NBER WORKING PAPER SERIES

RACIAL SORTING AND THE EMERGENCE OF SEGREGATION IN AMERICAN CITIES

Allison Shertzer
Randall P. Walsh

Working Paper 22077
http://www.nber.org/papers/w22077

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 2016

Support for this research was provided by the National Science Foundation (SES-1459847). Additional support was provided by the Central Research Development Fund and the Center on Race and Social Problems at the University of Pittsburgh. We are grateful to Brian Cadena, Terra McKinnish, 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 (DAE), ASSA Meetings, Carnegie Mellon, Michigan, Georgia State, Mississippi State, Colorado, and the University of Western Australia 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 Villarreal and the Union Army Project (www.uadata.org) 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. Antonio Diaz-Guy, Phil Wetzel, Jeremy Brown, Andrew O'Rourke, Aly Caito, Loleta Lee, and Zach Gozlan provided outstanding research assistance. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2016 by Allison Shertzer and Randall P. Walsh. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Racial Sorting and the Emergence of Segregation in American Cities
Allison Shertzer and Randall P. Walsh
NBER Working Paper No. 22077
March 2016
JEL No. J15,N32,R23

ABSTRACT

Residential segregation by race grew sharply in the United States as black migrants from the South arrived in northern cities during the early twentieth century. The existing literature emphasizes discriminatory institutions as the driving force behind this rapid rise in segregation. Using newly assembled neighborhood-level data, we instead focus on the role of "flight" by whites, providing the first systematic evidence of the role that prewar population dynamics played in the emergence of the American ghetto. Leveraging exogenous changes in neighborhood racial composition, we show that white departures in response to black arrivals were quantitatively large and accelerated between 1900 and 1930. Our preferred estimates suggest that white flight was responsible for 34 percent of the increase in segregation over the 1910s and 50 percent over the 1920s. Our analysis suggests that segregation would likely have arisen in American cities even without the presence of discriminatory institutions as a direct consequence of the widespread and decentralized relocation decisions of white urban residents.

Allison Shertzer
Department of Economics
University of Pittsburgh
4901 WW Posvar Hall
230 South Bouquet Street
Pittsburgh, PA 15260
and NBER
shertzer@pitt.edu

Randall P. Walsh
Department of Economics
University of Pittsburgh
4901 WW Posvar Hall
230 S. Bouquet St.
Pittsburgh, PA 15260
and NBER
walshr@pitt.edu

I. Introduction

Residential segregation by race remains one of the most salient and enduring features of American cities. There is substantial evidence 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; Ananat, 2011; Sharkey, 2013). Segregation has also been found to be negatively correlated with the intergenerational mobility of both black and white urban residents (Chetty, Hendren, Line, and Saez, 2014). In response to these disparities, both federal and state governments have enacted sweeping policies aimed at achieving more equitable access to housing and funding for public goods.

An initial wave of policies focused on dismantling structural barriers that prevented blacks from locating in white neighborhoods. The Federal Fair Housing Act of 1968 and California's Rumford Fair Housing Act of 1963 were both intended to limit the ability of whites to keep blacks out of their neighborhoods. The 1948 Supreme Court case *Shelley V. Kramer*, which ruled restrictive covenants unenforceable, also had the effect of reducing the number of legal mechanisms available to whites to maintain the color line.[2] The motivation behind these laws is consistent with the scholarly consensus that collective action by whites produced the American ghetto in the first half of the twentieth century (Massey and Denton, 1993; Cutler, Glaeser, and Vigdor, 1999).

More recent policies have instead focused on funding disparities across jurisdictions arising from the departure of wealthier white residents from central cities. Federal community development block grants, state-level school finance equalization schemes, and federal aid

[2] A recent empirical assessment of the *Shelley V. Kraemer* Supreme Court case found that striking down the enforceability of restrictive covenants had significant effects on black mobility and neighborhood racial composition (Kucheva and Sander, 2014).

targeted at low-income schools all aim to equalize public spending arising across cities and suburbs. These policies address the fact that segregation – and inequality – can arise as a consequence of uncoordinated choices in the housing market. Several studies have shown that over the postwar period, the willingness of white individuals to depart neighborhoods with rising black populations was an important mechanism through which racial segregation was perpetuated (Card, Mas, and Rothstein, 2008; Boustan, 2010). Moreover, related work has shown that these sorting patterns can become self-reinforcing through preferences for related public goods (Bayer, Ferreira, and McMillan, 2007). However, the potential contribution of sorting for the *emergence* of the American ghetto has been left largely unexplored.

In this paper we ask how segregation emerged in American cities, exploring in particular the importance of structural barriers versus white flight in the decades that saw the most rapid ghettoization. Figure 1 presents the aggregate trend in twentieth century segregation by race as computed by Cutler, Glaeser, and Vigdor (1999) for the ten largest northern U.S. cities (based on 1880 population).[3] The figure shows that 97 percent of the twentieth century increase in dissimilarity and 63 percent of the increase in isolation occurred between 1900 and 1930.[4] We thus focus on the first three decades of the twentieth century in our analysis of population sorting. White flight in this period can primarily be thought of as departures for neighborhoods outside the urban core but still within city boundaries. Thus, unlike the suburban destinations of postwar white flight (Boustan, 2013), the destination neighborhoods for whites fleeing black arrivals would have been similar to their origin neighborhoods in terms of spending on public

[3] 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.
[4] Isolation peaked in 1970, with isolation rising from .23 to .66 between 1900 and 1970. However, 63 percent of the overall increase had occurred by 1930. Dissimilarity 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. This sharp increase in northern urban segregation occurred against a backdrop of nationally rising segregation levels: recent work using a household-level measure finds that segregation levels doubled between 1880 and 1940 (Logan and Parman, 2015).

goods and tax base. Estimates of white flight in this period may thus provide a better gauge of racial distaste than those using postwar data.

To date, the capacity of scholars to rigorously investigate the mechanisms responsible for the emergence of segregation has been limited by a lack of spatial data covering racial composition in prewar neighborhoods in the United States. One contribution of our study is the construction of a fine-grained, spatially-identified demographic dataset covering ten major U.S. 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 time. We provide the first systematic analysis of sorting by white households in prewar America. We then go on to evaluate the relative importance of collective white action compared with white flight in explaining the emergence of segregation in northern cities.

We begin our empirical work with a simple nonparametric analysis of the demographic trends in our ten city sample. We identify patterns in the data that are consistent with the neighborhood “tipping" model of racial dynamics as first proposed by Thomas Shelling (1971). Taking this initial evidence as suggestive, we then adopt a more formal empirical strategy that identifies the causal link between black in-migration and white flight. We utilize exogenous changes in neighborhood-level black populations that we isolate by interacting variation in the state-level outmigration rates of blacks with within-city cross-neighborhood variation in the state of origin of early black arrivals. This strategy is similar in spirit to the approach taken in the immigration shock literature although we leverage variation across different neighborhoods within given cities rather than variation across cities.[5]

[5] See for instance, Altonji and Card (1991) and Card (2005).

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. The nonparametric analysis finds that over the 1910s, relative declines in white population were 30 percentage points higher in neighborhoods that were over 10 percent black. During the 1920s, these relative declines in white population grew to the order of 40 percentage points, consistent with a model of abrupt racial turnover in urban neighborhoods.

Our causal analysis confirms the existence and acceleration of white flight over the early twentieth century. Results from a naive OLS analysis find one black arrival in the preceding decade associated with .9 and 1.5 white departures during the 1910s and 1920s, respectively. Of course, these OLS results fail to account for endogeneity concerns and could for instance be explained solely by the one-for-one replacement of white movers by black migrants in an environment with inelastic housing supply. However, our instrumental variables analysis, which assigns estimated state-level black outflows to northern cites according to black settlement patterns prior to the Great Migration, indicates that one exogenous black arrival was associated with 1.9 white departures in the 1910s and 3.4 white departures by the 1920s. These IV results suggest that OLS estimates were biased against a finding of flight, likely due to both white and black settlement being drawn to generally growing neighborhoods.

In the final portion of our analysis, we construct a series of counterfactuals aimed at understanding how much of the observed increase in segregation over the 1900 to 1930 period can be attributed to white flight from black arrivals as opposed to institutional barriers constructed by whites. Our most striking finding is the sharp increase in the contribution of flight in each subsequent decade. While our preferred estimates suggest that white flight was inconsequential during the aughts, we estimate that flight was responsible for 34 percent of the increase in segregation (as measured by dissimilarity) over the 1910s and 50 percent of the increase over the 1920s. The impact of flight in the latter decade is particularly important given that the 1920s saw the largest increase in segregation of any decade in the twentieth century.

Our finding that sorting by whites out of neighborhoods with growing black populations was a quantitatively important phenomenon decades before the postwar opening of the suburbs is novel. This analysis suggests that segregation would likely have arisen even without the presence of discriminatory institutions as a direct consequence of the widespread and decentralized relocation decisions of white individuals. Whites could simply have responded to policies that reduced barriers to black settlement in their vicinity by accelerating their departure for neighborhoods at lower risk of "encroachment." Policies that reduce barriers faced by blacks in the housing market may thus not prevent or reverse segregation as long as white households have the ability and desire to avoid black neighbors.

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

A. The Rise of Segregation in the United States

We begin by confirming 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 highly sensitive to changes in the overall group share. For each year we compute:

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

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). This modified approach varies from the "standard" isolation index which simply computes the average percentage of a group member's neighborhood composed of members of her own group. We utilize this modified approach in order to control 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. While not completely delinking population size and isolation, this adjustment makes the measure less dependent on a group's share of the overall population.

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

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

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 are 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 segregation indices presented in Figure 1 were constructed using ward-level data for censuses prior to 1940 (this is the year when census tract data became widely available) and tract-level data in later decades. To make the ward and tract-level data comparable, Cutler et al estimate the relationship between tract-level and ward-level indices in 1940 and then use the estimated 1940 relationship to rescale the ward-level estimates in earlier years. Using our new enumeration district level data (discussed below in Section III), we compute these same segregation measures over the 1900 to 1930 time frame at both the enumeration district and ward level and report the results in Figure 2. As expected given their smaller scale, enumeration district-level segregation indices are markedly higher than those computed at the ward level (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 the racial segregation of neighborhoods. African Americans migration to northern cities began to accelerate 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). This combination of push and pull factors led to unprecedented out-migration from the South: 525,000 blacks came to the North in the 1910s while 877,000 came in the 1920s (Farley and Allen, 1987).

B. Scholarly Consensus and Context

The existing literature argues 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. In their seminal work on the emergence of segregation, Massey and Denton (1993) vividly describe 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 this view. Analyzing rental and house price data from 1940, they conclude that "in the mid-twentieth century, segregation was a product of collective actions taken by whites to exclude blacks from their neighborhoods." 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). Scholarly analysis of the role of individual sorting behavior has focused on the postwar period, concluding that white flight was critical to maintaining racial segregation (Boustan, 2011). We are unaware of any systematic empirical analyses of prewar population dynamics and segregation.

Of importance to our analysis 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.[6] 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.

III. Enumeration District Data for 1900 to 1930

The analysis in this paper is based on a new enumeration district-level spatial dataset spanning the years 1900 through 1930.[7] There are two major components to this data: 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.

[6] This computation uses the center city status variable from IPUMs samples for 1880 to 1930.
[7] A detailed description of the construction of this data can be found in Shertzer, Walsh, and Logan (2015).

The microdata 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 combination of cities and years that we study. Wards, which are still in use in some cities today, are large political units used to elect city council members while enumeration districts were small administrative units used internally by the census to coordinate enumeration activities prior to the shift to mail surveys in 1960. Each individual record in the Ancestry.com dataset includes 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 household 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 create digitized versions of census enumeration district maps based on two types of information available from the National Archives. We first employ written descriptions of the enumeration districts that are available on microfilm from the National Archives and have been made available online due to the work of Stephen P. Morse.[8] Second, we utilize a near complete set of physical enumeration district maps for our census-city pairs 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.[9] Figure 3

[8] website: http://stevemorse.org/ed/ed.php

[9] These street files can now be found at the Union Army Project's website (www.uadata.org). We used 1940 street maps produced by John Logan at the Spatial Structures in the Social Sciences at Brown University for Detroit, Cleveland, and St. Louis.

provides an illustration of this process which generated maps of more than 35,000 distinct enumeration districts. Here the shaded regions in panel D represent the digitized enumeration districts.

Analyzing demographic change over time within neighborhoods requires neighborhood definitions that are constant across census years. Using these data to form such neighborhoods is challenging 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. To address this challenge, we employ a hexagon-based imputation strategy. The strategy is illustrated in Figure 4. It involves covering the enumeration district maps (Panel 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 balanced panel comprised of all hexagons that were at least 95 percent covered by enumeration districts from the respective census in each year from 1900 to 1930, also trimming at the 1st and 99th percentile of both white and black population change for each decade to eliminate outliers from the sample. In Table 1 we provide summary statistics for the balanced sample of 1,975 hexagon neighborhoods. The neighborhoods have an average population of

3,160 individuals in 1910 and 4,216 in 1930, with the increase in density reflecting the rise in urban population density that occurred over this period. By 1930 the neighborhoods are thus roughly similar in population to modern-day census tracts. The average white population growth is positive in all years but declined from 650 over the 1900s to 282 over the 1920s, with much of this slowdown due to declining immigration from Europe after World War I and passage of the Immigration Restriction Act of 1921. The average black percent increased from 2.2 to 4.5 percent over the 1900 to 1930 period.

IV. Empirical Strategy

The objective of our 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. For example, newly arriving blacks may choose locations that were already being abandoned by white natives for reasons unrelated to race, leading to upwardly biased estimates of white flight responses in a naïve estimation framework. Conversely, blacks and whites could both be drawn to neighborhoods whose populations are growing due to other factors unrelated to race, leading to a downward bias in flight response estimates. To address this concern, we utilize an instrumental variables approach which leverages endogenous sources of variation in black population size at the neighborhood level. We begin by describing trends in our data nonparametrically.

A. Non-Parametric Analysis of Demographic Trends

To begin our analysis we consider neighborhood dynamics that are evident in the raw data. We use as our neighborhood definition the 800 meter hexagons described above and utilize a local polynomial smoothing approach to evaluate the non-parametric relationship between changes in neighborhood white populations and baseline black shares. 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:

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

where ΔWP_{ij}^{t1-t0} is the de-meaned (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.

B. Instrumental Variables Estimates of White Flight

While the non-parametric analysis is an effective way to provide visual evidence of the overall population dynamics, it cannot provide direct evidence on the causal relationship between black arrivals and the sorting behavior of whites. Our core estimation strategy addresses the causality of white flight by directly utilizing exogenous variation in neighborhood racial composition that arose as the result of heterogeneous state-level black outmigration shocks. Our analysis is in the spirit of the immigration shock literature (Altonji and 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 decadal change in black populations to the change in white populations:

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

where ΔW_{ij}^{t1-t0} (ΔB_{ij}^{t1-t0}) is the change in the number of whites (blacks) in a neighborhood over a decade and η_j is a city fixed effect. The coefficient of interest from this first differences strategy, β, relates the change in the number of blacks to the change in the number of whites in a particular neighborhood over the same decade with the city-level average captured by the fixed effect.[10]

In recent work there has been a growing concern that inappropriate model specification can lead to biased estimates in models of native displacement (Peri and Sparber, 2011; Wright et al., 1997; Wozniak and Murray, 2012). We implement a change in levels specification because it facilitates the implementation of our counterfactual analysis and provides the most parsimonious implementation for our IV strategy. This approach also does well in Peri and Sparber's Monte Carlo simulations of specification bias in displacement models and makes our results more directly comparable to work in the post-war period by Boustan (2010). One potential remaining concern is that a levels-based model will implicitly place a higher weight on more heavily populated neighborhoods. This concern motivates our decision to trim the sample at the 1st and 99th percentiles of black and white population changes.[11] As a further robustness check, in Appendix Table I, we demonstrate that our results are robust to stratification of the sample by population quartile.

While informative about general patterns in the data, due to a host of endogenity concerns, it would be inappropriate to draw causal inferences from estimates associated with equation (2). The following cases highlight a number of the potential sources of bias. First, consider the case where neighborhood choice is solely driven by unobserved neighborhood characteristics and is completely independent of race. If neighborhood-level housing supply is perfectly inelastic then any randomly driven increase (decrease) in a neighborhood's black population must be offset one for one with a decrease (increase) in its white population. Thus, a

---

[10] Note that because our neighborhoods (hexagons) are all of identical size, changes in population are equivalent to changes in population density.
[11] We also trim at the 1st and 99th percentiles of black and white head of household changes to facilitate the robustness check in Table 3.

highly inelastic housing supply will bias estimates downward towards -1 in cases where the actual causal relationship implies a value of β equal to 0. Conversely, if the supply of housing is perfectly elastic and whites and blacks are subject to the same neighborhood-specific demand shocks, on average blacks and whites would sort into neighborhoods at the same relative rates and we would expect $\beta > 0$. The exact relationship will be driven both by within city relocations and in-migration. If all population changes are driven by in-migrants, β will capture the relative increase in group populations. In our sample, for the 1920 to 1930 decade, this would imply an upwardly biased estimate of β that would be approximately equal to 2 when the true causal relationship implies β equal to 0. Finally, if supply is elastic and the neighborhood level demand shocks experienced by blacks and whites are negatively correlated, for instance due to low-income blacks being differentially attracted to low price neighborhoods that are being systematically vacated by higher income whites, then the OLS estimates will be biased downward.

Supply elasticity estimates are not available for our sample neighborhoods. However, the magnitude of population growth in our fixed-border neighborhoods (in terms of both individuals and households) suggests that housing supply was quite elastic during this period. As a result, we do not generally expect negative coefficients to arise purely as a result of supply inelasticity. Regardless, the above discussion highlights the likely problem of bias in these simple OLS regressions. Shared sorting on neighborhood characteristics will impart upward bias to OLS estimates of β (away from flight). While OLS estimates of β will be biased in a negative direction (towards flight) if black arrivals were settling in neighborhoods already being abandoned by whites either due to inelastic supply or negatively correlated tastes for other unobserved neighborhood characteristics.

To overcome this bias concern, we leverage exogenous variation in contemporary state-level black outmigration rates in combination with pre-1900 patterns of black settlement in our sample of northern cities. Particularly, we construct an instrument for ΔB_{ij}^{t1-t0} using the universe of historical census records, digitized versions of which were recently made available by Ancestry.com, to estimate black outflows from each state in each decade (1900 through 1930) and settlement patterns established by African Americans who came to the North before the Great Migration and were thus living in our sample cities by 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 through 1930 and count, for each state, 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 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 *t1*, we assign the estimated outflows according to the base year population for each cell and sum over each cell:

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

where $black\_basepop_{cgs}^{1900}$ is the national sum of all black out-migrant individuals in the cell in 1900.[12] Our instrument for ΔB_{ij}^{t1-t0} is thus $pred\_\Delta\_black_i^{t1-t0}$.

Our approach departs from much of the literature on the impact of immigration on local labor markets, where previous papers measure ***actual*** inflow rates across origin sources. Because there is no systematic data on internal migration in the United States prior to 1940, we need to instead work with ***estimated*** outflows. However, we are able to observe a rich set of characteristics of black migrants living outside their birth state, in particular year of birth and gender, enabling a close approximation to the true size of outflows in each decade. These two approaches are thus in principal very similar. Following other papers in this literature, our instrument relies on the fact that blacks departing their states of birth (primarily in 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.[13] However, we differ from this extant literature because we leverage cross-

---

[12] 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.

[13] See Grossman (1989, pp. 66-119) for a discussion of the importance of rail routes for black migration to the North.

neighborhood variation in the source states of early migrants. Because of this source state variation, we can control for baseline neighborhood-level black population in our analysis.

For our 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 illustrate the presence of variation in this dimension, Figure 5 provides city-level scatter plots showing by neighborhood the share of black men aged 20 to 29 in 1900 who were born in two exemplar pairs of source states. Panel A shows that for instance neighborhoods within Boston, Brooklyn, Chicago, 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 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, we also require 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.[14] 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

[14] These thirteen states represent between 87 and 92 percent of total black outflows in the years we study.

In this section we present the results from both the nonlinear and causal models of white flight proposed in the previous section.

A. Descriptive Evidence

Results from the nonparametric regressions are presented in Figure 7. In Panel A we present 1900 to 1910 neighborhood-level relationships between the change in the total white population (de-meaned decennial percentage change) and the baseline (t_0) black population share. The data exhibit relatively large but imprecisely measured relative declines in the white populations in neighborhoods with a 1900 black share in excess of 10 percent.[15] Beginning in the 1910 to 1920 decade, the relationship becomes clearer, increasing in both magnitude and the precision with which it is measured. On average, neighborhoods that were more than 10 percent black in 1910 experienced white population declines between 1910 and 1920 that were roughly 30 percentage points larger than that of their city's average neighborhood. In the final decade of our sample, the period of greatest segregation increase in our ten cities, the magnitude of flight appears to have accelerated even further. Between 1920 and 1930, neighborhoods that were more than 10 percent black in 1920 lost white population at a rate that was 40 percentage points higher than that of a typical city neighborhood.

B. OLS and IV Estimates of White Flight

Our descriptive evidence suggests that white flight behavior became an increasingly salient determinant of neighborhood segregation over the first three decades of the twentieth century; however, the inference is somewhat indirect. For more direct evidence we turn to a reduced form instrumental variables analysis of the presence and magnitude of white flight behavior.

[15] This lack of precision is most likely the result of the very limited number of neighborhoods in our sample with a 1900 black share in excess of 10 percent.

We begin with OLS estimation of equation (2). Results from this analysis are presented in Table 2. Here we follow the literature and consider changes in numbers rather than percentages (controlling for the city average change in white population with city fixed effects).[16] Between 1900 and 1910 we find that one black arrival has no statistically significant effect on white population dynamics. By the second decade (1910-1920), one black arrival is associated with a statistically significant .9 decline in the number of whites. This estimated relationship increases in precision and magnitude by our sample's final decade (1920-1930), with one black arrival now associated with the loss of 1.5 whites. Given the concerns about endogenity raised above, it would be inappropriate to directly interpret the OLS results for the later decades as evidence of flight behavior. However, they are suggestive. The final decade coefficient estimate is of a magnitude that exceeds that which could be explained solely through the assumption of a perfectly inelastic neighborhood-level housing supply.

To further consider these issues, we turn to the instrumental variables results also presented in Table 2. The IV estimate is -.9 and insignificant in the 1900s but grows to -1.9 in the 1910s before reaching -3.4 in the 1920s. The latter two coefficient estimates are both highly significant and in all three cases F-tests demonstrate an extremely robust first stage. 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 in the OLS regressions, likely due to positive correlation between neighborhood-level demand shocks experienced by both blacks and whites.[17]

---

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

[17] This result stands in contrast to that of Boustan (2010) ,who finds OLS coefficients that are negative in all years (1940-1970) and generally similar in magnitude to IV results from an estimation strategy similar to ours when measuring flight from the center city to the suburbs.

One potential concern with our approach is that spatial dependency across neighborhoods may cause our standard errors to be understated. Table 2 also presents standard errors computed using the GMM methodology proposed by Conley (1999) for addressing spatial clustering. The average ratio of the Conley standard error to the baseline IV standard error (estimated using LIML) is 1.57, indicating that spatial standard errors are roughly 60 percent larger than those estimated under the assumption of spatial independence. To further investigate the extent of spatial correlation in our data, we also run our specification on spatially independent subsamples, each comprising 25 percent of the overall sample. Appendix Figure I presents a visualization of a subsample for Pittsburgh.[18] In table 2 we report the results from 100 bootstraps of 25 percent spatially independent subsamples. Our coefficient estimates are essentially unchanged and, while the smaller sample size is associated with higher standard errors, they remain highly significant for the latter two decades. It is also interesting to note that if we adjust for the impact of the bootstrap sample size on standard error magnitude, both the Conley approach and the spatially independent subset approach suggest roughly the same level of attenuation in the uncorrected standard errors due to spatial dependence. Given this finding, except where noted, in the remaining analysis we report Conley standard errors.[19]

A second potential concern is the validity of our IV approach. The exogeneity of our instrument hinges on two critical assumptions. First, state-level black outmigration rates must

---

[18] These subsamples are constructed one city at a time by a simple select and reject algorithm. The algorithm randomly selects a candidate neighborhood for the subsample and tests for adjacency with the current elements of the subsample. If the candidate neighborhood is adjacent to a current subsample member it is dropped. Otherwise it is added to the sample. This process is repeated until a 25 percent subsample has been obtained.

[19] As noted above, an additional concern with our basic approach is the potential for a small number of very large population communities to drive our coefficient estimates. This concern motivates our decision to trim the sample at the 1st and 99th percentile of population. However, as a further robustness check we reran our analysis on subsets of our sample associated with the lowest quartile, highest quartile, and interquartile range of population. These results (presented in Appendix Table I) show no qualitative difference between results in the three subsamples and our results for the entire sample. The largest point estimate occurs on the interquartile subsample for the 1920 to 1930 decade, allaying concerns about our results being driven by a few highly populated neighborhoods.

not be influenced by differences in within-city cross-neighborhood pull factors that are systematically related to the origin state of early black settlers. Consider for example the fact that during the 1920s, more blacks left Virginia than Texas. It cannot be the case that this state-level differential in out-migrants arose (at least partially) because during the 1920s levels of economic opportunity were higher in Chicago neighborhoods that received large numbers of Virginian blacks before 1900 than in Chicago neighborhoods that received large numbers of Texan blacks. Second, because by construction our instrument will predict higher black population growth in neighborhoods that had relatively higher numbers of black residents in 1900, we need to generally assume that there are no systematic differences between these neighborhoods and low or no black neighborhoods that could potentially have a persistent confounding impact on migration patterns.

While we believe the first assumption to be quite defendable, the second is a potential concern. In 1900, even in those neighborhoods where they were most concentrated, blacks were generally a substantial minority. However, these neighborhoods were typically located in the urban core and hence may differ systematically in other potentially important dimensions. Fortunately, this concern is quite straight forward to address by controlling for the size of each neighborhood's 1900 black population in our IV analysis. In doing so we essentially guarantee that we are identifying the flight effect based solely on variation in the pre-1900 source state composition of these neighborhoods' black populations, independent of the overall size of their black populations.

This concern is the first issue we address in Table 3 which presents a number of robustness checks. We control for percent black in 1900 in the first set of checks and show our results are essentially unchanged (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.

As a further robustness test, we also show our results with the 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 in magnitude 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 suggesting that, as expected, 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 an exaggerated appearance of "flight" when a white family is replaced by a black family. Using the relationship to the head of household variable, we created an alternate dataset using only heads of household in the census and replicated our analysis at the household level.[20] The results from the 1920s indicate that the arrival of one black household led to the departure of 3.5 white households, strongly suggesting that differences in household composition are not driving our findings.

VI. How Important was Flight for the Rise of Segregation in U.S. Cities?

[20] 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 99th percentile of the ratio of white to white household heads as well as black to black household heads to remove these outliers in both the head of household dataset and the main dataset.

In this section we use our best causal estimates of white flight to construct a series of counterfactuals aimed at understanding how much of the observed increase in segregation over the 1900 to 1930 period can be attributed to population sorting as opposed to discriminatory institutions. We begin with a simple exercise focusing on the 1920 to 1930 decade to demonstrate the link between our coefficient estimates and the underlying population dynamics for whites and blacks. Next, we employ a range of assumptions on the sorting behavior of newly arrived black residents in each city – representing the extent of institutional barriers constraining where black families could live – and then apply our estimates to predict neighborhood-level white population changes associated with the resulting distribution of black in-migrants. This counterfactual exercise allows us to roughly decompose the relative contribution of white flight and housing market discrimination on the growth in segregation in each decade.

A. An Illustration for the 1920 to 1930 Decade

We begin with a simple exercise in Table 4 to demonstrate the link between our coefficient estimates from the instrumental variables analysis and underlying population dynamics. Focusing on the 1920 to 1930 decade, we use the complete set of coefficient estimates (i.e. including the full set of city fixed effects) to predict each neighborhood's change in white population as a function of its 1900 black share and its observed change in black population between 1920 and 1930.[21] These neighborhood level predictions are then aggregated to yield a sample-wide average.

The results for the full sample are presented in the first column of Table 4. The mean white population in 1920 across the sample is 3663 and the mean black population is 133. The

[21] We use the estimates presented in the second row of Table 3 that include controls for the percent black in 1900 as we believe this to be our most robust specification. The standard errors presented in this table are from the baseline IV specification that assumes spatial independence because of the difficulty of obtaining spatial standard errors for the smallest subsamples.

predicted average change in neighborhood white population based on our simple prediction exercise is 283 individuals. This result illustrates the fact that while neighborhoods with larger numbers of black in-migrants were losing whites relative to those with few black in-migrants, on average, across the entire sample, white populations were increasing. This relationship is captured in the city-level fixed effects.

Of course, the nonlinear regressions presented in Figure 7 suggest that the level 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 more than 10 percent black share by the 1920 to 1930 decade. We note that because we are generally seeing larger numbers of black in-migrants into these neighborhoods, this result does not necessarily require that the causal relationship between the number of black in-migrants and the number of white out-migrants differs across neighborhoods with differing black shares.

In the remaining columns of Table 4 we partition the sample by 1920 black share and rerun our specification for neighborhoods with 0 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 coefficient declines as 1920 black share increases, the implied average 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 on average 13 percent of their white population. For the two largest share black subsamples, our model predicts even larger white population losses. In particular, the -2.2 white flight coefficient for the over 20 percent black share subsample implies a loss of 37 percent of a neighborhood's white population.

B. Assessing the Relative importance of Institutional Barriers and White Departures

Finally, we leverage our empirical results to estimate the relative importance of white flight, as opposed to institutional barriers on the locational choices of black households, in explaining the observed rise in segregation over our study period. We focus exclusively on the dissimilarity measure of segregation because, unlike isolation measures, dissimilarity measures are not sensitive to proportional changes in relative population sizes. Furthermore, nearly all of the increase in dissimilarity in large cities occurred by 1930 (see Figure 1).

To identify the relative importance of white flight compared with institutional barriers, we must first identify a counterfactual baseline estimate of what segregation levels would have been if new black migrants had sorted based solely on their own preferences. In this counterfactual world, black arrivals from the South could have sorted into neighborhoods without facing institutional barriers or triggering white flight. Because of the inherent difficulty of this exercise, we produce three sets of counterfactual estimates that we believe span the range of possible outcomes.

Having established a baseline, we can compare dissimilarity measures under these "no institutions/no flight" counterfactuals to a set of "institutions/no flight" counterfactuals that hold white location choices fixed and allocate new black entrants based on the pre-existing black location choices. This comparison allows us to estimate the impact of institutions on segregation.[22] Next, using our empirical estimates of flight behavior to adjust the location choices of whites in the "institutions/no flight" counterfactual to reflect the role of white location decisions, we can measure the increase in segregation when both the barriers and flight mechanisms are in place ("institutions/flight"). Finally, we compare the "institutions/flight" outcomes from our constructed counterfactual to the actual observed level of segregation. The

[22] Another option for computing the "institutions/no flight" counterfactuals is to allocate the decadal inflow of blacks based on their actual location choices. The results presented below in Table 5 are essentially unchanged if we use this method instead of what is presented in the table.

residual from this comparison gives a sense of how well our model predicts the actual levels of segregation.

The most challenging part of this process is identifying the "no institutions/no flight" baseline. Our first approach is to allocate the net increase in each city's black population in a pattern consistent with the distribution of a European immigrant group that did not experience intense discrimination in the housing market. We choose Italians as our benchmark because this ethnic group was roughly similar in size to the black population and arrived in northern cities at approximately the same time.[23]

We consider two possible benchmark years, 1910 and 1930. Several factors lead us to conclude that 1910 likely provides an upper bound on the level of black segregation that would have arisen based solely on the preferences of black immigrants. The decade preceding 1910 represented the peak decade of Italian immigration into the United States. Unlike black immigrants from the southern United States, these recent Italian immigrants faced significant language barriers and thus had heightened incentives to locate in enclaves of native Italian speakers. Furthermore, while there is no evidence that Italian immigrants experienced housing discrimination at the level experienced by blacks, there is a large historical record suggesting that Italians experienced significant animus and ethnic prejudice during this era of mass

---

[23] To visualize the relative concentrations of blacks and Italian immigrants across these two target periods, Appendix Figure II presents the distribution of the both groups in Brooklyn (Panels A and B) and Cleveland (Panels C and D) in both 1910 and 1930. Neighborhoods are ordered according to their share of the city's respective minority population. For example, the 163rd neighborhood of Brooklyn had 3.3 percent of the city's Italian immigrants in 1910 and 3.0 percent of the city's black residents in 1930 (panels B and D show the top minority neighborhoods for each city only). These two cities are generally representative of the patterns we observe across the sample. In Brooklyn, Italians and blacks had similar distributions in 1910, but by 1930 blacks were more concentrated and Italians less so. In Cleveland, there were two black and two Italian enclave neighborhoods in 1910 that contained between 15 and 40 percent of the respective minority population. By 1930, both minorities had expanded beyond these enclaves, but blacks were still more concentrated than Italians.

immigration.[24] It is likely that this animus was associated with some forms of institutional housing restrictions. Thus, the level of Italian immigrant segregation observed in 1910 likely was above that which would have occurred based solely on the preferences of Italian immigrants.

With the rise of hostilities in Europe in 1914, the flow of Italian immigrants dropped by nearly an order of magnitude.[25] The National Immigration Acts of 1921 and 1924 served to make this reduction in immigrant flow permanent. Thus, by 1930, the vast majority of Italian immigrants in the United States had had more than a decade to assimilate, likely weakening the language-driven motivation for enclave formation. In addition, with the end of large-scale Italian immigration, the anti-immigrant imperative for anti-Italian prejudice was greatly attenuated and we can find no documented evidence of discrimination against Italians in housing markets by this point. As a result, the ethnic sorting of Italians in 1930 may provide a better benchmark for our "no institutions/no flight" counterfactual.

Thus, sorting like Italians in 1910 likely provides a reasonable upper bound for a "no institutions/no flight" black segregation counterfactual while an approach based on the groups distribution in 1930 may provide a more appropriate approximation to a true "no institutions/no flight" counterfactual. Finally, our third "no institutions/no flight" provides a lower bound by considering the segregation that would have occurred if all recent (over the previous ten years) black in-migrants sorted into neighborhoods in a way that reflected the pre-existing distribution

---

[24] As an example, in their 1947 evaluation of racial covenants on properties in St. Louis and Chicago, Long and Johnson present no evidence that Italian heritage was ever included as a condition for denying the transfer of a deed. For a detailed overview of anti-Italian animus, see ***Wop!: A Documentary History of Anti-Italian Discrimination in the United States*** by Salvatore John LaGumina.

[25] During the five year period from 1910 to 1914, 1.1 million Italians immigrated to the United States. Over the following five years (1915-1919), only around 125,000 Italians immigrated to the United States. Source: U.S. Department of Commerce, Bureau of the Census, A Statistical Abstract Supplement, Historical Statistics of the United States from Colonial Times to 1957, pp. 56-57.

of the entire population, making no distinction by race or ethnicity. Further details on how the counterfactuals were constructed can be found in the appendix.[26]

We present a summary of the dissimilarity results obtained from our counterfactual exercise in Table 5. Panel A presents counterfactual estimates of dissimilarity under each of our constructed scenarios. Actual dissimilarity increased from .532 to .666 between 1910 and 1930 in the sample, with the largest increase occurring over the 1920s. Our preferred approach to estimating the "no institutions/no flight" baseline for segregation is presented in the first column of Panel A (assigning black inflows to match Italian settlement in 1930). Comparing the three "no institutions/no flight" counterfactuals to the "institutions/no flight" counterfactual allows us to estimate the contribution of institutions that constrained where blacks could live to the growth in segregation over each decade. These estimates are presented in the first three columns of Panel B. Comparing the "institutions/no flight" and "institutions/flight" counterfactuals allows us to estimate the contribution of white flight (presented in the fourth column of Panel B).

Focusing on our preferred baseline, the most striking finding is the sharp increase in the contribution of flight in each subsequent decade (presented in Panel C). While the counterfactual results suggest that the flight effect was relatively small during the aughts, we estimate that flight was responsible for 34 percent of the increase in segregation (as measured by dissimilarity) in our model over the 1910s with institutions responsible for 66 percent of the

[26] We note that one potential concern is that by taking the previous decade's level of segregation as fixed and then building our counterfactuals based solely on the sorting of new in-migrants (and white responses to these new in-migration) we may have biased our baseline counterfactuals ("no institutions/no flight") upward (towards finding higher levels of segregation). However, this concern is mitigated by the following three factors. First, these new migrants make up a substantial portion of the overall black population (well over 50 percent over the critical 1920s). Second, given the rapid rise in segregation observed over each decade in our sample, bias imparted by producing a baseline distribution that incorporates the location decisions of blacks that were made in earlier decades will be contaminated by much lower levels of institutional constraints. Lastly, to the extent that a bias survives these first two points, it will be embedded in all three types of counterfactuals and should wash out in relative comparisons between the role of flight and institutional barriers.

total.[27] Over the 1920s, the decade of greatest increase in segregation, white flight was responsible for 50 percent of the increase. The residual, presented in the fifth column of Panel B, represents the difference between the observed level of segregation and our prediction. It is negligible for the 1930 decade. The residual is larger in the earlier two decades, particularly so in the 1910s, suggesting the emergence of new forms of discrimination in the housing market such as bombings or attempts at racial zoning ordinances that are not captured in our model. Institutions, on the other hand, made declining relative contributions to segregation in our baseline counterfactual over the 1900 to 1930 period, ranging from 73 percent in the 1900s to 50 percent in the 1920s.

As discussed above, assigning black inflows to match Italian settlement in 1910 provides a lower bound for the institutional effect and likely overstates the amount of segregation that would have arisen solely as a consequence of black preferences. Accordingly the results under this baseline, presented in the second column of Panel B, find that institutions played a very small role. Under this assumption, white flight explains at least 75 percent of observed segregation in each of the three decades. At the other extreme, assigning black inflows to match the overall population distribution arguably provides an upper bound on the role of institutions. However, even under this conservative approach where we essentially assume that blacks had no true preference for living near one another, white flight still is predicted to account for 23 percent of the rise in segregation during the 1920s (column 3 of Panel C).

The results from this counterfactual exercise demonstrate that decentralized sorting behavior by whites had a quantitatively important and increasing impact on the rise of residential segregation between 1900 and 1930. Our findings suggest that the transition from institutional

[27] The calculation for the role of flight over the 1910s is .026/(.026+.051) = .34. The calculation for the 1920s is analogous.

barriers to white flight as the driving force behind segregation in U.S. cities began several decades earlier than previously thought. Although the Fair Housing Act and other legislative and legal remedies have greatly reduced (without fully eliminating) the barriers faced by blacks in the housing market, white flight from black neighbors is an individual behavior that cannot be limited by local or federal government agencies. Thus, a key takeaway from this exercise is that segregation could have emerged even in the absence of discriminatory barriers in the housing market through the mechanism of population sorting.

VII. Conclusion

In this paper we asked why racial segregation emerged in American cities, providing the first empirical analysis of white flight and its role in the emergence of the black ghetto. Leveraging a new dataset, our empirical analysis identified 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. We asked to what extent white departures in response to black arrivals can account for the rise of segregation in American cities. Because restrictive covenants and racial zoning ordinances are no longer legal and racial violence and housing discrimination are less severe in the present day, our analysis to some extent investigates whether segregation could have emerged in the current institutional and legal environment.

Our analysis suggests 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 three white individuals. The robustness of these findings and the way in which they vary across time suggests that changes in white animus were 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.

An important question raised by the findings of this paper is what led to the accelerated white flight effect observed over the 1900 to 1930 period. Moving forward, understanding why white Americans fled black neighbors at increasing rates and where they settled subsequently is crucial to understanding why American cities became and remain sharply segregated by race. The failure of racial zoning ordinances and the expectation of continued migration of blacks to northern cities coupled with improvements in urban transit infrastructure are explanations that warrant further investigation.

BIBLIOGRAPHY

Altonji, J. and D. Card. “The Effects of Immigration on the Labor Market Outcomes of Less-skilled Natives. Immigration”, ***Trade and the Labor Market***. J. Abowd and R. Freeman. Chicago, University of Chicago Press, 1991, pp. 201-234.

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.

Bayer, Patrick, Fernando Ferreira, Robert McMillan. “A Unified Framework for Measuring Preferences for Schools and Neighborhoods,” Journal of Political Economy, 115(4), 2007, pp. 588-638.

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. “Racial Residential Segregation in American Cities,” in the *Handbook of Urban Economics and Planning*, eds. Nancy Brooks, Kieran Donaghy, and Gerrit Knaap. Oxford University Press, 2011.

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. "Is the new immigration really so bad?" *The Economic Journal*, 2005 pp. 300-323.

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.

Conley, Timothy G. "GMM Estimation with Cross Sectional Dependence." *Journal of Econometrics,* 92(1), 1999, pp. 1-45.

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.

Kucheva, Yana and Richard Sander. "The Misunderstood Consequences of *Shelley v. Kraemer*." *Social Science Research*, 48, 2014, pp. 212-233.

LaGumina, Salvatore John. *Wop!: A Documentary History of Anti-Italian Discrimination in the United States*. No. 32. Guernica Editions, 1999.

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.

Logan, Trevon and John Parman. "The National Rise in Residential Segregation." NBER Working Paper 20934, February 2015.

Long, Herman H. and Charles Johnson. *People vs. Property: Race Restrictive Covenants in Housing.* Nashville: Fisk University Press, 1947.

Massey, Douglas S. and Nancy A. Denton. *American Apartheid: Segregation and the Making of the Underclass.* Cambridge: Harvard University Press, 1993.

Peri, Givanni and Chad Sparber. "Assessing inherent model bias: An application to native displacement in response to immigration." Journal of Urban Economics, Volume 69, Issue 1, 2010, pp. 82-91.

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.

Shertzer, Allison, Randall P. Walsh, and John R. Logan. "Segregation and Neighborhood Change in Northern Cities: New Historical GIS Data from 1900 to 1930." 2016. *Historical Methods*, forthcoming.

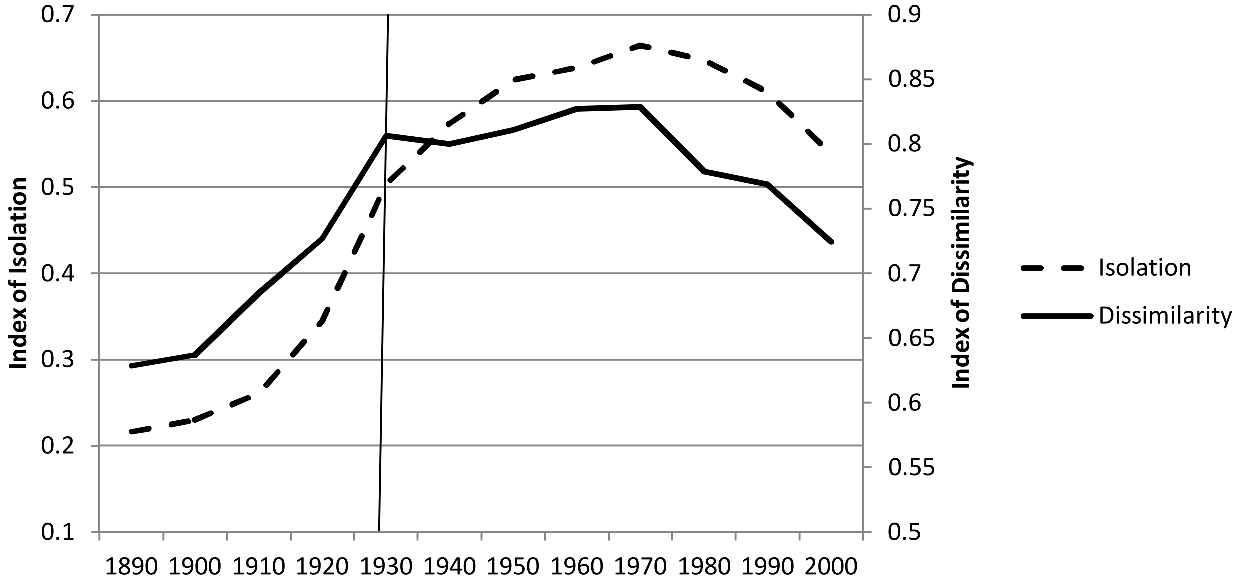
Willcox, Walter F. *Statistics of Migrations, National Tables, United States*. 1929. Retrieved from http://www.nber.org/chapters/c5134.

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

Wozniak, Abagail and Thomas J. Murray. "Timing is everything: Short-run population impacts of immigration in US cities." *Journal of Urban Economics*, Volume 72, Issue 1, July 2012, Pages 60-78.

Wright, Richard, Ellis Mark and Reibel Michael "The linkage between immigration and internal migration in large metropolitan areas in the United States," *Economic Geography*, 73 (2) (1997), pp. 234–254

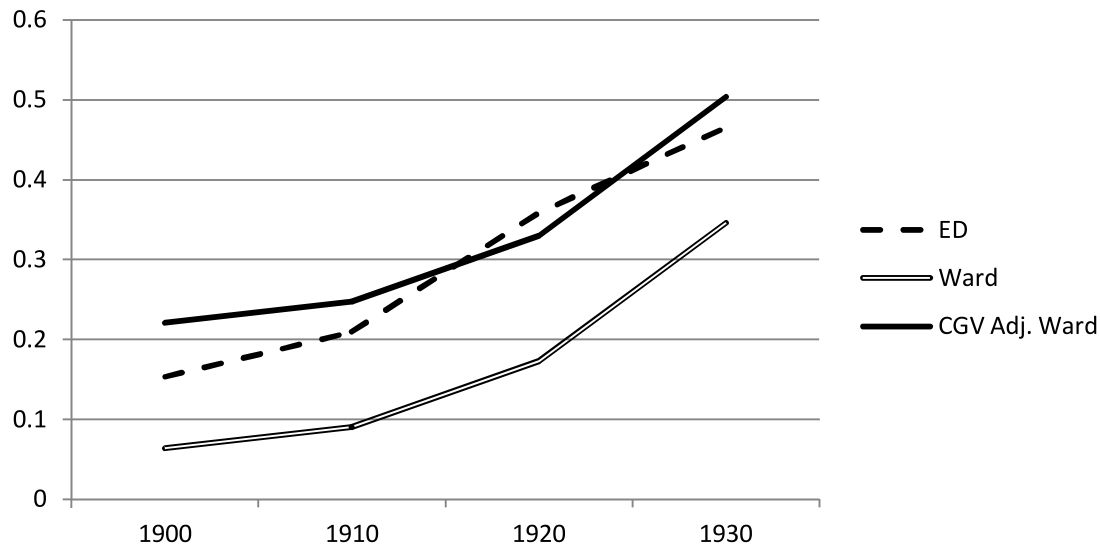
Figure 1. Segregation Trends in the Largest Ten American Cities, 1890-2000



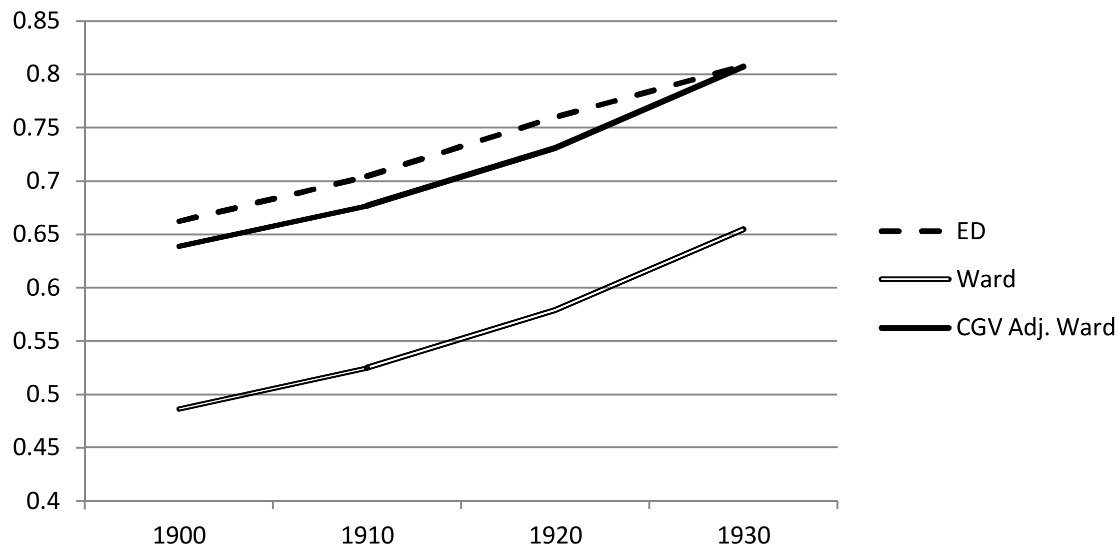
Notes: Data are taken from the dataset used in Cutler, Glaeser, and Vidgor (1999) and show the average segregation indices across Baltimore, Boston, Brooklyn, Chicago, Cincinnati, Cleveland, Detroit, Manhattan, Philadelphia, Pittsburgh, and St. Louis. We employ their adjustment factor to make the ward-level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

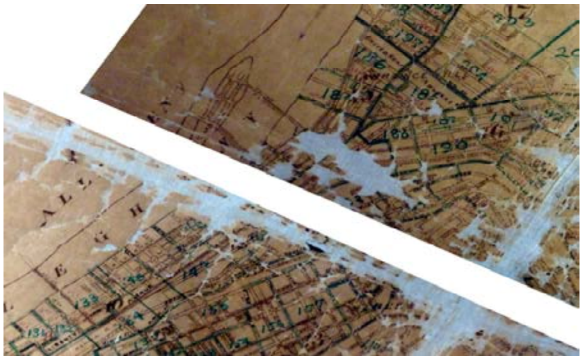


B. Dissimilarity



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

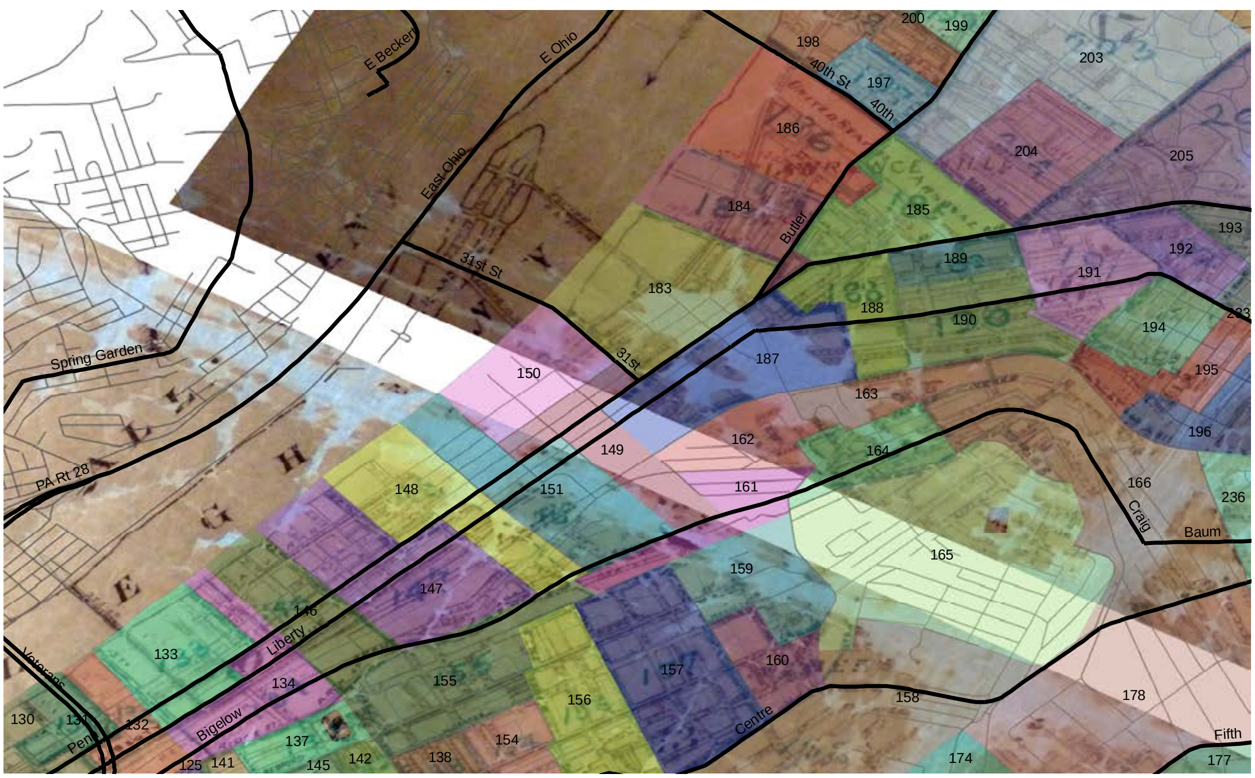


A. Enumeration District Map

B. Digitized Street Map

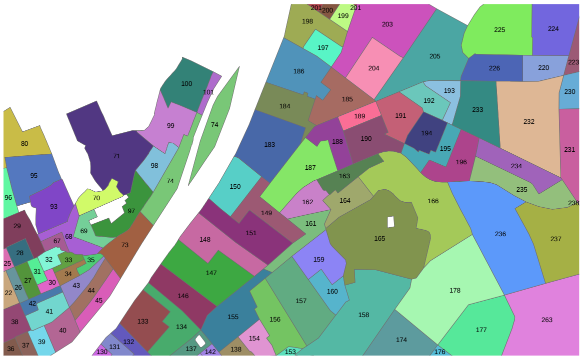
150	Pittsburg City, 12th Ward, Pct 5, bounded by Allegheny River, 31st, Smallman, 28th

C. Enumeration District Descriptions

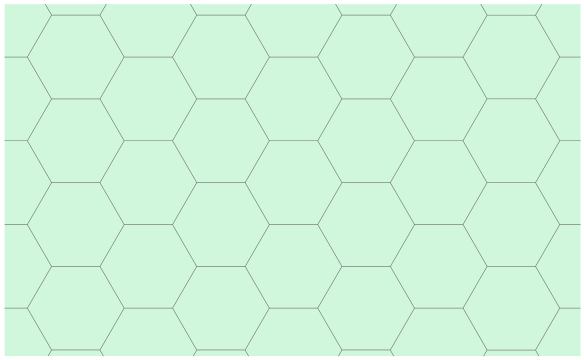


D. Digitized Enumeration District Map (ArcMap)

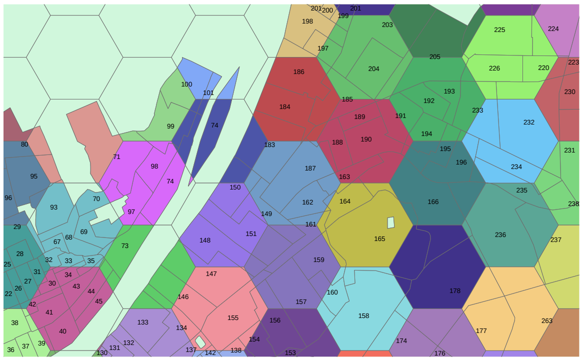
Figure 4. Constructing Hexagon Neighborhoods from Enumeration District Maps



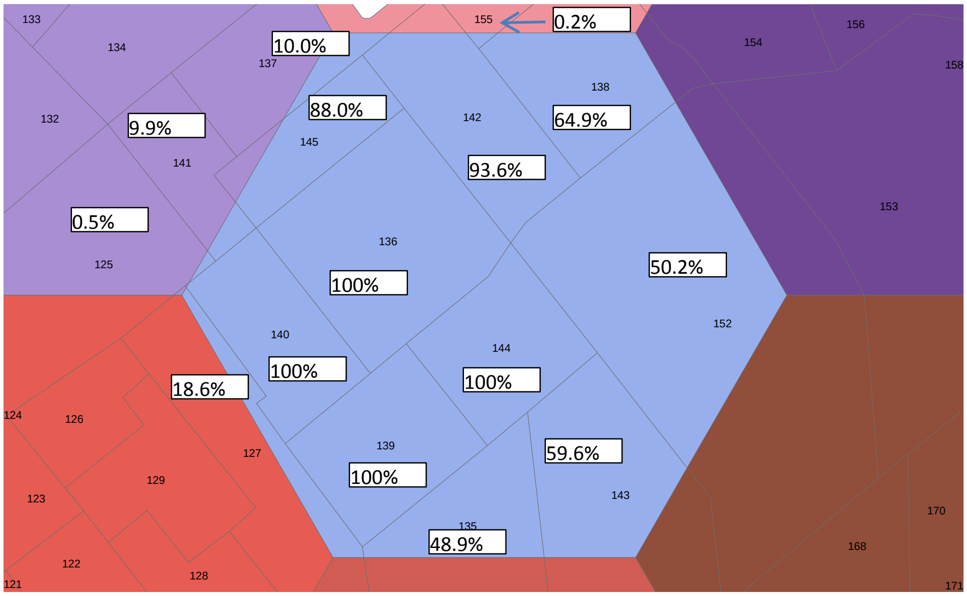
A. Enumeration District Map (1900)



B. Hexagon Grid (Constant across Decades)



C. Intersection between Enumeration Districts and Hexagons

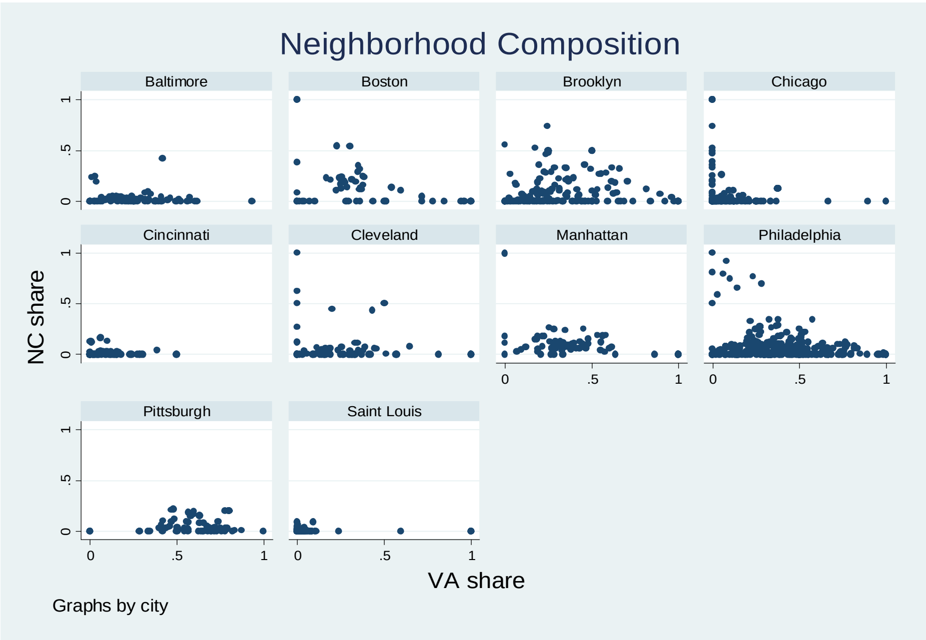


D. Allocating Enumeration District Count Data to Hexagon Neighborhoods

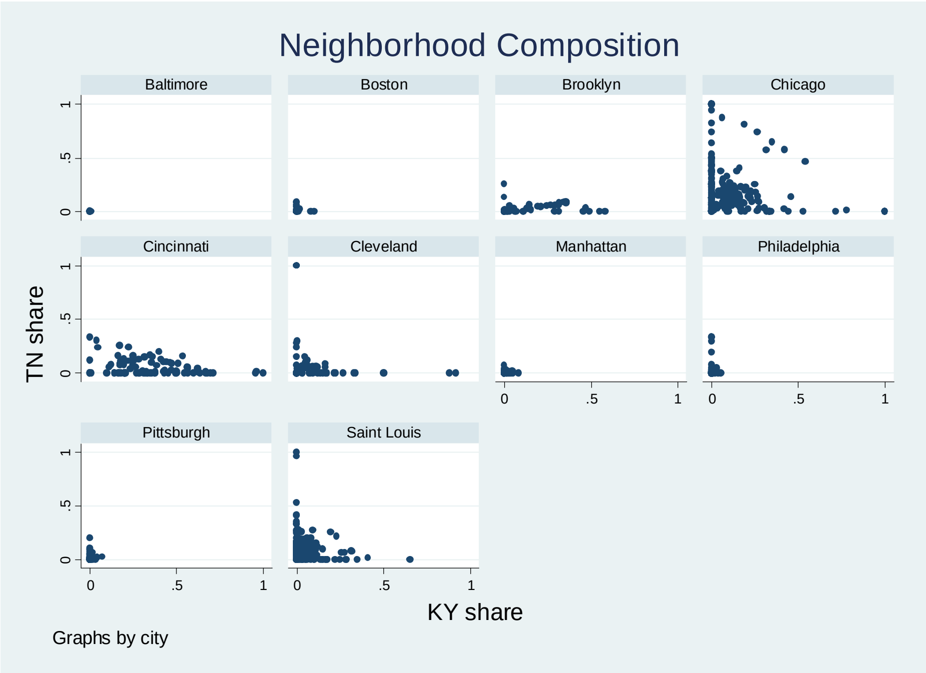
Notes: see Section III for details on the source of the maps and street files used to construct these images.

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

A. Virginia versus North Carolina

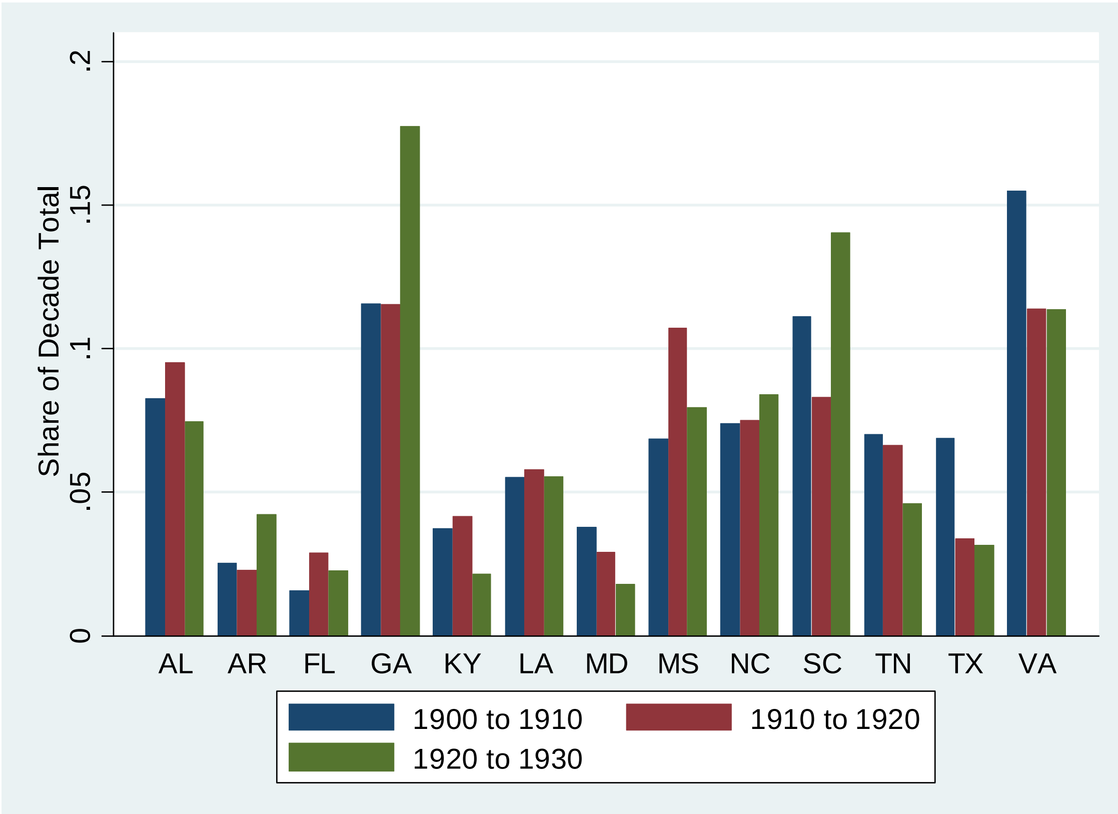


B. Kentucky vs Tennessee



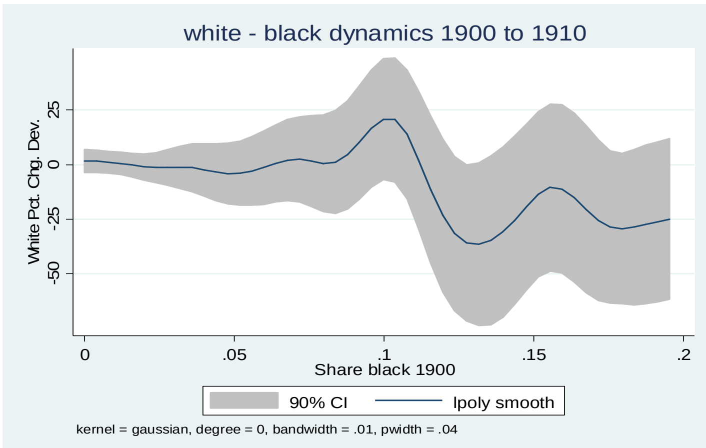
Notes: Scatterplots show the share of black men aged 20 to 29 born in each source state out of the total number of black men in the cohort in neighborhood. The shares are computed using the universe of census records with enumeration district identifiers from each city and the hexagon imputation strategy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

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

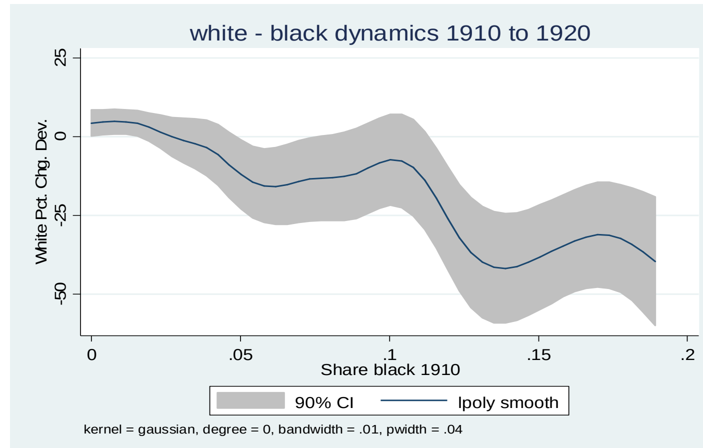


Notes: The data in this figure come from the universe of census microdata made available by Ancestry.com. Estimated outflows are computed by summing the change in the number of individuals in gender, state of birth, and birth cohort cells appearing outside their birth state in each census year.

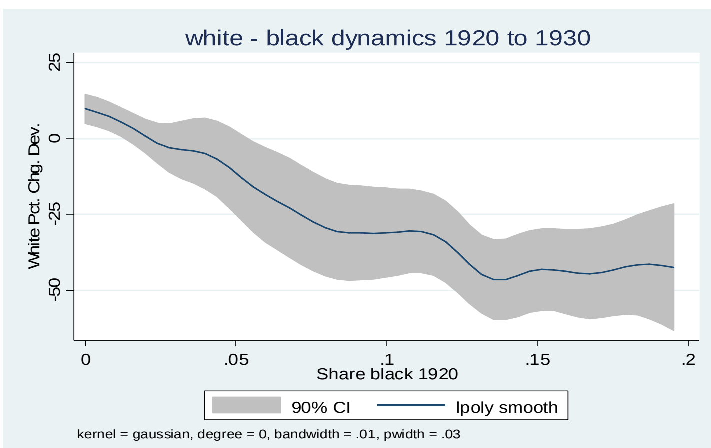
Figure 7: Panel A. Black and White Population Dynamics



Panel A. 1900-1910



Panel B. 1910-1920



Panel C. 1920-1930

Notes**:** Notes: All figures show the nonparametric relationship between share black and white population changes in the neighborhood over the next decade. All white population changes are de-meaned (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.

Table 1. Summary Statistics for Hexagon Panel Dataset

	1900	1910	1920	1930
Black Percent	2.24	2.25	2.74	4.54
	(3.86)	(4.28)	(6.45)	(11.78)
White 3rd Generation Percent	36.31	37.06	39.87	41.47
	(16.65)	16.74	(18.22)	(18.91)
White Second-Generation Percent	34.00	34.09	33.09	32.42
	(9.95)	(9.10)	(9.39)	(10.99)
White First-Generation Population	26.12	26.20	23.46	21.49
	(10.11)	(11.64)	(11.00)	(10.55)
Population	2504	3160	3802	4216
	(3857)	(4239)	(4343)	(3874)
Decadal Change in White Population		650.36	590.66	282.60
		(1147.63)	(1259.64)	(1741.58)
Decadal Change in Black Population		20.54	48.83	118.32
		(51.62)	(172.30)	(190.35)
Decadal Change in White 3rd Generation Population		206.60	323.05	186.10
		(484.36)	(540.35)	(657.94)
Decadal Change in White Second-Generation Population		217.03	172.87	121.90
		(470.84)	(503.40)	(696.04)
Decadal Change in White First-Generation Population		228.15	69.25	29.08
		(545.40)	(539.00)	(717.29)

Notes: 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 1st and 99th percentile of both white and black population change for each decade. We also trim at the 99th percentile of the ratio of white to white household heads and black to black household heads. The statistics presented cover the balanced panel of 1,975 hexagon neighborhoods that remain after these trims.

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
OLS Results			
Change in Black Population	0.189	-0.908***	-1.492***
	(0.264)	(0.122)	(0.075)
R-squared	0.088	0.139	0.258
IV Results			
Change in Black Population	-0.936	-1.886***	-3.389***
LIML Standard Errors	(0.577)	(0.227)	(0.246)
Conley GMM Spatial Standard Errors	(0.719)	(0.238)	(0.386)
Change in Black Population: Spatial Subsample	-0.871	-1.956***	-3.550***
Bootstrapped Standard Errors	(1.178)	(0.368)	(0.805)
First Stage			
Predicted Change in Black Pop.	0.918***	0.732***	0.878***
	(0.040)	(0.025)	(0.053)
F-test on First Stage	520.2	829.0	275.9
Observations	1,975	1,975	1,975

Notes: See Table 1 for sample and variable details. All regressions include city fixed effects. The instrumental variables regressions are estimated using limited information maximum likelihood estimation (LIML). The Conley (1999) spatial standard errors are estimated using GMM. The spatial subsample standard errors are generated using 25 percent spatially independent subsamples bootstrapped 100 times.

Table 3. 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	-0.936	-1.886***	-3.389***
(baseline)	(0.719)	(0.238)	(0.386)
Change in Black Population	0.703	-1.877***	-3.883***
	(0.939)	(0.379)	(0.554)
Percent Black in 1900	-41.15***	-0.556	39.89*
	(15.256)	(13.901)	(23.113)
Change in Black Population		-1.399*	-2.910***
		(0.906)	(0.644)
Number of Blacks in 1900		-0.249	-0.343
		(0.388)	(0.358)
Change in Black Population		-1.889***	-3.429***
		(0.314)	(0.524)
Percent Black in 1900		12.94	46.49**
		(10.828)	(23.895)
Pre-Trend in White Population		0.373***	0.389***
		(0.058)	(0.052)
Southern states IV	-0.749	-2.605***	-3.947***
	(1.437)	(0.561)	(0.636)
Observations	1,975	1,975	1,975

	dependent variable = change in white households		
	1900-1910 Decade	1910-1920 Decade	1920-1930 Decade
Change in Black Households	-0.625	-0.925***	-3.472***
	(0.859)	(0.178)	(0.482)
Observations	1,975	1,975	1,975

Notes: see Table 2 for sample and specification details. For the southern states IV only black outflows from Alabama, Arkansas, Florida, Georgia, Louisiana, Mississippi, North Carolina, South Carolina, Tennessee, Texas, and Virginia are used. Spatial standard errors are reported for all specifications.

Table 4. White Flight by Neighborhood Type

	1920 Black Share				
	Full	0-5%	5-10%	10-20%	>20%
Coefficient on black difference, 1920-1930	-3.389***	-7.632***	-4.435***	-3.887***	-2.159***
Standard error	(0.246)	(0.935)	(1.291)	(1.143)	(0.328)
Mean white population in 1920	3663	3632	3846	3560	4397
Mean black population in 1920	133	28	298	595	2138
Mean change in black population, 1920-1930	118	51	363	485	904
Implied change in white population	283	470	-506	-731	-1622
Implied percent change in white population	0.08	0.13	-0.13	-0.21	-0.37
N	1,975	1,680	134	109	52

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). The implied change in white population is predicted from the regression on each subsample.

Table 5. Role of White Flight and Institutions in Determining Segregation Growth

Panel A. Counterfactual Dissimilarity Levels

	No Institution No Flight Counterfactual			**Institutions No Flight Counterfactual**	**Institutions Flight Counterfactual**	**Actual Level of Dissimilarity**
	Basis for Counterfactual					
	$Italians_{1930}$	$Italians_{1910}$	Gen. Population			
1930	0.512	0.577	0.330	0.587	0.664	0.666
1920	0.479	0.534	0.326	0.530	0.556	0.587
1910	0.448	0.491	0.353	0.497	0.514	0.532

Panel B. Counterfactual estimates of Flight and Institution Effects

	Institution Effect			**Flight Effect**	**Residual**
	Basis for No-Institutions Counterfactual				
	$Italians_{1930}$	$Italians_{1910}$	Gen. Population		
1930	0.075	0.010	0.257	0.076	0.003
1920	0.051	-0.004	0.204	0.026	0.031
1910	0.048	0.005	0.144	0.018	0.018

Panel C. Flight Share in Counterfactual Dissimilarity

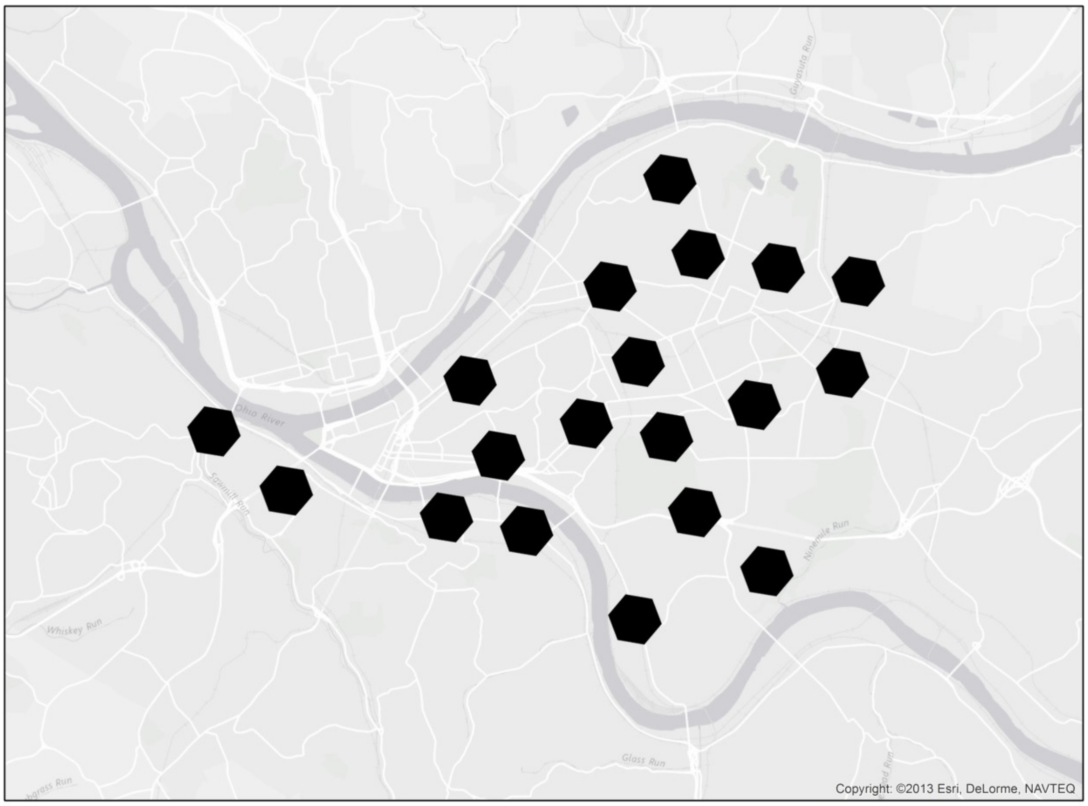
	Basis for No-Institutions Counterfactual		
	$Italians_{1930}$	$Italians_{1910}$	Gen. Population
1930	0.504	0.886	0.229
1920	0.339	1.195	0.113
1910	0.268	0.767	0.109

Notes: see the appendix for details on how each counterfactual was constructed. The “institution effect” in Panel B is the difference between the respective “no institutions/no flight” and “institutions/no flight” counterfactuals presented in Panel A. The “flight effect” is the difference between the “institutions/no flight” and “institutions/flight” counterfactuals. The residual is the difference between the “institutions/flight” counterfactual and the actual level of dissimilarity in the panel dataset. The flight share in Panel C is the share of segregation in the model explained by flight as a share of segregation explained by either flight or institutions.

Appendix

Supplemental Figures and Tables

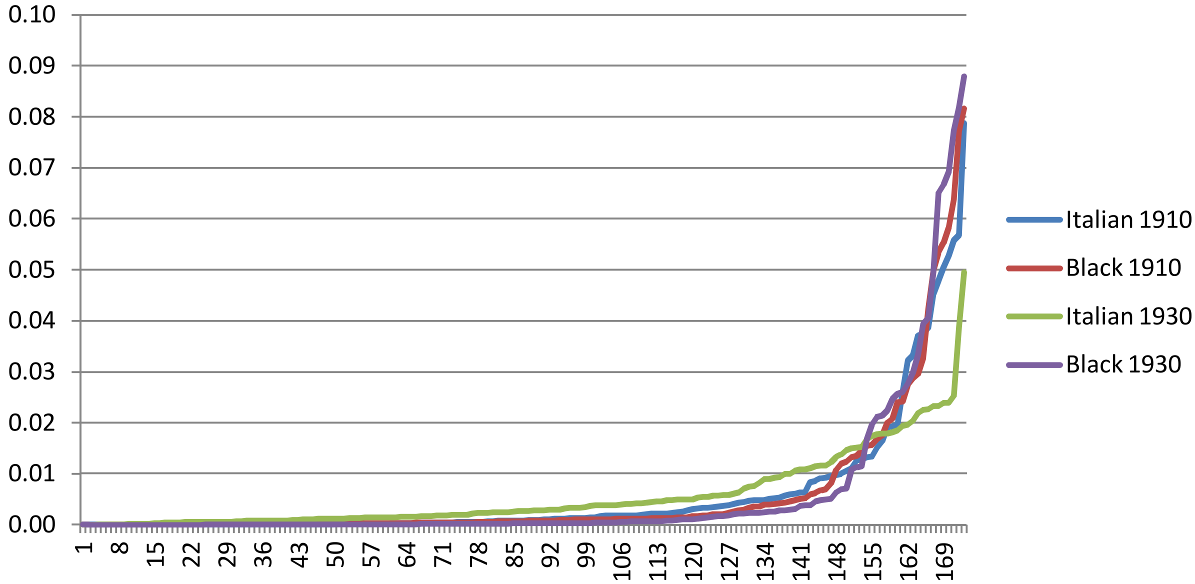
Appendix Figure I. Spatial Subsample for Pittsburgh



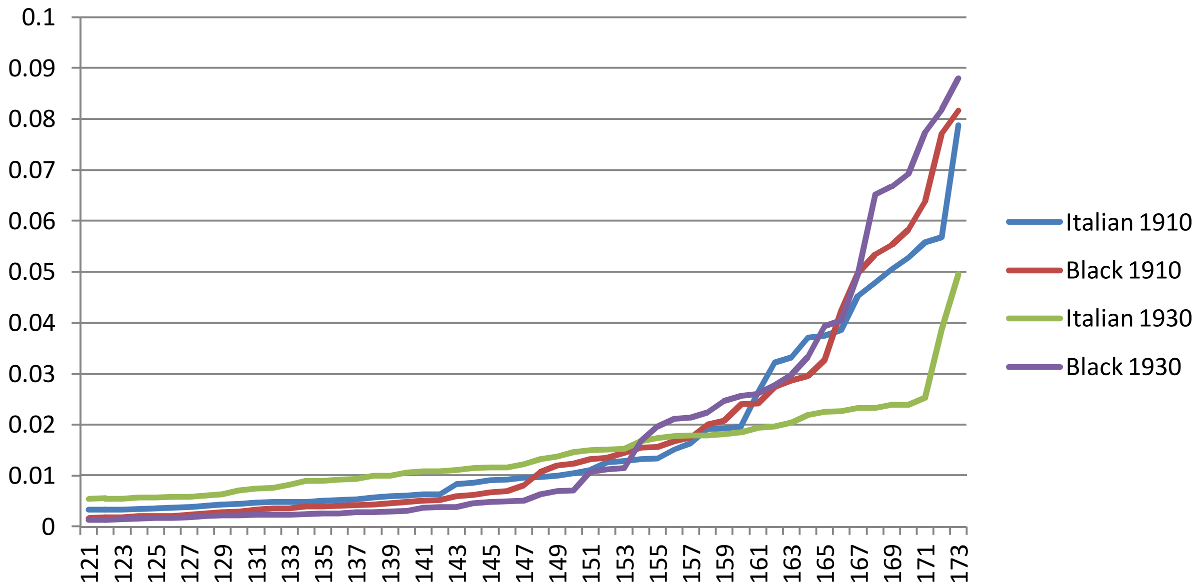
Notes: This image illustrates an independent spatial subsample comprising 25 percent of the overall sample for the city of Pittsburgh.

Appendix Figure II. Black and Italian Population Distributions

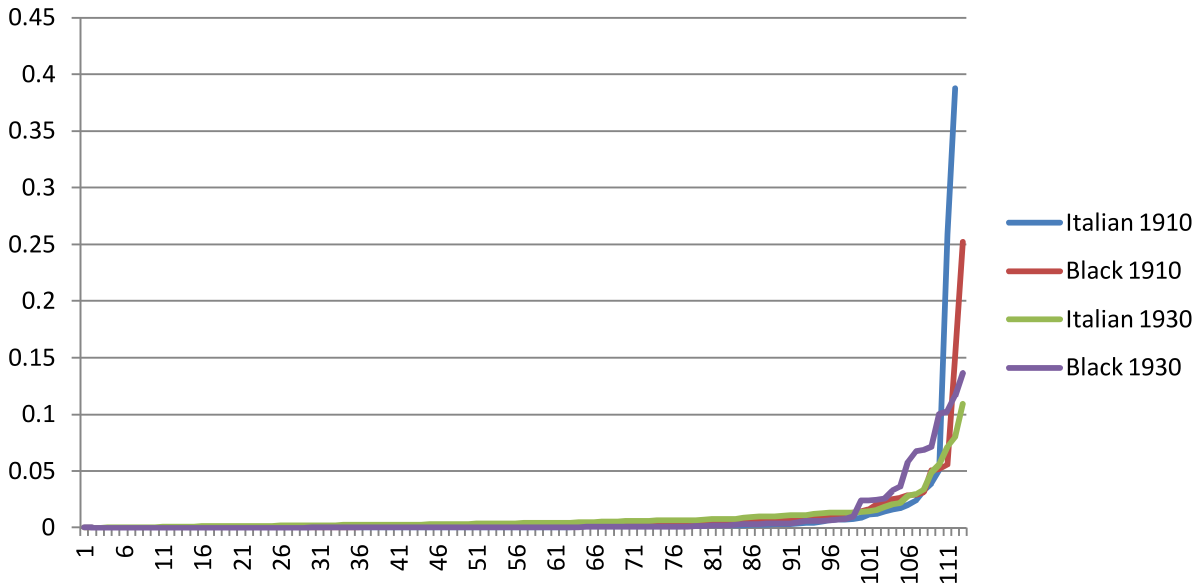
A. Brooklyn: Full Distribution



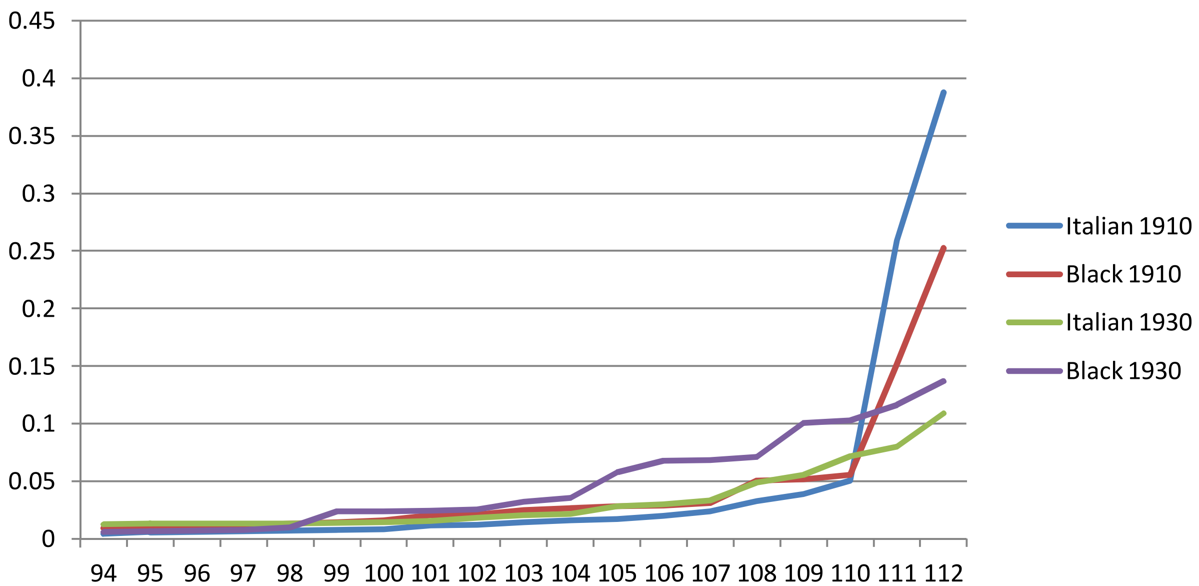
B. Brooklyn: Top Minority Neighborhoods



C. Cleveland: Full Distribution



D. Cleveland: Top Minority Neighborhoods



Notes: These figures show the distribution of the black and Italian populations in Brooklyn (Panels A and B) and Cleveland (Panels C and D) in both 1910 and 1930. Neighborhoods are ordered according to their share of the city's respective minority populations.

Appendix Table I

	Full Population	Percentile of Baseline Neighborhood Population		
		1st - 25th	25th - 75th	75th - 99th
1900-1910 Decade	0.703	2.766	-13.05	-3.257***
Change in Black Population	(0.939)	(10.37)	(9.348)	(0.875)
1910-1920 Decade	-1.877***	318.2	-1.198*	-2.096***
Change in Black Population	(0.379)	(14,345)	(0.712)	(0.396)
1920-1930 Decade	-3.883***	3.719	-4.988***	-2.389***
Change in Black Population	(0.554)	(5.928)	(1.055)	(0.414)
Observations	1975	494	990	494

Notes: See Table 1 for sample and variable details. All regressions include city fixed effects. The instrumental variables regressions are estimated using limited information maximum likelihood estimation (LIML). We present standard errors that have not been corrected for spatial correlation in this table because we did not have enough power for the Conley method on these subsamples.

Construction of the Segregation Counterfactuals

The details of our approach are as follows:

Step 1: Establish sample
We use the consistent panel of neighborhoods that appear in each decade from the primary dataset. This balanced panel has 1,975 neighborhoods across the ten cities in our sample. We use the actual 1900 neighborhood populations of both blacks and whites as fixed.

Step 2: Construct the black population distribution for the "no institutions" counterfactuals
We allocate observed city-level black population growth over the 1900 to 1910 decade to neighborhoods according to match that of the three allocation rules to model the presence of institutional barriers. These allocations are by Italians in 1930 (our baseline), by Italians in 1910 (our upper bound for preference-induced segregation), and by the total population (our lower bound for preference-induced segregation. These allocations are added to the base year populations for 1900. As an example, consider the case of the Italian 1930 counterfactual for Brooklyn. 4.9% of all Italian immigrants living in Brooklyn in 1930 were in the single neighborhood (hexagon) with the largest Italian population. Brooklyn received 5,502 new blacks between 1900 and 1910, so we allocate 4.9% of them or 270 of these black individuals to the neighborhood with highest 1900 total black population. Next we move to the neighborhood with the second-highest total black population. In 1930, 3.9% of all Brooklyn Italians lived in the neighborhood with the second-highest total population of Italians. So, we allocate 3.9% of the new black arrivals to the neighborhood with the second-highest number of blacks in 1900. We continue this process until all new black arrivals have been allocated and then repeat the process for the nine other cities in our sample.

Step 3: Construct the black population distribution for the "institutions" counterfactuals

To model the presence of discriminatory institutions, for this counterfactual we allocate observed city-level black population growth over the 1900 to 1910 decade to match black settlement at the start of the decade. For instance, a neighborhood that had 5 percent of the city's black population in 1900 would be assigned 5 percent of black population growth over the 1900 to 1910 decade.

Step 4: Construct the white population distribution for the "no flight" counterfactuals

We 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 5: Construct the white population distribution for the "flight" counterfactual

We apply the baseline instrumented equation (2) estimates to generate the predicted white population change over the 1900 to 1910 decade in each neighborhood.[28] This prediction takes into account both city-level fixed effects and the coefficient on the change in black population.

Step 6: Compute the counterfactual dissimilarity indices

We use the respective "no institutions" hypothetical black population from Step 2 and "no flight" hypothetical white population from Step 4 and compute the dissimilarity index to obtain the "no institutions/no flight" counterfactuals reported in Table 5. Next, we use the respective "institutions" hypothetical black population from Step 3 and "no flight" hypothetical white population from Step 4 and compute the dissimilarity index to obtain the "institutions/no flight" counterfactuals. Finally, we use the respective "institutions" hypothetical black population from Step 3 and "flight" hypothetical white population from Step 5 and compute the dissimilarity index to obtain the "institutions/flight" counterfactuals.

We proceed analogously for the 1910 to 1920 period and the 1920 to 1930 period to construct the rest of the counterfactual segregation levels in Table 5.

[28] We note that this approach allows the predicted white population to differ from the actual white population by some residual. We experimented with ways to assign the residual so that the predicted and actual white populations were the same, for instance by assigning the difference to neighborhoods proportionally according to their share of the city's population at the start of the respective decade. Our results were qualitatively unchanged. We ignore the residual in our baseline approach because assigning the residual effectively redistributes white population into neighborhoods with black residents, thus attenuating the increase in segregation that is generated in our counterfactual.