THE ECONOMIC IMPACT OF THE FIRST GREAT MIGRATION

by

Sijie Li

B.A. in Economics, Shanghai University of Political Science and Law, 2011

M.A. in Economics, Boston University, 2012

M.A. in Economics, University of Pittsburgh, 2015

Submitted to the Graduate Faculty of

the Dietrich School of Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2019

UNIVERSITY OF PITTSBURGH

THE KENNETH P. DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Sijie Li

It was defended on

April 3, 2019

and approved by

Allison Shertzer, Assistant Professor of Economics, University of Pittsburgh

Randall Walsh, Professor of Economics, University of Pittsburgh

Giovanni Osea Giuntella, Assistant Professor of Economics, University of Pittsburgh

Brian Beach, Assistant Professor of Economics, College of William and Mary

Dissertation Advisors: Allison Shertzer, Assistant Professor of Economics, University of Pittsburgh,

Randall Walsh, Professor of Economics, University of Pittsburgh

THE ECONOMIC IMPACT OF THE FIRST GREAT MIGRATION

Sijie Li, PhD

University of Pittsburgh, 2019

This dissertation studies the first Great Migration of African Americans from the rural South to Urban areas in the northern United States. While most existing research has focused on the experiences of the migrants themselves, I am focused on how this influx of rural black migrants impacted outcomes for African Americans who were already living in the north and had already attained a modicum of economic success. Common themes throughout this dissertation involve the use of the complete-count U.S. population census to link records across years. In the first chapter, I linked northern-born blacks from 1910 to 1930 to study how the arrival of new black residents affected the employment outcomes of existing northern-born black residents. I find that southern black migrants served as both competitors and consumers to northern-born blacks in the labor market. In the second chapter, my co-authors and I study the role of segregated housing markets in eroding black wealth during the Great Migration. Building a new sample of matched census addresses from 1930 to 1940, we find that racial transition on a block was associated with both soaring rental prices and declines in the sales value of homes. In other words, black families paid more to rent housing and faced falling values of homes they were able to purchase. Finally, the third chapter compares the rates of intergenerational occupational mobility by both race and region. I find that racial mobility difference in the North was more substantial than it was in the South. However, regional mobility difference for blacks is greater than any gap in intergenerational mobility by race in prewar American. Therefore, the first Great Migration helped blacks successfully translate their geographic mobility into economic mobility.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0 INTRODUCTION

Prior to the Great Migration of African Americans from the rural South to Urban areas in the northern United States, there were small communities of middle-class blacks living in the North. While research has investigated the importance of migration improving the fortunes of southern-born blacks, less is known about the impact of the Great Migration on existing black communities. This dissertation focused on how this influx of rural black migrants impacted outcomes for African Americans who were already living in the north and had already attained a modicum of economic success.

In the first chapter, I use individual U.S. census data to construct a new panel dataset of northern-born blacks from 1910 to 1930 to study how the arrival of new black residents shaped their economic fortunes. I exploit variation in the extent of in-migration across northern counties and instrument for black inflows by interacting pre-existing demographic patterns in the South with earlier black settlement patterns in the North. I find that, while in-migration decreased the likelihood for black northerners to be employed, it led them to have more high-status, better-paying jobs in 1930. Furthermore, these effects are not homogeneous throughout the northern black population. Since southern blacks served as both competitors and consumers to northern blacks, the economically disadvantaged northern blacks suffered from the competition pressure exerted by migrants while the economically advantaged northern blacks obtained the better occupations supported by in-migration.

Residential segregation accompanying the Great Migration made housing demand outpace housing supply for blacks in the northern cities. Blockbusting, or opening a new street to black settlement, was profitable because black families were desperate for housing and would outbid whites for apartments and homes outside of the existing ghetto. In the second chapter, my co-authors and I study the role of segregated housing markets in eroding black wealth in prewar America. Using a sample of matched addresses from prewar American cities, we find that rental prices and occupancy soared by about 40 percent in blocks that transitioned from all white to majority black. While, on these same blocks, home values fell on average by 10 percent and by a staggering 50 percent in major African American destinations such as Chicago, Philadelphia, and Detroit. These findings suggest that, because of the segregated housing market, black families faced dual barriers to wealth accumulation: they paid more in rent for similar housing while the homes they were able to purchase rapidly declined in value.

In the third chapter, I study the economic mobility of the existing population of northern-born blacks relative to that of southern-born blacks and northern-born whites. Previous studies examining the racial differences in intergenerational mobility in the early twentieth century have mainly focused on southern blacks. Since northern blacks began with a relatively better socioeconomic status than southern blacks, my prior is that northern blacks might have been better at capitalizing on their advantage in occupational attainment. Through building new samples of northerners with intergenerational linkages between 1910 and 1930, I show that northern blacks indeed had a different father-son occupational association from southern blacks; moreover, their father-son occupational association resembles more closely to that of northern whites than to that of southern blacks. While the black-white mobility gap in this period would have closed in the South if blacks had the same occupational opportunity as whites, it would persist in the North. These findings suggest that heterogeneity shaped by regional origin within and across racial groups should not be overlooked in addressing racial disparities in economic mobility.

2.0 COMPETITORS AND CONSUMERS: THE IMPACT OF THE GREAT MIGRATION ON EMPLOYMENT OUTCOMES OF BLACK NORTHERNERS

"Its imprint is everywhere in urban life. The configuration of the cities as we know them, the social geography of black and white neighborhoods, the spread of the housing projects as well as the rise of a well-scrubbed black middle class, along with the alternating waves of white flight and suburbanization - all of these grew, directly or indirectly, from the response of everyone touched by the Great Migration."

-Isabel Wilkerson, The Warmth of Other Suns

2.1 INTRODUCTION

Scholarship tends to view the Great Migration, a movement of approximately 6 million blacks from the agricultural South to the industrial North, as the key mechanism in black economic progress over the twentieth century (e.g., Margo (1995); Smith and Welch (1989); Farley and Allen (1987)). Compared to their southern counterparts, blacks in the North were economically and socially better-off at the turn of the century, and some even achieved middle-class status (see Figure 1). For example, in 1910 the median annual earning for southern blacks was \$235, 62 percent less than their white southern counterparts, while the comparable earnings disadvantage for northern blacks, whose median annual earning was \$633, was only 37 percent.[1] Thus poor southern blacks could obtain significant economic returns by migrating north. Though economists have made recent advancements in understanding the importance of the Great Migration on southern-born blacks (e.g., Alexander et

[1]The earnings disadvantage is measured as the difference in the median occupational earnings score between blacks and whites using the 1910 complete-count census data.

al. (2017); Collins and Wanamaker (2017); Collins and Wanamaker (2014)), the economic experience of black northerners, who had longstanding roots in the North, with respect to the significantly increased local black populations, remains largely unexplored.[2]

Between 1910 and 1930, the number of black workers in the North almost tripled.[3] The anticipated effect of such a large inflow of migrants on black northerners' employment outcomes is theoretically ambiguous. On the one hand, if black workers from any region are seen as substitutes in the labor market, migration would act as a supply shock, decreasing wages or employment of northerners. Labor market crowding could have also reinforced the traditional racial norms and inflamed already existing racial tensions, further disadvantaging northern black workers. On the other hand, the inflow of southern blacks could have created a new demand for black-owned establishments and so northern blacks, who had more wealth than southern blacks, might have responded to this demand by becoming more entrepreneurial, ultimately increasing their occupational standing. This paper finds support for both impacts.

A natural strategy for studying the effect of in-migration on the labor market opportunities of black northerners is to examine the relationship between employment outcomes and migrant inflow across migration destinations. However, such exercise is subject to three sources of bias. First, migrants may have been attracted to places with economic opportunity, which would cause OLS estimates for both employment rates and earnings to be biased upwards. By contrast, given the discrimination and information barrier faced by migrants in the northern labor market, they may have only been able to settle in declining areas where there are fewer occupational upgrading opportunities, causing OLS estimates for earnings to be biased downwards. Second, northerners may relocate in response to their expected labor market opportunities, causing sample composition bias when analyzing at the geographic aggregation level. Lastly, omitted variables in the cross-sectional data could also cause the OLS estimates to be biased.[4]

[2]Several papers explore competition between first and second waves southern blacks in the postwar period, e.g., Derenoncourt (2019); Boustan (2009). However, the first and second waves southern blacks are more similar to each other and this paper will specifically look at blacks who already lived in the North and who see themselves as a distinct group.

[3]Calculation is based on counting the black population aged 10 and above in the northern labor force for both census years (Ruggles et al. (2017)).

[4]For example, Chetty and Hendren (2016) shows that the neighborhoods in which children grow up shape their adult labor market outcomes; Hout (1984)) finds that black men who experienced occupational upgrading during the

Some of these obstacles can be overcome by constructing an individual-level panel dataset of black northerners (Foged and Peri (2016)). The availability of the digitized complete-count census data containing personally identifiable information in the prewar era made it possible to follow individuals over time. Using the supervised machine-learning approach introduced in Feigenbaum (2016), I build a new linked sample of northern-born blacks from 1910 to 1930 where I can observe a broad set of pre-migration sociodemographic characteristics to minimize the omitted variable bias. The benefits of having a linked sample also include the elimination of composition bias by comparing the outcomes of black northerners who already lived in migrant-receiving places in 1910 and the accessibility of critical demographic characteristics for investigating the potential mechanisms.

I address the endogeneity of migration inflow by taking advantage of the historical fact that current migrants tend to settle at the same place where early migrants from the same birth state already live, due to social networks and access to railroads.[5] Thus, a county's share of southern-born black residents before the migration is used as an instrument for the current settlement patterns of migrants from that particular southern state (Card (2001)). To address the concern that the aggregate inflow of black migrants is also correlated with local labor demand shocks because migrants cluster geographically in certain northern cities, I substitute the total migrant inflow from each southern state by its cohort size of blacks at risk to migrate but still living in the South in 1910. Therefore, the flow component of my instrument rests on the spatial and age distribution of blacks in the South prior to the migration.

I find that in-migration (from 1910-1930) resulted in significantly less employment but better occupational attainment for black northerners in 1930. For both outcomes, my 2SLS estimates are larger in magnitude than the OLS estimates. On the one hand, it is consistent with the historical view that migrants tended to move to places where jobs were relatively more abundant; on the other hand, it suggests that these places migrants chose offered fewer occupational upgrading opportunities to the new black workers.[6] Since my measure

1960s tended to come from advantaged social backgrounds in terms of their fathers' occupations.

[5] Carrington et al. (1996): "While this chain migration centered on kinship ties, the connections extended beyond family lines. For example, Southern 'migration clubs' would often finance a sequence of trips to the North for their members, and young males were usually the first to go."

[6] From the perspective of southern black migrants, any place in the North might have been better in terms of earnings and safety compared to their southern conditions. In the South, wages were paid in some form other than

of migration inflow compares locations where northern blacks lived in 1910, the observed effects could have come from two sources the creation of competition and new employment opportunity by migrants and the displacement of employment opportunity by the relocation of northerners between 1910 and 1930. To gain a better understanding of the source of the effects, I restrict the sample to stayers who lived in the same county in both census years. If the effects are indeed due to in-migration, I expect to find similar effects on the stayer sample. The results confirm my expectation; moreover, it suggests that out-migration of northerners is helpful only in securing employment, but not so much in regards to the quality of their employment.

I hypothesize that competition is the channel through which in-migration lower the employment of northern blacks. To investigate, I examine the relationship between in-migration and employment status for both northern-born whites, and northern-born blacks from different economic backgrounds. Since black and white workers were not considered as close substitutes in production, I expect in-migration to have no effect on the likelihood of employment for northern-born whites. Moreover, since migrants were often low-skilled sharecroppers prior to migrating, I expect the competition to be concentrated in the unskilled occupations. My estimates show that in-migration does not have a negative effect on whites regardless of their skill level, and in-migration exerts competitive pressure only on the unskilled northern-born blacks, a closer substitute to southern black migrants. Moreover, I show that in-migration did not disproportionately discourage northern black workers from participating in the labor force. Together, these findings suggest that competition with southern black migrants leads to lower employment for black northerners.

I close this chapter by documenting the mechanism through which in-migration increases northern-born blacks' occupational standing. The increasing black population in the segregated black communities created skilled employment opportunities for blacks. Northern-born blacks, especially those from more advantaged backgrounds, were able to reallocate their labor from unskilled to skilled occupations because they have the comparative advantage concerning access to capital and white support.

---

cash once a year. By contrast, the wages paid in cash by the week or month in the North seemed "fabulous sums promising speedy wealth and success." Therefore, the most immediate concern for most migrants was whether they could find a job outside the South. Moreover, the ultimate gain from moving to the highest paid place could have been lower than other places when considering the transportation cost.

2.2 HISTORICAL BACKGROUND

2.2.1 The Great Migration

From the introduction of slavery until the Civil War, black Americans mainly resided in the slave-owning South. However, all of the blacks living in the North before the Civil War were free, and some of them had achieved middle-class status. After Emancipation, the deprivation associated with share tenancy farming and the racial oppression of Jim Crow caused northward movement of southern blacks in search of a better life. However, northern cities in the late nineteenth century, the portals of industrial America for millions of European immigrants, had fewer opportunities for African Americans. Low-wage and low-skilled jobs were fulfilled by white Europeans.

Before the Great Migration, the small northern black population gained their livelihood mainly in domestic and personal service (Frazier (1957)). Despite restrictions with employment opportunities in industry and white-collar occupations, a few of them were found in a wide range of skilled occupations, such as physicians, dentists, journalists, attorneys, clergymen, and proprietors (e.g., Brady (1996)). Though residentially segregated from the whites, they did not suffer much discrimination in utilizing public institutions (Logan et al. (2015); Massey and Denton (1993)). The leading blacks were mainly supported by the white community, both economically and politically. They also maintained close social and professional relationships with whites (Katzman (1975)).

The outbreak of World War I (1914-1918) led to increased demand for manufactured goods throughout the industrial North. The curtailment of cheap immigrant labor by war and later the enactment of the Immigration Acts made the northern employers turn to domestic workers for factory employment. At the same time, destructive conditions in southern agriculture such as the boll weevil and devastating floods pushed large numbers of black laborers out of cotton farming. As a result of these factors, southern blacks began migrating to the North.[7] This massive movement later became known as the Great Migration, which was slowed down by the Great Depression, then resumed in the 1940s, and lasted until 1970.

[7] For example, the bleak job prospects and physical violence in the Jim Crow regime, the lure of higher wages in the North, the industrial recruitment campaigns, and the discontent spread by black press such as the *Chicago Defender* (DeSantis (1998)).

2.2.2 How the Great Migration Changed the North

A direct consequence of the Great Migration is that the size of the urban black population in the North significantly increased after 1910. For example, the black population in Wayne County, Michigan, increased from 7,241 to 131,836 between 1910 and 1930, with southern black migrants account for 76% of the change. "With generally minuscule black populations before the Great Migration," noted in Tolnay (2003), "Northern and Western cities had achieved a relatively stable state of race relations, albeit one characterized by distinct racial inequality. That situation began to change, however, as waves of migrants from the South produced extraordinary growth in local black populations."

The current view of how black employment in the North was affected is mixed. Gottlieb (1996) describes that "blacks in the North saw their occupations gradually compressed from a heterogeneous range of jobs into a comparatively narrow spectrum of employment." Blacks in the North were not only confined to the lowest-status jobs, but also consigned to black neighborhoods regardless of their social class and geographic origin, through racial violence, bombing, and white flight. Racial hostility increasingly blocked blacks access to a variety of amenities and necessities, such as theaters and health service (Trotter (1985)). By 1930, widespread interracial contact was no longer a phenomenon in American cities.

Despite the competition that underlaid the struggle for jobs and housing, the growing black population in the ghetto formed the demographic and financial foundation for the emergence of a larger black middle class between 1910 and 1930. The new black middle class was owners of small businesses as well as professional elite, who almost exclusively serve the needs of black clientele.[8] For instance, black ministers catered to black congregations, and black writers covered their social interests.[9] The number of blacks engaged in professional, business, and clerical occupations in the North doubled, increased from 55,586 in 1910 to 100,777 by 1930.

[8] Higgs (1977) wrote that "Discrimination outside the market sector had an important influence in determining the opportunities open to blacks inside the market sector. [...] A large, concentrated, black community allowed the emergence of a more complex and developed black community with its own merchants, craftsmen, and professional people."

[9] Urban black churches experienced significant growth during the migration era. For example, Abyssinian Baptist Church is founded in 1910 by Adam Clayton Powell, Sr., in Harlem. This church had become the largest in the nation by 1930, with over twenty thousand members. Other black churches, such as the Mount Olivet Baptist Church in Chicago and the Bethel AME Church (Mother Bethel) in Philadephia, all had a considerable number of members and became centers of religious activity.

Increased black populations in the North also granted blacks more political power, which yielded employment opportunities for blacks in the public sector.[10] For example, the Milwaukee County Board of Supervisors created the position of "Social Worker (For Colored People)" in 1930 in response to political pressure from blacks to serve the increasingly segregated black population. Police, school board, and other public departments followed suit (Trotter (1985)).

2.3 DATA AND MEASUREMENT

I use the regression-based, supervised machine-learning approach introduced in Feigenbaum (2016) to link individuals across censuses. Since it is a fairly standard approach, I briefly introduce my linking procedure here and the construction of key variables in this analysis. Details on the matching, representativeness, and sensitivity tests are provided in the Appendix. Lastly, I describe the black northerners in my sample.

2.3.1 Creating the Linked Census Sample for Black Northerners 1910-30

I begin with an initial sample from 1910 that includes all the northern-born black males, aged 0 to 17 inclusive, residing with their parent or step-parent in the North.[11] This restriction results in a total of 60,705 individuals who were young enough to be living with their fathers or household-head mothers in 1910 and were old enough to be participating in the labor force in 1930. They are also at an age where they can change their occupational and educational choices in response to the migrant inflow. I then find each individual in the complete-count 1930 census, limiting the set of possible matches for any given black in 1910 to be the black

[10] "There was no denial of voting rights in the North, and soon African American politicians were being increasingly elected to office. As early as 1915, blacks served on the city council in Chicago, as they did in 1919 in New York. By 1928, Oscar DePriest of Chicago had become the first northern black congressman and the first Black man to sit in Congress since 1901. In key industrial states like Ohio, New York, Illinois, and Pennsylvania black voters often swayed the balance of power between the Republican and Democratic parties, which gave them state and national influence far out of proportion to their numbers. To serve them a new urban black middle class emerged, out of the growth of black community services."

[11] Following previous literature, I define the following 14 states as the South: Texas, Oklahoma, Arkansas, Louisiana, Mississippi, Kentucky, Tennessee, Alabama, Florida, Georgia, South Carolina, North Carolina, Virginia, and West Virginia (Collins and Wanamaker (2015); Collins and Wanamaker (2014); Boustan (2009)). For convenience, all other parts of the contiguous United States are referred to as the North.

males in 1930 who meet the following criteria: born in the same state, born in the same year $\pm$ 3 years, have the first letter of first and last names match, and have a Jaro-Winkler string distance in the first and last names of greater than 0.8.[12]

The basic idea of the machine-learning approach is to have the algorithm learn the implicit rules that a careful and well-trained researcher uses to match records across historical samples and replicates these decisions for the full dataset (Feigenbaum (2016)). Therefore, I manually build a "ground truth" sample of approximately 5% of the initial sample to guide the algorithm.[13] In the case when an individual has more than one best match, he will have no match to be declared since it is impossible for me to decide which pairing is correct. Similarly, a person from 1930 cannot be matched to more than one individual from 1910.

For the algorithm to find the best match, I also need to choose a value between 0 and 1 for the absolute threshold and a value between 1 and infinity for the relative threshold. With the goal to minimize both the false match rate and the missed match rate, I set the absolute threshold at 0.36 and the relative threshold at 1.05.[14] The corresponding sample match rate is 40.68%, along with an average false match rate of 4.87% and an average missed match rate of 11.98%. These rates compare favorably with other studies linking across U.S. census files at the beginning of the twentieth century.[15]

2.3.2 Measuring the Occupational Earnings Score and Rank

Because wage data was not collected before 1940, scholars have developed approaches to estimate an individual's occupational earnings as a proxy for labor market outcomes (e.g., Collins and Wanamaker (2017); Ruggles et al. (2017)). Following the approach introduced by Collins and Wanamaker (2017), I construct an occupational earnings score for each occupation-race-gender-region cell based on its average 1940 income. This approach is similar in spirit to the

[12] Unlike contemporary dataset where there exists a unique personal identifier such as social security number, the linking of historical datasets have to rely on time-invariant variables such as names and ages. However, due to errors from census respondents, enumerators and digitization workers, information about names and ages are not always accurate. Feigenbaum (2016) examines the matched sample from IPUMS and finds that the linked individuals have birthyear difference less than 3 and name distance less than 0.2.

[13] Sampling from the initial sample to keep all possible matches for a given individual.

[14] False match rate represents the share of algorithm determined matches to be wrong; missed match rate represents the share of true matches failed by the algorithm. Both inaccurate matches and selective linkages could have significant effects on inference (Bailey et al. (2017)).

[15] A close comparison is Collins and Wanamaker (2017), where they use a different linking method to find southern-residing males aged 0-17, residing with their father or stepfather, from 1910 to 1930. The reported match rate is 27 percent.

OCCSCORE variable from IPUMS that is based on the median total income of all persons with that particular occupation in 1950 (Ruggles et al. (2017)). It has, however, more flexibility to reflect differences by location, race, gender, and farm ownership.[16] Most importantly, the 1940 income matches more closely to my sample period than the 1950 income.

Specifically, because income for self-employed workers is not recorded in the 1940 census, I first assign an imputed wage to self-employed as the average 1940 value for wage workers multiplied by the ratio of one to the other in 1960. Then I construct an earnings score for each occupation-race-gender-region cell by calculating the average earnings for everyone with income information in 1940, including both wage earners and self-employed workers, in that specific cell. Finally, I assign earnings scores to all individuals from both the 1910 census and the 1930 census based on their occupation, race, gender, and census division of residence.

Using the national earnings score distribution as a reference, I convert the assigned earnings score into percentile rank in each year (e.g., Collins and Wanamaker (2017); Chetty et al. (2014)). The 1910 national sample includes all household heads (black and white) with sons aged 0 to 17 in 1910. The 1930 national sample includes all black and white males aged 20 to 37 in 1930. Rank is used because it allows a direct comparison of one's economic status to its peers.

2.3.3 Measuring Migration Inflow 1910-30

The census records an individual's birthplace and place of residence at the census year. However, migration status is not available before 1940. Therefore, I define a "southern black migrant" as any black individual whose birthplace is in one of those 14 southern states and whose place of residence is in the North. Since states are too heterogeneous to reveal migration flows across labor markets, I examine migrant flows at the county level to focus more directly on the relevant labor markets. The *actual* migration inflow (MI) is then measured by the sum of changes in the total number of southern black migrants by state of

[16] Farmers' earnings are imputed by the product of the average 1940 value in farm laborer cell and the 1960 earning ratio of farmers over farm laborers. The ratio differs by gender and home ownership. Note that within cell differences or changes in earnings cannot be observed.

birth (s) in a northern county (c) between 1910 and 1930.[17]

$$MI_c = \sum_{s=1}^{14} (SB_{sc}^{1930} - SB_{sc}^{1910})$$

Between 1910 and 1930, the top 3 sending states are Georgia (195,999), South Carolina (133,518), and Virginia (131,117); the top 3 receiving counties are Cook County in Illinois (148,838), Wayne County in Michigan (94,672), and Philadelphia in Pennsylvania (90,741).[18] Figure 2 Panel A displays a map of all the receiving counties in the North. It is clear from the figure that migrants disproportionately settled in great urban centers.

2.3.4 Describing the Black Northerners in the Sample

While I have linked more than 24,000 individuals, only 12,623 were used in this analysis, all of whom participated in the labor force in 1930.[19] The sample counties are displayed in Figure 2 Panel B. Table 1 shows that the majority of black northerners stayed in their birth state in 1910, and very few of them lived on the farm. More than half of them were attending school; almost one-third of them were considered as light-skinned blacks (enumerated as "mulatto" in 1910 census); more than a fifth of them grew up in owner-occupied housing. They have a relatively stable family structure, and their parents were predominantly literate.

Table 2 splits the sample into two groups for comparison. High-migration includes those whose 1910 county of residence experienced above-median migration inflow. 1910 statistics are based on characteristics of the household head, while 1930 statistics are based on adulthood outcomes of the linked black northerners. There is positive selection into the migration destinations: both the employment and occupational earnings score were higher in high-migration counties in 1910. The high-migration counties are also less agricultural oriented in 1910: less than 2% of the black northerners in high-migration counties were farmers, while the share in low-migration counties is 16%. From 1910 to 1930, more black northerners en-

[17] I show in the Appendix that counties with migrants inflow were also destinations for northern black migrants. Their correlation is higher than 0.95 and is significant at 1%.

[18] The corresponding cities are Chicago, Detroit, and Philadelphia.

[19] First, I lose around 30% (7,430) of the linked individuals by restricting the sample to counties that had southern-born black residents in 1880 (discussed later) and migration inflow between 1910 and 1930. Then, I lose 3,998 individuals when restricting the sample to those in which they and their 1910 household head both had valid occupational information. Last, I lose 645 individuals by restricting the sample to counties where I have at least 10 observations.

gaged in the non-agricultural sector. Among all the occupation categories, the share of black northerners in the white-collar jobs saw the most significant expansion in high-migration counties while the most significant increase for the unskilled jobs happened in low-migration counties. Comparing between the two generations, an average black northerner's absolute economic standing is improving relative to its peers.[20] In 1930, high-migration counties were associated with lower employment and higher occupational earnings score for black northerners.

2.4 EMPIRICAL STRATEGY

In this section, I introduce the baseline estimating equation and construct an instrument for migration inflow.

2.4.1 Baseline Estimating Equation

To systematically examine how in-migration affects employment outcomes of black northerners, I estimate

$$y_{ics} = \alpha + \beta logMI_c + X_{ics}^{'}\gamma + \delta_s + \epsilon_{ics}$$

where y_{ics} is either the binary variable for employment status or the percentile rank of occupational earnings score in 1930; $logMI_c$ represents the natural log of migration inflow between 1910 and 1930 for northern county c at where individual i lived in 1910;[21] $X_{ics}^{'}$ is a vector of baseline controls measured in 1910 at the individual-, family-, and county-level (specified below); δ_s is the 1910 state of residence fixed effect. The coefficient of interest, β, thus relates the change in (log) migration inflow to the change in expected likelihood of employment or expected occupational standing of black northerners across counties within the same state. Standard errors are clustered at the county-level to reflect that northern-born blacks growing up in the same county experienced the same level of migration shock.

[20] Peer to household head is defined as all the household head (black and white) with sons aged 0 to 17 in 1910; Peer to sons is defined as all black and white males aged 20 to 37 in 1930.

[21] A binscatter plot of rank and migration inflow variable shows that the log transformation of migration inflow exhibits a linear relationship with rank and the residual is normally distributed.

The set of individual-level controls include skin color (light vs. dark skin), age, literacy (measured by both "can read" and "can write"), school, migration status before 1910 (measured by whether living in the birth state in 1910); family-level controls include his household head's occupation percentile rank, literacy, marriage status, homeownership, farm status, gender, age, and nativity; county-level controls include mean black earnings, total and black population.

2.4.2 Instrument for Migration Inflow

The Great Migration did not accelerate until the WWI led to a labor demand boom in the North. Since employment is fundamental to any migration decision, counties receiving more black migrants could have more jobs offered to black workers to begin with. As we saw in Table 2, high migration counties had better pre-migration labor markets. Therefore, the comparison across high- and low-migration counties will be biased against finding any effect on black northerners' employment rates. While the OLS estimate for occupational standing could be biased too, the likely direction of bias is unclear. It would be biased upward if migrants were attracted to growing counties with better employment options. Alternatively, the OLS results will be biased downward if migrants were selected into otherwise declining counties that weren't attracting foreign-born whites (Lemann (2011); Collins (1997)).

Although the inclusion of 1910 county characteristics listed above can account for selection to some degree, I construct a modified version of the shift-share instrument to further mitigate the endogeneity concern of migration choice (Boustan (2009); Card (2001); Altonji and Card (1991)). The instrument predicts the number of southern blacks received by northern counties between 1910 and 1930 by interacting 1880 settlements of southern blacks in the North with the pool of southern blacks "at risk" to migrate. This counterfactual measure indicates what the migration shock would have been if *all* the southern blacks aged 0-40 living in the South in 1910 had moved North. Literature suggests that younger blacks are more likely to migrate because they experience smaller psychic costs and can collect greater gains from migration over a longer period (Collins (1997); Higgs (1977); Long and Heltman (1975)). Moreover, this birth cohort is the freed generations born after the Civil

War who have developed a stronger sense of race consciousness and aspirations for social justice. The intuition behind this instrument is that (1) blacks departing the South tended to follow a settlement 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; (2) the stock of southern blacks from, for example, Mississippi in 1910 is unlikely to be influenced by the labor market condition in Chicago after 1910, while the *actual* national-level outflow from each southern state between 1910 and 1930 is more plausibly correlated with the local economic conditions in common migration destinations.

In mathematical form, MI_c from the previous equation is instrumented with

$$\hat{MI}_c = \sum_{s=1}^{14} [(\frac{SB^{1880}_{s,c}}{SB^{1880}_{s,North}}) * CS^{1910}_{s,South}]$$

where $SB^{1880}_{s,c}$ is the total number of blacks from a southern state s living in the northern county c in 1880; and $SB^{1880}_{s,North}$ is the total number of blacks from state s in the North in 1880. $CS^{1910}_{s,South}$ is the cohort size of blacks aged 0-40 born in state s and living in the South in 1910. Due to the log transformation, only the counties with a positive number of southern black residents in 1880 and with a positive migration inflow between 1910 and 1930 are considered in this paper to have consistent sample size across regression analysis.

2.4.2.1 IV Power The instrument constructed in equation above exploits two sources of variation in migration, which are not driven by other unobservable factors such as labor demand conditions. The first one is the variation in demographic patterns in the South before the migration (see Panel A in Figure 3), and the second one is the variation in the share of southern blacks from each southern state living in different counties within a northern state in 1880. Figure 3 Panel B presents an example of two primary receiving northern states (Illinois and Pennsylvania) and two primary sending states (Georgia and Virginia) to illustrate this variation.

2.4.2.2 First Stage Results and Instrument Validity For all specifications, the F-stat is very high, and the instrument has a large, positive, and significant effect on the actual migrant change.[22] It indicates that historical settlement patterns of southern blacks in the North are useful in predicting migration destinations more than 30 years later. Given the length of this lag, it is unlikely that the instrument is correlated with current labor demand conditions. Nevertheless, I include white (both immigrant and native) population change in the same county between 1910 and 1930 as a means of controlling for the potentially omitted labor demand shocks. One may still be concerned that counties with larger southern black migrants in 1880 were those having economic supremacy in 1880 and continued their reign in the subsequent decades (Jaeger et al. (2018); Goldsmith-Pinkham et al. (2018)). To alleviate this concern, I also include 1880 county characteristics (i.e., the size of total and black populations) in my preferred specification.

2.5 ESTIMATION RESULTS

In this section, I present my main results and a host of robustness checks.

2.5.1 Effects on Employment

In Panel A of Table 3, I study the effect of the Great Migration on black northerners' likelihood of employment. I start with OLS: column 1 includes state fixed effect and a host of individual-, family-, and county-level controls.[23] The point estimate on migration inflow is negative and statistically significant.

From column 2 onwards, I present 2SLS results. Column 2 replicates column 1, instrumenting the migrants change with the modified shift-share instrument introduced in Section 2.4.2. As expected, the coefficient is now larger in magnitude and remained statistically significant. Column 3 adds county characteristics in 1880 and the change of white popula-

[22]Full first stage results are reported in the Appendix.

[23]While it is a linear probability model, none of the predicted probabilities lie outside the unit interval. Horrace and Oaxaca (2006) shows that, if no (or very few) predicted probabilities lie outside the unit interval then the LPM is expected to be unbiased and consistent (or largely so).

tion between 1910 and 1930. Its point estimate implies that a one standard deviation (2.38) increase in the (log) migrant inflow decreases northern-born blacks' employment rates by 2.86 percentage points, or 3.3% of the 1930 mean for a typical recipient county. In the larger recipient county like Wayne, this effect amounts to 5.96 percentage points, or 6.87% of the 1930 mean when compared to its neighboring county Washtenaw.

To account for the possibility that in-migration may lead to an outflow of similarly skilled northern blacks, I run the same specification on the stayers. The result in column 4 remains qualitatively in line with those reported in column 3 but become larger in magnitude, confirming that displacement of northern blacks, if anything, attenuates the baseline results.

2.5.2 Effects on Occupational Standing

I next analyze the effect of the Great Migration on black northerners' occupational standing, measured by national percentile rank. As in Panel A, column 1 in Panel B reports OLS results, while subsequent columns present 2SLS estimates. In all cases, the point estimate is positive and statistically significant. Holding white population change and 1880 county characteristics constant in column 3 does not substantially change the result, suggesting that the different earnings created by the instrument flow were reasonably uncorrelated with the labor demand factors. The inclusion of these two additional controls, however, is a more rigorous way of accounting for potential spurious correlation between local economic activity and the Great Migration, and thus delivers relatively more conservative estimates.

According to my preferred specification, reported in column 3, a one standard deviation increase in (log) migration inflow raises northern-born working blacks' occupational standing by approximately 1.33 percentile ranks, or by 3.96% relative to 1930 mean.[24] By the construction of the occupational earnings score, the northern-born black males in my sample can increase their occupational earnings and thus percentile ranks by holding the same type of jobs but moving to a higher-paid census division, or by obtaining higher-paid jobs within the same division. The county where a northern-born black grew up does not necessarily correspond to the county he lived in as an adult when I measured his occupational earnings

[24]Since we cannot observe within-cell wage change, the possibility that in-migration leads to lower black wages, however, must still be borne in mind when interpreting any of the results.

in 1930. The result from column 4 using the stayer sample voids the concern that outflow of black northerners drives the impact of in-migration: the effect becomes substantially larger. Therefore, in-migration creates opportunities for black northerners to obtain higher-paid jobs.

The fact that 2SLS estimates are larger in magnitude than the OLS in both Panel A and B implies that migrants were moving to counties with more jobs but with fewer opportunities for skill upgrading.

2.5.3 Robustness Checks

I use an alternative measure OCCSCORE that were used in many other studies for occupational standing.[25] All persons with the same occupation share the same OCCSCORE regardless of his location. It is top coded at 80. The coefficients from a log-log regression of OCCSCORE on migration inflow in Panel A of Table 4 are statistically significant at the expected sign. The estimated elasticity is slightly larger than the one we obtained in Table 3 Panel C, probably because OCCSCORE fails to capture gaps by gender and home ownership status within an occupation.[26]

The construction of an occupational earnings score assigns the same score to all persons in the same occupational cell regardless of his employment status. Collins and Wanamaker (2014) finds that "black men in the North who were unemployed at the time of enumeration worked approximately 85 percent as many weeks in the previous year as those who were employed at the time of enumeration" in 1930 using the IPUMS data. Therefore, as a robustness check, I assign $EarningsScore^b$ to everyone where the unemployed workers now has only 85% of the earnings of the employed ones in the same occupational cell. Similar results as to those in Panel A are obtained in Panel B of Table 4.

Rather than distributing the state-level outflow of southern blacks, I use the national outflow of blacks from the South to the North to construct a different instrument flow measure that is purely based on the variation in the share of southern blacks in the North in 1880. The estimated effects, presented in the Appendix, are qualitatively similar and slightly

[25]Examples see Table 1 in Saavedra and Twinam (2018), where a long list of published studies using OCCSCORE is provided.

[26]The elasticity in the linear-log model is calculated as $\frac{\beta}{Y}$

larger in magnitude.

While the linked sample represents the initial linking sample overall, a few variables exhibit statistically significant influence on the probability of linkage. In the Appendix, I show that their presence did not bias my estimates: the weighted regressions deliver almost the same results as the unweighted ones in Table 3. Therefore, all other results in this paper are estimated based on the unweighted sample.

2.6 MECHANISMS

In this section, I offer support to the idea that black migrants serve as both competitors and consumers to black northerners in the labor market. I also explore the heterogeneous effect of the Great Migration, in an attempt to distinguish between winners and losers within black northerners.

2.6.1 Competitors

Did counties that experienced a more substantial migrant influx have lower employment rate for blacks to begin with? By comparing the employment status of their fathers in 1910, Figure 4 visually confirms that there is no significant difference in employment rates across these counties before the migration. So then, why did black northerners from high in-migration counties had significantly lower employment rate in 1930? Competition is one obvious explanation since southern black migrants tend to work in the same types of jobs as did the majority of the northern-born blacks when they enter the labor market in the North.

Given that southern black migrants were disproportionately unskilled workers, the competition should have concentrated in the unskilled occupations. Suggestive evidence in Table 5 supports this argument: the negative effect of in-migration on the likelihood of employment is imposed only on the unskilled northern black workers.[27] On average, northern blacks with skilled occupations have a higher employment rate than those who hold unskilled occupa-

[27]Note that a person in my sample is classified as an unskilled worker if IPUMS labeled his occupation as: operatives and kindred workers; private household workers; service workers (except private household); farm laborers and foremen; laborers (except farm and mine).

tions, and this discrepancy gets larger with in-migration since in-migration made only the unskilled northern blacks to experience a greater risk of unemployment. Additionally, given that the labor market was largely segregated by race, black migrants should have exerted little to none competitive pressure on white workers. This is supported by the evidence provided in column (5) to (8) of Table 5.[28]

The even higher unemployment rate for the stayers suggests that out-migration of black northerners alleviate the competition pressure. Surprisingly, however, Figure 5 shows that unskilled northern black workers from high in-migration counties are no more likely than their counterparts from low in-migration counties to move out, reflecting the power of economic opportunities in the high in-migration counties.

According to the simple supply and demand model, an influx of substitutable workers constitutes an increase in the supply of labor, causing wages to fall for black workers with similar skills. One potential difference between northern-born and southern-born blacks is that they might have different reference points for fair pay. For example, southern migrants tended to compare wages to what they would have expected in the South and accepted jobs not taken by the northern-born blacks (Long and Heltman (1975); Broom and Glenn (1967)). While wage changes within an occupation cannot be observed, northern-born blacks from high in-migration counties should have been less likely to participate in the labor force if the new wage scheme indeed discouraged them from working for what is not "respectable" pay. However, I find no evidence in Table 6 that in-migration affects black northerners' labor force participation rates when expanding the sample to include those not in the labor force.[29]

2.6.2 Consumers

The positive effect on occupational standing estimated in Table 3 is consistent with a recent body of the literature that documents a positive impact of immigrants on natives' occu-

[28] I linked northern-born whites using the same approach described in the data section. Wright (1986) argues that decades of occupational and industrial segregation had resulted in black and white workers occupying "noncompeting jobs," even at the entry level. Boustan (2009)) also provides evidence on racial segregation in employment within skill groups.

[29] According to IPUMS, the labor force consists of all persons defined as "employed" or "unemployed". Discouraged workers are considered not in the labor force. The 1930 census indicates whether or not a person filled out an unemployment schedule. Every person in the EMPSTAT "unemployed" category completed an unemployment schedule, regardless of whether or not they reported an occupation. Persons who did not report occupation and did not fill out an unemployment schedule are in the EMPSTAT "not in the labor force" category.

pational mobility (e.g., Tabellini (2017); Sequeira et al. (2017); Peri and Sparber (2009)). Their intuition is that the complementarity between immigrants and natives induced the latter to reallocate their labor from unskilled to skilled occupations, where they might have a comparative advantage. In line with this, Table 7 shows that black northerners from high-migration counties are more likely to become white-collar workers and less likely to become unskilled workers.

What kind of white-collar jobs did they hold? In numbers, clerical workers take the lead, followed by business and professional people and salesmen.[30] How did they obtain these white-collar jobs? One channel of entry, as mentioned in Section 2.2, is through exerting political pressure on the public sector to create positions for blacks to serve the new black communities.[31] Another channel of entry is through conducting business that caters to the black communities. Jobs such as funeral home owners and real estate agents were accessible to potential entrepreneurs before the migration, but their economic variability relies on its demographic base and intra-racial unity. The influx of migrants increased the number of black consumers; at the same time, with migration further intensifying racial discrimination, the ideology of black metropolis, racial solidarity, and self-help (black support of black enterprise) was resurgence among the black elite. Therefore, both the *number* and the *types* of black businesses and professionals are larger in high-migration counties.

Though migration expanded the demand for skilled labor, they did not remove all the barriers that had previously thwarted black business development, such as lack of capital and white support.[32] Therefore, black northerners from the better-off economic background would be more likely to be successful in these ventures. To test this hypothesis, I split the sample by their skin color (light-skinned vs. darker-skinned) and by their fathers' socioeconomic status. Light-skinned blacks were more likely to have family resources, white connections, and the ability to attract customers of both races, all of which would translate

[30] I group black business and professional people together because the line between them can sometimes be obscured. For example, a dentist could be a professional who also owns the business. This group is what Talcott Parsons has called the "market-oriented business group" (Trotter (1985)).

[31] However, they remained restricted to the least profitable and prestigious public positions such as mail carriers and post office clerks.

[32] For example, C. L. Johnson in 1931 wanted to employ unemployed black men by establishing a car parking service in downtown Milwaukee, but his plan was blocked by a hostile reaction from potential white customers. This example highlighted the importance of white assistance in developing black enterprises as well as the narrowing opportunities to build businesses catering mainly to whites (Trotter (1985)).

into greater success in seizing the employment opportunities created by in-migration.[33] A person is classified into the low SES group if his father's rank in 1910 below the median. Results from Table 8 suggests that, indeed, in-migration had greater positive impact on the occupational standing of mulattos and those from high SES families.

2.7 CONCLUSION

Peter Gottlieb (1987, p. 2) in his book, *Making Their Own Way: Southern Blacks' Migration to Pittsburgh, 1916-30*, wrote that "Southern blacks movement presaged lasting changes for the migrants, and for racial practices and habits of mind at every level of American society. Given the importance bestowed on blacks' northward migration, it is surprising how little we know about it." Today, 30 years later, our knowledge of the Great Migration's social impact is still far from complete.

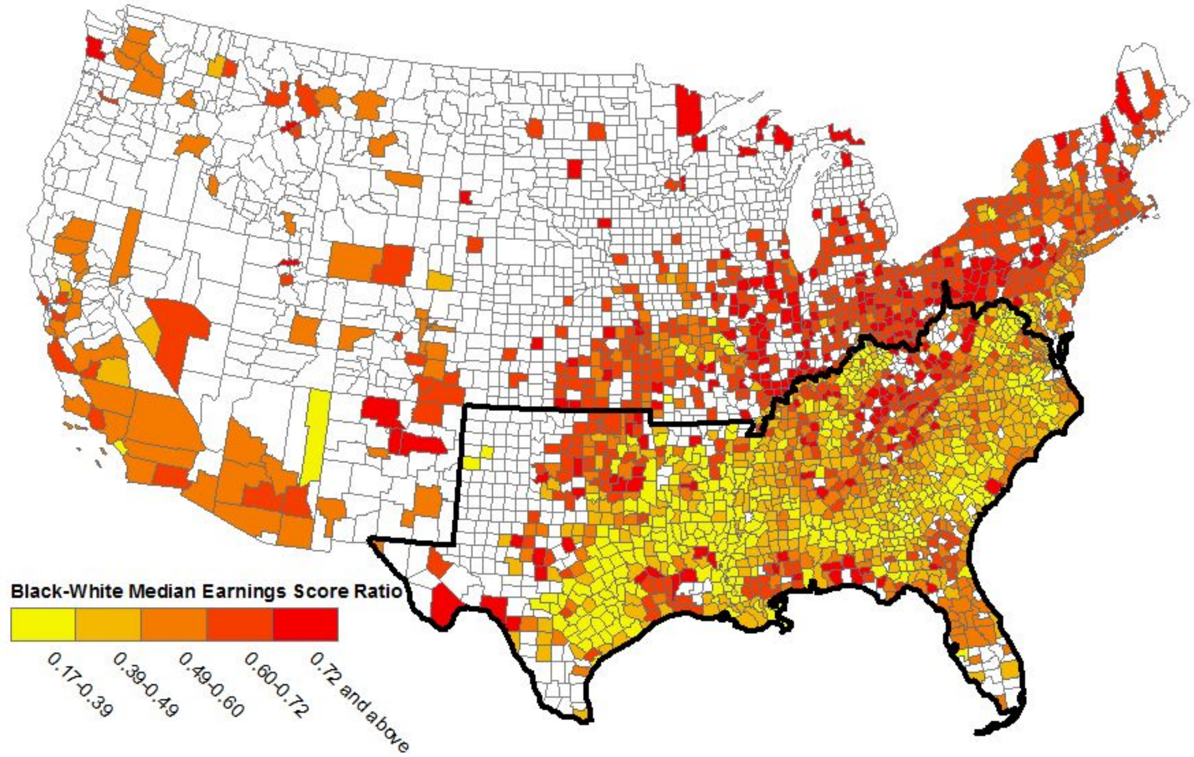
This paper contributes to a better understanding of how the Great Migration affects the labor market outcomes of blacks who had been longstanding residents of the North. Most importantly, this paper paints a nuanced picture, in which economically advantaged northern blacks benefited from better opportunity while disadvantaged northern blacks saw higher unemployment.

In contemporary America, blacks, as well as other minority groups, are often treated as a monolithic or homogeneous group. The vast majority of the existing literature makes the erroneous assumption that all blacks share the same experience and background. Little studies explain how within-group heterogeneity affects the experiences of blacks and their interactions with whites. This paper strengthens the importance of distinguishing blacks by their regional origin, socioeconomic status, family backgrounds, and skin tone. To better understand the experience of black Americans, future studies should not overlook critical within-group differences, particularly those examining the link between race and disparities in education, income, health, etc.

[33] Census in 1910 allowed individuals to be enumerated as black, white, or mixed. This trichotomous racial classification was abandoned in 1930 (Shertzer (2018)).

2.8 FIGURES AND TABLES

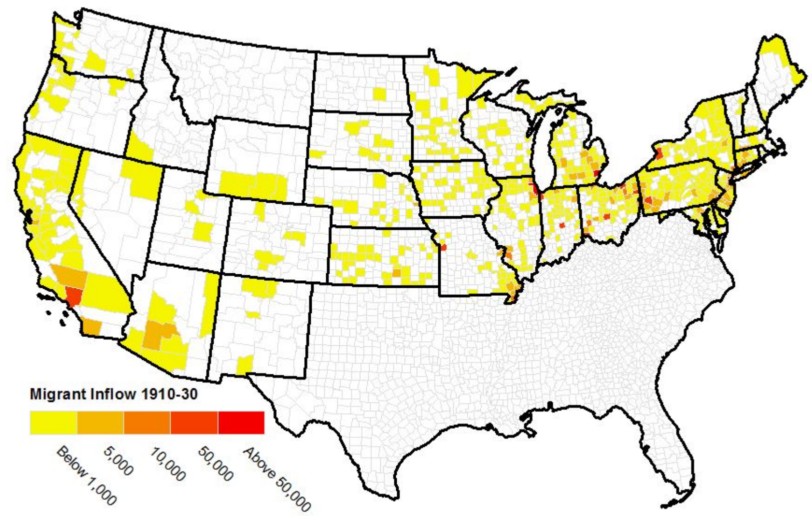
Figure 1: Black and White County Median Occupational Earning Ratio in 1910 (Before the Great Migration)



Notes: Occupational earnings is constructed by averaging 1940 income based on occupation-race-gender-region cell (Collins and Wanamaker (2017)). States with thick border are defined as the South. Only counties with at least 50 blacks and whites are considered. Data divided into five quantiles with the top quantile further divided by value 1. Source: Author's calculations using IPUMS data (Ruggles et al. (2017)).

Figure 2: Migration Inflow in Northern Counties Between 1910 and 1930

Panel A. All Counties



Panel B. Sample Counties

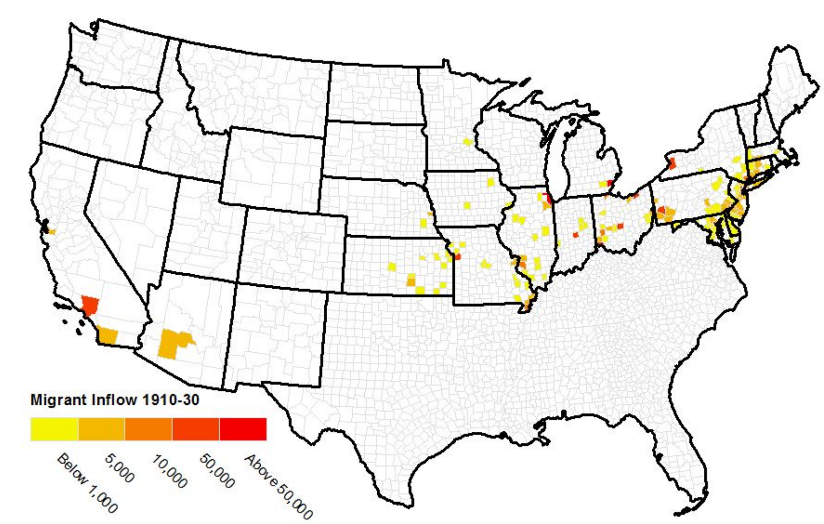
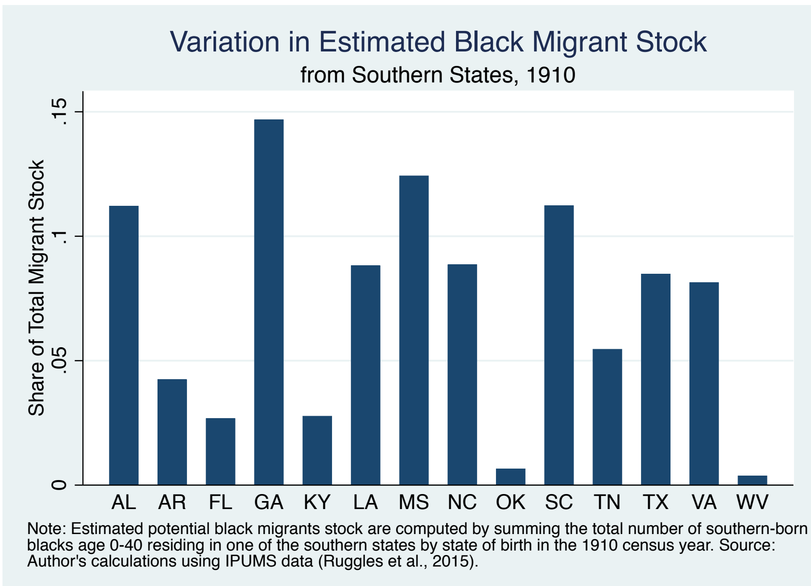


Figure 3: IV Power

Panel A



Panel B. Variation in Southern Black Settlement Within a State, 1880

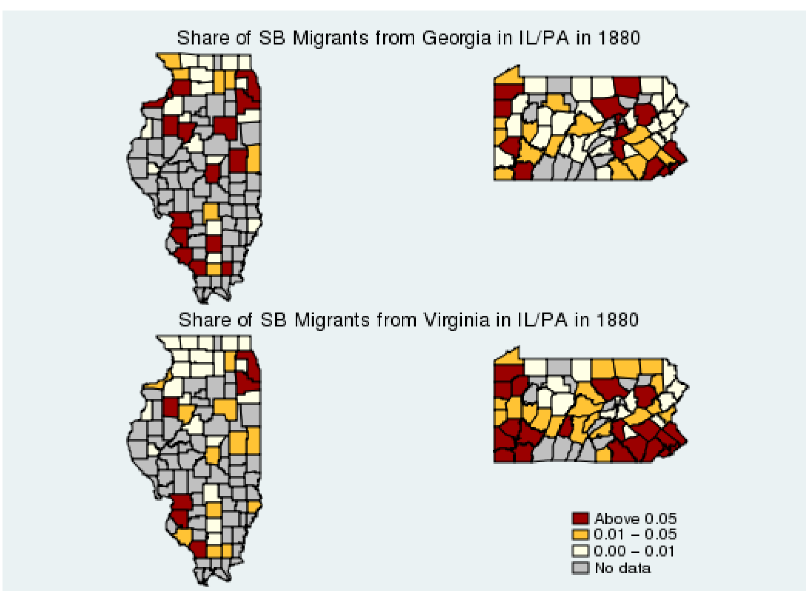
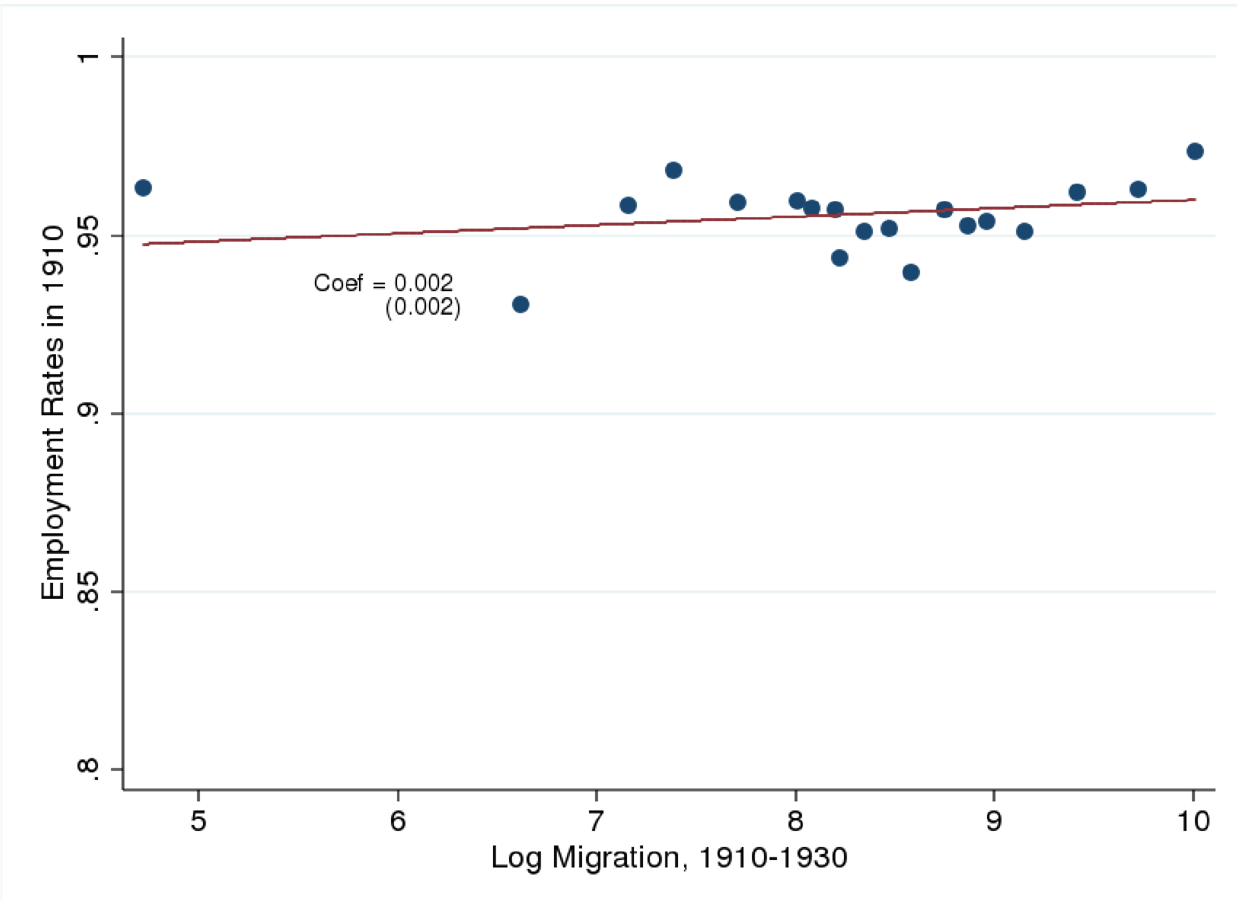
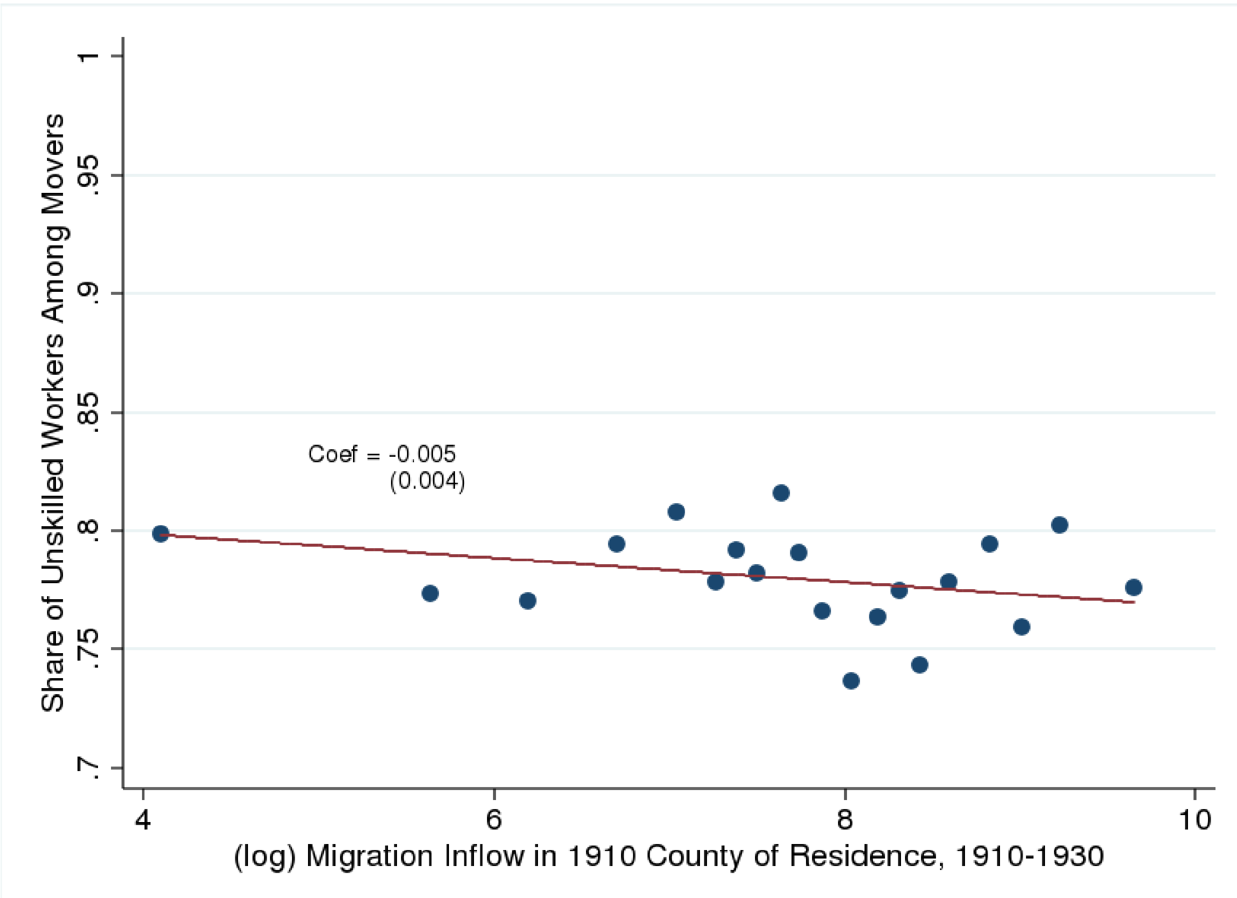


Figure 4: Employment Rates and Migration Inflow



Notes: The figure presents a binned scatter plot, grouping northern-born blacks in the sample in (log) migration inflow percentiles and plotting the share employed in the corresponding year. County-level characteristics and 1910 state of residence dummies are controlled. According to IPUMS, persons were considered employed if they were at work on the reference day, which is the day the census was taken for 1910 and the previous regular working day for 1930.

Figure 5: Share of Unskilled Movers vs. Migration Inflow



Notes: The figure presents a binned scatter plot, grouping movers in the sample in (log) migration inflow percentiles and plotting the share of movers who are unskilled workers within each group. Individual-, family-, county-level characteristics, and 1910 state of residence dummies are controlled.

Table 1: Descriptive Statistics of Black Northerners, 1910

VARIABLES	Obs.	Mean	Std. Dev.	Min	Max	Median
	Panel A. Individual Characteristics					
Age	12,623	7.755	5.068	0	17	7
Mulatto	12,623	0.300	0.458	0	1	0
Attending School	12,623	0.538	0.499	0	1	1
Stayed in Birth State	12,623	0.936	0.245	0	1	1
In Owner-occupied Housing	12,623	0.229	0.420	0	1	0
Live on Farm	12,623	0.085	0.278	0	1	0
	Panel B. Head of Household Characteristics					
Age	12,623	40.15	9.451	18	91	39
Literate	12,623	0.848	0.359	0	1	1
Married	12,623	0.908	0.289	0	1	1
Mulatto	12,623	0.261	0.439	0	1	0
Female	12,623	0.091	0.287	0	1	0

Table 2: Summary Statistics by the Degree of Migration Inflow

VARIABLES	All		High-Migration		Low-Migration	
	1910	1930	1910	1930	1910	1930
Employed	0.956	0.867	0.957	0.857	0.954	0.876
Occupational Earnings Score	694.1	777.7	731.2	817.3	657.4	738.4
Percentile Rank	29.58	33.49	32.25	36.20	26.94	30.81
Occupation Categories						
White-collar Jobs	0.072	0.129	0.098	0.169	0.046	0.089
Blue-collar Jobs	0.053	0.073	0.057	0.075	0.049	0.070
Unskilled Jobs	0.786	0.783	0.828	0.750	0.744	0.816
Farmers	0.089	0.015	0.017	0.006	0.161	0.025
Obs.	12,623		6,280		6,343	

Table 3: Migration and Black Northerners' Employment

	(1)	(2)	(3)	(4)
	OLS	2SLS	2SLS	2SLS
	Panel A. Prob. of Employment			
(log) Mig. Inflow	-0.006**	-0.013*	-0.012**	-0.027***
	(0.003)	(0.007)	(0.006)	(0.010)
Mean DV	0.867	0.867	0.867	0.864
	Panel B. Percentile Rank			
(log) Mig. Inflow	0.421***	0.603**	0.556**	1.293***
	(0.143)	(0.275)	(0.252)	(0.460)
Mean DV	33.49	33.49	33.49	33.97
	Panel C. (log) $EarningsScore^a$			
(log) Mig. Inflow	0.008***	0.014**	0.013**	0.030**
	(0.003)	(0.006)	(0.006)	(0.012)
Mean DV	6.614	6.614	6.614	6.624
Controls:				
White Pop Change			X	X
County Characteristics 1880			X	X
Stayer Only				X
Observations	12,623	12,623	12,623	7,234
County	151	151	151	149
Kleibergen Paap F-stat		30.96	41.59	34.42

county cluster-robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Note: All regressions include individual-level, family-level and county-level controls, as well as state fixed effects.

Table 4: Alternative Measures for Occupational Earnings

	(1)	(2)	(3)	(4)
	OLS	2SLS	2SLS	2SLS
	Panel A. (log) Occscore			
(log) Mig. Inflow	0.013***	0.022**	0.022***	0.039***
	(0.004)	(0.008)	(0.008)	(0.014)
Mean DV	2.953	2.953	2.953	2.961
	Panel B. (log) $EarningsScore^b$			
(log) Mig. Inflow	0.007**	0.012**	0.011**	0.025**
	(0.003)	(0.006)	(0.006)	(0.011)
Mean DV	6.592	6.592	6.592	6.601
Controls:				
White Pop Change			X	X
County Characteristics 1880			X	X
Stayer Only				X
Observations	12,623	12,623	12,623	7,234
County	151	151	151	149
Kleibergen Paap F-stat		30.96	41.59	34.42

county cluster-robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Note: All regressions include individual-level, family-level and county-level controls, as well as state fixed effects.

Table 5: Migration and Prob. of Employment by Skill Level

	Northern-born Blacks				Northern-born Whites			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	OLS	2SLS	2SLS	2SLS	OLS	2SLS	2SLS	2SLS
	Panel A. Unskilled Workers							
(log) Mig. Inflow	-0.009***	-0.015**	-0.015**	-0.032***	0.008*	0.006	0.006	0.015
	(0.003)	(0.007)	(0.006)	(0.010)	(0.004)	(0.007)	(0.008)	(0.013)
Mean DV	0.860	0.860	0.860	0.853	0.896	0.896	0.896	0.887
Observations	9,918	9,918	9,918	5,718	3,908	3,908	3,908	2,242
County	150	150	150	148	128	128	128	118
Kleibergen Paap F-stat		33.19	44.02	35.86		26.06	37.92	31.95
	Panel B. Skilled Workers							
(log) Mig. Inflow	0.003	-0.003	-0.001	-0.010	-0.001	0.005	0.002	0.003
	(0.005)	(0.011)	(0.010)	(0.020)	(0.002)	(0.007)	(0.007)	(0.010)
Mean DV	0.891	0.891	0.891	0.903	0.920	0.920	0.920	0.917
Observations	2,705	2,705	2,705	1,516	6,822	6,822	6,822	3,916
County	145	145	145	123	128	128	128	128
Kleibergen Paap F-stat		18.22	28.86	19.52		22.44	29.94	23.90
Controls:								
White Pop Change			X	X			X	X
County Characteristics 1880			X	X			X	X
Stayer Only				X				X

county cluster-robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Note: Following definition in Ferrie (1996), unskilled includes operatives and kindred workers; private household workers; service workers, except private household; farm laborers and foremen; laborers, except farm and mine. All regressions include individual-level, family-level and county-level controls, as well as state fixed effects.

Table 6: Migration and Labor Force Status

	(1)	(2)	(3)	(4)
	OLS	2SLS	2SLS	2SLS
(log) Mig. Inflow	0.003	-0.006	-0.006	-0.004
	(0.002)	(0.004)	(0.004)	(0.005)
Mean DV	0.912	0.912	0.912	0.918
Observations	13,837	13,837	13,837	7,884
County	151	151	151	150
Kleibergen Paap F-stat		31.02	41.96	35.71
Controls:				
White Pop Change			X	X
County Characteristics 1880			X	X
Stayer Only				X

county cluster-robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Note: The same set of counties from Table 3 are used in the white sample. A few counties are dropped because they do not have linked white residents or did not pass the sample restriction. All regressions include individual-level, family-level and county-level controls, as well as state fixed effects.

Table 7: Probability of Black Northerners Holding Certain Occupation in 1930

	(1)	(2)	(3)	(4)
Dependent Variable	**White Collar**	**Blue Collar**	**Unskilled**	**Farmer**
	Panel A. OLS Estimates			
(log) Mig. Inflow	0.005**	-0.002	-0.005	0.001
	(0.003)	(0.003)	(0.004)	(0.002)
	Panel B. 2SLS Estimates			
(log) Mig. Inflow	0.012**	0.003	-0.013**	-0.002
	(0.005)	(0.004)	(0.006)	(0.003)
Mean DV	0.129	0.073	0.783	0.015
Observations	12,623	12,623	12,623	12,623
County	151	151	151	151
Kleibergen Paap F-stat	41.59	41.59	41.59	41.59
Controls:				
White Pop Change	X	X	X	X
County Characteristics 1880	X	X	X	X

county cluster-robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Note: Following definition in Ferrie (1996), **white collar** workers are professional, technical, and kindred workers; managers, officials, and proprietors, except farm; clerical and kindred workers; sales workers; **blue collar** workers includes craftsmen, foremen, and kindred workers; **unskilled** workers includes operatives and kindred workers; private household workers; service workers, except private household; farm laborers and foremen; laborers, except farm and mine; **farmer** includes farmers and farm managers (including farm tenants).

Table 8: Migration and Rank by Skin Color and Father's SES Status

	Skin Color		SES Status	
	Dark	Light	Low	High
	Panel A. OLS Estimates			
(log) Mig. Inflow	0.399**	0.525**	0.306**	0.552**
	(0.171)	(0.257)	(0.154)	(0.219)
	Panel B. 2SLS Estimates			
(log) Mig. Inflow	0.493*	0.972**	0.279	0.925**
	(0.286)	(0.435)	(0.258)	(0.450)
Mean DV	32.27	36.34	30.44	35.88
Observations	8,832	3,791	5,541	7,082
County	150	141	150	147
Kleibergen Paap F-stat	46.10	19.95	29.58	41.57
	Panel C. 2SLS Estimates (Stayer Only)			
(log) Mig. Inflow	1.073**	2.000**	0.896*	1.962***
	(0.504)	(0.830)	(0.483)	(0.608)
Mean DV	32.48	37.25	29.81	36.75
Observations	4,969	2,265	2,893	4,341
County	147	124	140	135
Kleibergen Paap F-stat	39.08	14.11	26.59	29.31

county cluster-robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Note: All regressions include individual-level, family-level and county-level (including white population change and 1880 county characteristics) controls, as well as state fixed effects.

3.0 RACIAL SEGREGATION IN HOUSING MARKETS AND THE EROSION OF BLACK WEALTH

(With Prottoy A. Akbar, Allison Shertzer, and Randall P. Walsh)

"Daisy and Bill Myers, the first black family to move into Levittown, Pennsylvania, were greeted with protests and a burning cross. A neighbor who opposed the family said that Bill Myers was 'probably a nice guy, but every time I look at him I see $2,000 drop off the value of my house."'

-Ta-Nehisi Coates, We Were Eight Years in Power: An American Tragedy (2018)

"During the early nineteen twenties it is estimated that more than 200,000 Negroes migrated to Harlem... It was a typical slum and tenement area little different from many others in New York except for the fact that in Harlem rents were higher... Before Negroes inhabited them, they could be let for virtually a song. Afterwards, however, they brought handsome incomes."

-Frank Boyd, American Life Histories Manuscripts (WPA Federal Writers' Project, 1938)

3.1 INTRODUCTION

Housing is the most important asset for the vast majority of American households and a key driver of racial disparities in wealth (Blau and Graham (1990), Wolff (2014), Aladangady et al. (2017)). Social scientists have long hypothesized that racial income inequality reproduces itself in housing wealth, with minority groups who face discrimination in the labor market less able to build equity in their homes. This process yields impoverished neighborhoods that impede the health, educational attainment, and upward mobility of the next generation of black children (Ananat (2011); Wilson (2012); Chetty et al. (2014)). In this narrative, segregated neighborhoods harm the socioeconomic standing of their residents through a complex system of interrelated disadvantages.

Recent scholarship and public discourse has focused on how discrimination in credit markets served to exacerbate forces operating in labor markets and educational systems to limit black accumulation of housing wealth.[1] Far less attention has been paid to the direct role that segregated housing markets played in sapping black households of resources and inhibiting their ability to accumulate assets. Using a novel dataset of rents, home values, and the racial composition of city blocks in prewar American cities, we examine the distinct disadvantages faced by black households due to the establishment and expansion of segregated neighborhoods. We find that rental prices and occupancy rates soared when city blocks transitioned from all white to majority black. Further, when black families were able to escape these escalated rents through the purchase of their own home, these same market dynamics led to the erosion of their homes' value.

The history of racial transition in twentieth century American neighborhoods contains many references to "blockbusting," the process by which the black ghetto expanded into formerly white neighborhoods (Massey and Denton (1993)). Black families, who were desperate for quality housing, would outbid whites for apartments and homes outside but typically adjacent to the existing ghetto. Furthermore, because blacks were willing to pay a premium relative to whites for even substandard housing, formerly single-family homes could be subdivided to accommodate more residents. As highlighted by the two passages that begin this paper, the fact blacks paid higher rents stands in stark contrast to an expectation that housing prices would fall as the block transitioned. Popular accounts of the day focus extensively on the latter process and the concerns of white homeowners who expected their property to lose substantial value once black families began settling nearby.[2] These concerns fueled a white exodus to neighborhoods safe from the threat of black encroachment (Mehlhorn (1998), Boustan (2010)).

At first glance, the notion that racial transition could be associated with both increases

[1] A key focus of this work has been the role of redlining by both private lenders and federal agencies, namely the Federal Housing Authority. For instance, see Aaronson et al. (2017). Ta-Nehisi Coates cites redlining as an example of the type of government policy that deliberately inhibited the wealth accumulation of black families in his well-known essay "The Case for Reparations" in the Atlantic (2014).

[2] A recent theory paper (Ouazad (2015)) models blockbusting as a problem in which fee-motivated real estate agents trade off a decline in property values associated with black entry with an increase in commissions due to higher turnover. We cannot observe the actions of real estate agents in our setting and thus restrict our attention to prices in the housing market.

in aggregate rental prices and decreases in property values poses a puzzle. Indeed, to our knowledge, no existing empirical papers argue that the arrival of black residents caused the prices of owned and rented housing to diverge. To understand the mechanisms that could generate such a finding, we consider the capitalization rate that is implied by a no arbitrage condition between rents and home values (Kearl (1979); Poterba (1992)). Here, we formalize the capitalization rate as the percent of a propertys value that must be received as rent each year to make an investor indifferent between holding the asset and receiving rent and selling the property at its market value. Under the no arbitrage condition, all else equal, a real estate investor must be compensated by higher rents today if she is to purchase a property whose price is expected to fall in future periods. Similarly, expectations of high maintenance costs or more rapid rates of physical deterioration would require higher rents as well.

The causal effect of racial transition on rental and home prices is difficult to identify because the ghetto tended to expand endogenously into areas populated by older residents and with lower quality housing. Accordingly, our empirical approach addresses the concern that black families may have been allowed to move into blocks where home prices would have fallen even in the absence of racial transition. To facilitate the identification of the causal impact of racial transition on prices, we match the universe of household addresses from ten major northern cities across the 1930 and 1940 federal censuses to create a panel dataset of single-family homes and apartment buildings. The 1930 and 1940 censuses were the first to ask about home values and rents, and the expiration of census confidentiality rules enables us to observe the same address in both years, along with a reported rent or valuation and the race of the occupants.

This panel dataset provides several avenues for causal identification of the racial transition effect on home prices. First, to minimize the potential for omitted variable bias related to where black families already lived, our baseline sample consists of city blocks that were all white in 1930. We further restrict our attention to addresses that were single-family, owner-occupied homes at the start of the decade. Importantly, with our linked sample, we can use 1930 price as a proxy for all time-consistent unobserved address-level characteristics in a 1940 cross-sectional analysis. Linking also allows us to control for the 1930 occupancy rate and a set of city block-level characteristics such as the rental rate. In our preferred specification, we also include fixed effects for the enumeration district (a geographic area typically comprising less than four city blocks). Identification thus relies on variation in block-level racial transition within these narrowly defined neighborhoods that cannot be predicted by 1930 housing value, occupancy rate or city block-level rental share.

Impacts of racial transition are large. We find that rental prices and occupancy soared by 40 percent in blocks that transitioned from all white to majority black. In contrast, home values fell by 10 percent relative to blocks that remained all white. Further analysis suggests that increases in occupancy in homes on transitioning blocks were a key mechanism underlying these declines in value. When we focus our analysis solely on homes that saw both increases in occupancy and block-level racial transition, the estimated decline in price increases to 28 percent. In contrast, we observe sharp increases in rental prices for all houses on transitioning blocks, whether occupancy increased or not. Summarizing the impact of racial transition on the relationship between rents and prices, we estimate that the capitalization rate on blocks that became majority black was about 17 percent, compared with 11 percent on blocks that remained white.

Black families bore the brunt of both shifts in prices. When we decompose the fall in home prices by the racial composition of the block and race of the homeowner, we find that the prices fell the most during the beginning of the transition process, when black arrivals tended to be better off and were more likely to buy their home. On the other hand, rental prices increased the most after the block attained majority black status and black arrivals were typically poorer renters. Furthermore, rents only increased for black families, not white families who remained during the transition process. Finally, we document significant heterogeneity across cities with those cities that saw the largest inflows of black migrants (e.g. Chicago, Philadelphia and Detroit) experiencing the largest rent and price impacts from racial transition. In these cities, capitalization rates in blocks that transitioned to majority black exceeded 20 percent and home values on these blocks dropped by a staggering 50 percent over the first decade of transition.

Our findings have important implications for understanding racial disparities in wealth. Because of the segregated housing market in American cities, black families faced dual barriers to wealth accumulation: they paid more in rent for similar housing while simultaneously experiencing declining values of the homes they were able to purchase. The economically significant amount of wealth erosion endured by black families in the northern housing market nuances our understanding of the Great Migration, which is a key channel through which African Americans were able to improve their economic standing in prewar America (Myrdal (1944), Collins and Wanamaker (2014)). The cumulative gains in occupational standing and earnings achieved by black families may have been largely canceled out by segregated housing markets.

Meanwhile, real estate investors faced such large demand from black renter households desperate for housing outside of the already dilapidated ghetto that they were able to charge sufficiently high rental prices to overcome the expected losses due to future declines in the value of their capital. Investor pessimism was highly racialized, and we see no similar declines in housing values on blocks that saw large increases in occupancy yet remained white. The large gaps in capitalization rates that we document in this paper provide strong evidence that as of 1940, the market was already anticipating that homeowners in black neighborhoods would see their wealth severely eroded by housing market forces over the coming years.

3.2 BACKGROUND AND RELATED WORK

3.2.1 Historical Background

The Great Migration saw millions of African Americans leave the poverty and oppression of the Jim Crow South for better lives in northern cities. However, they soon discovered that the North maintained its own system of racial segregation, particularly in housing markets. Black families found themselves largely restricted to homes in the existing black ghetto through a mixture of threats, actual violence, and discriminatory real estate practices. The narrative history emphasizes collective action taken by whites to maintain the color line, which shifted over time from angry mobs in the early days of the Great Migration to the later establishment of genteel neighborhood "improvement" associations (Massey and Denton (1993)). Such associations were created in part to lower the costs of adopting restrictive covenants, which were deed provisions prohibiting the sale of a house to a black family. Such covenants had effect until 1948 when the Supreme Court struck down their enforcement in *Buchanan v. Warley*.

Still, the color line was not inviolate. The 1920s and 1930s saw significant expansions of the ghetto in most northern cities. Urban historians underscore the desperation of black families for better housing and their tendency to outbid whites for homes near the ghetto. On the other hand, real estate professionals and academics were united in their belief that black entry would harm home values.[3] Such expectations made banks reluctant to underwrite a mortgage for a "pioneer" black family entering a white neighborhood where the lending institution already held loans. One historian summarized the dichotomy thusly:

> "One of the most interesting points made in the [real estate] broker comments is the recurring theme that while sellers may not get their price from whites (who are reluctant to consider an area undergoing racial transition), they probably can from nonwhites. This is quite different from the unqualified prediction that all prices in an 'invaded' area fall" (Laurenti (1960), p. 20).

The fact that the ghetto expanded even though black families tended to have fewer assets to use for a down payment suggests that some banks did in fact underwrite mortgages for them. While banks were typically reluctant to initiate racial transition on a block, they appear willing to have made loans in neighborhoods "destined" to turn. Surveys of real estate brokers from the period suggest that the first family to enter a white neighborhood often sought a mortgage from a distant bank that did not have exposure to the area in question (Schietinger (1953), p. 172). The narrative history on the issue of mortgage terms is mixed, with some surveys finding blacks and whites received similar terms (Rapkin and Grigsby (1960), p. 77) and other scholars arguing that African American borrowers were steered towards installment contracts where they could be lose possession of their home if they were late on a single payment (Satter (2009), p. 4).

Of course, not all black families bought their own home. As we discuss below, we find that the proportion of renters increased throughout the transition process. The question of who owned properties rented out to black families is thus important for interpreting our

[3] Some social scientists had a more nuanced view of the process. For instance, Gunnar Myrdal argued in An American Dilemma that white racism was the primary cause of drops in home values as a block began transitioning and that prices should recover once the neighborhood was majority black (p. 623).

results. The census does not allow us to observe the identity of property owners in the case where the occupants are renters. We thus turn to the narrative history, which suggests white investors purchased properties in the black ghetto with the perhaps self-fulfilling expectation that their investment would sharply depreciate over time.[4] Real estate brokers believed that houses that were converted to multi-family rentals would lose value over time and were generally unwilling to make loans for the purchase of such properties (McEntire (1960), ch. xiii). It would thus be necessary to buy these properties with cash. It is also likely the case that some landlords were simply household heads who decided to rent out the former family home instead of selling. Both considerations underscore the fact that in our setting the owners of rental properties were most-likely white.

3.2.2 Related Literature in Economics

A large body of work in economics and related fields seeks to understand the causes and consequences of segregated housing markets. Of particular interest is the question of how preferences for racial residential segregation is manifested in housing prices. The consensus in the literature is that segregation that arises from constraints on black housing supply will result in black families paying higher prices for similar housing relative to whites. Indeed, most papers that examine racial housing price disparities between 1940 and 1970 have argued that blacks paid such a premium (King and Mieszkowski (1973); Yinger (1975); Schafer (1979)). The passage of the Fair Housing Act in 1968 reduced the tools available to white families to maintain the color line, and most papers working with data from after 1970 argue that segregation was maintained by whites paying a premium to avoid black neighbors (Follain and Malpezzi (1981); Chambers (1992)).

Yet establishing that black and white families paid different amounts for the same quality of housing is extraordinarily difficult. Much of the research on such differentials in the years after the Fair Housing Act was passed necessarily compares housing in very different neighborhoods because so many whites had already moved to the suburbs. The seminal paper on this topic is Cutler et al. (1999), which proposes an indirect empirical test of the

[4] See for instance United States Congress House Committee on the District of Columbia, 1935, Rent Commission: Hearings before the subcommittee on Fiscal Affairs on H.R. 3809, p. 7.

hypothesis that segregation generates price premia. They note that the black main effect on rental price is negative in every period they study, from 1940 to 1990, likely due to unobserved differences in housing quality. The authors thus draw inference from the interaction between black household and measured racial segregation in a particular city: a positive interaction term is then interpreted as evidence that blacks paid more for housing in segregated cities, hinting at the existence of supply constraints, while a negative interaction suggests that whites pay a premium. Interpreting estimates of these interaction terms, Cutler Glaeser and Vigdor conclude that blacks paid a premium in the 1940s and whites a premium by the 1990s.

In any case, the finding that the typical black family paid a premium for housing circa 1940 is difficult to square with the anecdotal literature on the impact of racial transition on property values in the early to mid-twentieth century. The history of the Great Migration makes many references to the supposed deleterious impact of black arrivals on home values in northern cities. The FHA underwriting manual emphasized maintaining the racial composition of neighborhoods for this reason (FHA, 1936). In any case, it remains necessary to reconcile the potential drop in property values associated with pre-Fair Housing Act black in-migration with the black rental premium found in other work.

Economists have recently dedicated a great deal of attention to government involvement in housing markets that may have had a discriminatory impact, particularly "redlining" in mortgage insurance (for instance, see Aaronson et al. (2017)). Beginning in 1934, the Federal Housing Authority initiated underwriting mortgages and imposed policies that would disadvantage black neighborhoods in central cities. However, FHA underwriting was still a nascent process during our sample period, particularly so in the extant neighborhoods that we study. As of the end of 1940, the FHA had underwritten only 60,339 mortgages on existing homes in the metropolitan areas of the cities we study in this paper.[5] Further, federal urban renewal policies did not begin until the 1949 Housing Act (Collins and Shester (2013); LaVoice, 2018). It is thus unlikely that federal government policies can explain the

[5] These numbers come from the FHA's Annual report for 1940 (FHA, 1941). We have been unable to identify exactly how large the metropolitan areas were for this reported data. However, as an example, the FHA reported more homes insured in the New York City Metropolitan area than it reported for the entire state of New York, suggesting that they used broad metropolitan area definitions. Thus, this number should likely be viewed as a conservative upper bound. In which case, FHA penetration into our sample would still have been quite limited as of 1940 (likely on the order of 2 to 4 percent).

findings of this paper. Instead, the FHA and subsequent federal policies likely served to institutionalize and reinforce the private market dynamics that we document.

Our paper is also related to the literature on white flight from central cities. A number of researchers have investigated the population dynamics associated with neighborhood racial turnover starting with the Schelling (1971) model of neighborhood tipping. Card et al. (2008) explored the existence of neighborhood tipping in the late twentieth century. Other scholars have explored white flight from black arrivals more generally, for instance Boustan (2010) for the postwar period and Shertzer and Walsh (2016) for the prewar period. We diverge from this literature by considering the economic toll of white flight on the black families who were left behind.

3.3 DATA

For this paper we construct a novel dataset composed of the universe of addresses in ten major cities matched across the 1930 and 1940 censuses. The sample cities are Baltimore, Boston, the Brooklyn and Manhattan boroughs of New York, Chicago, Cincinnati, Cleveland, Detroit, Philadelphia, Pittsburgh, and St. Louis. To create the set of addresses matched over time, we have developed an algorithm in the spirit of the individual matching literature (Long and Ferrie (2013); Feigenbaum (2016); Bailey et al. (2017)). However, while similar in many ways to the process of matching individuals across time, in matching addresses we are also able to leverage three additional sources of information to improve our accuracy: the structure of the census manuscripts, digitized historical street files and neighborhood geography. Our basic approach is as follows:

1. We first assign every individual living in one of our sample cities in either 1930 or 1940 an address that is consistent across all household members. If an address is missing, we impute it using another member of the household (households with inconsistent addresses are dropped).
2. We standardize street names to deal with variations of directional prefixes and typical suffixes ("First" vs. "1st", "st" vs. "Street"). We cross-reference street names using a

digitized street file for each city: if there is no corresponding street in the neighborhood in the spatial data, we drop everyone with an address on that street from the census data.

3. We conduct a series of consistency checks to identify the types of errors and omissions that are common in the address field, including making sure neighbors on the same street have addresses that change monotonically as we move down a manuscript page.
4. We retain only observations on streets that pass our quality checks and have no address inconsistencies.
5. We merge across the 1930 and 1940 census on our standardized street names and house numbers, yielding a sample of both single-family homes and apartment buildings.

Our algorithm is conservative in that we discard everyone associated with a particular address and everyone associated with an adjacent address on the manuscript when there is a potential problem with the census data, minimizing the risk of missing true occupants of a particular address in our final dataset. Because we wish to examine both occupancy rates and prices in our matched sample, developing an accurate count of household members is essential. Further details of the address data construction can be found in the Data Appendix B. Our final sample contains 591,780 unique addresses that could be located in both 1930 and 1940 from about 100,000 city-blocks across the sample cities (see Appendix Tables B1 and B2).[6] We have on average 10 to 15 addresses per city-block, depending on the city. We compare addresses that could be cleaned and matched to the universe of addresses in Appendix Table B3. There is some evidence that addresses with fewer occupants were more likely to be matched although the differences are economically small.[7]

In previous work, we constructed fine-grained, spatially-identified demographic data for neighborhoods in ten of the largest northern cities for 1900, 1910, 1920, and 1930 (Shertzer et al. (2016)). For this project we have expanded this data forward to 1940, and, using GIS software, created neighborhoods that are comparable over time across these two years.

[6] To obtain the final address-level dataset, we trim outliers that are likely transcription errors or records associated with institutionalized individuals. In particular, we drop any households with more than 10 members, any household with more than three heads, any addresses with monthly rent greater than \$100, and any addresses with a value greater than \$20,000.

[7] For instance, there were 7.51 individuals per address in the universe of addresses compared with 6.81 individuals on average in our matched addresses. Because of the large sample size, nearly every difference in Appendix Table B3 is statistically significant.

We are thus able to measure a relatively broad set of neighborhood characteristics at a small unit of geography, specifically at the level of the 1930 census enumeration district (typically around four city blocks in urban areas). Using our address data, we are further able to measure racial composition and other key variables at the city-block level. Blocks are delineated using postal service convention with street number intervals in the hundreds.

For purposes of identification, our empirical work relies primarily on a sample of single-family, owner-occupied homes located on blocks that were at least 95 percent white in 1930. We thus present summary statistics for this sample in Table 9, subdividing the sample by whether the block had begun undergoing racial transition or not (defined as having at least 10 percent black population in 1940). We first note the enormous drop in nominal home prices that accompanied the Great Depression, with homes in all blocks losing about 40 percent of their value between 1930 and 1940. Blocks that transitioned started with slightly lower average values relative to homes on blocks that did not transition ($5999 versus $6296, respectively).

The basic findings of this paper are evident in the 1940 values of homes that remained owned and rents of homes that switched being rental properties. Although homes on blocks that transitioned were cheaper in 1930 and lost proportionally more value over the next decade, average rents on these blocks were higher relative to homes on blocks that remained white ($38.95 versus $35.44, respectively). At the same time, homes on blocks that transitioned gained more occupants while homes on blocks that remained white actually saw occupancy decrease between 1930 and 1940.

Finally, to preview our basic findings and provide a visual treatment of the underlying dynamics in our data, Figures 7 and 8 present the semiparametric relationship between racial transition and rents and home prices estimated using the Robinsons double residual method (Robinson (1988)). The figures are based on our baseline matched sample of homes that were single family, owner occupied and located on a block that was at least 95 percent white in 1930. They visualize the non-parametric relationship between the level of racial transition as of 1940 (horizontal axis) and rent or price in 1940 (vertical axis), controlling parametrically for a full set of controls including the homes value in 1930.[8] We begin by

[8]That is, we estimate $lnPrice_{i40}$ = X'β + ($Black\ share_{i40}$) + ϵ_i where X_i includes controls for occupancy at the

showing the relationship between 1940 black share and log rent in Figure 7. Log rent (for houses that switched to being rentals) is flat for low black shares but then swells 30 log points, reflecting an increase of 35 percent, between 50 and 90 percent black share.

We now turn from rents to sales prices. If racial market dynamics are driven solely by supply pressures in the market for black housing, i.e. owing to the enforcement of segregated neighborhoods, we would expect the value of owner-occupied homes in black neighborhoods to experience similar increases in valuations upon racial transition. Yet, Figure 8 shows that home values in transitioned neighborhoods actually declined by 10 log points, or about 11 percent. When we decompose this relationship by the race of the owner, we see that both racial groups saw their homes lose value but that black families saw a larger decline over the range of black share. The primary goals of our empirical work are to ascertain whether these relationships are causal and to better understand the divergence between rents and owner occupied housing prices.

3.4 EMPIRICAL APPROACH

To these goals, our empirical framework models the relationship between rents, property values, and the racial composition of neighborhoods from the perspective of an arbitraging real estate investor. To fix ideas, we denote the price (rent or own) of and individual building as follows:

$$P_i = \begin{cases} annual\ rent_i, & \text{if tenure = r} \\ sale_i, & \text{if tenure = 0} \end{cases}$$

For a given owner occupied house, its price in year t is given by:

$$P_{it} = c_t * \rho_t * Q(Z_i)$$

where ρ_t is the city-specific price level at time t, Z_i is a vector of housing and neighborhood characteristics that are particular to the given house, $Q(.)$ is a quantity function that maps

address level, share renters and total number of addresses at the block level, and share black, share immigrant, share laborer, mean age, median home value, median rent, and median occupational score at the neighborhood level.

these characteristics into a unidimensional measure of service flow, and c_t is a capitalization rate that captures the equilibrium relationship between sales price and annual rent.

We follow Poterba (1992) in conceptualizing the capitalization rate as follows:

$$rent_{it} = c_t * sale_{it}$$

where c_t = i + τ_p + risk+maintenance + depreciation - appreciation and i is the interest rate and τ_p captures the relevant property taxes. Intuitively, the real estate investor scales the value of the property asset by a factor that takes into account the interest rate, property taxes, the risk premium associated with owning the property, costs for maintaining the property, expected depreciation, and appreciate net of the overall inflation rate. The capitalization rate thus captures the overall impact of these factors.

By combining the two equations above, we can derive a unified expression for P_{it}:

$$P_{it} = \rho_t * c_t^{I_{rent}} * Q(Z_i)$$

where I_{rent} is an indicator variable which equals 1 if the house is rented. Taking logs of both sides yields the following:

$$lnP_{it} = ln\rho_t + lnc_t * I_{rent} + q(Z_i)$$

where, $q(Z_i) = lnQ(Z_i)$. For convenience, we will refer this equation as (1) below. In our application, we don't directly observe characteristics Z_i, but we do observe prices in both 1940 and 1930 and can use this information to effectively control for these unobserved characterizations.

Solving the 1930 iteration of equation (1) for $q(Z_i) = lnP_{it} - ln\rho_t - lnc_t * I_{rent}$. Assuming that Z_i is time invariant, limiting our sample to houses that were owner occupied in 1930 (we relax both of these restrictions later) and substituting this expression back into the 1940 version of equation (1) yields the following expression for 1940 prices:

$$lnP_{i40} = ln\rho_{40} - ln\rho_{30} + lnP_{it} + lnc_{t40} * I_{rent40}$$

Thus, ignoring neighborhood racial transition for the moment, we have the following model:

$$lnP_{i40} = \alpha + \beta * I_{rent40} + \gamma lnP_{i30} + \epsilon_i$$

we can interpret the key coefficients in this equation as follows: α is the difference in the (logged) price levels between 1940 and 1930 and β is the logged capitalization rate in 1940. Further, inclusion of the 1930 house price effectively controls for all time-invariant house and neighborhood characteristics.[9]

To build on this basic empirical specification, we begin by limiting our sample to houses located on city blocks that were less than 5 percent black in 1930. We then generate an indicator variable for racial transition that we set equal to 1 if the block was more than 10 percent black in 1940. We then add the transition variable and its interaction with the rent indicator to the previous equation yielding our basic specification:

$$lnP_{i40} = \alpha + \beta_{trans} * I_{trans_i} + \beta_{rent} * I_{rent_i} + \beta_{transXrent} * I_{transXrent_i} + \gamma lnP_{i30} + \epsilon_i$$

In this specification, $\exp(\hat{\beta}_{trans})$ provides an estimate of the percent difference in sales prices between blocks that transitioned and those that did not. Further, $\exp(\hat{\beta}_{trans}+\hat{\beta}_{transXrent})$ provides an estimate of the percent difference in rental prices across transitioning and non-transitioning blocks.[10]

Finally, we note that certain characteristics of houses (or their neighborhoods) might change in systematic ways between 1930 and 1940. We control for this possibility in two separate ways. First, we directly include controls for a number of 1930 characteristics at the address, block, and neighborhood level that may be predictive of these systematic changes. Specifically, we control for the occupancy at the address level, share renters and total number of addresses at the city-block level, and finally at the neighborhood level we control for share black, share immigrant, share laborer, mean age, median home value, median rent, and median occupational score. Second, we drop the neighborhood-level controls (keeping the house and city-block-level controls) and instead include ED-level fixed effects. These fixed effects will absorb any time changing characteristics that are shared at the ED-level (recall that EDs in our sample are typically approximately four city blocks).

[9] One could restrict the coefficient γ to be equal to 1. However, not doing so allows for the possibility that price deflation between 1930 and 1940 varied across the distribution of housing quantities.

[10] By including additional indicator variables and interaction terms, equation (7) can be extended to provide a richer characterization of market dynamics across a broader range of racial transitions. Further, $\hat{\beta}_{rent}$ and $\hat{\beta}_{transXrent}$ allow for the recovery of effective capitalization rates in transitioned an un-transitioned neighborhoods.

3.5 RESULTS

3.5.1 Baseline Results

We begin our analysis of blockbusting by relating the change in the block-level racial composition to changes in housing prices over the 1930s. First, we illustrate the relationship between black share in 1930 and 1940 in Figure 6. Any non-zero black population share in 1930 was associated with large increases in black population share over the next decade, suggesting that any "tipping point" (Schelling (1971); Card et al. (2008)) in this context is very low. Thus, to reduce concern about omitted variable bias arising from neighborhoods that had already transitioned, in all subsequent work we restrict our sample to blocks that were still at least 95 percent white in 1930.

For our baseline specification, we consider the impact of city block-level racial change as measured by a variable that equals one if a formerly white block became majority black by 1940 and 0 otherwise. Column (1) of Panel A in Table 10 reports an empirical estimate of equation (7), restricting the sample to single-family, owner-occupied homes and controlling only for price and occupancy in 1930. The second column adds neighborhood-level controls and the third incorporates both neighborhood fixed effects as well as block-level controls for share renters and number of households in 1930. While results are generally consistent across specifications, the model presented in column (3) is the most robust in terms of controls. We therefore view it as our preferred specification.

The coefficient on the rent indicator measures the log of the capitalization rate for blocks that did not experience racial transition, which we find to be 11 percent.[11] Thus, in white neighborhoods the annual rent that a real estate investor should have expected to receive was about 11 percent of the value of the property. The coefficient on the racial transition variable implies that houses on blocks that saw an influx of blacks lost 11 percent of their value relative to blocks that remained white. Meanwhile, rents on these blocks increased by 37 percent relative to non-transitioning blocks (the exponent of the sum of the transition main effect and the interaction between rented and transition).

[11] Consistent with our discussion above, this estimate is consistent with the expectation that our study periods high levels of deflation would lead to capitalization rates that are in general larger than current-day capitalization rates that tend to center in the neighborhood of 6 or 6.5 percent (see for instance: Davis et al. (2008)).

Finally, the exponent of the sum of the racial transition coefficient and the interaction of transition and rented gives us the capitalization rate in transitioning neighborhoods, which is approximately 17 percent. Although we prefer to restrict our attention to single-family, owner-occupied homes for the purpose of identification of the transition effect, we also present results for a larger sample of addresses in column 4. Specifically, we also include buildings that were rented in 1930.[12] Our estimates are quantitatively similar, suggesting that 1930 prices and occupancy together with ED fixed effects together control for housing characteristics reasonably well.

All specifications indicate that racial transition was accompanied by falling home values, sharply increasing rents, and a substantially higher capitalization rate. The finding that rents and valuations diverged on transitioned blocks, while perhaps surprising at first, can be rationalized by investors having exceedingly pessimistic expectations regarding housing price depreciation or maintenance costs. One channel through which maintenance or depreciation costs could have been higher for buildings on blocks undergoing racial transition is through the impact of subdividing single-family housing into multiple rental units. Managing contracts with multiple households could have imposed direct costs, while the associated increased occupancy itself could have led to more rapid physical depreciation.

We explore this notion further in panel B of Table 10, which repeats the estimations from panel A with the log of aggregate occupancy as the outcome variable. The results in all four columns are again quantitatively similar. First, we note that houses that switched from being owned to rented saw increases in their aggregate occupancy of approximately 20 percent even on blocks that remained white. This increase was particularly pronounced in blocks that transitioned. The estimates from column (3) indicate that rental occupancy soared by 45 percent in homes that switched to being rented on such blocks.[13] Interestingly, the main effect of racial transition (i.e. in owner-occupied housing) is very small or negative in all specifications, suggesting that higher-income families who did not need to add members to their household to afford payments were the primary purchasers of homes during the transition process. This finding is consistent with the narrative evidence that

[12]This specification requires additional controls for tenure status in 1930. We do not include mixed-tenure or multiple owner addresses in this analysis because it is unclear how to aggregate a mix of valuations or valuations and rents into an address-level price.

[13]That is, exp(.186-.021+.204) = 1.45.

higher socioeconomic status black families were the first to arrive on a transitioning block and bought their homes rather than renting them (e.g. Massey and Denton (1993)).

These occupancy results suggest that the observed increase in capitalization rates on transitioned blocks, and the associated rent spikes, could simply be the direct result of increases maintenance or physical depreciation costs arising from higher density habitation. To examine this issue directly, in Table 11 we consider how capitalization rates varied with both occupancy and 1940 racial composition, replicating our preferred baseline log-price specification and splitting the sample between houses that experienced increased occupancy rates and those that experienced decreased or unchanged occupancy rates. The results suggest that while occupancy rates had a small impact on capitalization rates on blocks that remained white, the magnitudes are too small to explain the bulk of the rent hikes experienced in transitioning blocks. Even if we focus on an extreme subsample comprised only of addresses that gained at least four members between 1930 and 1940, the capitalization rate on blocks that remained white never exceeds 11.8 percent (see Table 16 which we discuss below). We can thus reject the notion that our results are driven entirely by density. Instead, the interaction of racial transition and higher density appears to have been uniquely associated with the divergence in the price of owned and rented housing. Thus, black families shouldered the burden of the segregated housing market on their own.

3.5.2 Heterogeneity Across Cities

The overall effects reported above mask significant heterogeneity across city types. To explore this heterogeneity, we split the sample as follows. First, we aggregate the neighborhoods in Baltimore, Cincinnati, and St. Louis into the category of "border" cities. These three cities differ from the rest of our sample in that they already had relatively large black populations as early as 1900. We split the remaining neighborhood sample in two based on the rate at which southern black families migrated into cities during the first wave of the Great Migration. Thus, Boston, Brooklyn, and Pittsburgh are categorized as low-migration cities while Chicago, Cleveland, Detroit, Manhattan, and Philadelphia are identified as high-migration cities.

Table 12 reports results for our preferred specification (with ED fixed effects) for these subgroups. The decomposition demonstrates that the drop in home values in transitioned neighborhoods was the largest in the high-migration cities, with houses on blocks that became majority black losing a staggering 54 percent of their value relative to houses on blocks that remained white. On the other hand, we find that while rental premia were also higher in high- migration cities than in low-migration in cities, black families renting on transitioning blocks in border cities paid the largest rent premia. In these cities, rented housing was 86 percent more expensive on blocks undergoing transition. These two sets of results (price and rental premia) together led the capitalization rates for housing on majority black blocks to exceed 20 percent in both border and high-migration cities. The heterogeneity in occupancy impacts are less striking, with addresses that both switched to being rentals and were located on transitioning blocks seeing the smallest occupancy increases in low-migration cities.

3.5.3 Decomposing Transition

Our initial results are based on a relatively simple characterization of the racial transition of city blocks: having moving from less than five percent black in 1930 to majority black in 1940. To develop a richer understanding of the underlying process, we explore the impact of racial transition on prices and occupancy over the full range of 1940 black share. This approach echoes our semiparametric analysis and provides insight into price dynamics on blocks that were at different stages of racial transition. Specifically, we partition our sample of blocks that were white in 1930 into four groups: those that remained white, those that had between 1 and 10 percent black population in 1940, those that had between 10 and 50 percent black population in 1940, and those that had over 50 percent black population in 1940. The results from this analysis are presented in columns (1) and (4) of Table 13. For ease of interpretation, we also present the results visually in Figures 9 and 10.

Figure 9 presents the overall effect of racial transition on prices, rents and capitalization rates. Prices (rents) are expressed relative to the 1940 level of prices (rents) on blocks that remained all white in 1940. Recall that our preferred specification includes neighborhood (ED) fixed effects along with block-level controls. Thus, identification comes from variation in block-level racial composition from within a very small neighborhood and beyond that which can be predicted by residential density and rental share. Turning first to prices, relative to houses on blocks that did not transition, houses lost little of their value in the 10 to 50 percent black range. A stark difference occurs above 50 percent black where houses lose 10 percent of their value. The same regression indicates that, relative to blocks that remained all white, rents rose by 11 percent on blocks that were 10 to 50 percent black in 1940 an effect that grows to 39 percent on blocks that transitioned to over 50 percent black. Similarly, relative to blocks that remained all white, the capitalization rate rises slowly until blocks switch to being majority black, where it exceeds 16 percent.

In Figure 10 we summarize the results from column (3) of Table 11 for occupancy, normalizing to the owner-occupancy rate in owner-occupied housing on blocks that did not transition. Consistent with our baseline analysis, we find that houses that switched from being owner occupied to rented and experienced no racial transition had on average 20 percent more occupants than did owner occupied houses on similar blocks. Houses that remained owner occupied actually saw slight declines in the number of residents as they transitioned. Conversely, rental units saw occupancy grow quickly as blocks experienced racial transition. Relative to owner occupied housing on blocks that did not transition, aggregate occupancy was 35 percent greater in rentals on blocks that were 10 to 50 percent black and 45 percent greater on majority black. These results reinforce our basic finding that subdividing of houses into high-occupancy rental units was a key component of the overall transition process.

3.5.4 Selection

One potential concern with our empirical approach is that neighborhoods that were already destined to experience declining values (or higher rents) were differentially targeted for racial expansion, even after controlling for price in 1930. Perhaps most concerning is the role played by proximity to the existing ghetto. The historical record, and our data, clearly document that proximity to the existing ghetto was a strong predictor of racial transition. If these neighborhoods were also destined to see systematic departures from price trends, for instance because of reduced city services or other forms of disinvestment, our results could be biased. The inclusion of enumeration district (ED) fixed effects in our preferred specification is largely a response to this concern as they will control for all factors effecting prices and that are constant over small neighborhoods. However, it is still possible that even differences in ghetto proximity across a few city blocks could lead to selection problems.

To investigate this possibility, we geocoded our sample of city blocks.[14] This geocoded subsample allows us to directly test the efficacy of our ED fixed effects in controlling for ghetto proximity. Appendix Figure B1 presents a visualization of our geocoded blocks for Detroit, which is typical of all of our sample cities. A limitation of our geocoding is that we were only able to geocode approximately 87 percent of our sample. One concern is that this subsample will vary systematically from our main sample as addresses that were targeted for urban renewal and demolition in the sixties and seventies may be overwhelming represented in the set of addresses that could not be geocoded. Columns (2) and (5) of Table 13 replicate the models of columns (1) and (4) on the geocoded subsample, and showing this concern to be valid. While qualitatively similar to the full-sample estimates, in the geocoded sample the rental premia on majority black blocks are smaller. Thus, it is important to focus within the geocoded subsample when assessing the impact of controls for distance to the ghetto on our coefficients of interest. Columns (3) and (6) add a control for distance to the nearest ghetto (defined as miles to an enumeration district that was at least 15 percent black) to the model. Comparing these results to those in columns (2) and (5) demonstrates that while ghetto distance is negatively associated with price, all other coefficient estimates are virtually identical, suggesting that the inclusion of enumeration district fixed effects provides sufficient controls for this source of selection bias.

Distance to the ghetto is not the only potential source of concern. The historical record suggests other factors also predicted selection into blockbusting. Table 14 presents average 1930 characteristics for our sample of blocks that were less than 5 percent black and had at least one owner-occupied single-family home by various stages of racial transition in 1940. While distance to nearest ghetto is by far the best predictor of racial transition, other sources of selection are also evident. For instance, the average age of household heads in 1930 is

[14] See the Data Appendix B for a description of this process.

two years higher in blocks that transitioned (47.5 versus 45.5), which is consistent with the narrative on blockbusting. In contrast to the literature, we do not find that per person rents were ex ante lower in blocks that would end up transitioning in 1930. This finding could be due to our sample restrictions, however, which exclude blocks comprised entirely of rentals or apartments. Appendix Figure B2 visualizes selection into racial transition again in the city of Detroit. While the majority of blocks that transition are near existing majority-black blocks, it is not true for all of them.

As a final test of our fixed effects strategy, we investigate whether any of these factors is also predictive in our baseline empirical approach. Table 15 presents the results of a block-level estimation of the determinants of racial transition for blocks that had at least one owner-occupied single-family home and were less than five percent black in 1930. We include ED fixed effects in addition to controls for household head age, share laborer, foreign-born share, average rent per person, homeownership share, and distance to nearest ghetto (the latter for geocoded blocks only). None of the reported predictors is economically and statistically significant in either the full or geocoded samples, whether we measure black share continuously (columns 1-3) or with an indicator for majority black (columns 4-6). We take the results of Table 15 as strong evidence that our price and occupancy results are driven by racial transition and not by other factors.

3.6 DISCUSSION

We have thus far shown that black arrivals on a block caused rents and home values to diverge, with increases in occupancy in addresses that became rentals. An important question for understanding the implications of these results is how much of the drop in home values was borne by black versus white households. For instance, if pioneer black families mainly rented their homes and waited to buy until prices had fallen, the decline in values associated with racial transition could have had a "silver lining" for black homeownership (Boustan and Margo (2013)). However, black families would still have faced high rents and declining public and private services in their neighborhoods.

We explore the question of homeownership by race in Figure 11 where we reproduce our race-specific semiparametric regressions with homeownership as the dependent variable. Panel A shows that black households were far more likely to buy houses if they arrived early in the transition process: about three-quarters of black residents were homeowners on blocks that were ten percent black but only a third were homeowners on blocks that were majority black. Meanwhile, about 70 percent of white families residing on the average block owned their homes throughout the entire transition process (Panel B). Because large declines in home values did not occur until blocks were majority black (see Figure 8), these results suggest that proportionally more homes switched hands from white to black families prior to the drop in values that accompanied racial transition. Rents did not sharply increase until a block was more than 50 percent black, when Figure 11 shows that most black families were renters. As a result, black families appear to have borne the worst of both the spike in rents and the fall in home values.

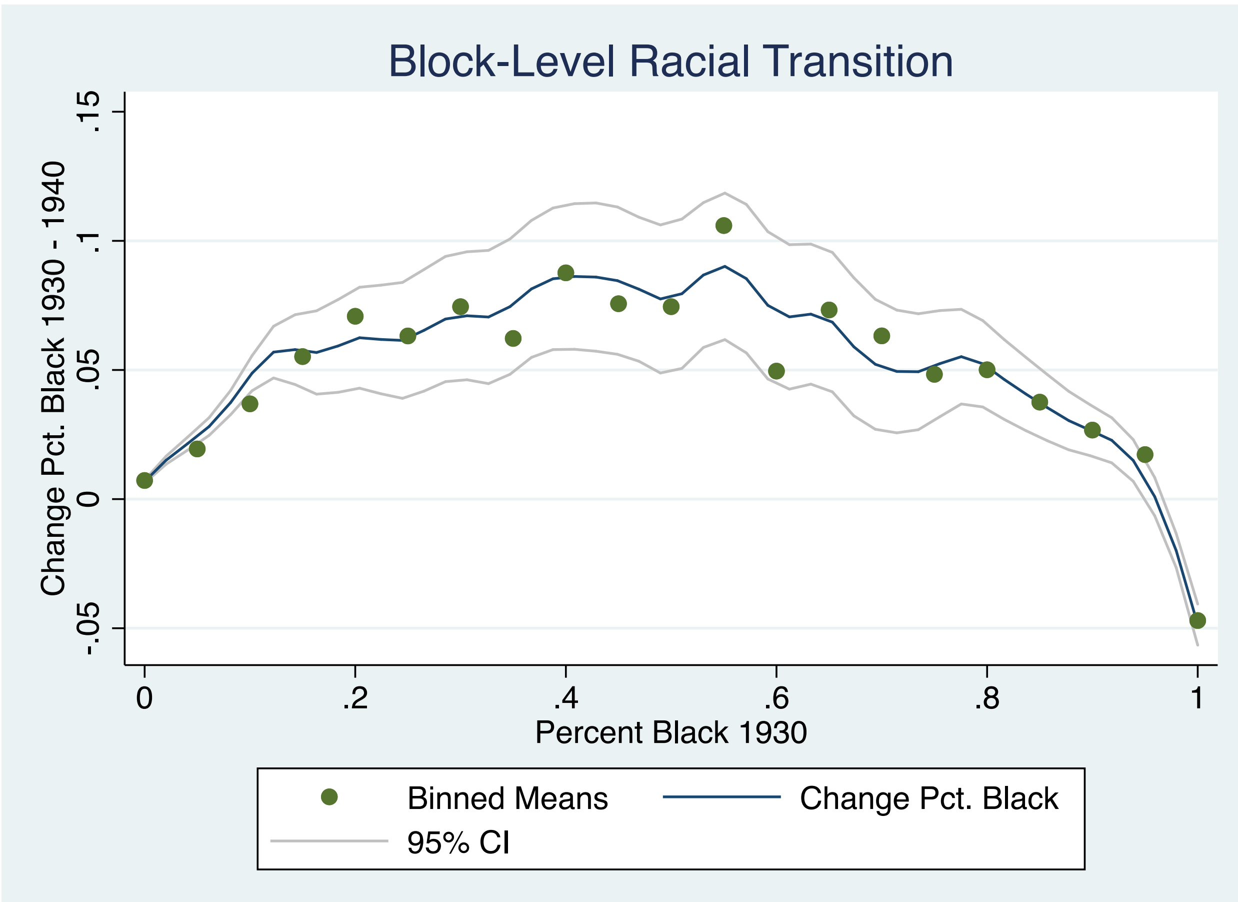
The question of who owned the rental housing in black neighborhoods is important for understanding our findings given that only 30 percent of families were homeowners once the racial transition process was complete. Census data do not allow us to observe the landlord of rented buildings. However, we can speculate on the incentives facing real estate investors by considering the degree to which the relatively low income of black families, which required them to live more densely in both owned and rented housing, can explain investor pessimism. Specifically, we explore in more detail the role of density in determining capitalization rates, focusing solely on blocks that remained white in 1940.

In Table 16, we replicate our basic specifications from Table 10 on subsamples of addresses that saw increases in occupancy but remained white through 1940. Here, we only report the coefficient on the "rented" indicator variable. Even if we limit the analysis to addresses that gained at least four members between 1930 and 1940, the capitalization rate implied by these coefficients never exceeds 11.8 percent. Clearly, even large density increases in homes on whites blocks did not translate into the same high capitalization rates observed in homes on transitioning blocks. We thus reject the notion that our results are all driven by density. Investor pessimism was highly racialized, and drops in prices only occurred when increases in density were accompanied by the transition from white to black occupants.

3.7 CONCLUSION

Our evidence suggests a blockbusting process that is consistent with the stylized facts from the narrative history. First, racial turnover was associated with rapid and dramatic falls in home values. Real estate investors, faced with the costs of creating and maintaining rental units that were going to depreciate in value and a ready supply of black households desperate for housing outside of the already underserved ghetto, were able to charge high enough rental prices to make their investment worthwhile. These processes overlapped and fueled the blockbusting process, which saw entire sections of cities transition from being all white to majority black over a relatively short period, with devastating results for black household wealth.

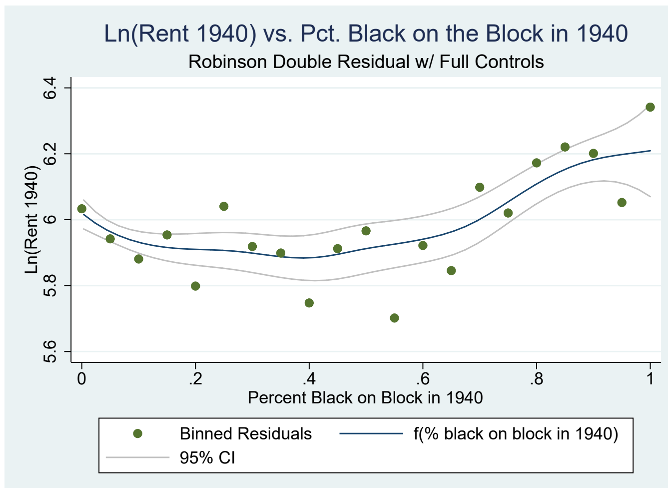
Figure 6: Relationship between 1930 Black Share and Black Population Growth



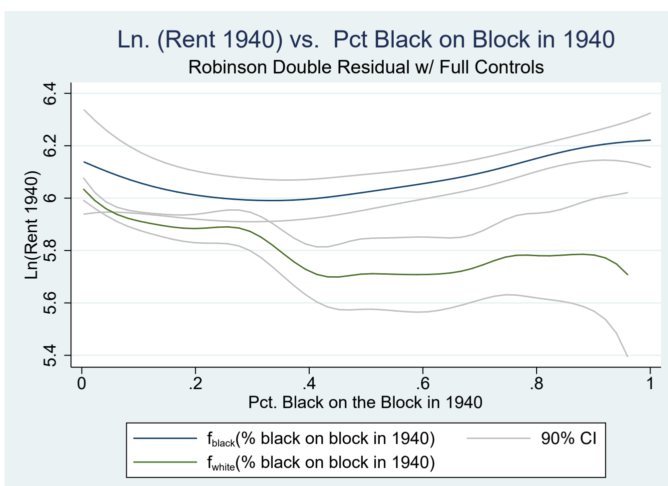
Notes: The figure presents a local polynomial smooth of block-level percent black in 1930 against the percent change in black share on the block over the next decade for every block in our sample.

Figure 7: Semiparametric Relationship Between Percent Black and Rents

Panel A. Baseline Sample



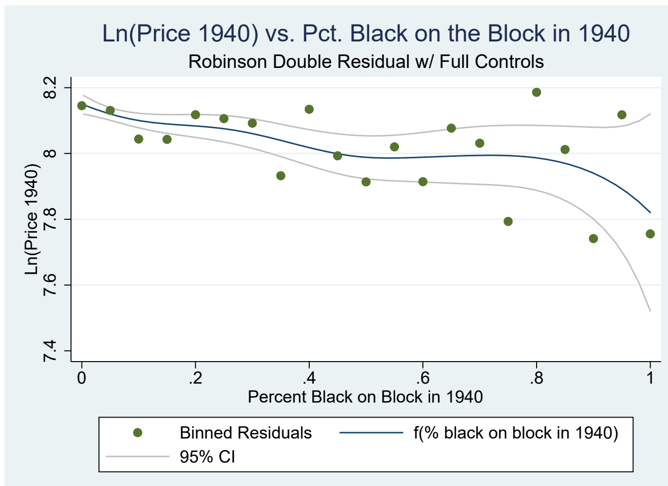
Panel B. Black and White Households Separately



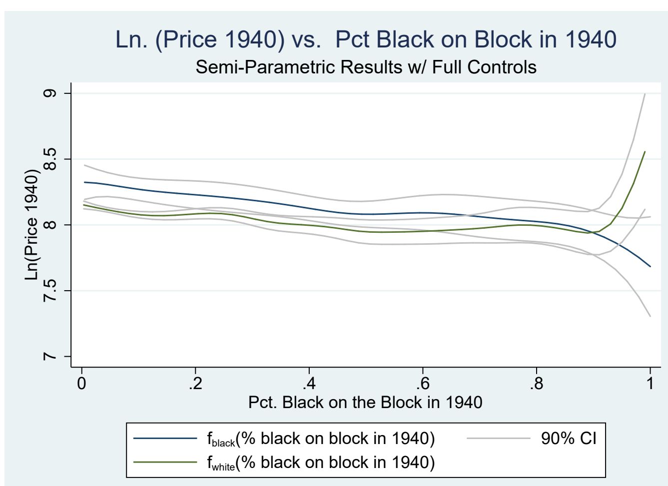
Notes: These figures show the semiparametric relationship between percent black on the block in 1940 (independent variable) and log rent in 1940 (dependent variable) on our baseline sample of homes that were single family, owner occupied, and located on a block that was at most 5 percent black in 1930. Controls are included for 1930 price and occupancy at the address level, share renters and total number of addresses at the block level, and share black, share immigrant, share laborer, mean age, median home value, median rent, and median occupational score at the neighborhood level. The estimation method is Robinsons double residual method (1988). We also include binned residuals from the regression on each chart.

Figure 8: Semiparametric Relationship Between Percent Black and Home Values

Panel A. Baseline Sample

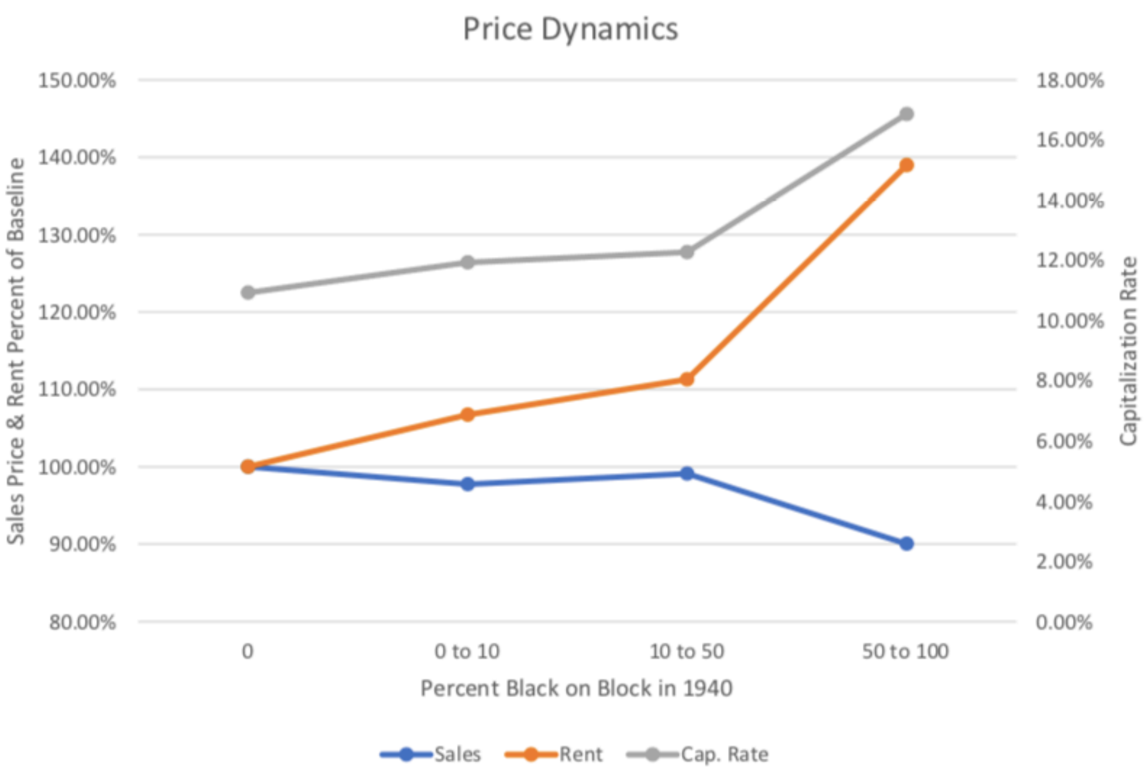


Panel B. Black and White Households Separately



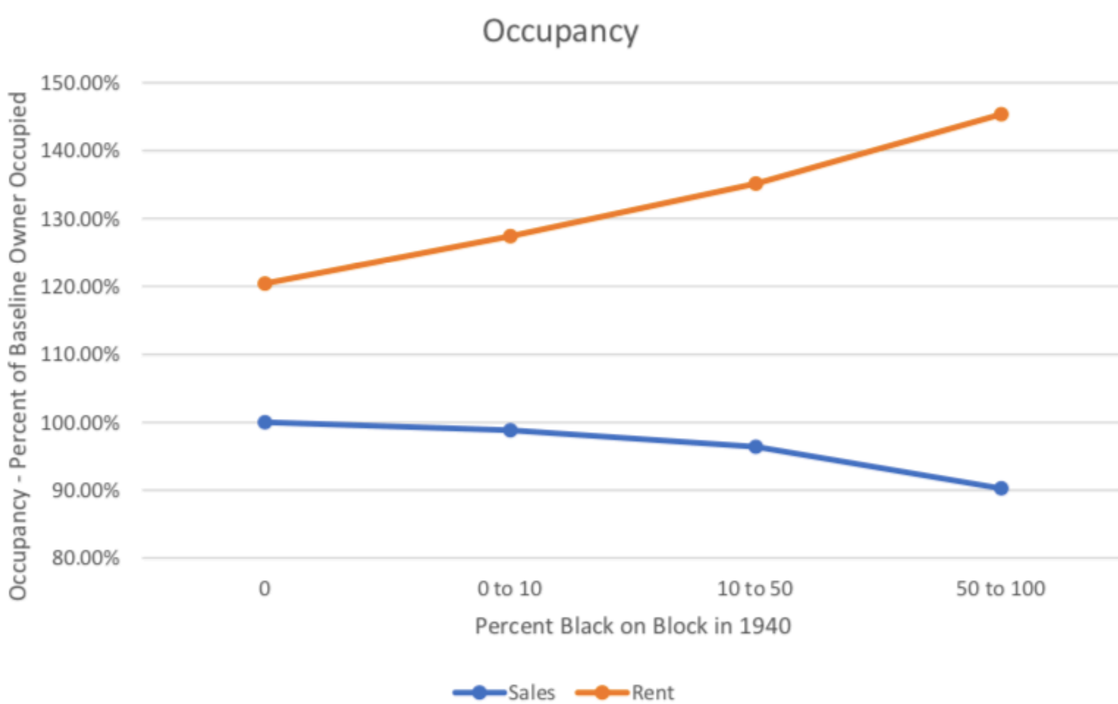
Notes: These figures show the semiparametric relationship between percent black on the block in 1940 (independent variable) and log home price in 1940 (dependent variable) on our baseline sample of homes that were single family, owner occupied, and located on a block that was at most 5 percent black in 1930. Controls are included for 1930 price and occupancy at the address level, share renters and total number of addresses at the block level, and share black, share immigrant, share laborer, mean age, median home value, median rent, and median occupational score at the neighborhood level. The estimation method is Robinsons double residual method (1988). We also include binned residuals from the regression on each chart.

Figure 9: Effect of Racial Transition on Prices



Notes: The figure shows the effects from the estimation of equation (7) presented in column (1) of Table 12 on our baseline sample of homes that were single family, owner occupied, and located on a block that was at most 5 percent black in 1930. The regression includes controls for the 1930 occupancy and price at the address level, block-level controls for number of households and share renters, and ED fixed effects. Both the overall price and rent effects are scaled relative to a house that remains owned on an all-white block.

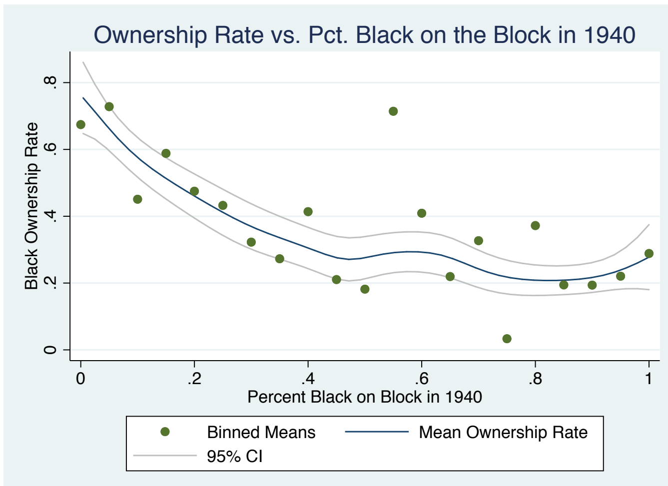
Figure 10: Effect of Racial Transition on Occupancy



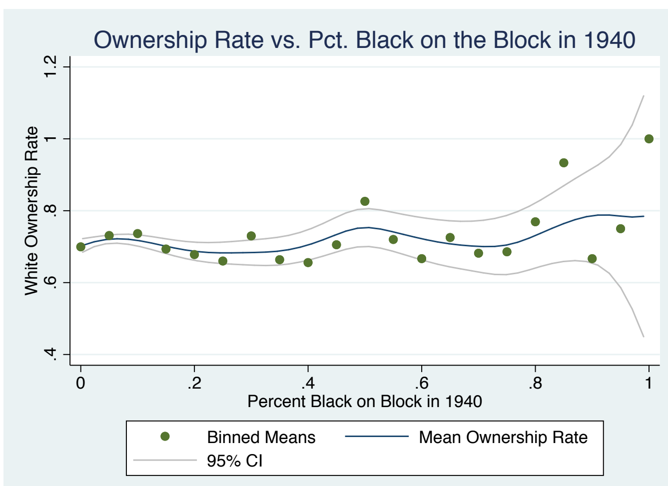
Notes: The figure shows the effects from the estimation of equation (7) presented in column (4) of Table 12 on our baseline sample single-family, owner-occupied homes located on blocks that were less than five percent black in 1930. The regression includes controls for the 1930 occupancy and price at the address level, block-level controls for number of households and share renters, and ED fixed effects.

Figure 11: Ownership Rate vs. Pct. Black on the Block in 1940

Panel A. Black Ownership Rate



Panel B. White Ownership Rate



Notes: These figures show the semiparametric relationship between percent black on the block in 1940 (independent variable) and log rent in 1940 (dependent variable) on our baseline sample of homes that were single family, owner occupied, and located on a block that was at most 5 percent black in 1930. Controls are included for 1930 price and occupancy at the address level, share renters and total number of addresses at the block level, and share black, share immigrant, share laborer, mean age, median home value, median rent, and median occupational score at the neighborhood level. The estimation method is Robinsons double residual method (1988). We also include binned residuals from the regression on each chart.

Table 9: Summary Statistics for Addresses in Baseline Matched Dataset

	Block did not transition		Block did transition	
	1930	1940	1930	1940
Nominal home value	6296.01	3901.78	5999.48	3591.37
	(3496.97)	(2336.39)	(3785.25)	(2355.28)
Aggregate monthly rent	-	35.44	-	38.95
	-	(24.79)	-	(31.47)
Aggregate occupancy	4.32	4.17	4.57	4.95
	(1.85)	(2.05)	(2.08)	(3.01)
Aggregate households	1.00	1.06	1.00	1.25
	(0.05)	(0.39)	(0.05)	(0.81)
Number of owner-occupied houses	256,471	194,633	2,726	1,560
Number of rental houses	-	61,838	-	1,166

Note: This table reports statistics on our baseline sample of homes that were single family, owner occupied, and located on a block that was at most 5 percent black in 1930 and could be matched across the 1930 and 1940 censuses. Transition is defined as having at least 10 percent black population in 1940.

Table 10: Main Results: Price and Occupancy

Panel A: Log price	No Controls	Controls	ED FE	All Obs FE
	(1)	(2)	(3)	(4)
Rented	-2.239***	-2.232***	-2.211***	-2.154***
	(0.003)	(0.003)	(0.003)	(0.002)
Racial Transition	-0.204***	-0.130***	-0.113***	-0.147***
	(0.034)	(0.033)	(0.038)	(0.032)
Rented x Transition	0.394***	0.416***	0.430***	0.327***
	(0.044)	(0.042)	(0.043)	(0.033)
Observations	248,560	247,896	248,560	414,223
R-squared	0.737	0.753	0.801	0.817

Panel B: Log occupancy	No Controls	Controls	ED FE	All Obs FE
	(1)	(2)	(3)	(4)
Rented	0.170***	0.185***	0.186***	0.189***
	(0.002)	(0.002)	(0.002)	(0.002)
Racial Transition	0.010	-0.018	-0.021	-0.047*
	(0.026)	(0.023)	(0.029)	(0.026)
Rented x Transition	0.248***	0.244***	0.204***	0.156***
	(0.033)	(0.030)	(0.033)	(0.027)
Observations	259,197	252,424	253,375	422,409
R-squared	0.025	0.212	0.252	0.380

Note: The table reports the OLS estimation of equation (7) on our baseline sample of homes that were single family, owner occupied, and located on a block that was at most 5 percent black in 1930. The first column controls only for price and occupancy of the address in 1930. The second column adds controls share renters and total number of addresses at the block level, and share black, share immigrant, share laborer, mean age, median home value, median rent, and median occupational score at the neighborhood level. The third column drops the neighborhood controls and includes ED fixed effects. The last column adds addresses that were rented in 1930 to the sample and uses the specification from column (3) with an additional control for tenure status in 1930. The "rented" variable is an indicator for whether the house switched to being a rental in 1940. The transition indicator is equal to one if the block became more than 50 percent black by 1940.

Table 11: Capitalization Rates by Occupancy Change

All Addresses			
Percent black on block in 1940	Sales	Rent	Cap. Rate
0	100.00%	100.00%	10.93%
0-10%	97.73%	106.72%	11.93%
10-50%	99.10%	111.29%	12.27%
50-100%	90.03%	138.96%	16.86%
Observations			248,560
Occupancy Increased			
Percent black on block in 1940	Sales	Rent	Cap. Rate
0	100.00%	100.00%	11.28%
0-10%	96.27%	110.52%	12.95%
10-50%	95.89%	112.30%	13.21%
50-100%	72.33%	116.18%	18.12%
Observations			73,905
Occupancy Decreased/Constant			
Percent black on block in 1940	Sales	Rent	Cap. Rate
0	100.00%	100.00%	10.33%
0-10%	99.60%	100.00%	10.37%
10-50%	102.02%	100.50%	10.18%
50-100%	100.30%	134.99%	13.90%
Observations			174,655

Note: The table reports the implied capitalization rates from an OLS estimation of equation (7) on our baseline sample of homes that were single family, owner occupied, and located on a block that was at most 5 percent black in 1930. Regressions include controls for price and occupancy of the address in 1930, share renters and total number of addresses at the block level, and ED fixed effects. See text for details on how to compute the capitalization rate from regression coefficients.

Table 12: Heterogeneity by City Type

Panel A: Log price	All Cities	Border	High Mig.	Low Mig.
	(1)	(2)	(3)	(4)
Rented	-2.211***	-2.098***	-2.252***	-2.254***
	(0.003)	(0.005)	(0.008)	(0.003)
Racial Transition	-0.113***	0.096	-0.766***	-0.202***
	(0.038)	(0.066)	(0.184)	(0.046)
Rented x Transition	0.430***	0.525***	0.707***	0.256***
	(0.043)	(0.073)	(0.178)	(0.055)
Observation	248,560	73,286	29,797	145,477
R-squared	0.801	0.757	0.785	0.826

Panel B: Log occupancy	All Cities	Border	High Mig.	Low Mig.
	(1)	(2)	(3)	(4)
Rented	0.186***	0.212***	0.264***	0.160***
	(0.002)	(0.004)	(0.006)	(0.003)
Racial Transition	-0.021	0.073	-0.057	-0.085**
	(0.029)	(0.047)	(0.141)	(0.039)
Rented x Transition	0.204***	0.226***	0.242*	0.162***
	(0.033)	(0.052)	(0.137)	(0.045)
Observation	253,375	76,887	30,543	145,945
R-squared	0.252	0.246	0.309	0.239

Note: The table reports the OLS estimation of equation (7) on our baseline sample of homes that were single family, owner occupied, and located on a block that was at most 5 percent black in 1930. Regressions include controls for price and occupancy of the address in 1930, share renters and total number of addresses at the block level, and ED fixed effects. The transition indicator is equal to one if the block became more than 50 percent black by 1940. Border cities are Baltimore, Cincinnati, and St. Louis. High-migration cities are Chicago, Cleveland, Detroit, Manhattan, and Philadelphia. Low-migration cities are Boston, Brooklyn, and Pittsburgh. See text for more detail on this classification.

Table 13: Results for Racial Transition and Proximity to Ghetto

	Dependent variable = log price			Dependent variable = log occupancy		
	All blocks	Geocoded	Geocoded	All blocks	Geocoded	Geocoded
	(1)	(2)	(3)	(4)	(5)	(6)
Rented	-2.214***	-2.222***	-2.222***	0.186***	0.185***	0.185***
	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)
Transition 1-10%	-0.023**	-0.018	-0.018	-0.012	-0.014	-0.014
	(0.011)	(0.012)	(0.012)	(0.008)	(0.009)	(0.009)
Rented x 1-10%	0.088***	0.070***	0.070***	0.068***	0.074***	0.074***
	(0.018)	(0.019)	(0.019)	(0.014)	(0.015)	(0.015)
Transition 10-50%	-0.009	-0.057**	-0.055**	-0.037**	-0.044**	-0.044**
	(0.019)	(0.022)	(0.022)	(0.015)	(0.018)	(0.018)
Rented x 10-50%	0.116***	0.109***	0.109***	0.152***	0.153***	0.153***
	(0.027)	(0.031)	(0.031)	(0.021)	(0.024)	(0.024)
Transition 50-100%	-0.105***	-0.233***	-0.231***	-0.026	-0.053	-0.053
	(0.038)	(0.049)	(0.049)	(0.029)	(0.038)	(0.038)
Rented x 50-100%	0.434***	0.353***	0.353***	0.214***	0.136***	0.136***
	(0.043)	(0.056)	(0.056)	(0.033)	(0.044)	(0.044)
Distance to Near. Ghetto			0.102***			-0.008
			(0.012)			(0.010)
Observations	248,560	217,124	217,124	253,375	221,472	221,472
R-squared	0.801	0.804	0.804	0.248	0.251	0.251

Note: The table reports the OLS estimation of equation (7) on our baseline sample of homes that were single family, owner occupied, and located on a block that was at most 5 percent black in 1930. Regressions include controls for price and occupancy of the address in 1930, share renters and total number of addresses at the block level, and ED fixed effects.

Table 14: Selection into Racial Transition in Baseline Sample

	Share Black on Block in 1940		
Mean on block in 1930:	<10%	10-50%	>50%
Average age of heads of HH	45.53	47.36	47.66
Share laborer heads of HH	0.07	0.09	0.06
Share foreign born heads of HH	0.38	0.45	0.39
Average rent per person	4.32	5.01	5.46
Ownership share	0.09	0.06	0.09
Distance to nearest ghetoo	1.45	0.29	0.07
Share black in 1930	0.00	0.01	0.01
N	41378	415	190

Note: This table reports statistics on our baseline sample of homes that were single family, owner occupied, and located on a block that was at most 5 percent black in 1930 and could be matched across the 1930 and 1940 censuses.

Table 15: Predicting Racial Transition in Baseline Sample

	Percent Black in 1940			Percent Black in 1940 >50%		
Block characteristics in 1930:	(1)	(2)	(3)	(4)	(5)	(6)
Average age of heads of HH	-0.001**	-0.000	-0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Share laborer heads of HH	0.000	0.000*	0.000*	0.001**	0.001***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Share foreign born heads of HH	0.001**	0.000	0.000	0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Average rent per person	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Ownership share	-0.000	0.000	0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Distance to nearest ghetto			-0.005*			-0.002
			(0.002)			(0.003)
Sample	All	Geo.	Geo.	All	Geo.	Geo.
Observations	41,968	35,248	35,248	41,968	35,248	35,248
R-squared	0.635	0.665	0.665	0.602	0.623	0.623

Note: This table reports OLS estimations of selection into racial transition using our baseline sample of blocks that had at least one owner-occupied, single-family home and were at most 5 percent black in 1930.

Table 16: Rental Indicators for Addresses on Blocks that Remain White

	No Controls	Controls	ED FE	All Obs FE
	(1)	(2)	(3)	(4)
All addresses	-2.240***	-2.234***	-2.213***	-2.156***
	(0.003)	(0.003)	(0.003)	(0.002)
Addresses that gained at least 2 members	-2.172***	-2.160***	-2.168***	-2.102***
	(0.009)	(0.009)	(0.009)	(0.007)
Addresses that gained at least 3 members	-2.150***	-2.143***	-2.170***	-2.093***
	(0.013)	(0.012)	(0.015)	(0.011)
Addresses that gained at least 4 members	-2.116***	-2.105***	-2.141***	-2.067***
	(0.019)	(0.018)	(0.024)	(0.017)

Note: The table reports the OLS estimation of equation (7) on our baseline sample of homes that were single family, owner occupied, and located on a block that was at most 5 percent black in both 1930 and 1940. The first column controls only for price and occupancy of the address in 1930. The second column adds controls share renters and total number of addresses at the block level, and share black, share immigrant, share laborer, mean age, median home value, median rent, and median occupational score at the neighborhood level. The third column drops the neighborhood controls and includes ED fixed effects. The last column adds addresses that were rented in 1930 to the sample and uses the specification from column (3) with an additional control for tenure status in 1930. The table reports the coefficient on the "rented" variable, which is an indicator for whether the house switched to being a rental in 1940. The exponent of this coefficient yields the relevant capitalization rate.

4.0 INTERGENERATIONAL MOBILITY IN PREWAR AMERICA: A COMPARISON BY RACE AND REGION

4.1 INTRODUCTION

Either geographical or racial differences in the rates of intergenerational mobility in the United States has been well documented in the existing literature (Collins and Wanamaker (2017); Chetty et al. (2014); Mazumder (2014); Hertz (2009)). However, studies that compare the rates of intergenerational mobility by both race and region are rare. The only one that I am aware of is Davis and Mazumder (2018), in which they use the 1979 National Longitudinal Survey of Youth data to investigate the relationship between geographic differences in intergenerational mobility and the geographical distributions of different racial groups. They find that gaps in intergenerational mobility by race are significantly larger than those by region.

But the 1970s marks the end of the Great Migration (1910-1970), during which millions of blacks flee their homeland in search of better opportunities for themselves and their children. Black populations have been significantly re-distributed from rural South to the urban North by 1970. Recent research by Derenoncourt (2019) shows that racial composition changes during the peak of the Great Migration (1940-1970) increased the upward mobility gap between blacks and whites in the North.

Therefore, in this paper, I ask whether the relationship between race and region in intergenerational mobility found at Davis and Mazumder (2018) holds at the beginning of the Great Migration. To do so, I compare the rates of intergenerational occupational mobility across four groups between 1910 and 1930: (1) northern blacks, (2) northern whites, (3) southern blacks, and (4) southern whites.

Due to northern blacks' small size in population, early studies that used census sample datasets from IPUMS have mainly focused either on southern blacks and their mobility relationship with southern whites (Collins and Wanamaker (2017)) or on *ALL* blacks in the North (Derenoncourt (2019); Boustan (2009)). With the release of digitized complete-count decennial census manuscripts, I was able to distinguish northern blacks, who have been settled in the North for generations before the Great Migration, from black migrants from the South in 1910. To my knowledge, this paper is the first one studying intergenerational mobility of these northern blacks and the racial mobility gap in the North in the prewar (World War II) era. To facilitate comparison, I applied a consistent linking method in sample construction where I link fathers in 1910 to their sons in 1930 for those four groups mentioned above. In these datasets, I observe self-reported occupations for both fathers and sons, allowing me to analyze the degree to which a father's occupation helped his son in accessing a given profession.

Following previous literature, I use Altham statistics to measure such "advantage" (Pérez (2019); Long and Ferrie (2013)). In a society with only two occupations available, mobility can be captured by the cross-product ratio of the occupational transition matrix. When more than two occupations present, Altham statistics aggregate the differences in cross-product ratios into one single statistics, thus providing us a simple measure to assess whether and by how much the father-son association in one group differs from that in another group.

I find that northern blacks had the highest mobility level among all groups, followed by northern whites, southern blacks, and southern whites. It makes sense to see that southerners have lower mobility levels than northerners. Both whites and blacks had limited economic opportunity in the South, including retarded industrial development and an agricultural system based heavily on the labor of landless farm tenants and sharecroppers (Alexander et al. (2017)). While the North was praised as the "promised land" for blacks, I find that racial mobility gap in the North is larger than it is in the South. Unlike the finding in Davis and Mazumder (2018) for the 1970s, both racial mobility gaps by region are smaller than regional mobility gap for blacks.

When splitting the sample of southern blacks into migrants (moved between 1910 and 1930) and stayers, I show that the mobility level of southern black migrants is closer to the one for northern blacks, much higher than the one for southern black stayers. However, when combining these southern black migrants with northern blacks, I see an overall lower joint mobility level for blacks and an even more significant racial gap in the North, echoing the conclusion of Derenoncourt (2019) that postwar migration destinations have become the mobility traps for blacks in the North. These results suggest that migration helped southern blacks through improving their absolute economic mobility, rather than their relative mobility with whites.

The paper proceeds as follows. Section 2 introduces the data. Section 3 describes the methodology. Section 4 presents the result. Section 5 discusses and concludes.

4.2 DATA

4.2.1 Linking

The study of intergenerational mobility requires dataset that contains labor market outcomes for both parents and children. I use the complete-count U.S. census from 1910 and 1930 to construct father-son linked datasets for northern blacks, northern whites, southern blacks, and southern whites.

In each of these datasets, I limit the initial sample to native-born males, aged 0 to 17, residing in their region of birth with their native-born father or step-father in 1910. The age restriction ensures that these boys in 1910 were young enough to be living with their parents and would be old enough to be participating in the labor force in 1930.[1] I can then compare their fathers' labor market outcomes in 1910 to their adult labor market outcomes in 1930. In this linking exercise, I cannot link women because they tend to change their maiden names upon marriage. Historical records linking, due to the lack of unique numerical identifier, has to rely on names and other time-invariant variables, such as place of birth and year of birth, to trace people across datasets. For the northern black and northern white samples,

[1] Region of birth is either in the South or in the North. Following previous literature, I define the following 14 states as the South: Texas, Oklahoma, Arkansas, Louisiana, Mississippi, Kentucky, Tennessee, Alabama, Florida, Georgia, South Carolina, North Carolina, Virginia, and West Virginia (Collins and Wanamaker (2015); Collins and Wanamaker (2014); Boustan (2009)).

I also require individuals to be urban residents, since the majority of the northern blacks at that time resided in the urban area, such restriction will ensure a fair comparison between northern blacks and northern whites.

As briefly mentioned above, I relied on race, the exact first and last names, self and parents' birthplaces, birth year to find individuals across censuses. Unfortunately, these variables do not always uniquely identify a person. Moreover, they are prone to errors from enumeration and transcription. A consistent linking methodology can help alleviate some of the selection concerns from linking. In this paper, I applied a procedure that is first developed by Ferrie (1996) and later advanced by Abramitzky, Boustan, and Eriksson (e.g., Abramitzky et al. (2012)). The basic steps are described below.

First, I clean the names by removing any non-alphabetic characters and by correcting common misspellings and nicknames. For example, Mike, Michael, and Micheal would be recognized as the same name. Second, I restrict the datasets in both 1910 and 1930 to individuals who are unique by their set of linking variables. In other words, exact matches can only be one-to-one matches. Third, for those individuals who do not have exact matches found, I search for matches within ± 3 years and only allow single matches. Overall, I started with more than 12 million unique records in 1910 and was able to identify around 14.5% of these records in 1930. The match rate breaks down by race and region is shown in Table 17.

4.2.2 Occupational Transition Matrices

Using these linked samples, I build intergenerational occupational transition matrices for northern whites, northern blacks, southern whites, southern blacks in Table 18.[2] Columns in these matrices represent the adulthood occupations of the boys in 1930, and rows represent the occupations of their fathers in 1910. Each cell in the matrix represents the counts of sons with occupation X_s conditional on having a father with occupation X_f (share in the parenthesis, sum to 100 within rows). An individual with a non-civilian occupational response for himself or his father is excluded.

[2] In terms of labor market outcomes, wage data was not collected by census until 1940, but the self-reported occupation was available for both 1910 and 1930.

Following previous literature, each father and son with occupational information was classified into one of the four categories: *White Collar*, *Farmer*, *Skilled/Semiskilled*, and *Unskilled* (Long and Ferrie (2013); Pérez (2019)).[3] Specifically, *White Collar* comprises professional, technical, and kindred; managers, officials, and proprietors; clerical and sales. *Farmer* comprises only farm owners and farm managers. *Skilled/Semiskilled* comprises craftsmen and operatives. *Unskilled* comprises service workers and laborers.

As expected, very few northerners (both blacks and whites) were farmers in either 1910 or 1930.[4] For all groups, there is a noticeable transition out of farming between 1910 and 1930. For example, more than 64% of the southern blacks and southern whites were employed in agriculture in 1910. However, the proportion of sons entering farming in 1930 is significantly lower in 1930: only 20% of the southerners remained as farmers.

From the fathers' generation to sons' generation, northern blacks had increased shares in white-collar and skilled blue-collar occupations, at the same time, decreased shares in farmers and unskilled occupations. It is consistent with my finding in a previous paper where I show that the Great Migration of southern blacks promoted northern blacks' occupational upgrading (Li (2019)).

4.3 METHODOLOGY

Sociological literature has long been using the transition matrices like Table 18 to study occupational mobility, with the diagonal cells represent immobility and thus mobility is measured by the proportion of people who fall in off-diagonal cells. However, such mobility rate is affected by the marginal distributions of a mobility table. For example, assuming the occupation of the son is *independent* of the occupation of the father in the following two groups:

[3]Using such broader occupational category also lessen the concern that the degree of intergenerational mobility would be affected by the earlier career stage of some of the sons.

[4]Due to the restriction to living in the urban area in 1910.

Group 1

$$\begin{bmatrix} & \multicolumn{2}{c}{Son} & \\ \cline{2-3} Father & Occ1 & Occ2 & MarginalDist.(Father) \\ Occ1 & 2500 & 2500 & 50\% \\ Occ2 & 2500 & 2500 & 50\% \\ MarginalDist.(Son) & 50\% & 50\% & \end{bmatrix}$$

Group 2

$$\begin{bmatrix} & \multicolumn{2}{c}{Son} & \\ \cline{2-3} Father & Occ1 & Occ2 & MarginalDist.(Father) \\ Occ1 & 625 & 1875 & 25\% \\ Occ2 & 1875 & 5625 & 75\% \\ MarginalDist.(Son) & 25\% & 75\% & \end{bmatrix}$$

Because they have different marginal distributions, the fraction of sons working on a different occupation than their father is 1/2 in Group 1 and 3/8 in Group 2. To avoid the confounding of marginal distributions, I followed the previous literature and used Altham statistics to measure mobility (Pérez (2019); Long and Ferrie (2013)).

First, Altham statistics remove the impact of marginal distributions by adjusting the marginal frequencies in both rows and columns to be consistent across two matrices. The fraction of sons working on a different occupation than their father in Group 2 above will have been 1/2 as well if we adjust the marginal frequencies in Group 2 to be the same as Group 1. Such adjustment is possible because we can multiply the rows and columns of a matrix by arbitrary constants without altering the underlying association between rows and columns in the matrix (Altham and Ferrie (2007)). Using a two by two matrix M to illustrate:

$$M = \begin{bmatrix} H_s H_f & L_s H_f \\ H_s L_f & L_s L_f \end{bmatrix}$$

the cross-product ratio is $\frac{(H_sH_f*L_sL_f)}{(H_sL_f*L_sH_f)}$. If we multiply each element in the first row by an arbitrary constant r_1, in the second row by an arbitrary constant r_2, in the first column by an arbitrary constant s_1, in the second column by an arbitrary constant s_2, we have

$$M' = \begin{bmatrix} H_sH_fr_1s_1 & L_sH_fr_1s_2 \\ H_sL_fr_2s_1 & L_sL_fr_2s_2 \end{bmatrix}.$$

So the cross-product ratio $\frac{H_sH_fr_1s_1*L_sL_fr_2s_2}{(H_sL_fr_2s_1*L_sH_fr_1s_2}$ is still equal to $\frac{(H_sH_f*L_sL_f)}{(H_sL_f*L_sH_f)}$.

Rearranging, we have $\frac{\frac{H_sH_f}{L_sH_f}}{\frac{H_sL_f}{L_sL_f}}$, which is the ratio of (1) the odds that sons of job H father get job H rather than job L to (2) the odds that sons of job L fathers get job H rather than job L. With perfect mobility, sons of job H fathers would have no advantage in getting job H relative to sons of job L fathers. The cross-product ratio would be 1. The more the cross-product ratio exceeds 1, the greater the relative advantage of having a job H father in getting job H.

With more than two occupational categories, there will be more than one cross-product ratio. Altham statistics aggregate the differences in cross-product ratios into one single statistics, allowing us to compare the full mobility patterns. Given two matrices **P** and **Q** of size r x s, the Altham statistic d(**P**, **Q**) measures whether and by how much the father-son association in matrix **P** differs from that in matrix **Q**:

$$d(\mathbf{P},\mathbf{Q}) = (\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{l=1}^{r}\sum_{m=1}^{s}[log(\frac{p_{ij}/p_{im}}{p_{lj}/p_{lm}}) - log(\frac{q_{ij}/q_{im}}{q_{lj}/q_{lm}})]^2)^{1/2}$$

Here, father's occupation is indexed by i and l, and son's occupation is indexed by j and m. p_{ij} denotes the probability of a son obtaining an occupation j given his father's occupation is i. The two-way odds ratio is $\theta_{ijlm} = log(\frac{p_{ij}/p_{im}}{p_{lj}/p_{lm}})$. Higher values of d(**P**, **Q**) imply larger differences in the row-column association. Therefore, if we compare each of the matrices in Table 18 to an independence matrix **J** (i.e., a matrix in which all the odds ratios are one), we can tell which group exhibits more mobility (i.e., closer to independence). In other words, if we observe that d(**P**, **Q**) $\neq 0$ and that d(**P**, **J**) $<$ d(**Q**, **J**), we can conclude that **P** has higher mobility than **Q**.[5]

[5]Note that Altham statistics do not require us to assume any occupational hierarchy. Movements across categories are treated symmetrically regardless of the origin and destination categories. Therefore, mobility could include a mix of upward and downward movements.

4.4 RESULTS

4.4.1 Main Results

Table 19 presents the results of calculating Altham statistic comparing each group down the rows to corresponding groups across the columns. For each pair, we can reject the hypotheses that their mobility patterns are the same. Each cell in the first column shows the distance between group i in the rows and the independence matrix **J**. Since the greater departure from independence implies less mobility, the statistics suggest that northern blacks had the highest mobility, followed by northern whites, southern whites, and southern blacks.

Comparing the mobility of the same race across regions, we see that northerners and southerners have different mobility patterns. The distance between northern blacks and southern blacks is 6.71, and it is 5.14 between northern whites and southern whites, both significantly different from zero.

Comparing the mobility of the same region across racial groups, we see that racial mobility difference in the North, which is 4.12, is greater than it is in the South, which is 1.98. Moreover, these racial mobility differences in either region are smaller than regional mobility differences for blacks, which is 6.71. In other words, the mobility pattern of northern blacks looks more similar to that of northern whites than to southern blacks.

With the Great Migration got underway in this period, I split the sample of southern blacks into migrants and stayers, where I define migrants by comparing their regions of residence between 1910 and 1930. Table 19 shows that the departure from independence is almost twice as large for southern black stayers relative to northern blacks. This pattern is consistent with mobility levels being higher in the "promised land" for blacks. Overall, migrants have a higher mobility level than stayers. Moreover, their mobility patterns are closer to northern blacks [d(**SB migrants**, **NB**) = 4.07***], rather than to southern black stayers [d(**SB migrants**, **SB stayers**) = 6.04***]. Indeed, as can be seen in Table 20, the mobility matrix for southern black migrants is quite similar to the mobility matrix for northern blacks, particularly the occupational distributions for sons. Compare to southern black stayers, both southern black migrants and northern blacks had higher barriers to enter the labor market as farmers for sons of non-farmers and a higher likelihood of leaving farming for sons of farmers. These patterns suggest that the similarity in mobility levels between northern blacks and southern black migrants is not simply a result of using the Altham statistics as the measure of mobility.

Not only did northern blacks have a mobility level that is more similar to northern whites than to southern blacks, southern black stayers also have a mobility level that is more similar to southern whites than to northern blacks, or even southern black migrants [d(**SB stayers**, **SW**) = 2.36***] < [d(**SB stayers**, **SB migrants**) = 6.04***] < [d(**SB stayers**, **NB**) = 7.21***]. Overall, geography has a greater impact than race on mobility for blacks in this period.

Lastly, I show that I can replicate the key findings in the postwar era by comparing *ALL* blacks in the North with northern whites and southern black stayers. Here, blacks in the North include both northern blacks and southern black migrants who moved North between 1910 and 1930. Table 21 presents the Altham statistics. We see that blacks in the North had a much lower mobility level than northern blacks alone [d(**All Blacks**, **J**) = 12.30*** > d(**N.Blacks**, **J**) = 6.98***]. Moreover, the racial gap in the North is larger than it was without adding migrants [d(**All Blacks**, **N.Whites**) = 7.87*** > d(**N.Blacks**, **N.Whites**) = 4.12***]. These findings corroborate with the conclusion in Derenoncourt (2019), that is, postwar migration destinations have become the mobility traps for blacks in the North. I also find that racial mobility gap in the North exceeds regional gap in mobility after including southern black migrants [d(**All Blacks**, **N.Whites**) = 7.87*** > d(**All Blacks**, **S.Black Stayers**) = 5.89***], consistent with the result in Davis and Mazumder (2018). Again, these results suggest that the mobility relationship between race and region in the prewar era is not driven by using Altham statistics to measure mobility but is driven by migration causing the demographic composition change in different regions.

4.4.2 Robustness Checks

The fact that there are no many farmers in the northerner samples could be a concern. To assess whether my results are driven by the low share of farmers in the North, I exclude farmers from all groups for both generations. Table C1 in the Appendix replicates Table 19, presenting the Altham statistic without farmers. We can see that, although the numbers are getting smaller, the ranking of groups in terms of their distance from independence does not change. All the observations listed in the main results above still hold.

The current classification for unskilled workers include both farm labors and other unskilled workers. I also re-estimate the Altham statistics separating farm labors from other unskilled workers. As expected, Table C2 in the Appendix displays a higher mobility level because five instead of four occupational categories are used. However, all the results are qualitatively similar to Table 19.

In the linking procedures, I allow age to vary ± 3 for those having no exact matches found. It could increase the chance of false matches, which have always been a concern in the study of intergenerational mobility. In Table C3, I show the robustness of my main results by restricting the sample to exact matches. Reassuring, we see a similar pattern as the main results.

4.5 DISCUSSION AND CONCLUSION

The U.S. North in the early twentieth century has been praised as the "promised land" for blacks. Millions of southern blacks left their hometowns and headed North into Detroit, Chicago, New York, and many other northern cities. Yet, there is no existing study providing us the mobility picture for northern blacks and how they fare with northern whites at the beginning of the Great Migration. Moreover, despite the immense amount of literature on black-white disparities, few studies have paid direct attention to the differences within the racial group shaped by region.

In this paper, I study intergenerational mobility for northern blacks, who settled in the North for generations before the arrival of southern migrants, and their relationship with northern whites, as well as southern blacks. I show that northern blacks had different mobility pattern from both the pattern for the northern whites and the pattern for southern blacks. Moreover, northern blacks and northern whites are closer in mobility level than do northern blacks and southern blacks. In fact, regional mobility gap for blacks is greater than racial mobility gaps in both the South and the North.

I show that migration helped southern blacks successfully translate their geographic mobility into economic mobility: southern black migrants have a mobility level that is closer to northern blacks than to southern black stayers. When including black migrants to make comparison with northern whites and southern black stayers, racial mobility gap was increased and become more substantial than the regional mobility gap among blacks. This finding is consistent with the prior literature using data from the 1970s, lending one support to the argument that the Great Migration reshaped the North.

4.6 TABLES

Table 17: Match Rate by Race and Region

		Initial Sample in 1910	Matched Sample in 1930	Match Rate %
Northern Whites		7,895,666	763,027	9.66
Northern Blacks		139,113	13,386	9.62
Southern Whites		3,281,052	813,046	24.78
Southern Blacks	Migrants	1,253,170	49,593	18.39
	Stayers		180,840	

Table 18: Intergenerational Occupational Mobility, by Race and Region

	Son's Occupation				
Father's Occupation	White Collar	Farmer	Skilled/semi-skilled	Unskilled	Row Total
Northern Whites					
White Collar	110578	2780	46626	67523	227507
	(48.60%)	(1.22%)	(20.49%)	(29.68%)	(100.00%)
Farmer	4971	1087	3840	5233	15131
	(32.85%)	(7.18%)	(25.38%)	(34.58%)	(100.00%)
Skilled/semi-skilled	100360	3351	113702	105813	323226
	(31.05%)	(1.04%)	(35.18%)	(32.74%)	(100.00%)
Unskilled	34638	1992	41318	47066	125014
	(27.71%)	(1.59%)	(33.05%)	(37.65%)	(100.00%)
Column Total	250547	9210	205486	225635	690878
	(36.27%)	(1.33%)	(29.74%)	(32.66%)	(100.00%)
Northern Blacks					
White Collar	285	5	236	614	1140
	(25.00%)	(0.44%)	(20.70%)	(53.86%)	(100.00%)
Farmer	29	0	58	122	209
	(13.88%)	(0.00%)	(27.75%)	(58.37%)	(100.00%)
Skilled/semi-skilled	298	4	627	1406	2335
	(12.76%)	(0.17%)	(26.85%)	(60.21%)	(100.00%)
Unskilled	990	32	1849	5648	8519
	(11.62%)	(0.38%)	(21.70%)	(66.30%)	(100.00%)
Column Total	1602	41	2770	7790	12203
	(13.13%)	(0.34%)	(22.70%)	(63.84%)	(100.00%)

Continued on next page

Table 18 (continued)

Father's Occupation	Son's Occupation: White Collar	Farmer	Skilled/semi-skilled	Unskilled	Row Total
Southern Whites					
White Collar	40571	6200	18830	27068	92669
	(43.78%)	(6.69%)	(20.32%)	(29.21%)	(100.00%)
Farmer	65960	133671	102109	182539	484279
	(13.62%)	(27.60%)	(21.08%)	(37.69%)	(100.00%)
Skilled/semi-skilled	21939	6200	36828	32782	97749
	(22.44%)	(6.34%)	(37.68%)	(33.54%)	(100.00%)
Unskilled	12325	10262	22915	36001	81503
	(15.12%)	(12.59%)	(28.12%)	(44.17%)	(100.00%)
Column Total	140795	156333	180682	278390	756200
	(18.62%)	(20.67%)	(23.89%)	(36.81%)	(100.00%)
Southern Blacks					
White Collar	421	373	659	2498	3951
	(10.66%)	(9.44%)	(16.68%)	(63.22%)	(100.00%)
Farmer	2607	36373	16598	86583	142161
	(1.83%)	(25.59%)	(11.68%)	(60.90%)	(100.00%)
Skilled/semi-skilled	570	1031	3004	8058	12663
	(4.50%)	(8.14%)	(23.72%)	(63.63%)	(100.00%)
Unskilled	1699	7803	8713	41838	60053
	(2.83%)	(12.99%)	(14.51%)	(69.67%)	(100.00%)
Column Total	5297	45580	28974	138977	218828
	(2.42%)	(20.83%)	(13.24%)	(63.51%)	(100.00%)

Note: Frequencies (Row Percent).

Table 19: Summary Measures of Altham Statistics

	J	**N.Blacks**	**N.Whites**	**S.Whites**	**S.Blacks** *Stayers*
J	.	.	.	.	.
N.Blacks	6.98***	.	.	.	.
N.Whites	11.99***	4.12***	.	.	.
S. Whites	12.77***	6.03***	5.14***	.	.
S. Blacks	12.77***	6.71***	6.23***	1.98***	.
Stayers	13.21***	7.21***	6.76***	2.36***	.
Migrants	11.90***	4.07***	6.62***	5.38***	6.04***

Note: **J** is an independence matrix in which all the odds ratios are one. Significance levels for the likelihood ratio chi-squared statistic G2.

*** Significant at the 1 percent level.

** Significant at the 5 percent level.

* Significant at the 10 percent level.

Table 20: Intergenerational Occupational Mobility, Blacks

	Son's Occupation				
Father's Occupation	White Collar	Farmer	Skilled/semi-skilled	Unskilled	Row Total
Southern Black Migrants					
White Collar	165	4	222	905	1296
	(12.73%)	(0.31%)	(17.13%)	(69.83%)	(100.00%)
Farmer	1042	182	5548	21571	28343
	(3.68%)	(0.64%)	(19.57%)	(76.11%)	(100.00%)
Skilled/semi-skilled	253	4	817	2599	3673
	(6.89%)	(0.11%)	(22.24%)	(70.76%)	(100.00%)
Unskilled	698	43	2730	10500	13971
	(5.00%)	(0.31%)	(19.54%)	(75.16%)	(100.00%)
Column Total	2158	233	9317	35575	47283
	(4.56%)	(0.49%)	(19.70%)	(75.24%)	(100.00%)
Southern Black Stayers					
White Collar	256	369	437	1593	2655
	(9.64%)	(13.90%)	(16.46%)	(60.00%)	(100.00%)
Farmer	1565	36191	11050	65012	113818
	(1.38%)	(31.80%)	(9.71%)	(57.12%)	(100.00%)
Skilled/semi-skilled	317	1027	2187	5459	8990
	(3.53%)	(11.42%)	(24.33%)	(60.72%)	(100.00%)
Unskilled	1001	7760	5983	31338	46082
	(2.17%)	(16.84%)	(12.98%)	(68.00%)	(100.00%)
Column Total	3139	45347	19657	103402	171545
	(1.83%)	(26.43%)	(11.46%)	(60.28%)	(100.00%)

Continued on next page

Table 20 (continued)

Father's Occupation	Son's Occupation: White Collar	Farmer	Skilled/semi-skilled	Unskilled	Row Total
Northern Blacks					
White Collar	285	5	236	614	1140
	(25.00%)	(0.44%)	(20.70%)	(53.86%)	(100.00%)
Farmer	29	0	58	122	209
	(13.88%)	(0.00%)	(27.75%)	(58.37%)	(100.00%)
Skilled/semi-skilled	298	4	627	1406	2335
	(12.76%)	(0.17%)	(26.85%)	(60.21%)	(100.00%)
Unskilled	990	32	1849	5648	8519
	(11.62%)	(0.38%)	(21.70%)	(66.30%)	(100.00%)
Column Total	1602	41	2770	7790	12203
	(13.13%)	(0.34%)	(22.70%)	(63.84%)	(100.00%)

Note: Frequencies (Row Percent).

Table 21: Summary Measures of Altham Statistics, All Blacks in the North

	J	**N.Whites**	**S.Black Stayers**
All Blacks in North	12.30***	7.87***	5.89***
N.Blacks	6.98***	4.12***	7.21***
S.Black Migrants	11.90***	6.62***	6.04***

Note: **J** is an independence matrix in which all the odds ratios are one. Significance levels for the likelihood ratio chi-squared statistic G2.

*** Significant at the 1 percent level.

** Significant at the 5 percent level.

* Significant at the 10 percent level.

APPENDIX A

COMPETITORS AND CONSUMERS: THE IMPACT OF THE GREAT MIGRATION ON EMPLOYMENT OUTCOMES OF BLACK NORTHERNERS

A.1 CREATING THE LINKED CENSUS SAMPLE FOR NORTHERN BLACKS 1910-30

In this section, I detailed my linking procedures and discussed the representativeness of my linked sample to the initial sample.

Initial linking sample in 1910 I started with the 1910 complete-count census and limited the sample to northern-born black males, aged 0 to 17 (inclusive of endpoints), residing with their parent or step-parent in the North. The age restriction is to ensure that these black sons were young enough to be living with their parents in 1910 and were old enough to be participating in the labor force in 1930. This generated an initial sample of 60,705 northern-born sons, some of them shared the same household heads.[1]

Final linking sample in 1930 Age-heaping is well documented in historical settings (A'Hearn et al. (2009)). Examining the matched samples produced by IPUMS, Feigenbaum (2016) shows that all the matched records have birth year differing by no more than 3 years. Allowing age to vary with a ± 3 year band, my final linking sample includes all the northern-born black males, aged 17 to 40.

Linking variables One of the challenges in linking historical records is the absence of

[1] Gender for the head of household could be either male or female.

unique personal identifiers such as the social security number. As a result, historical records linking relies on variables that are not supposed to change over time. They are names, race, gender, birth year, and birthplace. Note that parents' birthplaces were less frequently used in existing linking literature, because these variables are either not available or have been costly for scholars to link immigrants in the past studies: many immigrants reported very specific birthplace in one source ("Wuyishan, Fujian"), but only a general description in another ("China").[2] In theory, adding parents' birthplaces could help increase not only the accuracy of linkage but also the match rate.[3] Since the majority of blacks in my sample are domestic-born, and the state of birth variables are pretty clean and complete in the 1910 and 1930 censuses, I include parents' birthplaces to increase my match quality.

Linking method In reality, the presence of errors in these key identifying information posts another challenge. Feigenbaum (2016) developed a regression-based, supervised machine-learning approach that could help us tackle this challenge (discussed later).[4] It starts by building a "ground truth" sample. I manually identify possible matches for 2,696 individuals from 1910 initial linking sample (approximately 5%). After limiting candidate matches from 1930 to those with the same birthplace, race, sex, first letter of first and last names, age distance less than 3 and name Jaro-Winkler distance greater than 0.8, each person in 1910 from this training sample has around 5 candidate matches on average, with the maximum number of candidate matches equal to 78. In deciding the best match, I compare names (including middle name) and parents' birthplaces across candidate matches. When a person has more than one candidate matches sharing the same identifiable information, both pairs are determined as non-match since it is impossible for me to decide which pair should be the correct match. Table A1 provides a simple illustration of this manual process in determining a match.

I then use a probit regression to model these trade-offs and predict the score of a match

[2] Feigenbaum (2016) mentioned that only those respondents entered on the 14th or 29th line of each enumeration page in the 1940 complete-count census were asked additional questions on mother's and father's place of birth.

[3] For example, Ferrie's method with common names using phonetic name cleaning method has lower match rate because the phonetic name cleaning method took away variation from the identifying variables. Since all the candidate matches sharing the same identifiable information would be dropped, adding extra identifiable information reduces the number of common records. In another example, Bleakley and Ferrie (2016) linked sons of lottery winners from 1850 to the 1860-1880 censuses, and had a lower linking rate for 1860-1870 than for 1880 due to lacking data on parents' birthplaces (included for the first time in 1880).

[4] Other approaches to link individuals across historical datasets see Abramitzky et al. (2018)), Abramitzky et al. (2014), Goeken et al. (2011), and Ferrie (1996).

based on a series of characteristics deriving from the basic identifying variables (See Table A2). The best match, according to Feigenbaum (2016), should be the one with the highest score and the score should be higher than a threshold chosen by the researcher (i.e., absolute threshold), in addition, the ratio of the highest score to the second highest score should exceed another threshold selected by the researcher (i.e., relative threshold). The higher the thresholds are, the smaller the matched sample size would be.

The absolute threshold ranges from 0 to 1, as the predicted score cannot exceed 1. The relative threshold, however, ranges from 1 to ∞ because it is the ratio of two scores between 0 and 1. To find the appropriate thresholds, I loop over the absolute threshold with an increment of 0.01 each time between 0 and 1. At a given absolute threshold, I loop over the relative threshold with an increment of 0.05 each time between 1 and 2. To assess the match quality, I randomly select $\frac{2}{3}$ of the ground truth sample as the training set to train the algorithm and use the remaining $\frac{1}{3}$ of the ground truth sample as the test set to evaluate the algorithm performance at each combination of the thresholds. To avoid the possibility of getting low share of false matches due to random luck in selecting training and test sets, I performed 50 iterations on each combination of the thresholds to obtain the statistics of mean and standard deviation for both the share of false matches (Type I) and the share of missed matches (Type II). To illustrate, Panel A in Figure A1 plots average false match rate over the full range of absolute threshold given each level of relative threshold ranging between 1 and 2; Panel B in Figure A1 presents the average false match rate over relative threshold given each level of absolute threshold ranging between 0.2 and 0.39.[5] Figure A2 graph for the average missed match rate.

Ideally, I would like to minimize both the false match rate and the missed match rate. However, these two have a negative relationship: increasing both thresholds will decrease false match rate, but increase missed match rate.[6] In a two-dimensional world, for example, fixing the relative threshold at a certain level (see Figure A3), then we can compare the marginal benefit of increasing absolute threshold in decreasing false match rate with the marginal cost it has in increasing missed match rate. We can find out until which point of

[5]The pattern is the same for the full range of both absolute and relative thresholds.

[6]This negative relationship is consistent with the findings in Bailey et al. (2017) using IPUMS historical census data.

the absolute threshold that the marginal benefit would exceed the marginal cost. Applying this logic to every level of relative threshold, we can find the corresponding optimal absolute threshold. Similarly, we can also fix the absolute threshold at a certain level, then find the corresponding optimal relative threshold.[7] Figure A4 shows that there exist one and only one intersection point when the absolute threshold equals 0.36, and the relative threshold equals 1.05.

Representativeness It is important to ensure that selection into the linked sample is not biased; otherwise, it will prevent me from generalizing any result to the target population. Long and Ferrie (2013) argues that "as long as the matching process does not skew the sample, the set of matched individuals should also be representative of the national populations." They assess the representativeness of their linked sample by running probit regressions on an appended sample consisting of both the linked sample and the initial linking sample. The dependent variable is 1 for individuals from the linked sample and 0 for individuals from the initial linking sample. Each linked individual thus enter the regression twice to facilitate comparison with the weighted regressions. Following their approach, Table A3 column 1 evaluates the representativeness of my linked sample.[8] It is reassuring that the linked sample and the initial sample are very similar in the majority of baseline characteristics, though some variables exert a statistically significant influence on the probability of linkage. To eliminate these variables' impact, I construct weights to produce linked samples that would duplicate the marginal frequencies of the characteristics in the initial sample. After the weights are imposed on the linked individuals, they are statistically indistinguishable from the target population (see column 2 of Table A3).

Comparing with another matching approach Because of the cost of building a "ground truth" sample, many studies use the iterative matching approach to link individuals across censuses (both the details on how to implement this approach and STATA code are available online).[9] The basic idea of this approach is to link the unique records from both the initial sample and the final sample, assuming that all of the linking characteristics in

[7]This approach is essentially the same as minimizing $|| \, \Delta \mid$ false match rate - $\lambda \mid \Delta \mid$ missed match rate$|$ as we change the thresholds by assigning $\lambda = 1$. The optimal thresholds might change for a different λ.

[8]I dropped the observations living in counties without positive migration inflow from both the initial sample and the linked sample.

[9]https://people.stanford.edu/ranabr/matching-codes

the unique records are correctly measured (Abramitzky et al. (2012); Ferrie (1996)). Based on this approach, I build an independent linking sample where I avoid using phonetic name cleaning methods by requiring the exact match on raw names.[10] The sample match rate is 11.79%.[11]

Comparing the linked records across the two different samples that I have constructed, I find that (1) the best match decided by the machine learning approach may not be the one that exactly match on all identifying variables; (2) not all the pairs that exactly match on all identifying variables would be chosen as the best match (see Figure A5). It is easy to understand the first point, but why the pairs that exactly match on all are not determined as a match in the machine learning approach?

To illustrate, I use a made-up example in Table A4 where we have two people, "Hickly aged 10" and "Hickley aged 11" in 1910. Assuming they share the same first name, race, gender, birthplace, and parents' birthplace, the only difference here is their supposed last name and age. As mentioned, linking variables such as names and age presented in the data could have been mis-measured by individuals' misreporting, enumerators' misspelling, or census workers' mis-digitization. If everything is correctly recorded, "Hickly aged 10" from 1910 will be linked to "Hickly aged 30" from 1930, and "Hickley aged 11" from 1910 will be linked to "Hickley aged 31" from 1930, using the iterative matching approach. If error C occurs, however, "Hickly aged 10" from 1910 will be linked to "Hickley aged 30" from 1930, and "Hickley aged 11" from 1910 will be linked to "Hickly aged 30" from 1930, using the same approach. These types of one-to-one matches might not be made using the machine-learning approach because either Hickly or Hickley will have more than one candidate matches and their score difference might not be high enough to pass the relative threshold. So the supervised machine-learning approach is not only able to identify more matches but also less prone to this type of error.

Why is the sample match rate based on the supervised machine-learning approach (40.68%) so much higher than the sample match rate based on the iterative matching ap-

[10] A recent paper by Bailey et al. (2017) compares various matching algorithms and shows that (1) algorithms using NYSIIS and Soundex tend to increase Type I errors without reducing Type II error rates (2) erroneous matches biased the estimates for intergenerational mobility.

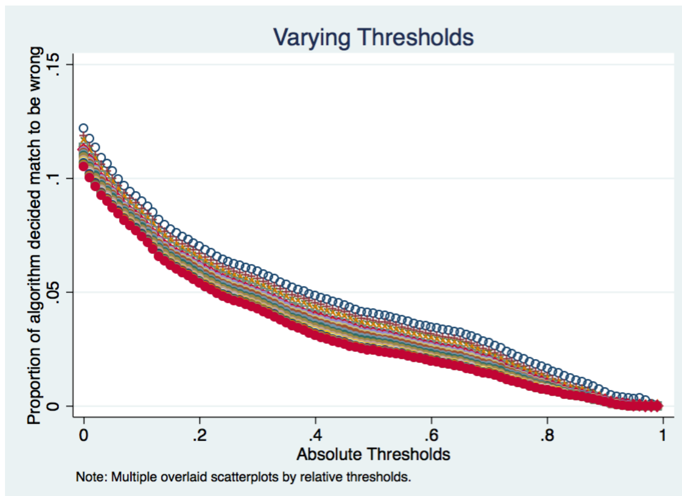
[11] comparing to a small ground truth sample, 21 out of 572 pairs that are identified as the match by the algorithm is determined as non-match by the researcher due to inconsistent middle names (i.e., false positive rate = 3.67%).

proach? The latter required unique records from both initial and final linking sample. For example, in Table A4, if error A occurs, "Hickly aged 10" will be dropped out of the initial linking sample due to non-uniqueness and we will not even try to find them in 1930. If error B occurs, "Hickly aged 30" will be dropped out of the final linking sample due to non-uniqueness and we will not be able to find a match for them. Another reason for the iterative matching approach to have a lower match rate is its requirement of exact matches on all linking characteristics. In the machine learning sample, many northern-born blacks from 1910 have only one candidate match from 1930 after blocking on birthplace, name distance, and age. These unique candidate matches do not always have the exact name string due to misspelling, but we could be confident that they are the appropriate matches under certain condition.[12] Under such circumstances, the cost of using the iterative matching approach is enormous. Therefore, the supervised machine-learning approach is better suited for the study conducted in this paper.

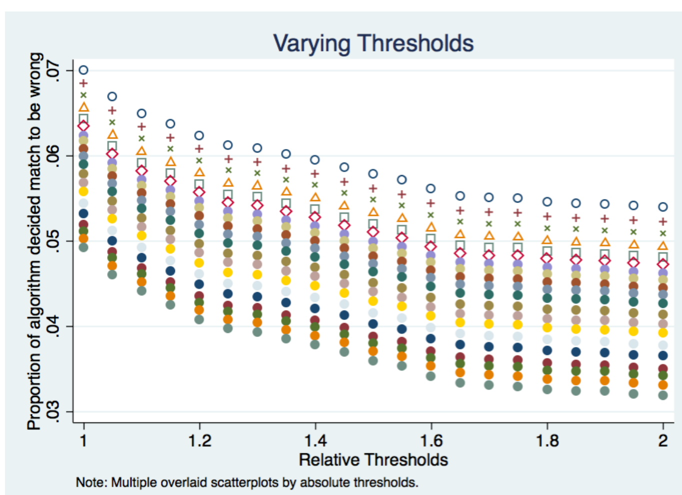
[12] 68.45% (16,905 out of the 24,696) of the matches made by the algorithm are those with unique candidate matches. Although these pairs are only subjected to the absolute threshold requirement, their mean score is 0.92 and median is 0.98, much higher than the score of the best match for 1 to N pairs (mean=0.75, median=0.79). The match rate will be 12.83% ($\frac{24696-16905}{60705}$) if we exclude these unique matches.

Figure A1: Match Quality Assessment Over Absolute/Relative Threshold: False Matches

Panel A. Absolute Threshold



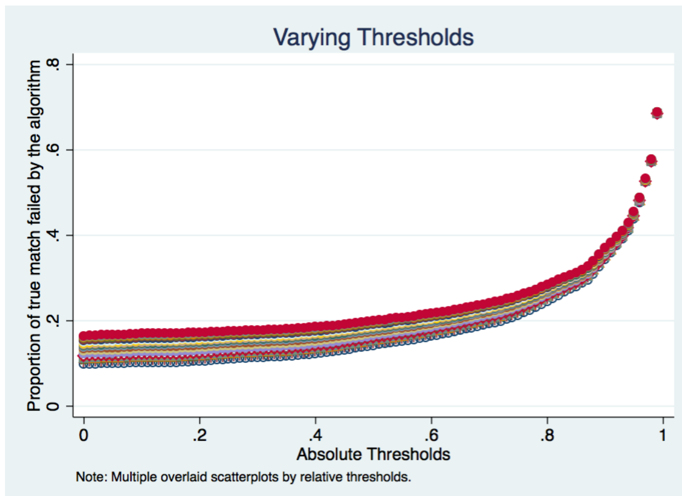
Panel B. Relative Threshold



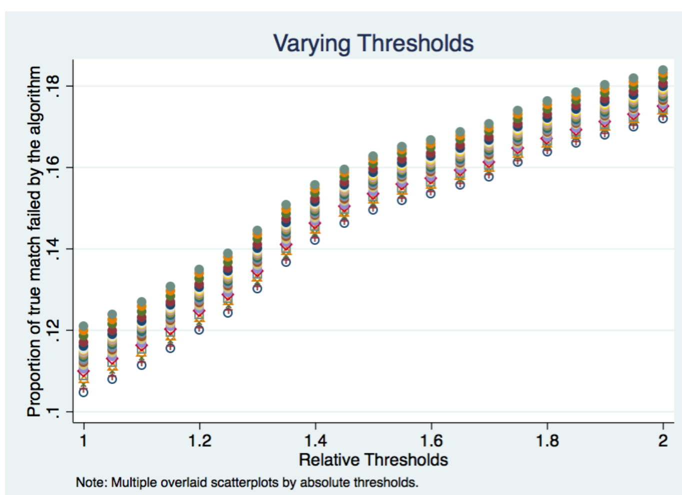
Notes: The line with blue color empty dot corresponds to relative (absolute) threshold equal to 1 (0.2) and the line with red color solid dot corresponds to relative (absolute) threshold equal to 2 (0.39). At a given absolute (relative) threshold, false match rate (i.e. the proportion of algorithm determined match to be wrong) is decreasing as the relative (absolute) threshold is increasing (i.e. moving from blue empty dot line to red solid dot line). For all levels of the relative (absolute) threshold, false match rate is decreasing as the absolute (relative) threshold is increasing.

Figure A2: Match Quality Assessment Over Absolute/Relative Threshold: Missed Matches

Panel A. Absolute Threshold

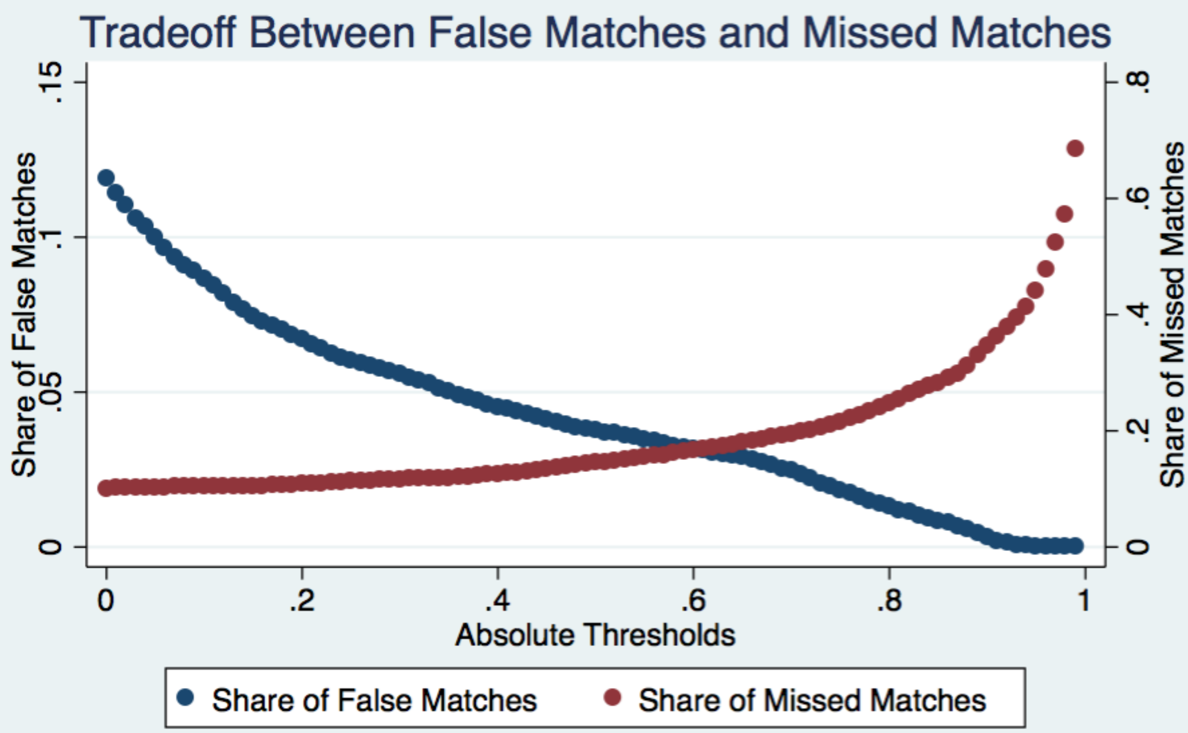


Panel B. Relative Threshold



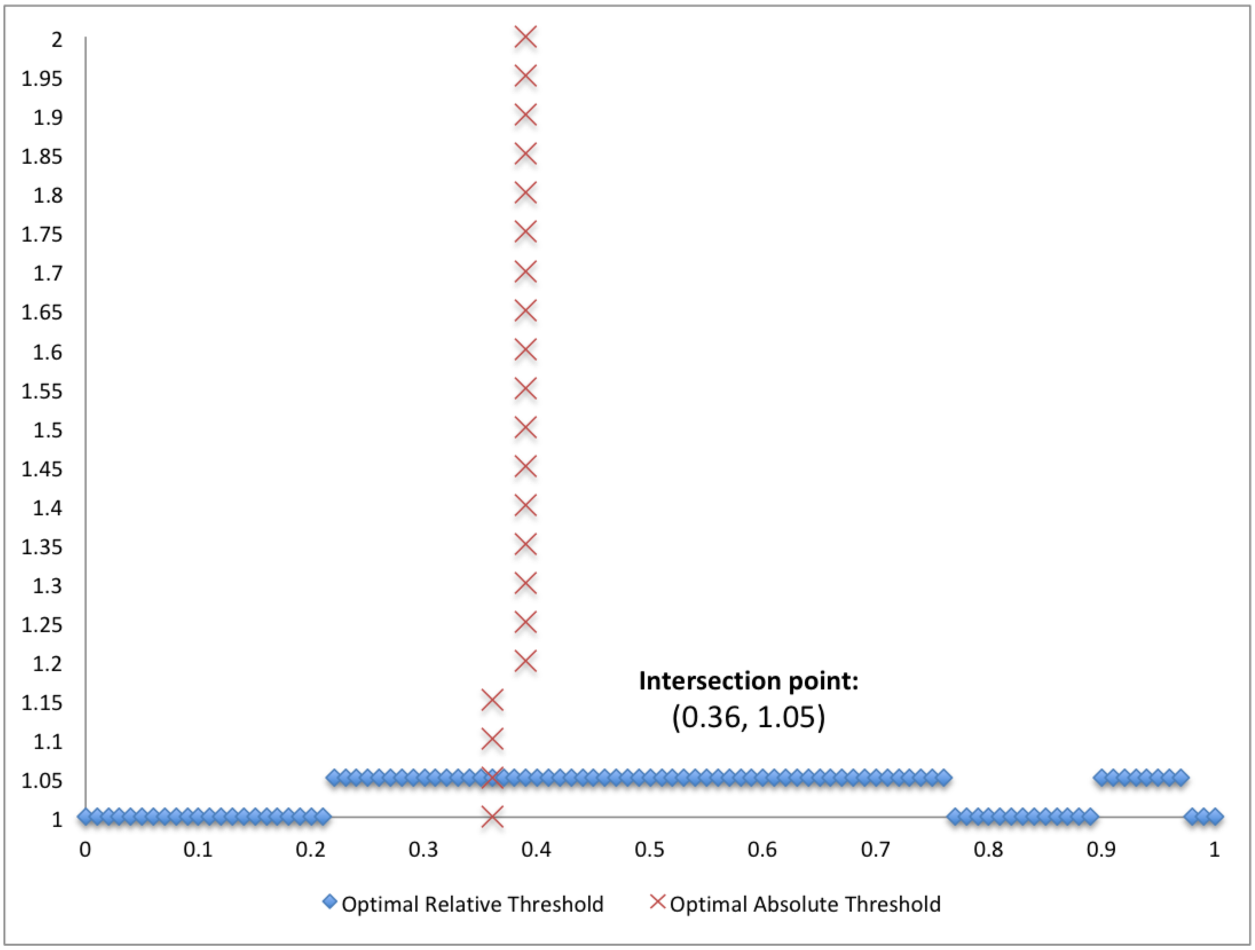
Notes: The line with blue color empty dot corresponds to relative (absolute) threshold equal to 1 (0.2) and the line on the top corresponds to relative (absolute) threshold equal to 2 (0.39). At a given absolute (relative) threshold, missed match rate (i.e. the proportion of true match failed by the algorithm) is increasing as the relative (absolute) threshold is increasing (i.e. moving from blue empty dot line to red solid dot line). For all levels of the relative (absolute) threshold, missed match rate is increasing as the absolute (relative) threshold is increasing.

Figure A3: Inverse Relationship Between False Match Rates and Missed Match Rates



Note: The share of false matches (or missed matches) at each threshold equal to the mean from 50 iterations. In each iteration, I took 2/3 of the ground truth sample as the training dataset and the remaining 1/3 of the ground truth sample as the test dataset to generate the share of algorithm determined match to be wrong (or the share of true match failed by the algorithm). Relative threshold is fixed at 1.05 in this example.

Figure A4: Optimal Thresholds: Northern-born Blacks



Notes: X-axis is the absolute threshold and Y-axis is the relative threshold.

Figure A5: Comparing Linked Samples Across Linking Approaches

Notes: The total linked records in machine-learning sample is 24,696, and the total linked records in iterative matching sample is 7,156. The number of pairs that were identified as linked in both approach is 5,968.

Table A1: A Simple Example in Building "Ground Truth" Sample

1910				1930				
Surname	**Given**	**fbpl**	**mbpl**	**Surname**	**Given**	**fbpl**	**mbpl**	**match**
Carter	Johnson	Missouri	Missouri	Carter	John	Missouri	Missouri	0
				Carter	John H	Missouri	Missouri	0
				Cade	John	Missouri	Missouri	0
				Carter	John William	Missouri	Missouri	0
				Carey	John	Missouri	Missouri	0
				Carter	John W	Missouri	Missouri	0
Jones	William	Virginia	New York	Johnson	William	Connecticut	Connecticut	0
				Jones	William M	South Carolina	South Carolina	0
				Jones	William A	Virginia	Connecticut	0
				Johnson	William	Connecticut	Connecticut	0
				Jones	William	Virginia	New York	1
				Jones	William S	Virginia	North Carolina	0

Table A2: Probit Regression on the Match Dummy in the Human Review Process

VARIABLES	match
First name distance, Jaro-Winkler	-4.143***
	(1.0129)
Last name distance, Jaro-Winkler	-8.406***
	(0.7912)
Absolute value difference in year of birth is 1	-0.527***
	(0.0882)
Absolute value difference in year of birth is 2	-0.862***
	(0.0954)
Absolute value difference in year of birth is 3	-1.117***
	(0.1030)
(log) Hits	-1.339***
	(0.0502)
First and last name match (including middle name if available)	0.336**
	(0.135)
More than one match for first and last name	-1.029***
	(0.1312)
Last letter of first name matches	-0.0867
	(0.1214)
Last letter of last name matches	0.201*
	(0.1120)
Middle initial matches, if there is a middle initial	0.305***
	(0.0804)
Father's birthplace matches	0.560***
	(0.0697)
Mother's birthplace matches	0.455***
	(0.0695)
First name Soundex match	0.470***
	(0.1503)
Last name Soundex match	0.696***
	(0.1205)
Constant	0.386*
	(0.2205)
Observations	4,433
Log Likelihood	-1044.5943
Akaike's information criterion	2123.189

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table A3: Probit Marginal Effects on Linkage

VARIABLES	(1) unweighted	(2) weighted
birthplace (New England Division)	-0.003	-0.001
	(0.085)	(0.083)
birthpalce (East North Central Division)	0.032	0.027
	(0.084)	(0.080)
birthplace (West North Central Division)	-0.016	-0.024
	(0.093)	(0.089)
birthplace (Mountain Division)	-0.026	-0.042
	(0.164)	(0.160)
birthplace (Pacific Division)	-0.089	-0.094
	(0.136)	(0.126)
residence division (New England)	0.017	0.017
	(0.085)	(0.082)
residence division (East North Central)	-0.010	-0.007
	(0.084)	(0.080)
residence division (West North Central)	-0.036	-0.029
	(0.093)	(0.089)
residence division (Mountain)	-0.191	-0.189
	(0.193)	(0.187)
residence division (Pacific)	0.108	0.114
	(0.131)	(0.122)
Son's age 6-11	-0.055***	-0.011
	(0.019)	(0.019)
Son's age 12-17	-0.091***	-0.007
	(0.018)	(0.018)
Mulatto (son)	0.032	0.033
	(0.023)	(0.023)
Urban Area	-0.028**	-0.004
	(0.012)	(0.012)
Live in Farm	0.047	0.039
	(0.032)	(0.032)
In School (son)	0.016	0.016
	(0.016)	(0.016)
Head's age 26-50	0.030	0.032
	(0.027)	(0.027)
Head's age 51-75	0.040	0.045
	(0.032)	(0.032)
Head's age 76-100	0.058	0.061
	(0.142)	(0.143)
Homeowner (head)	0.063***	-0.009
	(0.014)	(0.014)
Literacy (head)	0.058***	-0.010
	(0.016)	(0.016)
Speak English (head)	0.033	0.033
	(0.032)	(0.032)
Mulatto (head)	0.020	0.018
	(0.024)	(0.024)
Low White Collar Occupations (head)	0.058	0.059
	(0.043)	(0.043)
Skilled Blue Occupations (head)	0.008	0.007
	(0.035)	(0.035)
Unskilled Blue Occupations (head)	-0.023	-0.021
	(0.027)	(0.027)
Farmer (head)	-0.069*	-0.043
	(0.040)	(0.040)
Occupation Not Specified (head)	-0.037	-0.035
	(0.041)	(0.041)
Constant	-0.619***	-0.602***
	(0.052)	(0.052)
Observations	59,726	59,726
Pseudo R^2	0.002	0.001
Predicted Probability	0.289	0.289

Note: Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1 Omitted: Middle Atlantic Division for birth and residence place; son's age 0-5; head's age 0-25; high white collar jobs for head.

Table A4: A Made-up Example

	1910		1930	
	Surname	**Age**	**Surname**	**Age**
Without Error	Hickly	10	Hickly	30
	Hickley	11	Hickley	31
With Error A	Hickly	10	Hickly	30
	Hick**ly**	**10**	Hickley	31
With Error B	Hickly	10	Hickly	30
	Hickley	11	Hick**ly**	**30**
With Error C	Hick**ley**	10	Hickly	30
	Hick**ly**	**10**	Hickley	**30**

A.2 CONSTRUCTING OCCUPATIONAL EARNINGS SCORE

The needs for skills is shifting over time. For example, the earnings for workers in jobs that require higher levels of analytical skills have grown faster than jobs that require physical skills in recent decades. In the absence of income information in the 1930 census, it is better to use the occupational earnings from the census that is the closest to the sample year.

This consideration points me to the 1940 census, the first census year that I can observe both occupations and income for an individual. However, there are two shortcomings of using the 1940 wage data. Income information for self-employed workers and farmers is not available. Collins and Wanamaker (2017) addresses these two issues by first calculating the earnings ratio of self-employed workers over wage earners for all the occupations using the 5-percent sample of the 1960 census from IPUMS, and then scale up the average earnings of wage workers from 1940 by this ratio to impute earnings for self-employed workers in 1940.

Following their approach, I include in the calculation only those individuals aged 17-66 (uninclusive of endpoints) and had the number of weeks worked for profit, as well as income (either as self-employed workers "incbusfm" or as wage earners "incwage"), being positive. I then calculate the average 1940 earnings for all persons sharing the same 3-digit occupation code, race, gender, location, and working relationship. Both mulattos and darker-skinned blacks are considered as blacks. I use three different levels of geographical location to ensure that I have at least 50 observations per occupational cell in constructing the earnings score: census division is the most preferred, followed by census region and nationwide.

For self-employed workers, its earnings are scaled by the ratio of the average earnings for all the self-employed workers over the average earnings for all the wage earners in 1960 by gender. Again, if an occupation does not include both types of workers or its observations for either class is less than 50 in 1960, its self-employed vs. employee earning ratio would be imputed by the average earnings for the entire self-employed workers and wage earners instead. All these steps are done by gender separately.

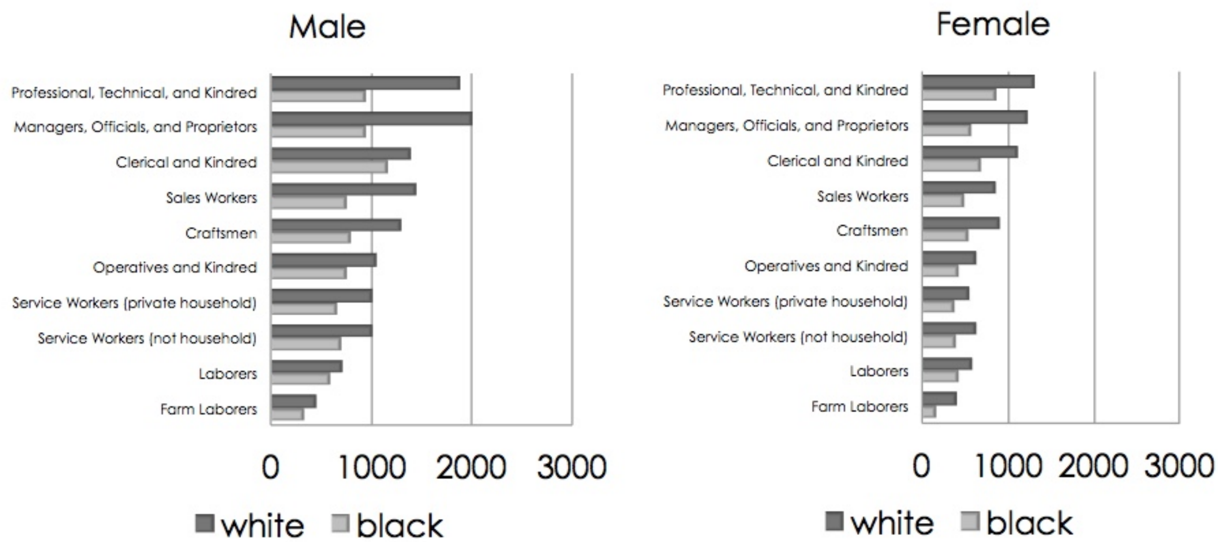
There are still a few occupations that do not have enough workers to calculate an imputed average to then scale. For this group, I assign earnings based on a broader occupational classification by converting the 3-digit occupational code to 1-digit. For example, both

"Optometrists (3-digit code=070)" and "Osteopaths (3-digit code=071)" will be assigned the average earnings for all the "Professional and Technical (1-digit code=0)" workers who share the same race, location, and the working relationship.

For farmers, they are further differentiated by homeowner status. I first estimate the earnings ratio of farmers over farm laborers in 1960, by gender and homeowner status. Then, I scale the earnings for farm laborers in 1940 by its corresponding 1960 ratio and assign this product to farmers of the same race, gender, and division.

Lastly, I want to learn how plausible is the constructed earnings score and how reliable it is to measure a person's socioeconomic standing. Figure A6 shows the average earnings score for ten broader occupational headers by gender and race. The information conveyed coincides with our prior, that white-collar occupations > blue-collar occupations > unskilled ones regarding average earnings; blacks earn less than whites and males earn more than females within the occupational category; racial earning difference is more prominent at a higher level of skill.

Figure A6: Sensitivity and Plausibility of the Occupational Earnings Score



Notes: The occupational earnings score assignment is based on the 3-digit IPUMS 1950 occupational classification system which contains hundreds of detailed occupational titles. For illustration purpose, I classified them into ten broader occupational headers shown in the vertical axis. The value in the horizontal axis reflects average occupational earnings score. Farmers are omitted from the graph because their earnings score is further differentiated by homeownership status in the income assignment.

A.3 ADDITIONAL FIGURES AND TABLES

Figure A7 plots the total black population change against migrants change between 1910 and 1930. Each dot represents a northern county that experienced migrants inflow. For almost all the counties, the points lie above the 45-degree line, suggesting that places attracted migrants were also destinations for black northerners if they move. Moreover, it reveals that southern black migrant influx was responsible for most of the increase in the local black population.

Figure A8 shows an advertisement from African-American newspaper *The Indianapolis Recorder* revealing that blacks were barred from amusement places run by whites. The demand from black consumers offered the business opportunities for black establishments.

Figure A9 displays the white-collar jobs held by black northerners in my sample in 1930. For illustration purpose, the detailed occupational titles were classified into 13 categories. **Entertainers** include *Actors and Actresses, Artists and Art Teachers, Entertainers (n.e.c.), Musicians and Music Teachers, Recreation and Group Workers*; **Teachers** include *Professors and Instructors that Subject Not Specified, Teachers (n.e.c.)*; **Managers, Officers, and Proprietors** include *Railroad Conductors, Building Managers and Superintendents, Public administration officials and administrators (n.e.c.), Managers, officials, and proprietors (n.e.c.)*; **Messengers and Mail Clerks** include *Express messengers and railway mail clerks, mail carriers, messengers and office boys, Newsboys*; **Clerks** include *Attendants at physician's and dentist's office, transportation baggagemen, bank tellers, bookkeepers, bill and account collectors, shipping and receiving clerks, telegraph operators, telephone operators, clerical and kindred workers (n.e.c.)*; **Agent & Brokers** include *Insurance agents and brokers, real estate agents and brokers*; **Salesmen** include *Advertising agents and salesmen, Demonstrators, Hucksters and Peddlers, Stock and bond salesmen, salesmen and sales clerk (n.e.c.)*; **Other Professionals, Technicians** include *Accountants and auditors, athletes, chemists, chiropractors, college presidents and deans, editors and reporters, civil engineers, electrical engineers, engineers (n.e.c.), photographers, social and welfare workers (except group), medical and dental technicians, therapists and healers (n.e.c.), veterinarians, professional technical and kindred workers (n.e.c.), officials lodge society union etc*

Figure A10 lists the top 20 white-collar occupations that have increasing share from 1910 to 1930. Blacks were able to hold trades in 1930 that none of their fathers had in 1910, such as engineers and telephone operators. While black professionals such as dentists, pharmacists, and funeral directors had significant growth, the most significant expansion within white-collar occupations lies on low-skilled and less-profitable ones, such as mail carriers and office boys.

Table A5 reports the full first stage results where the dependent variable is the (log) actual migration inflow and the regressor of interest is the (log) predicted in-migration. The first three columns correspond to the 2SLS specifications in Table 3 and Table 4. Column 4 to 9 correspond to the 2SLS specifications in Table 5 with column 4 to 6 for black sample and 7 to 9 for the white sample. Column 2, 5, and 8 present results after including the additional controls. Column 3, 6 and 9 report results for the sample of stayers. In all cases, the F-stat is very high, and the instrument has a large, positive, and significant effect on the actual migrant change. Overall, Table A5 suggests that the positive relationship between actual and predicted in-migration is robust to the use of different specifications.

There are five variables (age, literacy, homeownership, farmer and urban status) that have a statistically significant influence on the probability of linkage (see column (1) of Table A3). While their magnitude is relatively small, I construct weights to eliminate these variables' impact and evaluate whether they bias my estimates:

$$w = \frac{P_{initial}}{(1 - P_{initial})} \frac{(1 - P_{linked})}{P_{linked}}$$

where $P_{initial}$ (or P_{linked}) represents the group share of the characteristics that matter to the linkage selectivity in the initial sample (or in the linked sample).[13] I then replicate Table 3, with these weights imposed on the linked sample. Table A6 shows that coefficients estimated from the weighted sample are almost the same as their counterparts estimated from the unweighted sample.

Table A7 presents results using a different instrument variable where the national outflow of blacks from the South to the North (i.e., treating all the southern blacks as a group, as opposed to distinguishing southern blacks by birth state) is distributed by the earlier southern

[13]Note that the weights cannot eliminate the impact of unobservables on the linkage probability.

black settlement patterns in the North. The estimated effects are qualitatively similar to the results in main text but slightly larger in magnitude.

Figure A7: Destinations for Migrants and Black Northerners

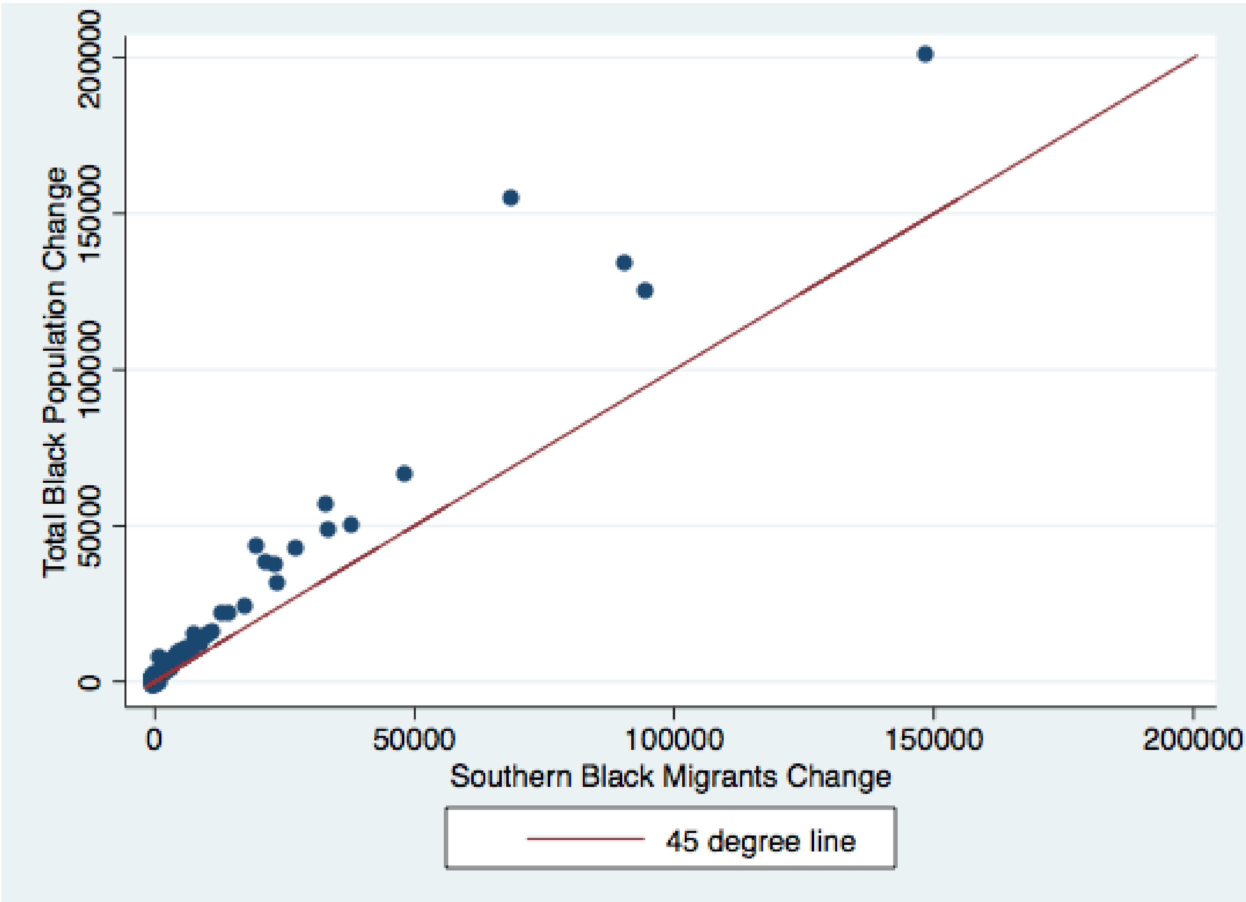
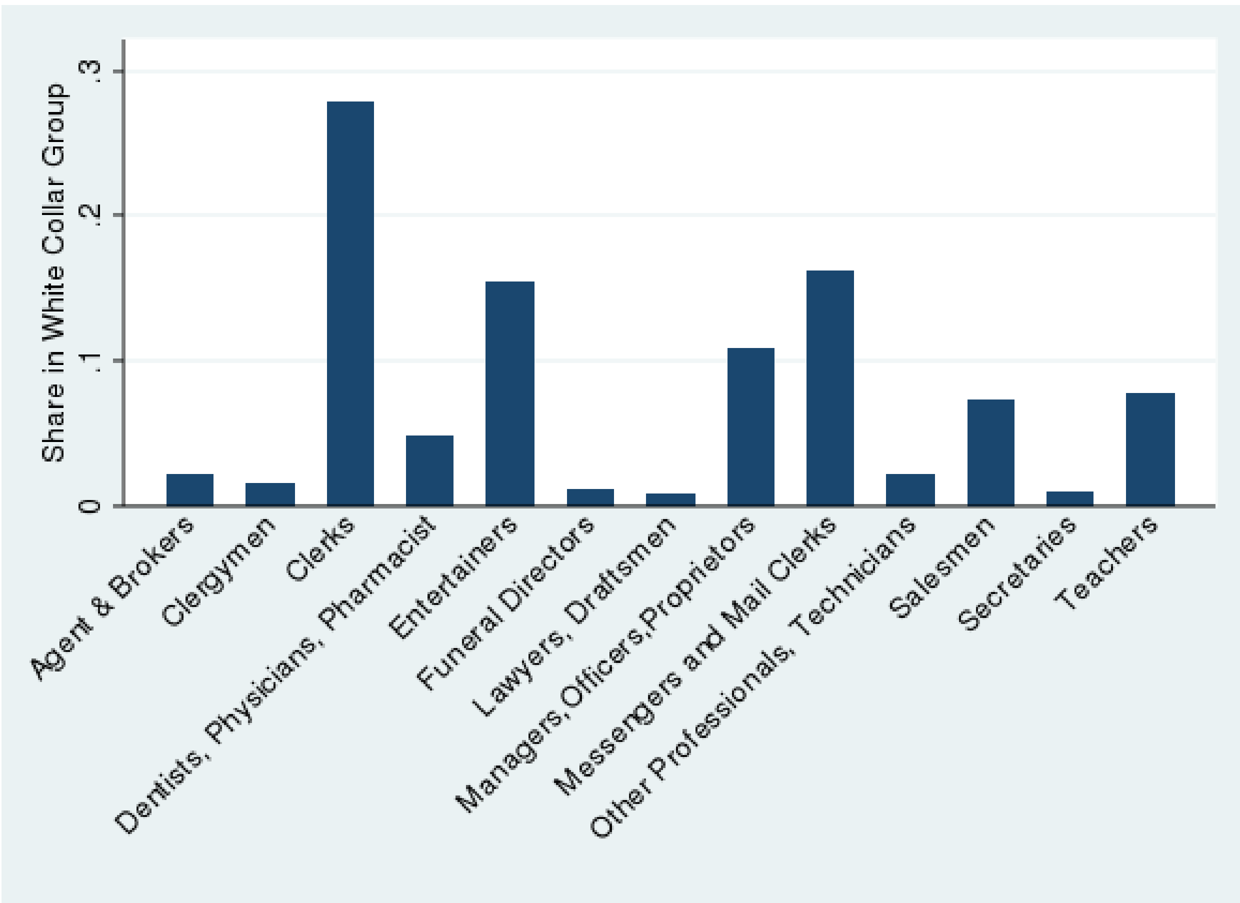


Figure A8: Advertisement from African-American Newspaper

New Columbia Theater.

With the opening of the new Columbia 5 ct. Theater, 524 Indiana ave., Saturday night to a crowded house, Indianapolis now boasts of two such amusement places run by colored men and both filliug a long felt need, the first one being the Manilla at 12th and West street, conducted by Mr. J. L. Lewis for some time. The Columbia has a capacity of about 175 persons; is seated with modern opera chairs and has the up to date conveniences and appointments of the older and higher priced places. It is clean, spic and span and every care is being taken for the comfort of its patrons. The shows will run every afternoon and evening and on Sundays the charity donation will alternate between the Alpha Home and tne Woman's club. Mrs. Tisha Lee has been engaged to sing at each performance and Miss Florence Griggsby will preside at the piano. The Recorder congratulates Mr. James D. Hill and his brother Mr. L. G. Hill upon the push and enterprise and bespeak for them a fiattering success.

Figure A9: White Collar Jobs Held By Northern-born Blacks in 1930



Notes: For the total 1,764 northern-born blacks who held white collar jobs, I classified them into 13 categories above.

Figure A10: Top 20 Occupations Among White Collar Group By Share Changes between 1910 and 1930

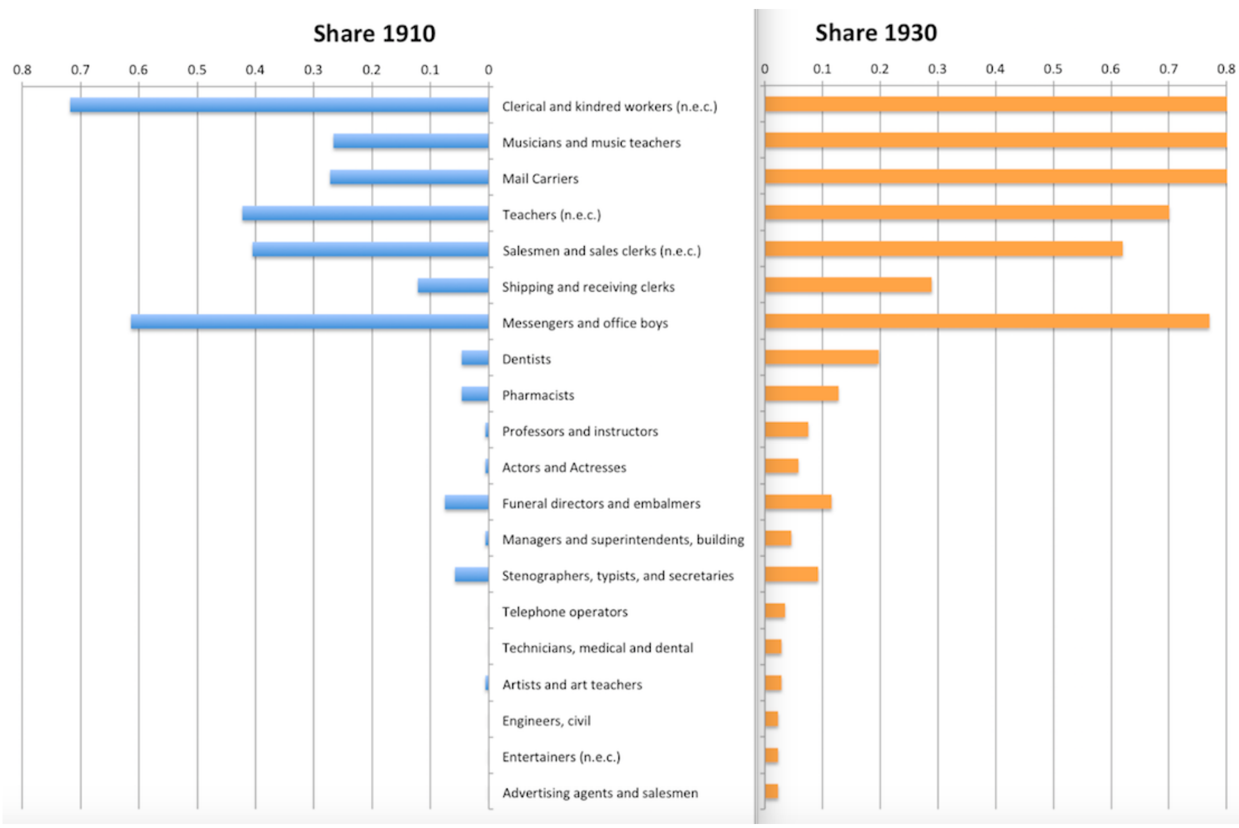


Table A5: First Stage

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	DV: (log) Mig. Inflow								
(log) Predicted Mig. Inflow	0.770***	0.821***	0.799***	0.781***	0.835***	0.828***	0.729***	0.741***	0.692***
	(0.138)	(0.127)	(0.136)	(0.139)	(0.128)	(0.136)	(0.148)	(0.128)	(0.129)
Kleibergen Paap F-stat	30.96	41.59	34.42	31.73	42.86	37.28	24.41	33.76	28.68
Controls:									
White Pop Change		X	X		X	X		X	X
County Characteristics 1880		X	X		X	X		X	X
Stayer Only			X			X			X
Observations	12,623	12,623	7,234	15,927	15,927	9,178	10,730	10,730	6,158

Note: All regressions include individual-level, family-level and county-level controls, as well as state fixed effects.

Table A6: Robustness on Weighted Sample

	(1)	(2)	(3)	(4)
	OLS	2SLS	2SLS	2SLS
	Panel A. Prob. of Employment			
(log) Mig. Inflow	-0.006**	-0.012*	-0.012**	-0.026**
	(0.003)	(0.007)	(0.006)	(0.010)
Mean DV	0.867	0.867	0.867	0.864
	Panel B. Percentile Rank			
(log) Mig. Inflow	0.424***	0.620**	0.571**	1.343***
	(0.144)	(0.277)	(0.254)	(0.468)
Mean DV	33.49	33.49	33.49	33.97
	Panel C. (log) $EarningsScore^a$			
(log) Mig. Inflow	0.008***	0.014**	0.013**	0.031**
	(0.003)	(0.006)	(0.006)	(0.012)
Mean DV	6.614	6.614	6.614	6.624
Controls:				
White Pop Change			X	X
County Characteristics 1880			X	X
Stayer Only				X
Observations	12,623	12,623	12,623	7,234
County	151	151	151	149
Kleibergen Paap F-stat		31.23	41.68	34.53

county cluster-robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Note: All regressions include individual-level, family-level and county-level controls, as well as state fixed effects.

Table A7: Robustness on A Different Instrument Variable

	(1)	(2)	(3)	(4)
	OLS	2SLS	2SLS	2SLS
	Panel A. Prob. of Employment			
(log) Mig. Inflow	-0.006**	-0.016**	-0.016***	-0.033***
	(0.003)	(0.007)	(0.006)	(0.010)
Mean DV	0.867	0.867	0.867	0.864
	Panel B. Percentile Rank			
(log) Mig. Inflow	0.421***	0.830***	0.780***	1.432***
	(0.143)	(0.296)	(0.273)	(0.439)
Mean DV	33.49	33.49	33.49	33.97
	Panel C. (log) $EarningsScore^{a}$			
(log) Mig. Inflow	0.008***	0.018***	0.018***	0.032***
	(0.003)	(0.007)	(0.006)	(0.012)
Mean DV	6.614	6.614	6.614	6.624
Controls:				
White Pop Change			X	X
County Characteristics 1880			X	X
Stayer Only				X
Observations	12,623	12,623	12,623	7,234
County	151	151	151	149
Kleibergen Paap F-stat		31.77	39.19	33.22

county cluster-robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Note: All regressions include individual-level, family-level and county-level controls, as well as state fixed effects.

APPENDIX B

RACIAL SEGREGATION IN HOUSING MARKETS AND THE EROSION OF BLACK WEALTH

B.1 CONSTRUCTING THE MATCHED ADDRESS SAMPLE

Each record in the census data represents an individual in a household. Each household has a head and related members who share the same address. An address is the combination of a house number and a street name. In an ideal world, we would know the number of individuals and households residing at a given address. However, either the house number or the street name entry for an individual could have been mis-recorded by the census enumerators or mis-digitized by the contemporary census digitization workers. Therefore, some households have incorrect or incomplete addresses, possibly leading to inaccurate counts of households in any building. This appendix describes the algorithm we used to construct a representative set of households for our sample cities in 1930 and 1940, focusing in particular on the challenge of assigning all individuals to the correct address.

We first need to make sure that no household is either missing an address or assigned more than one. We assume that the enumeration districts (EDs) and tracts reported in the census data were transcribed correctly. A tiny fraction of EDs and tracts from the census do not coincide with the list of EDs that we use to define our cities. We drop those EDs or tracts, as they are likely to be institutions that were given a separate ED number.

We have digitized 1930 enumeration district boundaries (Shertzer et al. (2016)) and obtained census tract boundary files from the National Historical Geographic Information

System (NHGIS). We cross-check census address data by "fuzzy" matching each census street name to a list of street names from the corresponding ED/tract obtained from the spatial datasets. We exclude addresses on streets that have either no reasonable match or too many potential matches among the digitized streets.

Census enumerators were instructed to survey households as they moved along a street, and thus we do not expect to see house numbers within a street jump around. Thus, the order in which households appear on the manuscripts should generally reflect their location within the ED relative to neighboring households.[1] To ensure that we have all the households living in each address in our sample, we also drop any address that shares a street-block (or the entire street-ED when the block cannot be identified) with an address that is potentially out of order on the manuscript. We provide further details of the process below.

B.2 DETAILS ON METHODOLOGY

We make sure that every household has exactly one address composed of a street name and house number. To begin, we assign the address information from the household head to everyone in his/her household. When the household head has partial (e.g. only a street name or only the house number) or no information on address, we fill in information from the household's non-head member. We perform a series of quality checks on these imputed addresses that are described below. If the household head is missing an address and household members disagree on either street name or house number, we impute the missing address information from those of households listed just before this one in the census manuscripts and flag these households.

In the case of multiple households sharing the same dwelling unit, we will have more than one household head. When these household heads disagree on the address, we compare each component of the addresses (the street names and house numbers) to those of adjacent households and keep the one(s) that matches that of the most number of neighbors. We flag all addresses imputed from adjacent households. A very small number of dwellings from the

[1] Our indicators of manuscript page and line numbers are not very reliable, so we use the household IDs assigned by IPUMS as proxy for the order in which households appear in the original census manuscripts.

1940 Census seem to have members belonging to different EDs/tracts. As with street names and house numbers, we assume the household heads ED/tract is the correct one. In the case of multi-family households, we compare each candidate ED/tract with those of households appearing immediately before and after on the census manuscripts, and only retain the EDs/tracts with the highest number of matches. We have a few households located at the intersection of EDs/tracts, and we flag these as well.

Then we standardize street names in the census, which are noisy and frequently riddled with typos. We first standardize all the directional prefix and street suffix, convert ordinal street numbers to their cardinal text forms, and remove any redundant information from street name (such as "Block A"). We then match these formatted street names to our digitized 1930 city streets to standardize further the names. We create a crosswalk of digitized street names, 1930 EDs, and 1940 tracts and fuzzy match them with the set of unique census street names by ED/tract (allowing some margin of error in the string match). We use STATA's reclink2 command for this task. If a census street matches to more than one digitized street (a "one-to-many" match) within an ED/tract, then we flag all the digitized streets that were a match. Eventually we drop all Census records where the street does not match a digitized street or matches one that is flagged as part of a one-to-many match. Note that the process is sensitive to the margin of error that we allow in our string match. A wide error margin means we will have more one-to-many matches and fewer non-matches, whereas with a narrow error margin, we will have more non-matches and fewer one-to-many matches. The former introduces false one-to-one matches that might otherwise stay unmatched, whereas the latter introduces false one-to-one matches that might otherwise be matched to many. Thus, a conservative approach is to allow a wide margin of error, but narrow enough that we are still left with a reasonably sized sample after dropping one-to-many and non-matches.

House numbers, like street names, are also prone to errors and typos. The next step is to standardize house numbers as best as we can across ED/tracts and census years. When the house number variable is just one clear number, we leave it as it is. When it is not (e.g., "945/6", "4531 667" or "1??2"), we try to identify a minimum and a maximum possible house number. For instance, when the reported house number is "4531 667", we treat it as ranging from 667 to 4531 and flag all addresses on the same street block and ED with house numbers in that range.[2] We assume a “?” can range from 0 to 9, so that house number “1??2” ranges from 1002 to 1992. We treat separators like “/”, “-”, “&”, “+”, “~” and “,” as spaces when identifying the range, while we ignore alphabets (treating “5a” as “5”) and other non-alphanumeric characters (e.g. parentheses and brackets). All problematic addresses are flagged.

We do not have digitized historical house numbers as with street names to validate our cleaning process. Instead, we perform a number of quality checks based on the ordering of households in the census manuscripts and flag households that fail to satisfy these checks. Failing one or more of these reality checks implies that the re-formatted and standardized addresses are unlikely to be correct. These flagged households include cases where:

1. the address differs from that of adjacent households on the manuscript when adjacent households share an address,
2. only the house number matches that of one adjacent household, and only the street name matches that of the other adjacent household,
3. the house number differs from adjacent house numbers by more than 10 along the same street,
4. the house number changes non-monotonically (and differs from adjacent house numbers by at least 4) along the same street, and
5. the address is (either partially or completely) imputed from that of the preceding household when adjacent street names differ.

We drop households in all addresses that were flagged in any of the previous steps. If a households address is flagged, the correct address is likely to be that of adjacent households on the manuscript, given our assumption on the path of the enumerators. To avoid undercounting the individuals in these adjacent addresses, we also drop all addresses adjacent to flagged addresses on the manuscript. Thus, we generate a sample of addresses that are

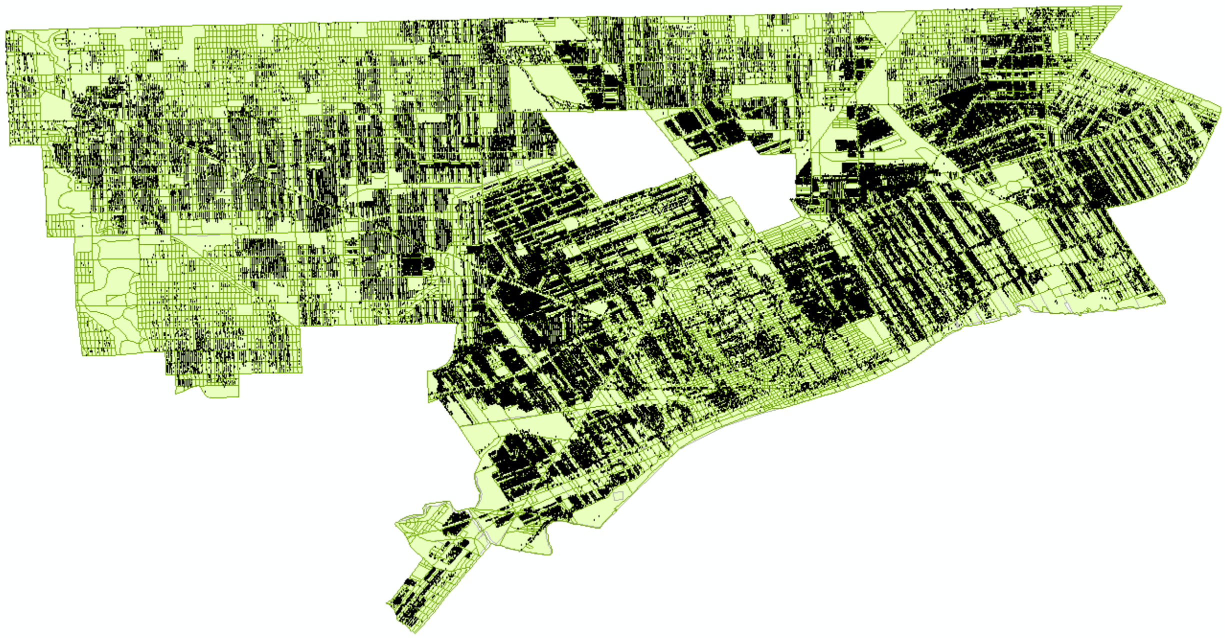
[2] There are alternative ways of interpreting a reported house number of “4531 467”. The second number might be an apartment number within the building, or the building might span house numbers 4531 to 4667. However, given that we eventually drop all street blocks intersecting this range of numbers, we believe our range assignment is the most conservative in dealing with such ambiguity. The street block is defined by the street name and the hundreds of the house number.

correct with a reasonable degree of accuracy that is our baseline.[3]

Finally, from each sample, we retain only the addresses that appear in both the 1930 and the 1940 Census. Since we have digitized 1930 ED boundaries and 1940 tract boundaries, we further make sure that the reported EDs (in 1930) and tracts (in 1940) corresponding to each address overlap spatially.

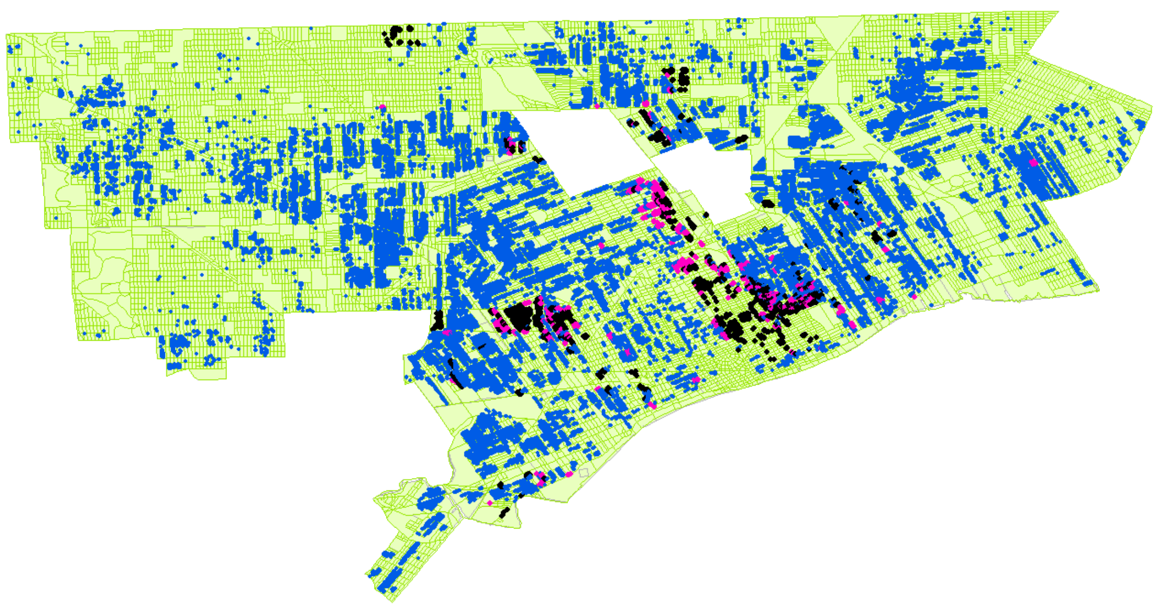
[3] If the street block of a flagged address cannot be identified credibly (e.g. when the house number is completely non- numeric or the range of house numbers is unrealistically large), we drop all addresses on the same street and ED.

Figure B1: Geocoded Detroit Addresses



Notes: the figure shows the addresses in our sample for the city of Detroit that could be geocoded against a map of 1940 enumeration districts produced by Logan and Zhang (2017).

Figure B2: Racial Transition in Geocoded Blocks in Detroit



Notes: the figure shows the addresses in our sample for the city of Detroit that could be geocoded against a map of 1940 enumeration districts produced by Logan and Zhang (2017). Blocks are color-coded as follows: blue blocks were less than 5 percent black in both 1930 and 1940, pink blocks were less than 5 percent black in 1930 and more than 5 percent black in 1940, and black blocks were over 5 percent black in both 1930 and 1940.

Table B1: Address Statistics for Block Sample

	All Households				Addresses		Blocks		Addresses	
	Total Households		Quality Address		Unique Addresses		Unique Blocks		per Block	
	1930	1940	1930	1940	1930	1940	1930	1940	1930	1940
Baltimore	193,979	245,862	147,962	132,680	118,741	97,264	8,249	7,831	14.4	12.4
Boston	182,090	211,731	132,944	135,944	62,913	61,052	4,090	4,051	15.4	15.1
Brooklyn	614,082	752,606	390,826	358,432	157,005	125,803	8,935	7,450	17.6	16.9
Chicago	845,436	1,025,731	545,383	437,973	278,694	198,297	20,530	17,766	13.6	11.2
Cincinnati	124,321	143,864	87,188	65,169	51,436	38,384	4,898	4,009	10.5	9.6
Cleveland	222,856	247,713	129,774	99,907	86,588	65,744	10,991	8,745	7.9	7.5
Detroit	370,556	451,198	225,457	219,961	168,955	163,406	18,380	18,169	9.2	9.0
Manhattan	470,552	614,786	188,258	191,471	25,178	20,876	1,854	1,856	13.6	11.2
Philadelphia	459,749	515,472	338,928	254,737	291,919	211,705	15,054	12,033	19.4	17.6
Pittsburgh	153,628	185,039	107,276	102,587	78,809	66,712	7,878	7,134	10.0	9.4
St. Louis	216,133	225,794	116,945	111,305	77,551	72,166	7,117	6,560	10.9	11.0
Total/Average	3,853,382	4,619,796	2,410,941	2,110,166	1,397,789	1,121,409	107,976	95,604	12.9	11.9

Note: The first two columns report the number of households reported in the census in each city. "Quality addresses" are the households for which we were able to assign an address that passed all quality checks described in the Data Appendix B. "Unique addresses" are addresses that both pass the quality checks and are unique with a street name, street number, and 1930 enumeration district. We use postal service convention and assign house numbers to blocks using hundreds within a given street name. "Unique blocks" are the number of unique blocks represented by our sample of unique addresses. The last column of the table reports the number of unique addresses per unique block. This is the sample of addresses we used to construct our block sample.

Table B2: Address Sample Statistics

	Households with address found in both census years						Addresses		Households	
	Total		Trimmed Sample		Quality Address		Unique Addresses		per Address	
	1930	1940	1930	1940	1930	1940	1930	1940	1930	1940
Baltimore	110,312	125,598	98,780	111,757	67,925	75,888	57,287	57,287	1.2	1.3
Boston	122,353	136,230	100,785	109,334	67,008	71,343	33,267	33,267	2.0	2.1
Brooklyn	365,589	413,796	254,723	286,483	144,116	159,787	62,108	62,108	2.3	2.6
Chicago	443,948	497,700	355,109	389,081	190,802	207,627	104,553	104,553	1.8	2.0
Cincinnati	78,245	85,719	67,293	72,601	33,132	35,112	20,967	20,967	1.6	1.7
Cleveland	124,151	135,182	111,170	118,948	48,676	51,126	34,843	34,843	1.4	1.5
Detroit	212,211	228,290	184,660	194,112	95,309	98,560	76,845	76,845	1.2	1.3
Manhattan	235,841	299,774	95,304	119,594	29,369	36,001	3,913	3,913	7.5	9.2
Philadelphia	227,479	244,202	206,716	218,856	145,313	152,650	131,469	131,469	1.1	1.2
Pittsburgh	84,028	94,428	73,731	81,806	43,172	47,767	32,289	32,289	1.3	1.5
St. Louis	141,183	148,756	124,771	130,522	48,185	50,361	34,239	34,239	1.4	1.5
Total/Average	2,145,340	2,409,675	1,673,042	1,833,094	913,007	986,222	591,780	591,780	2.1	2.3

Note: The "Total" columns report the number of households with addresses we were able to locate in both the 1930 and 1940 censuses. We trimmed this sample to eliminate transcription errors and institutions (we drop any households with more than 10 members, any household with more than three heads, any addresses with monthly rent greater than $100, and finally any addresses with a value greater than $20,000). The "Trimmed Sample" columns report the number of households without problematic census values in both 1930 and 1940. The "Quality Address" columns report the number of households without problematic census values that passed the address quality checks described in the Data Appendix B. The "Unique Addresses" columns report the number of addresses represented by this sample of households. This is the sample of addresses we used in our address-level analysis.

Table B3: Selection into Sample

	Year	All	Quality Address	Matched Address
Individuals	1930	15,591,308	9,894,466	4,495,743
	1940	15,729,224	7,560,898	4,345,911
Households	1930	3,845,617	2,406,975	1,082,691
	1940	4,610,562	2,106,438	1,180,009
Addresses	1930	2,077,442	1,407,878	659,688
	1940	2,217,640	1,125,845	659,688
Households per address	1930	1.85	1.71	1.64
	1940	2.08	1.87	1.79
Individuals per address	1930	7.51	7.03	6.81
	1940	7.09	6.72	6.59
Average household size	1930	4.40	4.39	4.39
	1940	3.87	3.96	3.99
Distance to CBD (tract centroid)	1930	4.43	4.32	4.30
	1940	4.61	4.47	4.30
Population density (tract)	1930	0.013	0.013	0.013
	1940	0.012	0.012	0.013
Percent black (tract)	1930	0.076	0.073	0.066
	1940	0.084	0.078	0.078

Note: The "All" column reports statistics for the full sample of census records across all ten cities. The "Quality Address" column reports statistics for census records that had an address that passed our quality checks as described in the Data Appendix. The "Matched Address" column reports statistics for the sample of quality addresses that could be matched across the 1930 and 1940 census. The distance to CBD is defined as the distance from the central business district to the centroid of the 1940 tract. All tract variables refer to the 1940 census tract.

APPENDIX C

INTERGENERATIONAL MOBILITY IN PREWAR AMERICA: A COMPARISON BY RACE AND REGION

C.1 ROBUSTNESS CHECKS

Table C1: Summary Measures of Altham Statistics, Excluding Farmers

	J	**N.Blacks**	**N.Whites**	**S.Whites**	**S.Blacks** *Stayers*
J	.	.	.	.	.
N.Blacks	3.63***	.	.	.	.
N.Whites	3.66***	1.32***	.	.	.
S. Whites	5.45***	2.16***	2.15***	.	.
S. Blacks	5.09***	1.66***	2.04***	0.69***	.
Stayers	5.79***	2.30***	2.77***	1.05***	.
Migrants	3.79***	1.09***	0.83**	1.82***	2.44***

Note: **J** is an independence matrix in which all the odds ratios are one. Significance levels for the likelihood ratio chi-squared statistic G2.

*** Significant at the 1 percent level.

** Significant at the 5 percent level.

* Significant at the 10 percent level.

Table C2: Summary Measures of Altham Statistics, Separating Farm Labors from Unskilled Workers

	J	**N.Blacks**	**N.Whites**	**S.Whites**	**S.Blacks** *Stayers*
J	.	.	.	.	.
N.Blacks	17.72***	.	.	.	.
N.Whites	18.44***	8.47***	.	.	.
S. Whites	20.60***	7.97***	6.12***	.	.
S. Blacks	19.92***	9.31***	8.53***	4.61***	.
Stayers	20.27***	9.73***	8.98***	5.12***	.
Migrants	17.16***	8.80***	10.53***	10.07***	8.95***

Note: **J** is an independence matrix in which all the odds ratios are one. Significance levels for the likelihood ratio chi-squared statistic G2.

*** Significant at the 1 percent level.

** Significant at the 5 percent level.

* Significant at the 10 percent level.

Table C3: Summary Measures of Altham Statistics, Exact Matches

	J	**N.Blacks**	**N.Whites**	**S.Whites**	**S.Blacks** *Stayers*
J	.	.	.	.	.
N.Blacks	10.43***	.	.	.	.
N.Whites	13.39***	7.75***	.	.	.
S. Whites	14.30***	9.79***	5.71***	.	.
S. Blacks	14.31***	10.94***	7.33***	2.78***	.
Stayers	14.65***	11.29***	7.74***	3.07***	.
Migrants	12.20***	5.88***	8.97***	8.79***	8.95***

Note: **J** is an independence matrix in which all the odds ratios are one. Significance levels for the likelihood ratio chi-squared statistic G2.

*** Significant at the 1 percent level.

** Significant at the 5 percent level.

* Significant at the 10 percent level.

BIBLIOGRAPHY

Aaronson, Daniel, Daniel A Hartley, and Bhashkar Mazumder, "The effects of the 1930s HOLC "redlining" maps," 2017.

Abramitzky, Ran, Leah P Boustan, and Katherine Eriksson, "Europe's tired, poor, huddled masses: Self-selection and economic outcomes in the age of mass migration," *The American Economic Review*, 2012, *102* (5), 1832–1856.

_ , _ , and _ , "A nation of immigrants: Assimilation and economic outcomes in the age of mass migration," *Journal of Political Economy*, 2014, *122* (3), 467–506.

_ , Roy Mill, and Santiago Pérez, "Linking Individuals Across Historical Sources: a Fully Automated Approach," 2018.

A'Hearn, Brian, Jörg Baten, and Dorothee Crayen, "Quantifying quantitative literacy: Age heaping and the history of human capital," *The Journal of Economic History*, 2009, *69* (3), 783–808.

Aladangady, Aditya, David Albouy, and Mike Zabek, "Housing Inequality," 2017.

Alexander, J Trent, Christine Leibbrand, Catherine Massey, and Stewart Tolnay, "Second-Generation Outcomes of the Great Migration," *Demography*, 2017, *54* (6), 2249–2271.

Altham, Patricia ME and Joseph Ferrie, "Comparing contingency tables tools for analyzing data from two groups cross-classified by two characteristics," *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 2007, *40* (1), 3–16.

Altonji, Joseph G and David Card, "The effects of immigration on the labor market outcomes of less-skilled natives," in "Immigration, trade, and the labor market," University of Chicago Press, 1991, pp. 201–234.

Ananat, Elizabeth Oltmans, "The wrong side (s) of the tracks: The causal effects of racial segregation on urban poverty and inequality," *American Economic Journal: Applied Economics*, 2011, *3* (2), 34–66.

Bailey, Martha, Connor Cole, Morgan Henderson, and Catherine Massey, "How Well Do Automated Linking Methods Perform in Historical Samples? Evidence from New Ground Truth," 2017.

Blau, Francine D and John W Graham, "Black-white differences in wealth and asset composition," *The Quarterly Journal of Economics*, 1990, *105* (2), 321–339.

Bleakley, Hoyt and Joseph Ferrie, "Shocking behavior: Random wealth in antebellum Georgia and human capital across generations," *The Quarterly Journal of Economics*, 2016, *131* (3), 1455–1495.

Boustan, Leah P, "Competition in the promised land: Black migration and racial wage convergence in the North, 1940–1970," *The Journal of Economic History*, 2009, *69* (3), 755–782.

_ **and Robert A Margo**, "A silver lining to white flight? White suburbanization and African–American homeownership, 1940–1980," *Journal of Urban Economics*, 2013, *78*, 71–80.

Boustan, Leah Platt, "Was postwar suburbanization "white flight"? Evidence from the black migration," *The Quarterly Journal of Economics*, 2010, *125* (1), 417–443.

Brady, Carolyn M, "Indianapolis at the Time of the Great Migration, 1900-1920," *Black History News and Notes*, 1996, *65*, 12–36.

Broom, Leonard and Norval D Glenn, *Transformation of the Negro American*, Vol. 117, Harper & Row, 1967.

Card, David, "Immigrant inflows, native outflows, and the local labor market impacts of higher immigration," *Journal of Labor Economics*, 2001, *19* (1), 22–64.

_ **, Alexandre Mas, and Jesse Rothstein**, "Tipping and the Dynamics of Segregation," *The Quarterly Journal of Economics*, 2008, *123* (1), 177–218.

Carrington, William J, Enrica Detragiache, and Tara Vishwanath, "Migration with endogenous moving costs," *The American Economic Review*, 1996, pp. 909–930.

Chambers, Daniel N, "The racial housing price differential and racially transitional neighborhoods," *Journal of urban Economics*, 1992, *32* (2), 214–232.

Chetty, Raj and Nathaniel Hendren, "The impacts of neighborhoods on intergenerational mobility II: County-level estimates," 2016.

_ **, _ , Patrick Kline, and Emmanuel Saez**, "Where is the land of opportunity? The geography of intergenerational mobility in the United States," *The Quarterly Journal of Economics*, 2014, *129* (4), 1553–1623.

Coates, Ta-Nehisi, *We were eight years in power: An American tragedy*, One World/Ballantine, 2018.

Collins, William J, "When the tide turned: Immigration and the delay of the great black migration," *The Journal of Economic History*, 1997, *57* (3), 607–632.

_ **and Katharine L Shester**, "Slum clearance and urban renewal in the United States," *American Economic Journal: Applied Economics*, 2013, *5* (1), 239–73.

_ **and Marianne H Wanamaker**, "Selection and economic gains in the great migration of African Americans: new evidence from linked census data," *American Economic Journal: Applied Economics*, 2014, *6* (1), 220–252.

_ **and** _ , "The Great Migration in Black and White: New Evidence on the Selection and Sorting of Southern Migrants," *The Journal of Economic History*, 2015, *75* (4), 947–992.

_ **and** _ , "Up from Slavery? African American Intergenerational Economic Mobility Since 1880," 2017.

Cutler, David M, Edward L Glaeser, and Jacob L Vigdor, "The rise and decline of the American ghetto," *Journal of political economy*, 1999, *107* (3), 455–506.

Davis, Jonathan and Bhashkar Mazumder, "Racial and Ethnic Differences in the Geography of Intergenerational Mobility," 2018.

Davis, Morris A, Andreas Lehnert, and Robert F Martin, "The Rent-price ratio for the aggregate stock of owner-occupied housing," *Review of Income and Wealth*, 2008, *54* (2), 279–284.

Derenoncourt, Ellora, "Can you move to opportunity? Evidence from the Great Migration," 2019.

DeSantis, Alan D, "Selling the American dream myth to black southerners: The Chicago Defender and the great migration of 1915–1919," *Western Journal of Communication (includes Communication Reports)*, 1998, *62* (4), 474–511.

Farley, Reynolds and Walter R Allen, *The color line and the quality of life in America*, Russell Sage Foundation, 1987.

Feigenbaum, James J, "A Machine Learning Approach to Census Record Linking," 2016.

Ferrie, Joseph, "A new sample of males linked from the public use microdata sample of the 1850 US federal census of population to the 1860 US federal census manuscript schedules," *Historical Methods*, 1996, *29* (4), 141–156.

Foged, Mette and Giovanni Peri, "Immigrants' effect on native workers: New analysis on longitudinal data," *American Economic Journal: Applied Economics*, 2016, *8* (2), 1–34.

Follain, James R and Stephen Malpezzi, "Another look at racial differences in housing prices," *Urban Studies*, 1981, *18* (2), 195–203.

Frazier, Franklin, *Black bourgeoisie*, Vol. 91058, Simon and Schuster, 1957.

Goeken, Ron, Lap Huynh, TA Lynch, and Rebecca Vick, "New methods of census record linking," *Historical methods*, 2011, *44* (1), 7–14.

Goldsmith-Pinkham, Paul, Isaac Sorkin, and Henry Swift, "Bartik Instruments: What, When, Why, and How," 2018.

Gottlieb, Peter, *Making Their Own Way: Southern Blacks' Migration to Pittsburgh, 1916-30*, Vol. 82, University of Illinois Press, 1996.

Hertz, Tom, "Rags, Riches, and Race," *Unequal chances: Family background and economic success*, 2009, *165.*

Higgs, Robert, *Competition and Coercion: Blacks in the American economy 1865-1914*, Vol. 163, CUP Archive, 1977.

Horrace, William C and Ronald L Oaxaca, "Results on the bias and inconsistency of ordinary least squares for the linear probability model," *Economics Letters*, 2006, *90* (3), 321–327.

Hout, Michael, "Occupational mobility of black men: 1962 to 1973," *American Sociological Review*, 1984, pp. 308–322.

Jaeger, David A, Joakim Ruist, and Jan Stuhler, "Shift-share instruments and the impact of immigration," 2018.

Katzman, David M, "Before the ghetto: Black Detroit in the nineteenth century," *Science and Society*, 1975, *39* (4), 503–506.

Kearl, James R, "Inflation, mortgage, and housing," *Journal of political economy*, 1979, *87* (5, Part 1), 1115–1138.

King, A Thomas and Peter Mieszkowski, "Racial discrimination, segregation, and the price of housing," *Journal of Political Economy*, 1973, *81* (3), 590–606.

Laurenti, Luigi, *Property values and race*, Univ of California Press, 1960.

Lemann, Nicholas, *The promised land: The great black migration and how it changed America*, Vintage, 2011.

Li, Sijie, "Competitors and Consumers: The Impact of the Great Migration on Employment Outcomes of Black Northerners," 2019.

Logan, John R, Weiwei Zhang, Richard Turner, and Allison Shertzer, "Creating the black ghetto: Black residential patterns before and during the Great Migration," *The ANNALS of the American Academy of Political and Social Science*, 2015, *660* (1), 18–35.

Long, Jason and Joseph Ferrie, "Intergenerational occupational mobility in Great Britain and the United States since 1850," *American Economic Review*, 2013, *103* (4), 1109–37.

Long, Larry H and Lynne R Heltman, "Migration and income differences between black and white men in the North," *American Journal of Sociology*, 1975, *80* (6), 1391–1409.

Margo, Robert A, "Explaining black-white wage convergence, 1940–1950," *ILR Review*, 1995, *48* (3), 470–481.

Massey, Douglas S and Nancy A Denton, *American apartheid: Segregation and the making of the underclass*, Harvard University Press, 1993.

Mazumder, Bhashkar, "Black–white differences in intergenerational economic mobility in the United States," 2014.

McEntire, Davis, *Residence and race: final and comprehensive report to the Commission on Race and Housing*, Univ of California Press, 1960.

Mehlhorn, Dmitri, "A requiem for blockbusting: Law, economics, and race-based real estate speculation," *Fordham L. Rev.*, 1998, *67*, 1145.

Myrdal, Gunnar, "An American dilemma: The Negro problem and modern democracy," 1944.

Ouazad, Amine, "Blockbusting: Brokers and the dynamics of segregation," *Journal of Economic Theory*, 2015, *157*, 811–841.

Pérez, Santiago, "Intergenerational Occupational Mobility across Three Continents," *The Journal of Economic History*, 2019, pp. 1–34.

Peri, Giovanni and Chad Sparber, "Task specialization, immigration, and wages," *American Economic Journal: Applied Economics*, 2009, *1* (3), 135–69.

Poterba, James M, "Taxation and housing: Old questions, new answers," Technical Report, National Bureau of Economic Research 1992.

Rapkin, Chester and William G Grigsby, *The demand for housing in racially mixed areas: A study of the nature of neighborhood change*, Univ of California Press, 1960.

Robinson, Peter M, "Root-N-consistent semiparametric regression," *Econometrica: Journal of the Econometric Society*, 1988, pp. 931–954.

Ruggles, Steven, Katie Genadek, Ronald Goeken, Josiah Grover, and Matthew Sobek, "Integrated Public Use Microdata Series: Version 7.0[dataset]," Minneapolis: University of Minnesota 2017.

Saavedra, Martin and Tate Twinam, "A Machine Learning Approach to Improving Occupational Income Scores," 2018.

Satter, Beryl, *Family properties: Race, real estate, and the exploitation of black urban America*, Macmillan, 2009.

Schafer, Robert, "Racial discrimination in the Boston housing market," *Journal of Urban Economics*, 1979, *6* (2), 176–196.

Schelling, Thomas C, "Dynamic models of segregation," *Journal of mathematical sociology*, 1971, *1* (2), 143–186.

Schietinger, Egbert Frederick, "Racial succession and changing property values in residential Chicago." PhD dissertation, University of Chicago, Department of Sociology 1953.

Sequeira, Sandra, Nathan Nunn, and Nancy Qian, "Migrants and the making of America: The short-and long-run effects of immigration during the age of mass migration," 2017.

Shertzer, Allison, "A Review Essay on Howard Bodenhorn's The Color Factor: The Economics of African-American Well-Being in the Nineteenth-Century South," *Journal of Economic Literature*, 2018, *56* (1), 206–16.

_ **and Randall P Walsh**, "Racial sorting and the emergence of segregation in American cities," 2016.

_ **,** _ **, and John R Logan**, "Segregation and neighborhood change in northern cities: New historical GIS data from 1900–1930," *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 2016, *49* (4), 187–197.

Smith, James P and Finis R Welch, "Black economic progress after Myrdal," *Journal of Economic Literature*, 1989, *27* (2), 519–564.

Tabellini, Marco, "Gifts of the Immigrants, Woes of the Natives: Lessons from the Age of Mass Migration," 2017.

Tolnay, Stewart E, "The African American "great migration" and beyond," *Annual Review of Sociology*, 2003, *29* (1), 209–232.

Trotter, Joe William, *Black Milwaukee: The making of an industrial proletariat, 1915-45*, University of Illinois Press, 1985.

Wilson, William Julius, *The truly disadvantaged: The inner city, the underclass, and public policy*, University of Chicago Press, 2012.

Wolff, Edward N, "Household wealth trends in the United States, 1962-2013: What happened over the great recession?," Technical Report, National Bureau of Economic Research 2014.

Wright, Gavin, *Old South, new South: Revolutions in the southern economy since the Civil War*, Basic Books (AZ), 1986.

Yinger, John, "The Black-White Price Differential in Housing: Some Further Evidence.," 1975.