The Dynamics of White Flight and Segregation in Urbanizing America

Allison Shertzer (University of Pittsburgh and NBER)
Randall P. Walsh (University of Pittsburgh and NBER)

[preliminary and incomplete, please do not cite]

July 2014

Abstract: Residential segregation by race first emerged in American cities in the early twentieth century when millions of black migrants from the South arrived in northern urban areas. In this paper we use newly assembled, fine resolution spatial data on city populations in prewar America to investigate why black-white segregation rose so sharply between 1900 and 1930, focusing on the role of decentralized white location decisions. Employing both nonlinear “tipping” and linear “flight” empirical approaches, we show that white departures in response to black arrivals were quantitatively large and accelerated over the early twentieth century. In contrast to the existing literature, which emphasizes centralized government policies as the driver of segregation, our results suggest that uncoordinated behaviors appear to have been a key mechanism behind the development of racially segregated cities in the United States.
I. Introduction

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

Decentralized actions by white homeowners have, on the other hand, been credited with reinforcing racial segregation later in twentieth century, particularly in the postwar era when highways were constructed that swiftly connected the central city to the newly opened suburbs (Baum-Snow, 2007). The existence of white “flight” from black arrivals has been rigorously documented in this period, and several studies have shown that the willingness of white individuals to depart neighborhoods with rising black populations and move to those with few black residents is a quantitatively important mechanism through which racial segregation has been perpetuated (Card, Mas, and Rothstein, 2008; Boustan, 2010). In contrast, the extant literature has not addressed the role that individual residential decisions played in the emergence
of segregation in the prewar era, a time period for which little detailed spatial demographic data previously existed.

In this paper we return to the question of why racial segregation emerged in American cities, revisiting in particular the scholarly consensus that racist institutions were essential for constructing the ghetto. We focus our empirical analysis on the residential response of native white individuals to the initial influx of rural blacks into the industrial cities of the North on the eve of the First World War, asking to what extent white departures in response to black arrivals can account for the particularly rapid increase in segregation prior to 1940. Our approach can thus be thought of as separately identifying the role of decentralized locational decisions from the institutions that have been the focus of previous work. Because restrictive covenants and racial zoning ordinances are illegal and racial violence and housing discrimination less severe in the present day, our analysis to some extent answers the question of whether segregation could have emerged in the current institutional and legal environment.

To undertake this study we required a fine-grained, spatially-identified demographic dataset with neighborhood comparability over time for the early twentieth century. We constructed the first such dataset using 100 percent of the population of ten major American cities in 1880, 1900, 1910, 1920, and 1930 using census microdata accessed from the genealogy website Ancestry.com. This dataset has several key advantages over previously existing sources of data on prewar urban populations in the United States. First, our base geographic unit is a census enumeration district, a small administrative unit used internally by the census. The average enumeration district in our sample contains just 1,400 individuals, so these units are smaller than contemporary census tracts. Second, we used the enumeration district data to construct a set of neighborhoods with consistent borders that are comparable over the half
century spanning 1880 to 1930, making our study the first capable of systematically examining the process of neighborhood demographic change over this period. Third, the expiration of census confidentiality rules allow us to observe age, race, place of birth, parents’ place of birth, and year of immigration (of the foreign born) for each of the approximately 60 million individuals in our dataset. This rich dataset allows us to make counts of specific demographic groups in small geographic areas, facilitating a rich array of empirical investigation.

We use two strategies to test for the existence of “white flight” from neighborhoods in our cities over the 1900 to 1930 decades. The primary difficulty in identifying this effect is that minorities do not exogenously arrive in neighborhoods, and we are particularly concerned that blacks and immigrants may locate in locations already being abandoned by whites. We use two different strategies to address this concern. First, we search for nonlinearities in white population share as a function of black population share that could indicate “tipping” behavior in the data. Second, we develop an instrumental variable for black population size in neighborhoods in the spirit of the immigration shock literature.

Both strategies provide evidence of white flight from blacks in the early twentieth century; moreover, the flight effect appears to accelerate over the three decades we study. We find robust evidence of nonlinear declines in white population consistent with tipping: loss of white population in neighborhoods is discontinuous between 10 and 15 percent black share in every decade and city. Neighborhoods with black share above the tipping point lost about 35 percent more whites (relative to the city average) than neighborhoods with few blacks in the 1910s. By the 1920s these neighborhoods with black share above the discontinuity lost about 50 percent more white population.
While the relationships between black share and white population share dynamics were nonlinear, they were also monotonic. In our second empirical approach we relate changes in black populations to changes in white populations, again considering each decade separately. The OLS results suggest one black arrival was associated with one white departure, a finding that would not indicate racial animus in location decisions. However, our instrumental variables strategy, which assigns estimated black outflows from southern states to northern cites according to black settlement patterns prior to the Great Migration, indicates that OLS estimates were likely biased against a finding of flight due to black settlement in generally growing neighborhoods. Our IV results indicate that one black arrival was associated with 1.8 white departures in the 1900s but 2.5 white departures by the 1920s. We show that part of the acceleration of the effect is due to immigrants behaving more like natives in their residential responses to black arrivals over time.

In future work, we will combine the two empirical approaches to develop estimates of flight at various points in the black share distribution and use these estimates to calibrate a model of white flight and segregation. Simulations of this model will allow us to estimate the extent to which uncoordinated market actions can account for the rise of segregation in prewar American cities.

Our results are not meant to imply that collective action and government policy had no impact on patterns of racial segregation. Centralized institutions such as restrictive covenants were no doubt successful in directing black settlement by raising the relative cost of settling in some neighborhoods. However, in contrast to previous work, our findings suggest that these institutions were unnecessary for achieving residential segregation by race. Furthermore, the evidence on the importance of decentralized market choices nuances our understanding of efforts
to reduce segregation in the postwar era. Government interventions have focused on breaking down institutional barriers to racial integration, most notably with the Fair Housing Act in 1968. Such approaches should perhaps not have been expected to reduce segregation if individual residential preferences were unchanged, and indeed segregation in the United States has remained high even after restrictive covenants were ruled unenforceable and discrimination in housing and lending markets declined.

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 discusses our findings.

II. Background on Segregation and Urbanization in the United States

Residential segregation by race has remained one of the most prominent and enduring features of American cities. Social scientists have linked segregation to a host of adverse minority outcomes, arguing that blacks living in more segregated cities have worse health, human capital accumulation, and labor market outcomes (Wilson, 1996; Cutler and Glaeser, 1997; Card and Rothstein, 2007; Sharkey, 2013). Ananat (2011) provides causal evidence of the impact of segregation on black poverty using historical railroads as an instrument for contemporary dissimilarity levels. Recent scholarship has also found that segregation is negatively correlated with the intergenerational mobility of the entire American urban population, including whites (Chetty, Hendren, Line, and Saez, 2014). While a broad literature studies the impact of segregation, less empirical work exists to explain the genesis of the phenomenon in the American cities.
In this paper we study the marked rise in racial segregation that occurred between 1900 and 1930 in American cities. We set the stage for our analysis by re-establishing the extant understanding of this rise in segregation levels using our newly constructed spatial data set. We measure segregation using the two most common indices of segregation: isolation and dissimilarity. In constructing isolation indices we follow Cutler, Gleaser, and Vidgor (1997) and compute a modified index which controls for the fact that under the standard approach there is a potential for the index to be sensitive to changes in the overall group share. For each year compute:

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

where \textit{population} refers to the population of the enumeration district (\textit{i} subscript) or city (\textit{total} subscript) and \textit{blacks} refers to the racial group’s enumeration district population (\textit{i} subscript) or city population (\textit{total} subscript). The “standard” isolation index simply computes the average percentage of a group member’s neighborhood composed of members of her own group. The modified measure controls for the fact that under random sorting groups with larger overall population shares will, by construction, experience neighborhoods with larger own group shares. The modification addresses this issue by expressing the average exposure share relative to the group’s overall share of the population. This relative measure is then rescaled (hence the numerator in Equation 1) so that it spans the interval from zero to one. Such an adjustment makes measured isolation less dependent on a group’s share of the overall population.

Our second measure is the dissimilarity index (Duncan and Duncan, 1955). For blacks and whites it is defined as:
Dissimilarity Index \[ \frac{1}{2} \sum_{i=1}^{N} \left| \frac{\text{black}_i}{\text{black}_{\text{total}}} - \frac{\text{whites}_i}{\text{whites}_{\text{total}}} \right| \] (2)

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

Figure 1 reports the average racial segregation in our sample cities across the past century taken from the dataset used in Cutler, Glaeser, and Vidgor (1999). The 1900 to 1930 decades saw the largest increase in measured segregation of any period in the twentieth century. Isolation peaked in 1970. Between 1900 and 1970, isolation rose from .23 to .66, with 63 percent of the overall increase having taken place 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.

The Cutler et al data was constructed using adjusted ward-level measures for isolation and dissimilarity prior to 1940, the year when census tract data became widely available. We compute the same measures using our enumeration district data discussed below in Section III for comparison and report these results in Figure 2. Segregation indices computed at the enumeration district are much higher than those computed at the ward level, which is to be expected given the smaller scale of these units (the average enumeration district had 1,400 individuals while wards could have as many as 100,000 residents in large cities). However, the trends in ward and enumeration district segregation are nearly parallel, showing a steep increase between 1900 and 1930. Furthermore, the Cutler et al adjusted ward measures are quantitatively similar to the enumeration district measures of both isolation and dissimilarity.
These stylized facts are not new. Scholars have long argued that the groundwork of the black ghetto was laid during the first decades of the twentieth century as black populations in northern cities grew, leading to the sharp increase in racial residential segregation. African Americans began to migrate to northern cities on the eve of World War I, an event that brought European immigration to a temporary halt while simultaneously increasing demand for industrial production. These wartime developments in the northern labor market coincided with the arrival of the Mexican boll weevil in Mississippi and Alabama (1913 and 1916, respectively), which devastated cotton crops and led to a decline in demand for black tenant farmers (Grossman, 1991). The combination of push and pull factors led to unprecedented out-migration from the South: 525,000 blacks came to the North in the 1910s and 877,000 came in the 1920s (Farley and Allen, 1987).

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

---

2 This computation uses the center city status variable from IPUMs samples for 1880 to 1930.
from later in the twentieth century when whites were leaving cities for different tax and public good combinations that were available in the suburbs.

III. Enumeration District Data for 1900 to 1930

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

The microdata we compiled for this paper represent a significant improvement over existing sources of data on early twentieth century urban populations. Ward-level tabulations published by the census are the smallest unit at which 100 percent counts were previously available for the years we study.³ Wards were large political units used to elect city councilmen.

³ Some census tract tabulations are available for the early twentieth century, but we prefer our approach because the enumeration district data is available for all cities and years in our sample and counts of any subpopulation can be created from microdata.
still in use in some cities today, while enumeration districts were small administrative units used internally by the census in this period. Every Ancestry.com record contains the enumeration district of the individual being surveyed, allowing us to make counts of any population at this level of geography. The other digitized variables from the 1900 to 1930 censuses include place of birth, father’s place of birth, mother’s place of birth, year of birth, marital status, gender, race, year of immigration (for foreign-born individuals), and relation to head of house in addition to place of residence (city, ward, and enumeration district) at the time of the respective census.

To place these individuals in urban space, we created digitized versions of the census enumeration district maps based on information available from the National Archives. We used written descriptions of the enumeration districts that are available on microfilm from the National Archives. The written descriptions from these microfilms has been digitized and made available on the web due to the work of Stephen P. Morse.\textsuperscript{4} We also employed a near complete set of physical maps for our census-city pairs located in the maps section of the National Archives. We took digital photographs of these maps as a second source for our digitization effort. Working primarily with geocoded (GIS) historic base street maps that were developed by the Center for Population Economics (CPE) at the University of Chicago, research assistants generated GIS representations of the enumeration district maps that are consistent with the historic street grids. Figure 3 provides an illustration of this process. Here the shaded regions in panel D represent the digitized enumeration districts.

In order to conduct analysis of change over time within neighborhoods, we require neighborhood definitions that are constant across census years. Forming such neighborhoods is challenging for this data because enumeration districts were redrawn for each decadal census and, unlike the case of modern-day census tracts, most changes were more complex than simple

\textsuperscript{4} website: http://stevemorse.org/ed/ed.php
combinations or bifurcations. In this paper we employ a hexagon-based imputation strategy. The strategy is illustrated in Figure 4. It involves covering the enumeration district maps (Panel A) with an evenly spaced temporally invariant grid of 800 meter hexagons (Panel B) and then computing the intersection of these two sets of polygons (Panel C). The count data from the underlying enumeration districts is attached to individual hexagons based on the percentage of the enumeration district’s area that lies within the individual hexagon. Panel D presents the allocation weights for a sample hexagon. In the example, 100 percent of four enumeration districts lies completely within the hexagon (136, 139, 140, and 144) while 11 enumeration districts are partially covered by the hexagon. For these partial enumeration districts, only fractions of their counts are attributed to the hexagon, ranging from a minimum of 0.2 percent (155) to 93.6 percent (142).

We form a panel of hexagons with at least 95 percent coverage by enumeration districts from the respective census in each year from 1900 to 1930. Our sample hexagons have average populations of 3,175 in 1900 and 4,892 in 1930, with the increase in density reflecting the rise in urban population density that occurred over this period. Thus, they are roughly similar in scale to modern-day census tracts. In Table 1 we provide summary statistics on the neighborhoods with growing black populations that we focus on in our empirical analysis. The average white population growth is positive in all years but declining from 972 over the 1910s to 188 in the 1920s. Only first-generation immigrants decline in absolute numbers by the 1920s. The average black percent increases from 3.8 to 8.7 percent over the 1910 to 1930 period.

IV. Empirical Strategy
The objective of the empirical work is to ascertain whether black arrivals had a causal impact on white population dynamics over the 1900 to 1930 period. The primary difficulty in identifying such an effect is that minorities do not exogenously arrive in neighborhoods, and we are particularly concerned that blacks and immigrants may locate in locations already being abandoned by white natives for other reasons. We use two different strategies to address this concern. First, we search for nonlinearities in white population share as a function of black population share that could indicate “tipping” behavior in the data. Second, we develop an instrumental variable strategy for black population size in our neighborhoods. In future work we will use our estimates of white flight to investigate the importance of uncoordinated market actions in the rise of segregation in American cities.

A. Nonlinear Evidence of White Flight

The intuition behind our first empirical approach is as follows: we wish to know if the presence of racial and ethnic minorities above some “tipping point” induces white natives to “flee” their neighborhood at an accelerated pace, giving rise to a nonlinear relationship between baseline black population share and the change in white population share. Our methodology is similar in nature to that of Card, Mas and Rothstein (2008) who study neighborhood tipping later in the twentieth century based on the classic Schelling (1971) model. However, we depart from their empirical framework in several ways. First, we use as our neighborhood definition the 800 meter hexagons instead of census tracts, which were not used by the census until 1910. Second, the criteria for identifying specific “tipping” points were less clear in our context.\(^5\) Thus, instead

\(^5\) We experimented with the different approaches employed by Card, Mas, and Rothstein in their work as well as other candidate approaches. While the shape of the non-parametric relationships is quite robust, the specific tipping points identified were quite sensitive to small changes in procedure.
of looking explicitly for tipping points based on specifically defined criteria, we present results of entire nonparametric regressions using local polynomial smoothing.

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

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

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

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

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

---

6 This assumption would hold if all potentially confounding relationships between black share and percent change in white population change were linear or if any confounding non-linear relationships were sufficiently small.
We begin this analysis by considering a simple OLS model relating the change in black populations to the change in white populations. Because we are interested in understanding the response of native whites to the influx of new black residents, we restrict our attention to neighborhoods that saw increases in the number of black residents. Such a model is:

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

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

In a world where neighborhood choice was unrelated to race, ordinary least squares (OLS) estimates of $\beta$ in equation (2) would expected to be positive because any neighborhood factor that encouraged increased black in-migration (relative to the city average) would also lead to increased white in-migration. Thus, shared sorting on neighborhood characteristics will bias OLS estimates of $\beta$ upwards (away from flight). Conversely, OLS estimates of $\beta$ will be biased in a negative direction (towards flight) if black arrivals are settling in neighborhoods already being abandoned by whites.

To overcome these bias concerns we leverage arguably exogenous variation in contemporary state-level outmigration rates in combination with pre-1900 patterns of black migration into our sample of northern cities. To do so, we construct an instrument using the universe of historical census records recently made available by Ancestry.com. Our instrument for $\Delta B_{ij}^{t1-t0}$ is constructed from two components: estimated black outflows from each state in
each decade (1900 to 1930) and settlement patterns established by African American who came to the North before the Great Migration and were thus living in our sample cities 1900.

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

\[
black_{outflow}^{t1-t0}_{cgs} = \sum_{k=1}^{51} black_{outmigrants}^{t1}_{icgs} - \sum_{k=1}^{51} black_{outmigrants}^{t0}_{icgs}
\]

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

For the 1900 base year component of the instrument, we count the number of black out-migrants in each birth cohort-gender-state of birth cell present in each neighborhood of our sample in 1900 to obtain \( black_{basepop}^{1900}_{cgs} \). 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:
where $\text{black_outmigrants}^{1900}_{cgs}$ is the national sum of all black out-migrant individuals in the cell in 1900.\footnote{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.} Our instrument for $\Delta B_{ij}^{t_1-t_0}$ is thus $\text{pred_black}^{t_1-t_0}_i$.

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

For this instrument to have power, two types of variation are needed. First, within a given city the distribution of blacks across neighborhoods must differ by state of origin. To give a sense of the existence of this type of variation, we provide city-level scatter plots in Figure 5 showing by neighborhood the share of black men aged 20 to 29 in 1900 who were born in two pairs of source states. Panel A shows that neighborhoods within Boston, Brooklyn, Chicago, Cleveland, and Philadelphia all exhibit rich variation in the share of black men from this cohort originating in North Carolina as opposed to Virginia (Panel A). Panel B shows the significant
variation across neighborhoods in Chicago, Cincinnati, and St. Louis in the share of the black population originating in Kentucky versus Tennessee.

In addition to differential within city sorting, for our instrument to have predictive power, it should also be the case that variation exists across sending states over time. Figure 6 shows the estimated outflows from the thirteen most important sending states for black men aged 20 to 29 across each of the decades we study in this paper.\textsuperscript{8} 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.

\section*{V. Analysis of White Flight in the Early Twentieth Century}

\subsection*{A. Nonlinear Evidence of Tipping}

In this section we present the results from both the nonlinear and causal models of white flight proposed in the previous section. The results from the nonparametric regressions are given in Figure 7 using local polynomial smoothing. We begin with the relationship between black neighborhood share and (de-meaned) change in white population share over the next decade in Panel A. For the 1900 to 1910 decade, a decline in white population relative to the city average is apparent for neighborhoods with more than 10 percent black share, but the effect is somewhat noisy due to the small number of blacks in northern cities in 1900. By the 1910 to 1920 decade, the relationship is much more precisely estimated, with a discontinuity in white share suggestive of tipping apparent between 10 and 15 percent black share. Neighborhoods with more than 20

\textsuperscript{8} These thirteen states represent between 87 and 92 percent of total black outflows in each year we study.
percent black population lost about 35 percent more white population than neighborhoods with less than 10 percent black population. By the 1920 to 1930 decade, this affect had accelerated, and neighborhoods above the 20 percent black population mark lost about 50 percent more white population than neighborhoods with small black population shares, with the discontinuity in white population change again readily apparent.

If neighborhoods in these northern cities did exhibit white flight when faced with growing numbers of black residents, a natural question is which whites were doing the “fleeing.” In Panel B we show the same local regression results for third-generation natives and in Panel C the results for first-generation immigrants (we show all three groups in the linear regression results in the next section). The declines in white population are similar in magnitude for natives across the range of black population in all three census years; however, the likely tipping point appears to decline from about 15 percent to 10 percent over the 1910 to 1920 and 1920 to 1930 decades.

The dynamics of white immigrant flight follow a different pattern (Panel C). In the 1900 to 1910 decade, nonlinearity in white immigrant percent change is muted and confidence interval nearly covers zero along the full range of black share. By the next decade, the effect is estimated much more tightly, and a sharp discontinuity emerges around 15 percent black share. The magnitude of the flight effect accelerates between the last two decades, and by the 1920 to 1930 decade, neighborhoods above the discontinuity are losing 50 percent more immigrant population than neighborhoods below the tipping point (less than 15 percent black share). Thus, some portion of the acceleration in the overall white flight effect appears to be due to immigrants behaving more like natives over time. The nonlinear results we document in the aggregated city sample are also apparent if we perform the nonparametric estimation on city at a time, as shown
in Panel D for the 1920 to 1930 decade. The discontinuity in white population change appears at the highest share black for Cincinnati and St. Louis (both around 20 percent black). The relationship for Baltimore is noisiest, likely due to the larger pre-existing black population in that city.

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

In our discussion of the nonlinear estimation results in Figure 7, we focused largely on evidence of tipping behavior. However, while the relationships between black share and white population share dynamics were generally nonlinear, they were also monotonic. We thus turn to the second empirical strategy, beginning with OLS estimation of equation (2). These results are presented in Table 2. Here we follow the literature and look at changes in numbers (controlling for the city average change in white population) rather than percentages. Across the three decades we study, an additional black resident in the neighborhood is associated with about one white departure, the outcome we would expect if no racial animus existed (Boustan, 2010). When we instrument for black population change, our estimates of white flight increase in magnitude to -1.8 in the first decade and -2.5 in the last decade, suggesting that blacks were heading to otherwise growing neighborhoods in our sample cities. We note that our estimates are similar in magnitude to what Boustan (2010) finds for the 1940 to 1970 period when measuring flight from central cities to the suburbs. The F-test on the first stage in over 50 in all three decades, indicating our instrument is strong.

As with the nonlinear approach, we decompose the flight effect by white population type, considering first-generation immigrants, second-generation immigrants, and third or more-generation natives using our IV specification. The relative size of each of these groups is roughly similar, facilitating comparison across the specifications presented in Table 3. We see
that an increase of one black resident causes a loss of 1.2 natives and .4 second-generation immigrants, with no impact on first-generation immigrants, over the 1900 to 1910 decade. Immigrants’ residential responses converge on that of natives over time, and by the 1920 to 1930 decade, an increase of one black resident causes a loss of .9 natives, .8 second-generation immigrants, and .7 first-generation immigrants (roughly summing to our overall white flight effect of -2.5).

In Table 4 we present a series of robustness checks to our main linear regression results. One concern with our instrumental variable is that percent black in 1900 is highly correlated with the urban core and neighborhoods that were set to both attract black population and lose white population. We control for percent black in 1900 in the first set of checks and show our results are slightly larger in magnitude. Our results are also robust to controlling for percent southern black in 1900 and the number of blacks in the neighborhood in 1900 (not shown). We also show our results with the standard inclusion of pre-trends in white population. Although this control 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 greater in magnitude than 1. When we include both black percent in 1900 and white pre-trends, our results are very similar to the baseline.

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

[under construction]

VII. Conclusion
Taken in total, our results suggest that the dynamics of white populations likely played a key role in the sharp increase in racial segregation over the 1900 to 1930 period. Our nonlinear analysis of white population dynamics showed “tipping” was occurring at lower black shares while white population loss after a neighborhood had passed the threshold was growing over the early twentieth century. Furthermore, our 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 2.5 white individuals. The robustness of these findings and the way in which they vary across time suggests that evolving changes in white animus was a key factor in rising racial segregation. White flight was not simply a response to deplorable ghetto conditions developed over decades of black migration to northern cities. Instead, whites appear to have been fleeing black neighbors per se in prewar America, and these decentralized location decisions had important impacts on the aggregate level of racial segregation in cities.
BIBLIOGRAPHY


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

![Graph showing segregation trends from 1890 to 2000](image)

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

C. Enumeration District Descriptions

D. Digitized Enumeration District Map (ArcMap)
Figure 4. Constructing Hexagon Neighborhoods from Enumeration District Maps

A. Enumeration District Map (1900)

B. Hexagon Grid (Constant across Decades)

C. Intersection between Enumeration Districts and Hexagons

D. Allocating Enumeration District Count Data to Hexagon Neighborhoods

Notes: see Section III for details on the source of the maps and street files used to construct these images.
Figure 5: Variation in Origin of Black Settlement across Neighborhoods in 1900

A. Virginia versus North Carolina

B. Kentucky vs Tennessee

Notes: Scatterplots show the share of black men aged 20 to 29 born in each source state out of the total number of black men in the cohort in neighborhood. The shares are computed using the universe of census records with enumeration district identifiers from each city and the hexagon imputation strategy discussed in Section III.
Figure 6. Variation in Estimated Black Outflows from Southern States by Decade

Notes: The data in this figure come from the universe of census microdata made available by Ancestry.com. Estimated outflows are computed by summing the change in the number of individuals in gender, state of birth, and birth cohort cells appearing outside their birth state in each census year.
Figure 7: Panel A. Black and White Population Dynamics

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

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

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

Notes:
D. By City: Relationship between Share Black and White Population Change, 1920-1930 Decade
Figure 4.D, con’t

Notes: All figures show the nonparametric relationship between share black and either total white or white subpopulation changes in the neighborhood over the next decade. All white population changes are de-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>
<thead>
<tr>
<th></th>
<th>1910</th>
<th>1920</th>
<th>1930</th>
</tr>
</thead>
<tbody>
<tr>
<td>Change in White Population</td>
<td>972.60</td>
<td>518.86</td>
<td>188.73</td>
</tr>
<tr>
<td></td>
<td>(2247.89)</td>
<td>(2068.33)</td>
<td>(3559.80)</td>
</tr>
<tr>
<td>Change in Black Population</td>
<td>85.41</td>
<td>236.99</td>
<td>378.44</td>
</tr>
<tr>
<td></td>
<td>(305.57)</td>
<td>(851.22)</td>
<td>(1403.48)</td>
</tr>
<tr>
<td>Change in White 3rd Generation Population</td>
<td>258.90</td>
<td>314.11</td>
<td>155.74</td>
</tr>
<tr>
<td></td>
<td>(640.78)</td>
<td>(694.21)</td>
<td>(942.25)</td>
</tr>
<tr>
<td>Change in White Second-Generation Population</td>
<td>312.44</td>
<td>190.02</td>
<td>133.37</td>
</tr>
<tr>
<td></td>
<td>(822.94)</td>
<td>(737.26)</td>
<td>(1371.89)</td>
</tr>
<tr>
<td>Change in White First-Generation Population</td>
<td>409.42</td>
<td>9.24</td>
<td>-2.60</td>
</tr>
<tr>
<td></td>
<td>(1472.91)</td>
<td>(1220.12)</td>
<td>(1700.33)</td>
</tr>
<tr>
<td>Black Percent</td>
<td>3.84</td>
<td>6.64</td>
<td>8.66</td>
</tr>
<tr>
<td></td>
<td>(7.06)</td>
<td>(13.00)</td>
<td>(18.47)</td>
</tr>
<tr>
<td>White 3rd Generation Percent</td>
<td>38.72</td>
<td>37.99</td>
<td>38.69</td>
</tr>
<tr>
<td></td>
<td>(16.25)</td>
<td>(18.15)</td>
<td>(20.41)</td>
</tr>
<tr>
<td>White Second-Generation Percent</td>
<td>32.22</td>
<td>31.43</td>
<td>31.32</td>
</tr>
<tr>
<td></td>
<td>(8.90)</td>
<td>(9.79)</td>
<td>(11.61)</td>
</tr>
<tr>
<td>White First-Generation Population</td>
<td>24.76</td>
<td>23.35</td>
<td>21.51</td>
</tr>
<tr>
<td></td>
<td>(11.47)</td>
<td>(11.46)</td>
<td>(11.97)</td>
</tr>
<tr>
<td>Population</td>
<td>4335.40</td>
<td>5417.05</td>
<td>5888.34</td>
</tr>
<tr>
<td></td>
<td>(6936.86)</td>
<td>(6332.53)</td>
<td>(5945.96)</td>
</tr>
</tbody>
</table>

Notes: Summary statistics cover the hexagons in our panel dataset with growing black populations (as measured relative to the preceding decade). Changes in population are also with respect to the previous decade’s value. All demographic variables were created using the universe of census records from Ancestry.com.
Table 2. Baseline OLS and IV Results for Effect of Black Arrivals on White Departures

<table>
<thead>
<tr>
<th></th>
<th>dependent variable = change in white population</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1900-1910 Decade</td>
</tr>
<tr>
<td>OLS Results</td>
<td></td>
</tr>
<tr>
<td>Change in Black Pop.</td>
<td>-0.906***</td>
</tr>
<tr>
<td></td>
<td>(0.172)</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.270</td>
</tr>
<tr>
<td>IV Results</td>
<td></td>
</tr>
<tr>
<td>Change in Black Pop.</td>
<td>-1.816***</td>
</tr>
<tr>
<td></td>
<td>(0.401)</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.256</td>
</tr>
<tr>
<td>First Stage</td>
<td></td>
</tr>
<tr>
<td>Predicted Change in Black Pop.</td>
<td>1.359***</td>
</tr>
<tr>
<td></td>
<td>(0.0758)</td>
</tr>
<tr>
<td>F-test on First Stage</td>
<td>321.3</td>
</tr>
<tr>
<td>Observations</td>
<td>1,412</td>
</tr>
</tbody>
</table>

Notes: See Table 1 for sample details. All regressions include city fixed effects. The instrumental variables regressions are estimated using limited information maximum likelihood estimation (LIML).
Table 3. White Flight Effect by White Subpopulations (IV)

<table>
<thead>
<tr>
<th></th>
<th>1900-1910 Decade</th>
<th>1910-1920 Decade</th>
<th>1920-1930 Decade</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dep. Var. = Change in White 3rd-Gen. Pop.</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Change in Black Population</td>
<td>-1.270***</td>
<td>-0.715***</td>
<td>-0.928***</td>
</tr>
<tr>
<td></td>
<td>(0.139)</td>
<td>(0.0718)</td>
<td>(0.0896)</td>
</tr>
<tr>
<td><strong>Dep. Var. = Change in Second-Gen. Pop.</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Change in Black Population</td>
<td>-0.406***</td>
<td>-0.442***</td>
<td>-0.760***</td>
</tr>
<tr>
<td></td>
<td>(0.154)</td>
<td>(0.0661)</td>
<td>(0.112)</td>
</tr>
<tr>
<td><strong>Dep. Var. = Change in First-Gen. Pop.</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Change in Black Population</td>
<td>-0.0439</td>
<td>-0.446***</td>
<td>-0.734***</td>
</tr>
<tr>
<td></td>
<td>(0.263)</td>
<td>(0.129)</td>
<td>(0.161)</td>
</tr>
<tr>
<td>Observations</td>
<td>1,412</td>
<td>1,060</td>
<td>1,326</td>
</tr>
</tbody>
</table>

Notes: see Table 2 for sample and specification details.
Table 4. White Flight Effect Robustness Checks (IV)

<table>
<thead>
<tr>
<th></th>
<th>1900-1910 Decade</th>
<th>1910-1920 Decade</th>
<th>1920-1930 Decade</th>
</tr>
</thead>
<tbody>
<tr>
<td>Change in Black Population (baseline)</td>
<td>-1.816***</td>
<td>-1.692***</td>
<td>-2.509***</td>
</tr>
<tr>
<td></td>
<td>(0.401)</td>
<td>(0.207)</td>
<td>(0.313)</td>
</tr>
<tr>
<td>Change in Black Population</td>
<td>-1.175</td>
<td>-2.089***</td>
<td>-3.566***</td>
</tr>
<tr>
<td></td>
<td>(0.838)</td>
<td>(0.360)</td>
<td>(0.483)</td>
</tr>
<tr>
<td>Percent Black in 1900</td>
<td>-24.95</td>
<td>29.86</td>
<td>130.7***</td>
</tr>
<tr>
<td></td>
<td>(24.35)</td>
<td>(21.44)</td>
<td>(38.91)</td>
</tr>
<tr>
<td>Change in Black Population</td>
<td>-1.464***</td>
<td>-1.744***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.195)</td>
<td>(0.247)</td>
<td></td>
</tr>
<tr>
<td>Pre-Trend in White Population</td>
<td>0.247***</td>
<td>0.566***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0319)</td>
<td>(0.0331)</td>
<td></td>
</tr>
<tr>
<td>Change in Black Population</td>
<td>-1.957***</td>
<td>-2.701***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.341)</td>
<td>(0.370)</td>
<td></td>
</tr>
<tr>
<td>Percent Black in 1900</td>
<td>37.88*</td>
<td>118.7***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(20.57)</td>
<td>(29.79)</td>
<td></td>
</tr>
<tr>
<td>Pre-Trend in White Population</td>
<td>0.258***</td>
<td>0.568***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0348)</td>
<td>(0.0410)</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>1,412</td>
<td>1,060</td>
<td>1,326</td>
</tr>
</tbody>
</table>

Notes: see Table 2 for sample and specification details.