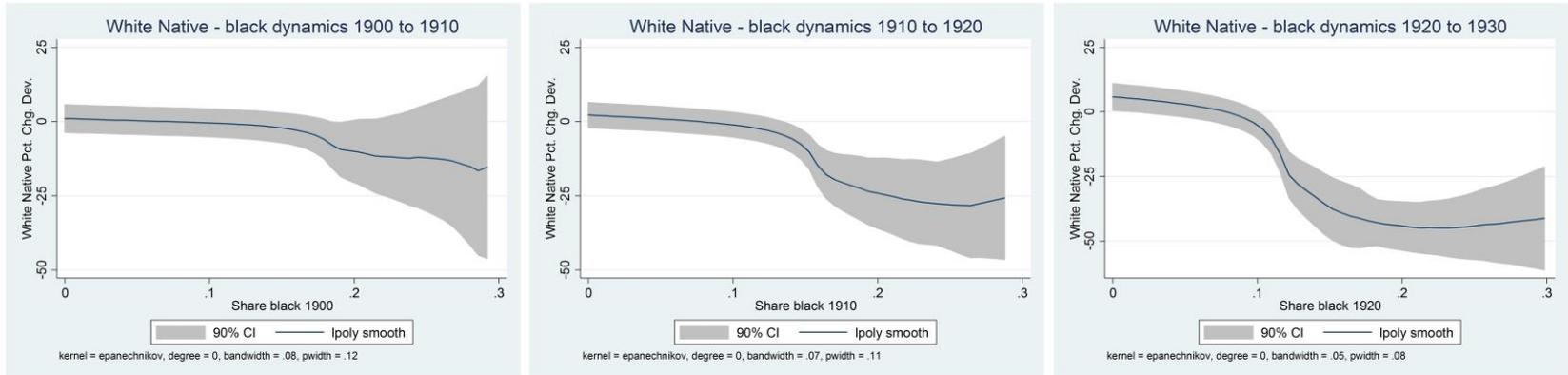
Notes: The data in this figure come from the universe of census microdata made available by Ancestry.com. Estimated outflows are computed by summing the change in the number of individuals in gender, state of birth, and birth cohort cells appearing outside their birth state in each census year.

**Figure 7: Panel A. Black and White Population Dynamics**

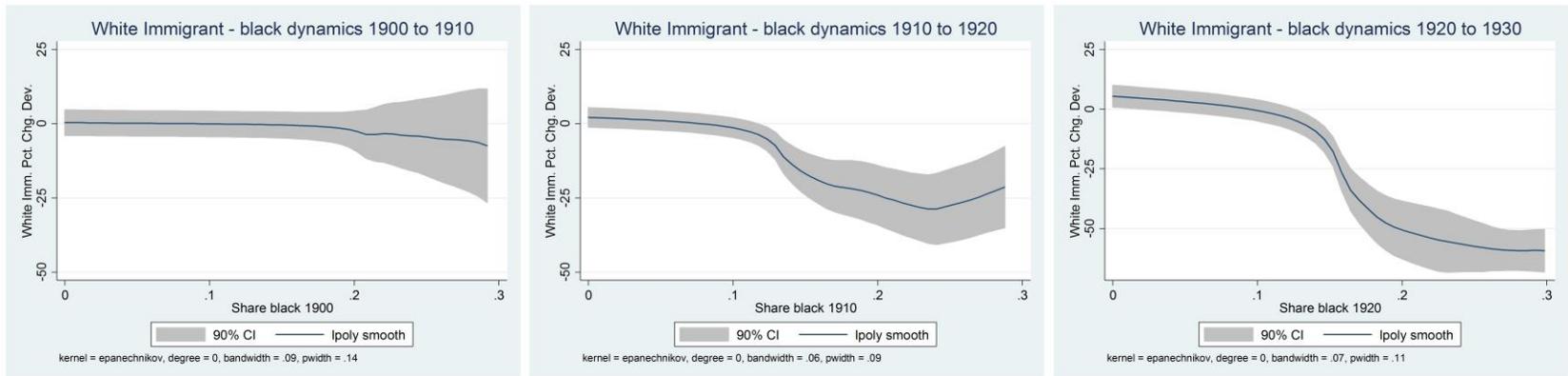
*A. Full Sample: Relationship between Share Black and White Population Change*



**B. Full Sample: Relationship between Share Black and White Native Population Change**

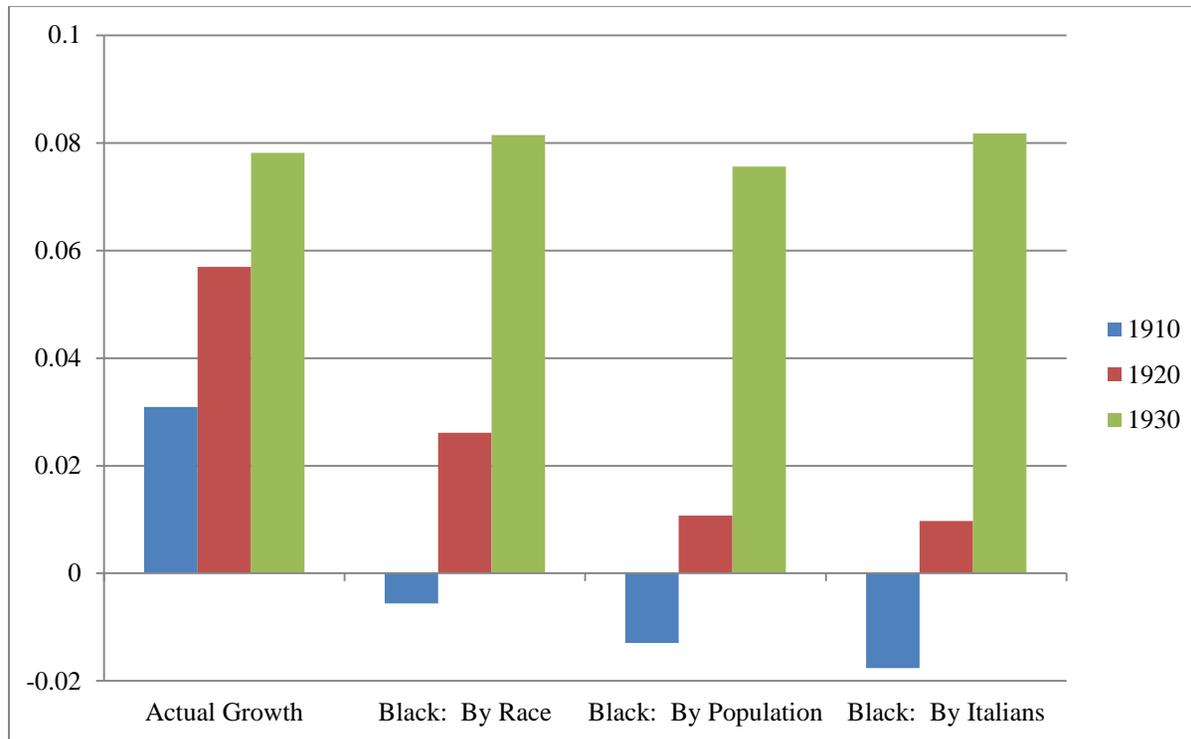


**C. Full Sample: Relationship between Share Black and White Immigrant Population Change**



Notes: All figures show the nonparametric relationship between share black and either total white or white subpopulation changes in the neighborhood over the next decade. All white population changes are de-meanded (at the city level) values. The demographic measures are computed from the universe of census records and the neighborhoods are the panel of 800 meter hexagons described in Section III.

**Figure 8. Flight vs. No Flight Calculation of Changes in Dissimilarity**



Notes: The chart shows the actual growth in dissimilarity across the balanced panel of neighborhoods in each decade. Black: By Race shows the difference in dissimilarity predicted by the “with flight” and “no flight” calculations from Section VI when blacks are assigned according to the black population distribution in each base year. Black: By Population shows the difference in dissimilarity when blacks are assigned according to the overall population distribution in each base year. Black: By Italians show the difference in dissimilarity when blacks are assigned according the matched Italian percentile allocation rule described in Section VI.

**Table 1. Summary Statistics for Hexagon Panel Dataset**

	1910	1920	1930
Change in White Population	683.87 (1246.37)	588.91 (1317.75)	264.70 (1954.33)
Change in Black Population	24.12 (104.25)	60.37 (261.06)	131.36 (491.04)
Change in White 3rd Generation Population	208.55 (507.19)	320.96 (556.33)	182.22 (704.77)
Change in White Second-Generation Population	219.49 (507.18)	188.21 (539.68)	119.97 (781.43)
Change in White First-Generation Population	256.24 (643.96)	61.45 (585.62)	16.60 (822.20)
Black Percent	2.40 (4.64)	3.18 (7.89)	5.25 (13.82)
White 3rd Generation Percent	37.74 (17.12)	40.28 (18.60)	41.30 (19.55)
White Second-Generation Percent	33.34 (9.22)	32.41 (9.69)	31.82 (11.23)
White First-Generation Population	25.99 (11.91)	23.37 (11.26)	21.56 (11.02)
Population	3349 (5098)	4047 (5069)	4493 (4376)

Notes: Summary statistics cover the hexagon neighborhoods in our panel dataset. Changes in population are also with respect to the previous decade's value. All demographic variables were created using the 100 percent sample of census records from Ancestry.com. Only hexagons with at least 95 percent coverage by enumeration districts from the respective census in each year are included in the panel. We also trim the sample at the 1<sup>st</sup> and 99<sup>th</sup> percentile of both white and black population change for each decade.

**Table 2. Baseline OLS and IV Results for Effect of Black Arrivals on White Departures**

	dependent variable = change in white population		
	1900-1910 Decade	1910-1920 Decade	1920-1930 Decade
<u>OLS Results</u>			
Change in Black Population	0.473** (0.232)	-0.856*** (0.0968)	-1.542*** (0.0709)
R-squared	0.110	0.133	0.265
<u>IV Results</u>			
Change in Black Population	-0.770 (0.548)	-2.247*** (0.220)	-4.219*** (0.306)
R-squared	0.307	0.042	0.127
<u>First Stage</u>			
Predicted Change in Black Pop.	0.814*** (0.0355)	0.420*** (0.0167)	0.516*** (0.0346)
F-test on First Stage	526.5	631.5	222.9
Observations	2,408	2,405	2,406

Notes: See Table 1 for sample details. All regressions include city fixed effects. The instrumental variables regressions are estimated using limited information maximum likelihood estimation (LIML).

**Table 3. White Flight Effect by White Subpopulations (IV)**

	1900-1910 Decade	1910-1920 Decade	1920-1930 Decade
<u>Dep. Var. = Change in White 3rd-Gen. Pop.</u>			
Change in Black Population	-1.675*** (0.243)	-0.869*** (0.0925)	-1.514*** (0.114)
<u>Dep. Var. = Change in Second-Gen. Pop.</u>			
Change in Black Population	-0.121 (0.225)	-0.564*** (0.0873)	-1.238*** (0.112)
<u>Dep. Var. = Change in First-Gen. Pop.</u>			
Change in Black Population	1.117*** (0.277)	-0.710*** (0.101)	-1.394*** (0.126)
Observations	2,408	2,405	2,406

Notes: see Table 2 for sample and specification details.

**Table 4. White Flight Effect Robustness Checks (IV)**

	dependent variable = change in white population		
	1900-1910 Decade	1910-1920 Decade	1920-1930 Decade
Change in Black Population (baseline)	-0.770 (0.548)	-2.247*** (0.220)	-4.219*** (0.306)
Change in Black Population	1.342* (0.756)	-2.316*** (0.344)	-4.585*** (0.340)
Percent Black in 1900	-43.20*** (9.133)	2.828 (10.21)	20.96 (13.16)
Change in Black Population		-1.645** (0.676)	-3.012*** (0.450)
Number of Blacks in 1900		-0.187 (0.210)	-0.525*** (0.191)
Change in Black Population		-2.278*** (0.328)	-4.100*** (0.323)
Percent Black in 1900		14.18 (9.797)	25.64** (12.12)
Pre-Trend in White Population		0.254*** (0.0171)	0.236*** (0.0286)
Southern states IV	-1.318** (0.629)	-2.129*** (0.219)	-5.053*** (0.429)
Observations	2408	2405	2406
	dependent variable = change in white households		
	1900-1910 Decade	1910-1920 Decade	1920-1930 Decade
Change in Black Households	-0.598 (0.466)	-1.362*** (0.159)	-4.038*** (0.327)
Observations	2296	2289	2283

Notes: see Table 2 for sample and specification details. The household-level regressions at the bottom of the table are also trimmed at the 99<sup>th</sup> percentile of the ratio of white to white household heads and black to black household heads.

**Table 5. White Flight by Neighborhood Type**

	1920 Black Share				
	Full	0-5%	5-10%	10-20%	>20%
Coefficient on black difference, 1920-1930	-4.585***	-8.006***	-4.882***	-5.232***	-2.677***
Standard error	(0.340)	(0.774)	(1.417)	(1.763)	(0.635)
Mean white population in 1920	3883	3853	3961	3715	4642
Mean black population in 1920	193	32	313	619	3013
Mean change in black population, 1920-1930	131	58	345	444	992
Implied change in white population	281	438	-640	-2011	-5588
Actual change in white population	265	485	-510	-829	-1889
Implied percent change in white population	0.07	0.11	-0.16	-0.54	-1.20
Actual percent change in white population	0.07	0.13	-0.13	-0.22	-0.41
N	2406	2040	152	122	92

Notes: All specifications include share black in 1900 as well as city fixed effects. See Table 1 for sample details. The instrumental variables regressions are estimated using limited information maximum likelihood estimation (LIML).