NBER WORKING PAPER SERIES

RACIAL SEGREGATION IN HOUSING MARKETS AND THE EROSION OF BLACK WEALTH

Prottoy A. Akbar
Sijie Li
Allison Shertzer
Randall P. Walsh

Working Paper 25805
http://www.nber.org/papers/w25805

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 2019

Support for this research was provided by the National Science Foundation (SES-1459847). We thank Werner Troesken, Arie Beresteneau, Dan Bogart, Ed Coulson, William Even, Thomas Davidoff, Leah Boustan, Hoyt Bleakley, Bill Collins, Ross Mattheis, Christian Dippel and seminar participants at the NBER (DAE), Arizona, Harvard, UC-Irvine, Miami of Ohio, Pittsburgh, and Tennessee for helpful comments. We thank John Logan for assistance with enumeration district mapping and Carlos Villarreal and the Union Army Project (www.uadata.org) for the 1930 street files. Antonio Diaz-Guy, Phil Wetzel, Jeremy Brown, Andrew O'Rourke, Aly Caito, Loleta Lee, Zach Gozlan, and Sai Konduru provided outstanding research assistance. Corresponding author's email: shertzer@pitt.edu (A. Shertzer). The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

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 They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2019 by Prottoy A. Akbar, Sijie Li, Allison Shertzer, and Randall P. Walsh. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Racial Segregation in Housing Markets and the Erosion of Black Wealth
Prottoy A. Akbar, Sijie Li, Allison Shertzer, and Randall P. Walsh
NBER Working Paper No. 25805
May 2019
JEL No. J15,N12,R31

ABSTRACT

Housing is the most important asset for the vast majority of American households and a key driver of racial disparities in wealth. This paper studies how residential segregation by race served to erode black wealth. Using a novel 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. However, home values fell on average by 10 percent over the first decade of racial transition 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.

Prottoy A. Akbar
University of Pittsburgh
Dietrich School of Arts and Sciences
Department of Economics
4901 Wesley W. Posvar Hall
230 South Bouquet Street
Pittsburgh, PA 15260
prottoyamanakbar@pitt.edu

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

Sijie Li
Department of Economics
University of Pittsburgh
4519 Wesley W. Posvar Hall
Pittsburgh, PA 15260
sijie.li@pitt.edu

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

"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* (2017)

"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)

I. 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, Albouy and Zabek 2016). 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 (Massey and Denton 1993; Wilson 2012; Chetty, Hendren, Kline, and Saez 2014).

In this paper we study housing market dynamics that arose due to the racial segregation of northern cities in the United States prior to World War II. The narrative history suggests that once established the urban color line moved because black families, desperate for better quality housing, outbid whites for the purchase of homes in neighborhoods just outside of the established ghetto (Mehlhorn 1998, Troesken and Walsh, forthcoming). In response to these new black arrivals, and at least in part compelled by concerns about falling home values and the quality of public services, white households subsequently fled these transitioning areas (Boustan 2010, Shertzer and Walsh 2016). Further, as black renters arrived on transitioning blocks, they

faced higher rents than those paid by their white counterparts for similar housing on nearby blocks. Highlighted here, and by the passages that begin our paper, is the tension between elevated prices in transitioning neighborhoods and the expectation that prices will fall as a result of black entry.

To systematically investigate these housing market dynamics, we construct a novel dataset of rents, home values, and the racial composition of city blocks in prewar American cities. We find that pioneering black families did indeed pay a purchase premium for housing in newly transitioning areas. As the block continued its racial transition rents increased while the values of homes, initially purchased at a premium by early black arrivals, fell below those of similarly situated housing on all-white blocks. The finding that racial transition led black families to face both *higher* rents and the *erosion* of home values poses a puzzle, which we unravel below. Further, in spite of the extant qualitative narrative, we are aware of no empirical work documenting this link between racial transition and housing market dynamics. Our analysis documents that segregation and the process of ghetto expansion left African Americans both living in declining neighborhoods and doubly poorer. The impacts of these dynamics still resonate today.

To understand the mechanisms that could generate our findings, we consider the capitalization rate that is implied by a no arbitrage condition between rents and home values (Kearl 1979; Poterba 1992). We formalize the capitalization rate as the percent of a property's 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, distaste for

black tenants, 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 rental rates. 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. Meanwhile, 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 to these declining prices, we observe sharp increases in rental prices for all houses on transitioning blocks, whether occupancy increased or not. 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.

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. We also 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, home values on a block dropped by a staggering 50 percent over the first decade of racial transition.

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 on blocks that were more than 50 percent black. And, during the beginning of the transition process, when black arrivals tended to be better off and were more likely to buy their home, these families paid a premium relative to

white owners. Turning to the rise in rental prices, we similarly find the largest impacts on blocks that attained majority black status. On these blocks black arrivals were typically poorer renters. Furthermore, rents only increased for black families, not white families who remained during the transition process. Black families thus bore the brunt of both shifts in prices: when black families tried to escape escalated rents through the purchase of their own home, these same market dynamics led to the erosion of their homes' value.

Our findings have important implications for understanding racial disparities in wealth. 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 losses in wealth associated with segregated housing markets.

Our work is also related to recent scholarship on how historical discrimination in the mortgage insurance market served to exacerbate forces operating in labor markets and educational systems to limit black accumulation of housing wealth.[2] However, we focus on private market dynamics that were in place prior to large-scale federal involvement in mortgage lending markets. In these prewar markets, real estate investors faced such large demand from black 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. Our findings suggest that the dynamics associated with

[2] 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, Hartley, and Mazumder (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).

racially segregated housing markets may have led to substantial erosion of black household wealth even in the absence of discriminatory federal government involvement in housing markets.

II. Background and Related Work

A. 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 *Shelley v. Kraemer*.

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 urban 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 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

[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).

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 former homeowners who decided to convert the house into a rental property instead of selling. Both considerations underscore the fact that in our setting the owners of rental properties were most likely white.

B. 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 1978; Schafter 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).

[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. The investors are described as follows: "It is a certain class of individuals in a great many cases that buys up these properties and gets as much out of them as they possibly can until the properties are condemned or fall down or are converted to some other use… In a great many other cases [the houses] have been in the family for years, and the family does not know how to get rid of it, so they just keep renting the house."

Yet establishing that black and white families paid different amounts for the same quality of housing is extraordinarily difficult, particularly in a historical context. 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, Glaeser, and Vidgor (1999), which proposes an indirect empirical test of the 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 supposedly 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, Hartley, and Mazumder 2017). Beginning in 1934, at the height of the Depression, 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 across the entire 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 exceedingly unlikely that federal government policies can explain the findings of this paper. Instead, the FHA and subsequent federal policies likely served to institutionalize and reinforce the private market dynamics that we document.

III. 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

[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 very 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).

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. 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 II and III).[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 III. There is some evidence that addresses with fewer occupants were more likely to be matched although the differences are economically small.[7]

We aggregate households in addresses with multiple units to obtain aggregate rents and occupancy. Addresses that report both an owner and a renter are dropped from the sample. One concern with using self-reported valuations as a measure of home prices is accuracy. For a sample of addresses in Pittsburgh that sold in 1940, we found the corresponding address in the county Recorder of Deeds office and obtained the actual sales price. We plot the differential between the census valuation and the sales price in Appendix Figure I. The figure suggests that there is no systematic bias.

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, Walsh, and Logan 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. 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

[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 III is statistically significant.

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 1, 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 to 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.

IV. Semi-Parametric Analysis

We begin with a discussion of the underlying dynamics in our data, Figures 1 and 2 present the semiparametric relationship between racial transition and rents and home prices estimated using the Robinson's double residual method (Robinson, 1998). 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 home's value in 1930.[8] We begin in Panel A of Figure 1 by showing the relationship between 1940 black share and the log rent for all houses that had switched to being rentals by 1940. Overall, rents are flat for low black shares but then swell by 30 log points between 50 and 90 percent black share, reflecting an increase of 35 percent. In Panel B, we decompose rents based on the race of the occupant. The figure shows that while blacks in general always paid a premium relative to whites to rent on the same block, this premium grew significantly along with the level of transition on the block above approximately 40 percent black. Thus, over this range white renters demanded, and landlords were willing to provide, a significant discount to remain on transitioned blocks. The most direct explanation for these differentials is racial animus.

We now turn from rents to valuations of home prices. If racial market dynamics were driven solely by supply restrictions in the market for black housing related 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, as is shown in Panel A of Figure 2, overall home values in fully transitioned neighborhoods actually declined by 10 log points (about 11 percent). In Panel B we decompose this relationship by the race of the owner. Here we find that both racial groups saw their homes lose value and there is overall less divergence based on the homeowner's race. Nevertheless, because early in the transition black pioneers actually paid a premium of approximately 10 percent for housing

[8] That is, we estimate $\ln Price_{i40} = X_i'\beta + f(Black\ share_{i40}) + \epsilon_i$ where X_i includes controls for 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.

relative to prices paid by white households, black homeowners saw a larger erosion of wealth over the process of transition. Finally, Figure 3 summarizes black and white ownership rates across differing levels of transition. Panel A documents the high ownership rates among early black pioneers. These high black ownership rates heighten the important role that falling home prices on transitioning blocks played in eroding black wealth.

We now turn to a more parametrized analysis. Here, we have two primary goals. First, we seek to ascertain whether the relationships we document in the semi-parametric analysis are causal. Second, we look to better understand the divergence between rents and owner-occupied housing prices.

V. Capitalization Framework and Parametric Analysis

Our parametric 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:

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

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

$$P_{it} = c_t * \rho_t * Q(Z_i) \quad (1)$$

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 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:

$$rent_{it} = c_t * sale_{it} \quad (2)$$

The capitalization rate, c_t can be decomposed as follows:

$$c_t = i + \tau_p + risk + maintenance + depreciation - appreciation$$

where i is the risk-free interest rate and τ_p captures the relevant property taxes. Intuitively, the real estate investor must receive a return on her investment equal to the risk free interest rate available in the broader market place. This risk-free rate of return is adjusted for additional costs and benefits associated with owning the property. In particular: property taxes, a risk premium associated with housing price uncertainty, costs for maintaining the property, physical depreciation, and appreciation net of the overall inflation rate, with all of these terms expressed as percentages of the properties values. We also note that the results presented in Panel B of Figure 1 suggest the inclusion of an additional term to account for landlord preferences over the race of potential tenants. The large divergence that we observe in the figure between the rents paid by black and white households, *holding block-level racial composition constant*, suggests the importance of accounting for racial preferences in our capitalization framework.

By combining equations (1) and (2), we can derive a unified expression for P_{it}:

$$P_{it} = \rho_t * c_t^{I_{rent}} * Q(Z_i) \quad (3)$$

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

$$\ln P_{it} = \ln \rho_t + \ln c_t * I_{rent} + q(Z_i) \quad (4)$$

where, $q(Z_i) = lnQ(Z_i)$. 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 (4) for $q(Z_i)$ gives: $q(Z_i) = \ln P_{it} - \ln \rho_t - \ln c_t * I_{rent}$. 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 (4) yields the following expression for 1940 prices:

$$\ln P_{i40} = \ln \rho_{40} - \ln \rho_{30} + \ln P_{i30} + \ln c_{t40} * I_{rent40} \, . \quad (5)$$

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

$$\ln P_{i40} = \alpha + \beta * I_{rent40} + \gamma \ln P_{i30} + \epsilon_i \, . \quad (6)$$

We can interpret the key coefficients in equation (6) 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 50 percent black in 1940. Finally, we add the transition variable and its interaction with the rent indicator to equation (6) yielding our basic specification:

$$\ln P_{i40} = \alpha + \beta_{trans} * I_{trans_i} + \beta_{rent} * I_{rent_i} + \beta_{transXrent} * I_{transXrent_i} + \gamma \ln P_{i30} + \epsilon_i \quad (7)$$

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]

[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 and un-transitioned neighborhoods.

One potential concern is 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).

As a final preliminary, in Figure 4 we illustrate the relationship between black shares in 1930 and 1940. The figure shows that 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, Mas, and Rothstein 2008) in this context was 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.

A. Baseline Results

We begin our parametric analysis by relating changes in block-level racial composition to changes in housing prices over the 1930s. 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 2 reports

the 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 qualitatively 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 (-2.211) captures the log of the capitalization rate for blocks that did not experience racial transition, which implies a capitalization rate of 11 percent. The Great Depression was associated with substantial housing price deflation, and thus we should expect capitalization rates that are in general larger than those from the current day, which tend to center around 6 percent (see for instance Davis et al. 2008). 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 (-.113) 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). Finally, the exponent of the sum of the rented 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.[11] 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. We note that the channel through which racial preferences impact prices and rents is less direct here where we are comparing average prices and rents across blocks at *different stages of racial transition* than it was in panel B of Figure 1 where we compared differences in rent paid by black and white households on blocks at *identical stages of transition*. In the former comparison, differentials likely arose directly from white landlords preferring to rent to white tenants and thus charging different rents for identical properties. Here, we are focused on how rents change with block-level transition, independent of the race of a home's resident. The primary channel through which racial preferences drive cross-block market dynamics are more likely white flight and related expectations about future price drops.

Independent of racial preferences, 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.

[11] 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.

We explore this notion further in panel B of Table 2, which repeats the estimations from panel A with the log of aggregate occupancy as the outcome variable. Results are generally similar across all four models. Houses that switched from being owned to rented saw increases in their aggregate occupancy of approximately 20 percent even on blocks that remained white. The occupancy increase was particularly pronounced in blocks that transitioned, however. The estimates from column (3) indicate that rental occupancy soared by 45 percent in homes that switched to being rented on such blocks.[12] 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 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 raise the possibility that observed increase in capitalization rates on transitioned blocks, and the associated rent spikes, could simply be the direct result of increases in maintenance or physical depreciation costs arising from higher-density habitation. To examine this issue directly, in Table 3 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.

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

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 8 which we discuss below). We can thus reject the notion that our results are driven mainly by occupation rates. 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.

B. 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 are proximate to the South and already had relatively large black populations by 1900. We then 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 4 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. Consistent with our arbitrage model, also we find that rental premia were also higher in high-migration cities than in low-migration in cities. Finally, black families renting on

transitioning blocks in border cities paid the largest rental 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. Heterogeneity in occupancy impacts are less striking, with addresses that both switched to being rentals and that were located on transitioning blocks seeing the smallest occupancy increases in low-migration cities.

C. Decomposing Transition

Our initial results are based on a relatively granular 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 5. For ease of interpretation, we also present the results visually in Figures 5 and 6.

Figure 5 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 6 we summarize the occupancy results from column (4) of Table 5, 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 blocks that were 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.

D. 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 affecting prices that are constant over very small neighborhood definitions. 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.[13] This geocoded subsample allows us to directly test the efficacy of our ED fixed effects in controlling for ghetto proximity. Appendix Figure II 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 1960s and 1970s may be overwhelming represented in the set of addresses that could not be geocoded. Columns (2) and (5) of Table 5 replicate the models of columns (1) and (4) on the geocoded subsample, 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)

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

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 6 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 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 III 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 7 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 7 as strong evidence that our price and occupancy results are driven by racial transition and not by other factors.

VI. 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 explored the question of homeownership by race in Figure 3 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 2), these results suggest that proportionally more homes switched hands from white to black families prior to the drop in values that accompanied racial transition. Note also that black families typically bought at a premium in early stages of transition (see Panel B of Figure 2). Similarly, rents did not sharply increase until

a block was more than 50 percent black, when Figure 3 shows that most black families were renters. As a result, black families appear to have borne the worst of both the escalation of rents and the fall in home values.

The question of who owned the rental housing in black neighborhoods is also important for understanding our findings given that only 30 percent of families on fully transitioned blocks were homeowners. 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 8, we replicate our baseline specifications from Table 2 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.

VII. Conclusion

In this paper we constructed a novel dataset of rents, home values, and the racial composition of city blocks in prewar American cities to systematically investigate the housing market dynamics associated with black entry into white neighborhoods. We find that racial

transition was associated with both increases in aggregate rental prices and decreases in property values. To our knowledge this is the first paper to demonstrate that black entry into a neighborhood caused the price of owned and rented housing to diverge, a finding that is consistent with much of the narrative history.

Impacts of racial transition were 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. We also found that cities that saw the largest inflows of black migrants experienced the largest rent and price impacts from racial transition. In these cities, home values on a block dropped by a staggering 50 percent over the first decade of racial transition. The impact of these market dynamics for racial wealth inequality were further exacerbated by our finding that pioneering black families paid a significant premium for homes on majority white blocks at the early stages of transition. Similarly rent discounts to white families that remained on transitioning blocks later into the process also further eroded black wealth relative to that of whites.

The dramatic decline in property values had important implications for city budgets and real estate investors alike. Rental property owners, faced with the costs of creating and maintaining rental units that were going to depreciate in value and with 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 reinforced each other, during which entire sections of cities transitioned from being all white to majority black over a relatively short period, with devastating results for black household wealth.

Our results underscore the importance of segregation and white flight in explaining racial disparities in wealth accumulation.

References

Aaronson, Daniel, Daniel A. Hartley, and Bhashkar Mazumder. "The Effects of the 1930s HOLC 'Redlining' Maps." Manuscript, 2017.

Albouy, David, and Mike Zabek. *Housing inequality*. No. w21916. National Bureau of Economic Research, 2016.

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* 3.2 (2011): 34-66.

Bailey, Martha, Connor Cole, Morgan Henderson, and Catherine Massey. *How Well Do Automated Methods Perform in Historical Samples? Evidence from New Ground Truth*. No. w24019. National Bureau of Economic Research, 2017.

Blau, Francine D., and John W. Graham. "Black-white differences in wealth and asset composition." *The Quarterly Journal of Economics* 105.2 (1990): 321-339.

Boustan, Leah Platt. "Was postwar suburbanization "white flight"? Evidence from the black migration." *The Quarterly Journal of Economics* 125.1 (2010): 417-443.

Boustan, Leah P., and Robert A. Margo. "A silver lining to white flight? White suburbanization and African–American homeownership, 1940–1980." *Journal of Urban Economics* 78 (2013): 71-80.

Card, David, Alexandre Mas, and Jesse Rothstein. "Tipping and the Dynamics of Segregation." *The Quarterly Journal of Economics* 123.1 (2008): 177-218.

Chambers, Daniel N. "The racial housing price differential and racially transitional neighborhoods." *Journal of Urban Economics* 32.2 (1992): 214-232.

Chetty, Raj, Nathaniel Hendren, 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* 129.4 (2014): 1553-1623.

Collins, William J., and Katharine L. Shester. "Slum clearance and urban renewal in the United States." *American Economic Journal: Applied Economics* 5.1 (2013): 239-73.

Collins, William J., 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* 6.1 (2014): 220-52.

Cutler, David M., Edward L. Glaeser, and Jacob L. Vigdor. "The rise and decline of the American ghetto." *Journal of political economy* 107.3 (1999): 455-506.

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 54.2 (2008): 279-284.

Feigenbaum, James J. "Automated census record linking: A machine learning approach." (2016).

Follain Jr, James R., and Stephen Malpezzi. "Another look at racial differences in housing prices." *Urban Studies* 18.2 (1981): 195-203.

Kearl, James R. "Inflation, mortgage, and housing." *Journal of Political Economy* 87.5, Part 1 (1979): 1115-1138.

King, A. Thomas, and Peter Mieszkowski. "Racial discrimination, segregation, and the price of housing." *Journal of Political Economy* 81.3 (1973): 590-606.

Laurenti, Luigi. *Property Values and Race*. University of Berkeley Press, 1960.

John R. Logan and Weiwei Zhang. 2017. "Developing GIS Maps for U.S. Cities in 1930 and 1940" In Don Lafreniere, Ian Gregory, and Don Debats (editors), *The Routledge Handbook of Spatial History. Routledge: UK.*

Long, Jason, and Joseph Ferrie. "Intergenerational occupational mobility in Great Britain and the United States since 1850." *The American Economic Review* 103.4 (2013): 1109-1137.

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

McEntire, Davis. *Residence and race*. Univ of California Press, 1960.

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

Myrdal, Gunnar. *An American dilemma; the Negro problem and modern democracy*. Harper and Bros., 1944.

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

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): 931-954.

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* 6.2 (1979): 176-196.

Schelling, Thomas C. "Dynamic models of segregation." *Journal of mathematical sociology* 1.2 (1971): 143-186.

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

Shertzer, Allison, and Randall P. Walsh. *Racial sorting and the emergence of segregation in American cities*. No. w22077. National Bureau of Economic Research, 2016.

Shertzer, Allison, Randall P. Walsh, 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* 49.4 (2016): 187-197.

Troesken, Werner, and Randall P. Walsh. "Collective Action, White Flight, and the Origins of Racial Zoning Laws." Journal of Law, Economics and Organization (Forthcoming).

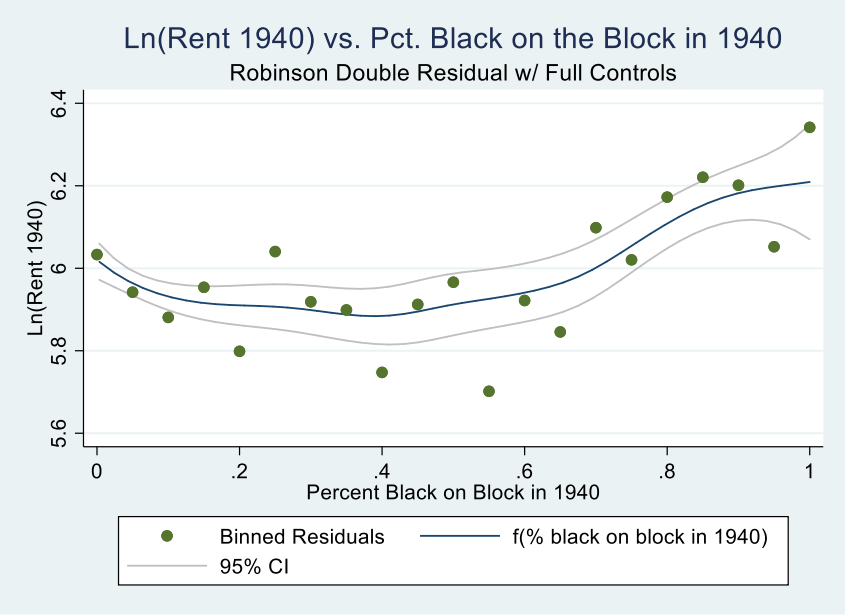
Wilson, William Julius. *The truly disadvantaged: The inner city, the underclass, and public policy*. University of Chicago Press, 2012.

Wolff, Edward N. *Household Wealth Trends in the United States, 1962-2013: What Happened over the Great Recession?*. No. w20733. National Bureau of Economic Research, 2014.

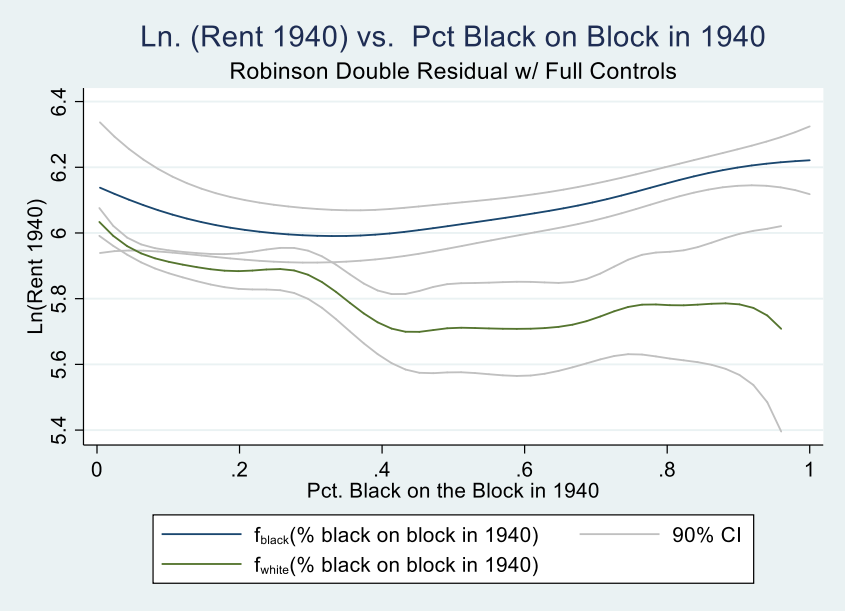
Yinger, John. "The black-white price differential in housing: some further evidence." *Land Economics* 54.2 (1978): 187-206.

Figure 1. Semiparametric relationship between Percent Black and Rents

Panel A. Baseline Sample



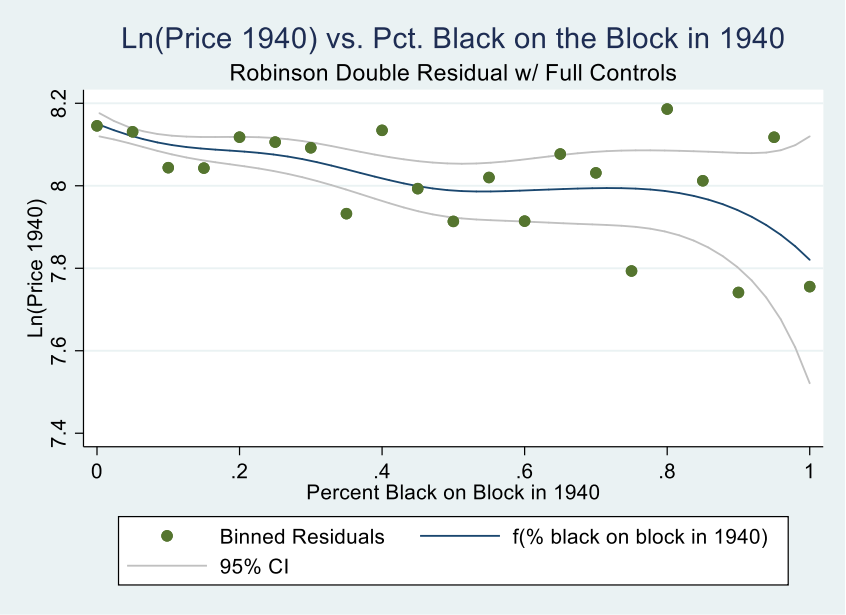
Panel B. Black and White Households Separately



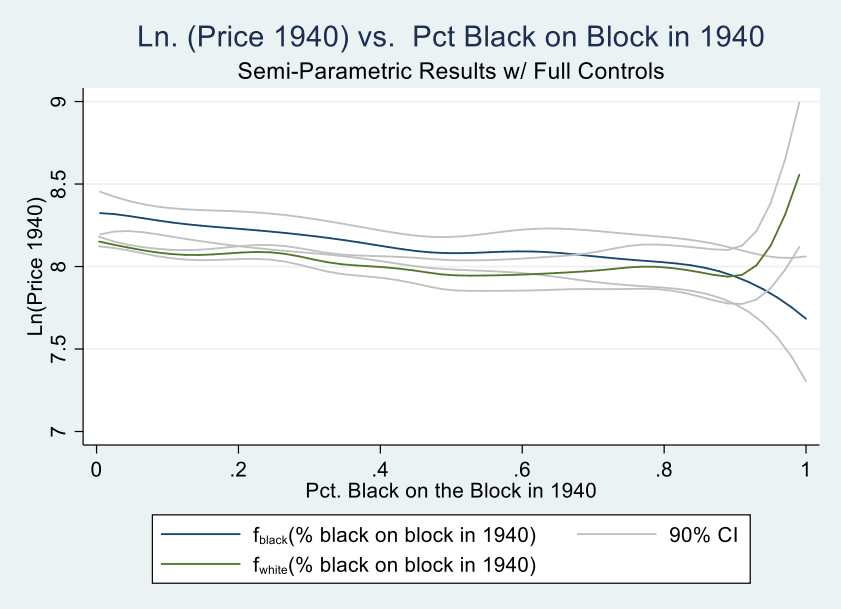
Notes: These figures show the semiparametric relationship between percent black on the block in 1940 (independent variable) and log rent in 1940 (dependent variable) on our baseline sample of homes that were single family, owner occupied, and located on a block that was at most 5 percent black in 1930. 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 Robinson's double residual method (1998). We also include binned residuals from the regression on each chart.

Figure 2. Semiparametric relationship between Percent Black and Home Values

Panel A. Baseline Sample



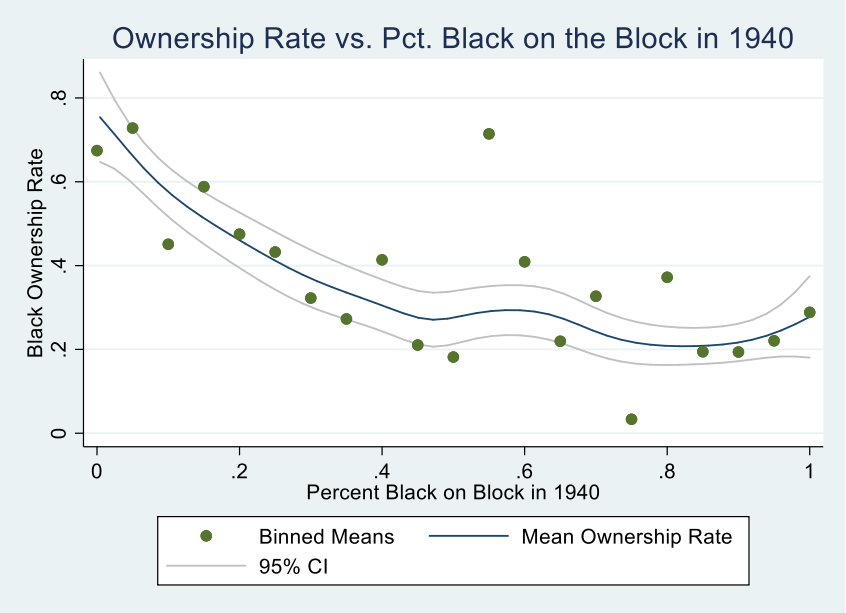
Panel B. Black and White Households Separately



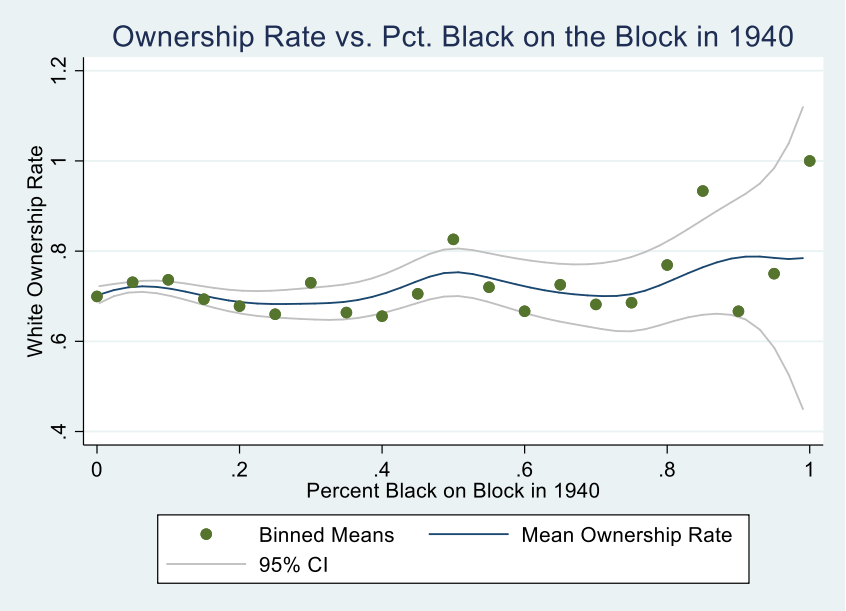
Notes: These figures show the semiparametric relationship between percent black on the block in 1940 (independent variable) and log home price in 1940 (dependent variable) on our baseline sample of homes that were single family, owner occupied, and located on a block that was at most 5 percent black 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 Robinson's double residual method (1998). We also include binned residuals from the regression on each chart.

Figure 3. The Relationship between Percent Black and Ownership Rates

Panel A. Black Ownership Rate

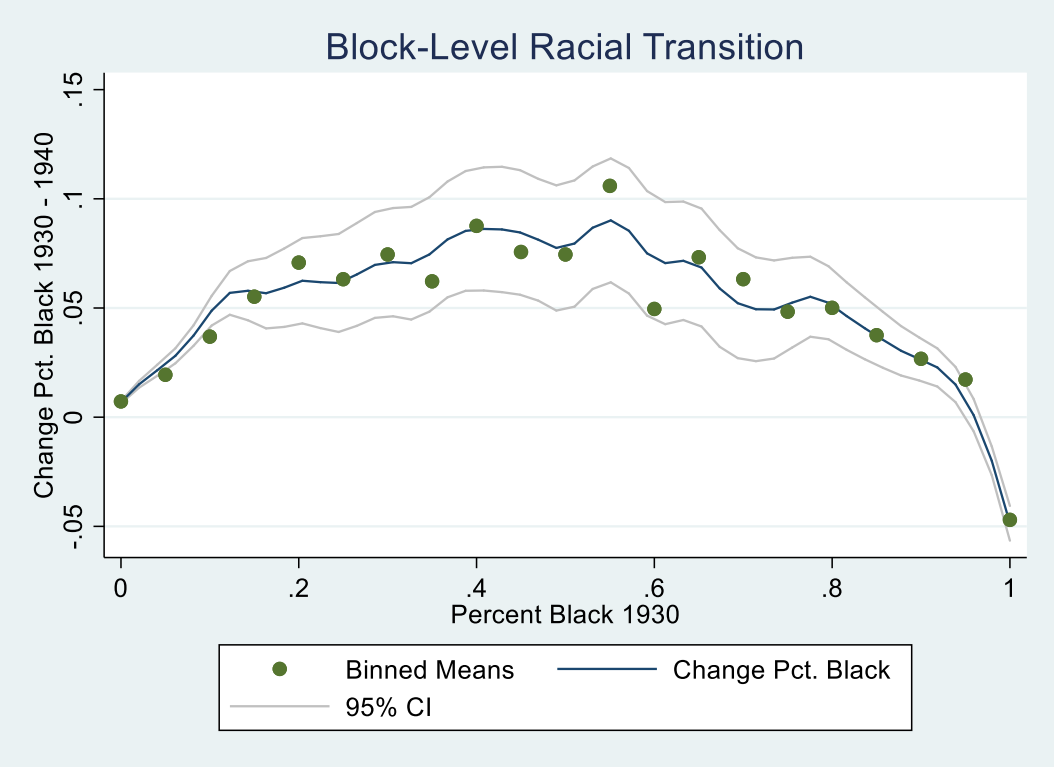


Panel B. White Ownership Rate



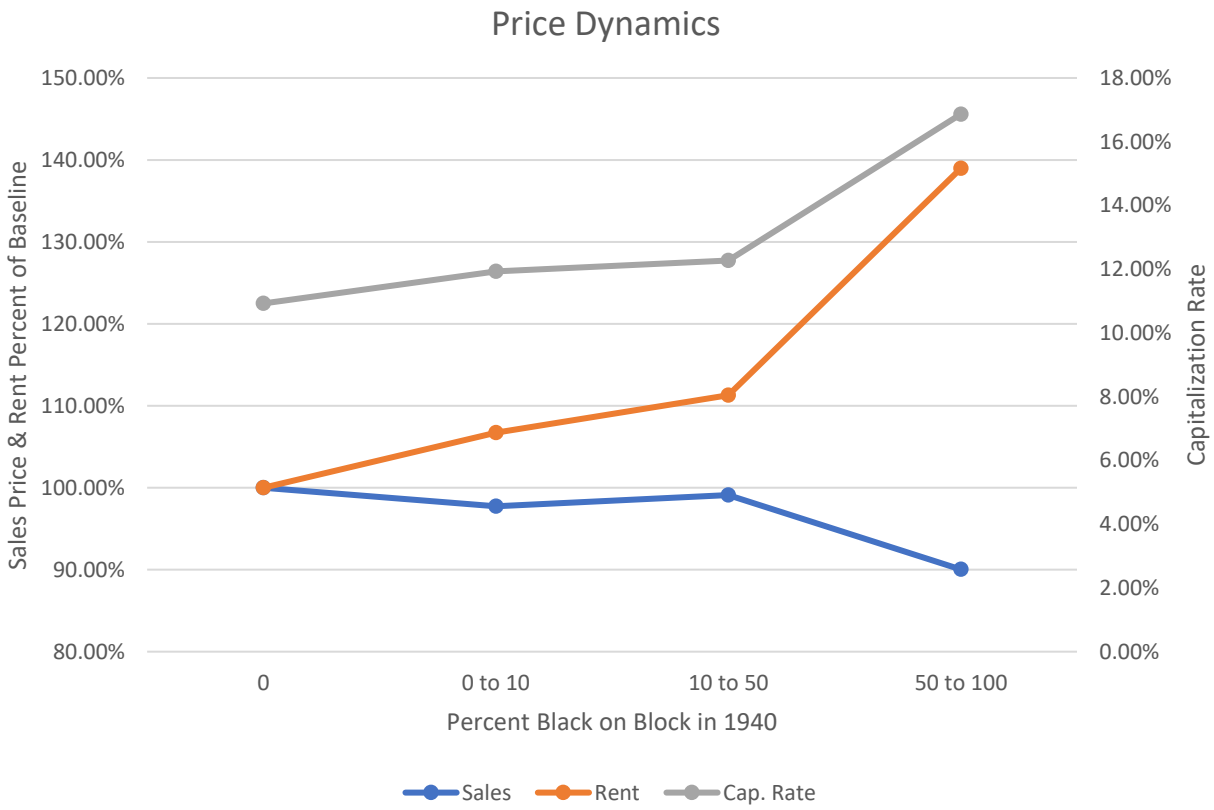
Notes: These figures show the semiparametric relationship between percent black on the block in 1940 (independent variable) and log rent in 1940 (dependent variable) on our baseline sample of homes that were single family, owner occupied, and located on a block that was at most 5 percent black in 1930. 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 Robinson's double residual method (1998). We also include binned residuals from the regression on each chart.

Figure 4. Relationship between 1930 Black Share and Black Population Growth



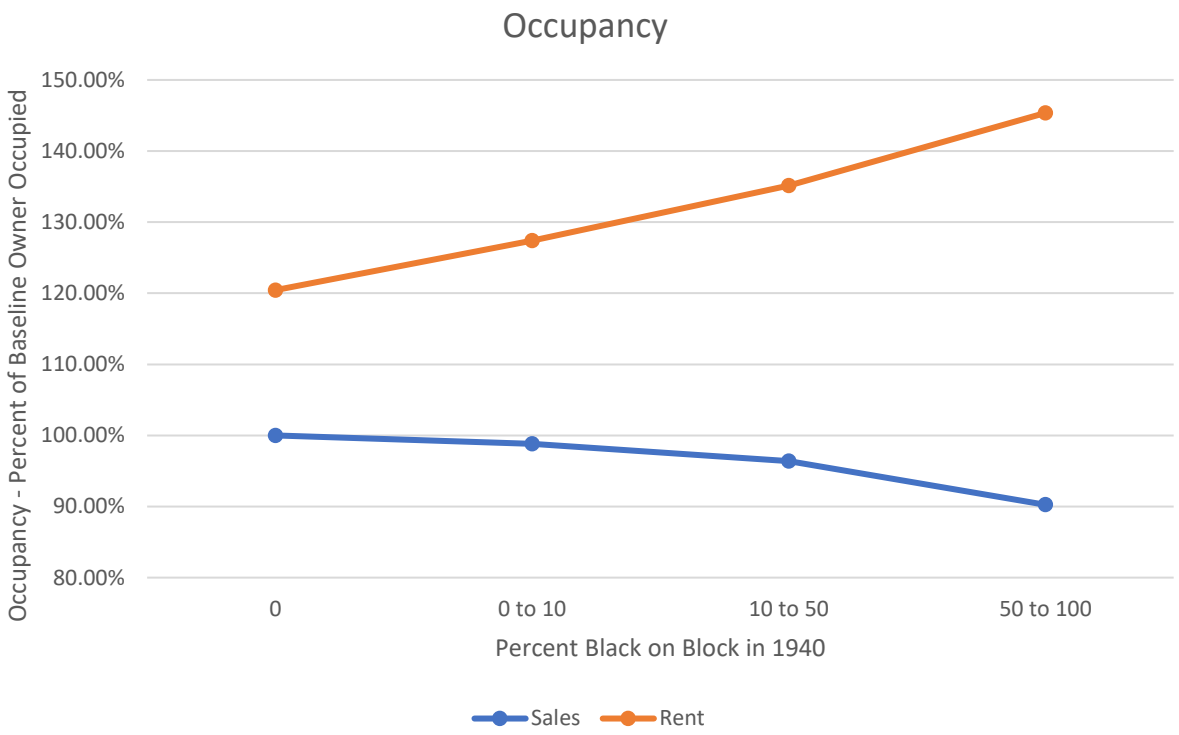
Notes: The figure presents a local polynomial smooth of block-level percent black in 1930 against the percent change in black share on the block over the next decade for every block in our sample.

Figure 5. Effect of Racial Transition on Prices



Notes: The figure shows the effects from the estimation of equation (7) presented in column (1) of Table 4 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 6. Effect of Racial Transition on Occupancy



Notes: The figure shows the effects from the estimation of equation (7) presented in column (4) of Table 4 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.

Table 1. 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 rented houses	-	61,838	-	1,166

Notes: 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 2. Main Results: Price and Occupancy

Panel A: Log price	No Controls (1)	Controls (2)	ED FE (3)	All Obs FE (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 (1)	Controls (2)	ED FE (3)	All Obs FE (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

Notes: 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 3. 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

Notes: 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 4. 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)
Observations	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)
Observations	253,375	76,887	30,543	145,945
R-squared	0.252	0.246	0.309	0.239

Notes: 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 5. 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

Notes: 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 6. Selection into Racial Transition in Baseline Sample

Mean on block in 1930:	Share Black on Block in 1940 <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 ghetto	1.45	0.29	0.07
Share black in 1930	0.00	0.01	0.01
N	41378	415	190

Notes: 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 7. 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

Notes: 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 8. 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)

Notes: 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.

Data Appendix

I. 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.[14] 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.

II. Details on Matching 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

[14] 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.

addresses imputed from adjacent households. A very small number of dwellings from the 1940 Census seem to have members belonging to different EDs/tracts. As with street names and house numbers, we assume the household head's 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.[15] 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,

[15] 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.

(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 household's 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 correct with a reasonable degree of accuracy that is our baseline.[16]

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.

III. Geocoding Addresses

We geocode all formatted address strings on Google Maps' Directions API. We include all 1930 and 1940 addresses with non-missing street names and house numbers, including those we

[16] 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.

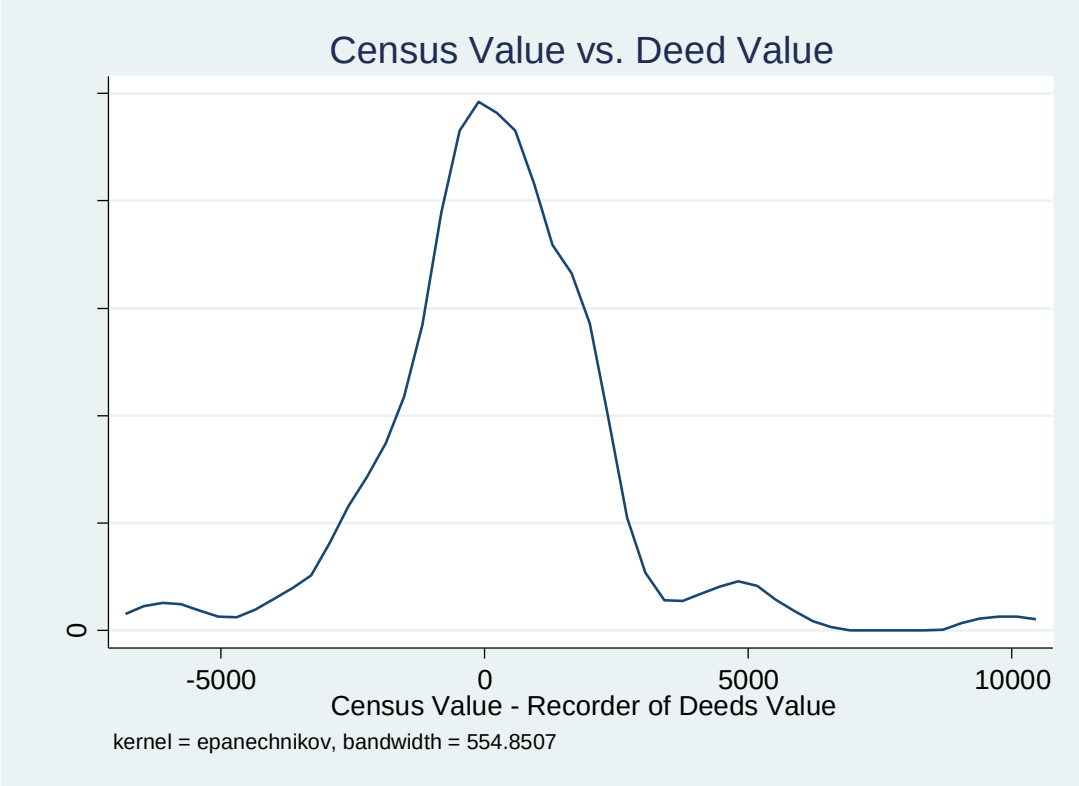
have flagged as potentially erroneous. The Directions API does a fuzzy name match of our input strings with addresses on Google's database and returns none, one or multiple location matches. For each location match, the API returns the geographic coordinates, the level of precision of the geocoding (e.g. "street address", "route", "intersection", "ward", etc.), and any administrative/political areas that the geocoded location falls within (e.g. the county, city, state, postal code or other well-defined "neighborhoods"). We drop any matches where the precision of the geocoding is an administrative area (e.g. a ward, a neighborhood, a city, etc.) or if the state differs from that of our city.

We then map each geocoded location to our 1930 ED and 1940 tract boundaries, and drop any location matches that do not coincide with either the ED or the tract associated with the input address. From each remaining geocoded location matches, we compute straight-line distances to the nearest ghetto. If there are still multiple location matches for an address, we keep the location match whose distance to ghetto is closest to the average distance from all location matches of addresses in the same block. Finally, in a small number of cases, when location matches for an address are tied in their deviation from the average distance to ghetto in the block, we pick the location match that appears first in the Directions API's sorting of results.[17]

To compute block-level distances to the nearest ghetto, we take the average of distances from each address in the block. As long as a block includes at least one address that is not flagged as problematic, we exclude distances from flagged addresses.

[17] The sorting reflects the "prominence" of the location, which is Google's measure of how likely the location is to be the result of a search.

Appendix Figure I. Self-Reported Value vs. Deed Value from County Records

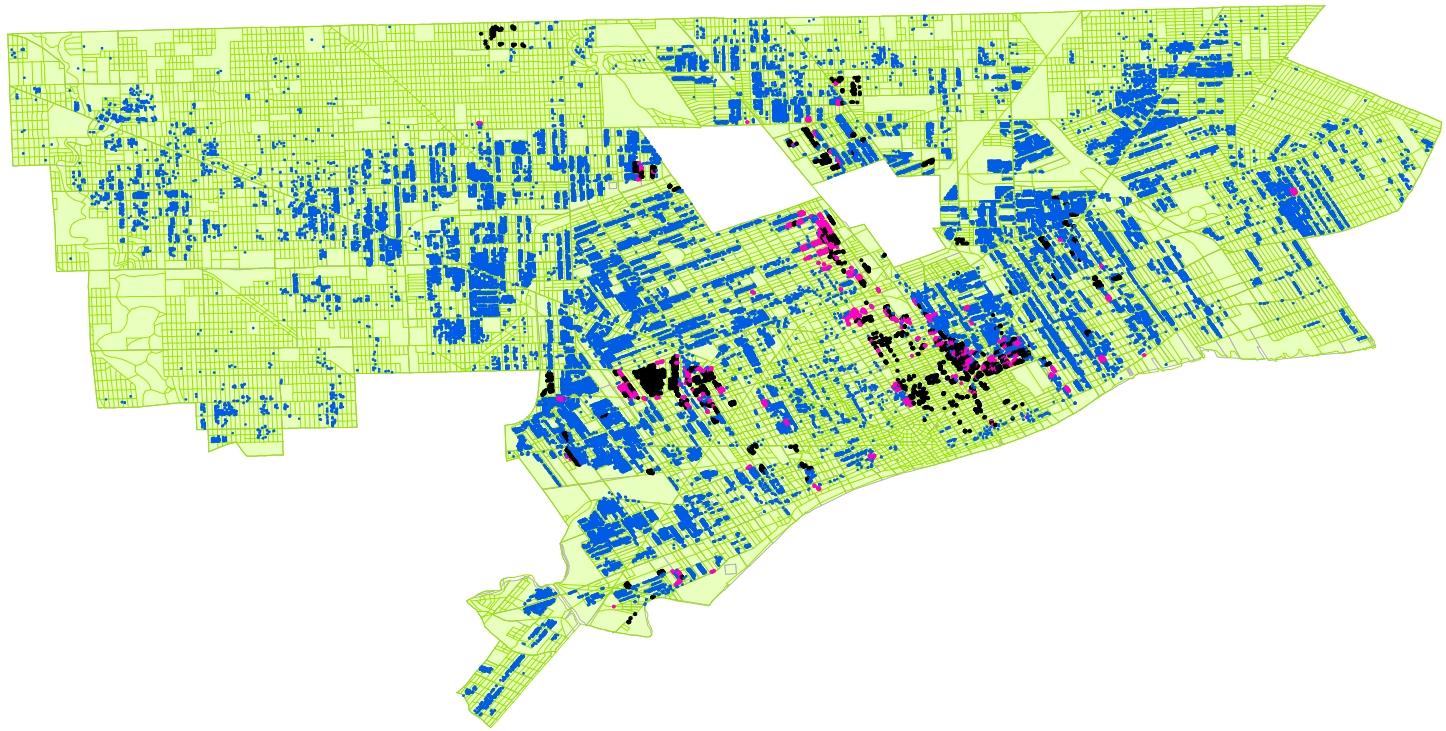


Notes: the figure presents a Kernel Density Estimate of the PDF of differences between self-reported home valuations as recorded in the decennial census and sales amounts as recorded by the Allegheny County Recorder of Deeds for a sample of 404 owner-occupied homes in the city of Pittsburgh. The data were constructed by identifying homes in the recorder of deed's records that were sold in either 1930 or 1940 and then hand matching them to the appropriate individual census record based on the home's address.

Appendix Figure II. 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).

Appendix Figure III. 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.

Appendix Table I. Address Statistics for Block Sample

	All Households				Addresses		Blocks		Addresses per Block	
	Total Households		Quality Address		Unique Addresses		Unique Blocks			
	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

Notes: 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. "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.

Appendix Table II. Address Sample Statistics

	Households with address found in both census years						Addresses		Households per Address	
	Total		Trimmed Sample		Quality Address		Unique Addresses			
	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

Notes: 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. 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.

Appendix Table III. 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

Notes: 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.