American Economic Journal: Economic Policy 2021, 13(4): 355–392
https://doi.org/10.1257/pol.20190414

The Effects of the 1930s HOLC "Redlining" Maps†

By Daniel Aaronson, Daniel Hartley, and Bhashkar Mazumder*

This study uses a boundary design and propensity score methods to study the effects of the 1930s-era Home Owners Loan Corporation (HOLC) "redlining" maps on the long-run trajectories of urban neighborhoods. The maps led to reduced home ownership rates, house values, and rents and increased racial segregation in later decades. A comparison on either side of a city-level population cutoff that determined whether maps were drawn finds broadly similar conclusions. These results suggest the HOLC maps had meaningful and lasting effects on the development of urban neighborhoods through reduced credit access and subsequent disinvestment. (*JEL* G21, J15, N32, N42, N92, R23, R31)

Social scientists have long been interested in the link between place and socioeconomic success, and there is now growing recognition that where you grow up may causally affect your socioeconomic outcomes (e.g., Chetty, Hendren, and Katz 2016). Moreover, striking racial gaps in outcomes have led to a large literature examining the role of residential segregation in explaining geographic disparities (e.g., Boustan 2011). We focus on credit access, a potential channel that could drive both place- and race-based disparities. Neighborhoods inappropriately deprived of credit could suffer from insufficient investment and become magnets for an array of social problems related to poverty.[1]

* Aaronson: Federal Reserve Bank of Chicago (email: daaronson@frbchi.org); Mazumder: Federal Reserve Bank of Chicago (email: bhash.mazumder@gmail.com); Hartley: Federal Reserve Bank of Chicago (email: Daniel.Hartley@chi.frb.org). Matthew Shapiro was coeditor for this article. The authors thank participants at the spring 2016 Federal Reserve System Meeting on Applied Microeconomics, the fall 2016 Federal Reserve System Meeting on Regional Economics, DePaul University, UIC, Marquette University, Vanderbilt University, the University of Pittsburgh, the University of Bergen, the University of North Carolina, the Cleveland Fed, the Federal Reserve Board of Governors, the Atlanta Fed, the Minneapolis Fed, the Philadelphia Fed, the NBER DAE meetings, the University of Chicago Law School, the University of Edinburgh, the University of Essex, the University of Sussex, UCL, Cambridge, and the Policy History Meetings. We also thank Leah Boustan, Bill Collins, Price Fishback, Ben Keys, Jeff Lin, Raven Molloy, Derek Neal, Jonathan Rose, Allison Shertzer, Tate Twinam, Randall Walsh, James Greer, Ladale Winling, and three anonymous referees for helpful comments. We are especially grateful to Stephanie Grove, Avi Moorthy, Leah Plachinski, and Joey Reiff for outstanding research assistance, the Digital Scholarship Lab at the University of Richmond for the digitized HOLC maps, and both the Digital Scholarship Lab and Price Fishback for the Area Description Files. The views expressed here do not necessarily represent the views of the Federal Reserve Bank of Chicago or the Federal Reserve System.

† Go to https://doi.org/10.1257/pol.20190414 to visit the article page for additional materials and author disclosure statement(s) or to comment in the online discussion forum.

[1] A few examples of studies that examine access to credit include: Cameron and Taber (2004) on skill investment, Black and Strahan (2002) on entrepreneurship, Carroll (2001) on consumption, and Breza and Kinnan (2018) on economic activity.

Our study is set in the aftermath of the Great Depression, when the federal government undertook dramatic reforms to limit foreclosures and stabilize the housing market. One seemingly innocuous initiative was the overhaul of property appraisal practices. The Home Owners Loan Corporation (HOLC), a now-defunct federal agency, drew maps for over 200 cities as part of its City Survey Program to document the relative riskiness of lending across neighborhoods. Neighborhoods were classified based on detailed risk-based characteristics (e.g., housing age and price). However, nonhousing attributes such as race, ethnicity, and immigration status were influential factors as well. Since the lowest-rated neighborhoods were drawn in red and often had the vast majority of African American residents, these maps have been associated with the so-called practice of "redlining" in which borrowers are denied access to credit due to the demographic composition of their neighborhood. Credit was also potentially restricted to neighborhoods scoring in the next lowest neighborhood grade marked in yellow, which has received much less public and academic attention.

We address one particularly important effect of the City Survey Program, namely, how the maps impacted narrow local areas surrounding the HOLC's demarcation of neighborhoods. We mainly focus on these border areas to leverage a research strategy that can plausibly deliver causal effects. The effects of the maps on other parts of redlined cities and the aggregate effect of the program are clearly important but much more challenging to identify. As we discuss below, we study aggregate effects for only a set of smaller cities that were close to the population cutoff used by the HOLC in the 1930s to determine inclusion in the City Survey Program.

Since our analysis is nonexperimental, our methodology must address confounding factors for valid inference. A key concern is that the maps may have simply reflected and codified preexisting differences in neighborhoods but not actually caused changes in credit access. We address this concern through a multipronged approach. We begin by considering changes over time in the difference in outcomes between neighbors that lived on either side of an HOLC boundary within a tightly defined geographic band, typically a few city blocks. Comparisons of spatially proximate neighbors address some confounding factors like access to labor markets, public transportation, or other local amenities that might differentially influence neighborhood growth. However, a border design on its own is insufficient since, as we document, there were preexisting differences and differential trends even among nearby neighbors.[2]

To address this problem, our main strategy compares "treated" boundaries with a set of comparison boundaries using propensity score weighting. The weighting ensures that preperiod patterns in race and housing characteristics in our weighted comparison boundaries are virtually identical to the treated boundaries. Our second strategy limits our sample to a subset of the HOLC borders that are least likely to have been drawn based on our propensity score analysis. We hypothesize that the demarcation of many of these borders reflected idiosyncratic factors. For example, they

[2] Bayer, Ferreira, and McMillen (2007) similarly find substantial demographic differences between households on the opposite sides of school attendance zones. Dhar and Ross (2012) find even larger neighborhood quality differences across school district boundaries.

may have been drawn simply to close a polygon and therefore do not reflect a gap in creditworthiness. Reassuringly, this sample of low-propensity-score borders exhibits *no* preexisting differences or trends across the two sides of the boundaries and therefore eliminates the need for a comparison group.

Our analysis uses 149 geocoded city maps from the Digital Scholarship Lab at the University of Richmond (Nelson et al. 2016) merged with (i) address-level data from the 100 percent count of the 1910 to 1940 US decennial censuses (Minnesota Population Center and Ancestry.com 2013), (ii) census tract-level data from the 1950 to 1980 censuses, and (iii) block- and block-group-level data from the 1990 to 2010 censuses (IPUMS, US NHGIS-A 2020). This provides a century of data on neighborhood characteristics, including race, home ownership, house values, rents, and population.

We find that areas that received a lower grade on the HOLC maps experienced worse housing market outcomes with respect to home ownership, house values, rents, and vacancy rates over subsequent decades. This suggests that there was significant and persistent housing disinvestment in the wake of restricted credit access. We also find that the maps affected the degree of racial segregation as measured by the fraction of African American residents on each side of a neighborhood boundary. Areas graded "D" (most risky) became more heavily African American than nearby C-graded areas over the twentieth century. This gap rises steadily from 1930 until about 1970 or 1980 before declining thereafter.

Moreover, we find a markedly similar temporal pattern in "C" neighborhoods that bordered "B" neighborhoods. Indeed, the housing effects were larger and a bit more persistent along the C–B boundaries than along the D–C boundaries. The C–B result is particularly noteworthy given there were virtually no Black residents in either C or B neighborhoods prior to the maps. These results reveal the importance of "yellow-lining" as a historical phenomenon. We consider some plausible explanations for the additional impact along C–B boundaries but are unable to come to any firm conclusions. We also show that our boundary results are robust to several reasonable modifications to measurement, sampling, and estimation approaches.

As noted earlier, a limitation of our main analysis is that we only focus on highly localized effects. However, there could have been other broader or "general equilibrium" effects of the HOLC maps that are not captured in our analysis. Other parts of cities may have been helped or harmed by the HOLC maps. For example, poverty may have become more concentrated in redlined areas but fallen in other areas. Partly to address this issue, we also use a discontinuity strategy that exploits the HOLC's decision to limit maps to cities with a population of 40,000 or more in the 1930s.[3] We compare the outcomes of cities with a population between 30,000 and 39,999 to cities with a population between 40,000 and 49,999 in 1930. We find that cities with HOLC maps experienced a relative decline in home ownership, house values, and rents accompanied by a rise in the share of African Americans. No such impact arises at placebo population cutoffs, such as 25,000. These findings suggest that, at least for this subsample of smaller cities, our main estimates based

[3] We thank William Collins for first suggesting this idea to us. Related work by Anders (2019) also uses the 40,000 population cutoff and finds that the maps led to higher rates of crime in redlined cities.

on borders do not appear to be offset by changes occurring in other parts of cities and therefore might reflect aggregate effects as well.

An added benefit of this alternative empirical strategy is that it uses a highly credible identification strategy where we do not rely on comparisons across borders. This addresses a concern that real estate professionals who created the HOLC maps were able to perfectly forecast the trajectory of US neighborhoods over subsequent decades and that the maps themselves played no direct role.

To put our results in perspective, if we look across *entire neighborhoods*, our findings suggest that the maps could account for between 15 and 30 percent of the D–C gaps in the share African American and home ownership and 40 percent of the gap in house values from 1950 to 1980. The maps account for roughly half of the home ownership and house value gaps along the C–B borders over the same period.

After 1970, many of our estimates wane, perhaps indicating that federal policies around lending, including the 1968 Fair Housing Act, the 1974 Equal Credit Opportunity Act (ECOA), and the 1977 Community Reinvestment Act (CRA), may have helped reverse the influence of the HOLC maps. Glaeser and Vigdor (2012) document a similar pattern in aggregate measures of segregation and likewise speculate that housing policies may have contributed to the decline in segregation post-1970. However, other factors undoubtedly contributed to the reversal of these trends and understanding their roles is an important topic for future research.[4]

I. Background and Related Literature

The HOLC and the City Survey Program.—After the Great Depression, house prices fell precipitously and a foreclosure crisis ensued (White 2014).[5] To address this devastating situation, the Roosevelt Administration initiated a series of federal programs intended to alter the nature of housing finance. These policies shifted mortgages from short duration loans with balloon payments to fully amortized higher loan-to-value mortgages with 15- to 20-year durations. The Federal Housing Administration (FHA) introduced mortgage insurance, and a secondary loan market was created by the Federal National Mortgage Agency.[6]

In 1932, the Federal Home Loan Bank Board (FHLBB) was created to charter and oversee federal savings and loan associations. An important new agency, operating at the direction of the FHLBB, was the HOLC.[7] Our study focuses on an initiative undertaken by the HOLC at the behest of the FHLBB: to introduce a systematic

[4] Other factors include restrictive covenants (Sood, Speagle, and Ehrman-Solberg 2019), zoning regulations (Shertzer, Twinam, and Walsh 2018), the location of highway construction (Brinkman and Lin 2019), urban renewal policies (Collins and Shester 2013), the urban riots of the 1960s (Collins and Margo 2007), public housing location (Hunt 2009), and Federal Housing Administration policies (Rothstein 2017). Some of these forces conceivably interacted with and were a part of the long-run reduced form effect of the HOLC maps.

[5] For example, foreclosure rates in New York City rose from essentially zero in the 1920s to as high as 7 percent in 1935 and averaged about 2 to 3 percent per year during the early and mid-1930s (Ghent 2011).

[6] Several studies describe the residential real estate environment at the time and evaluate the effectiveness of HOLC and FHA initiatives to deal with the foreclosure crisis (Wheelock 2008, White 2014, Fishback et al. 2011, Rose 2011, Ghent 2011, and Fishback et al. 2017). Fishback et al. (2017) emphasize complications in the mortgage market that slowed the 1930s housing recovery.

[7] The HOLC was initially tasked with issuing bonds to buy and refinance mortgages at more favorable terms to borrowers. By 1936, the HOLC had refinanced roughly 10 percent of nonfarm mortgages (Fishback et al. 2011).

appraisal process that included neighborhood-level characteristics when evaluating residential properties. The FHLBB was concerned about the long-term value of real estate now owned by the federal government as well as the health of the lending industry, which was devastated by the foreclosure crisis (Hillier 2005, Nicholas and Scherbina 2013). Using the new appraisal system, the HOLC drew residential "security" maps for 239 cities between 1935 and 1940 and completed more than 5 million appraisals. The maps and the appraisal process were intended to solve a coordination problem that would help ensure the continued stability of property values.[8]

The maps were based on the input of thousands of local brokers and appraisers, as well as surveys containing information on housing markets and demographic and economic characteristics. Neighborhoods were graded on a scale of A (least risky/most stable) to D (most risky/least stable). Hillier (2005) quotes the 1937 FHLBB Appraisal Manual in describing neighborhood grades as follows:[9]

- Grade A = "homogeneous," in demand during "good times or bad"
- Grade B = "like a 1935 automobile—still good, but not what the people are buying today who can afford a new one"
- Grade C = becoming obsolete, "expiring restrictions or lack of them" and "infiltration of a lower grade population"
- Grade D = "those neighborhoods in which the things that are now taking place in the C neighborhoods, have already happened"

The term "redlining" is thought by many to derive from the red shading that demarcated the lowest-ranked D neighborhoods. There is clear evidence that the racial makeup of neighborhoods were explicit factors that were often pivotal in assigning grades to neighborhoods. [10] A statistical analysis in the online Appendix confirms the importance of race and other economic and housing characteristics in determining HOLC grades.

How Were the HOLC Maps Used?—There is an active debate among historians about the degree to which lending was influenced by the HOLC maps. Hillier (2003) stresses that access to the maps was not widespread despite high demand among private lenders.[11] Woods (2012), in contrast, argues that the FHLBB widely distributed

[8] From Hillier (210), citing an FHLBB document: "[HOLC] experts believe that since its interest is duplicated by that of all home-financing and mortgage institutions, a program can be evolved which will reclaim large residential areas which are doomed unless some concerted action is taken. Those experts believe that a joint program of Government agencies and private capital can save millions of dollars in property values now being wasted each year. If such efforts are undertaken in the future, the HOLC will be able to contribute surveys made of more than 300 cities throughout the United States—an accumulation of real estate and mortgage data never before available."

[9] A complete description of the security grades from an HOLC document for Baltimore made available to us by Jonathan Rose is shown in the online Appendix.

[10] This is apparent in the area description files (ADFs) that accompanied the HOLC maps. Online Appendix Figure A1 shows an example of an ADF for a D-graded area in Tacoma, Washington, where it is stated that "This might be classed as a 'low yellow' area if not for the presence of the number of Negroes and low class Foreign families who reside in the area." In numerous other examples highlighted in the online Appendix, race appears to be pivotal.

[11] Hillier (2003) argues that the FHLBB preserved their confidentiality as a matter of policy and allowed only a limited number of copies (50 to 60) of each map to be made and that there is little historical record of the use of the maps prior to researchers discovering them in the National Archives.

HOLC appraisal practices and fostered close communication between the private sector and government institutions and that these interactions had a profound and lasting influence on creating a uniform appraisal process.[12] Woods (2012) further claims that as a matter of regulatory policy, banks were required to construct their own maps describing their geographic lending patterns that essentially replicated the HOLC maps.[13]

Anecdotes suggest that some lenders accessed the maps. Jackson (1980), citing evidence from an FHLBB survey of New Jersey bankers and the participation of local realtors as consultants in constructing the St. Louis maps, argues that "private banking institutions were privy to and influenced by the government security maps" (430). Hillier (2003) cites an example of a Chicago real estate official who wrote the following to the City Survey program director: "I hope to be able to 'borrow' a map from your portfolio when you are not looking during your journey in Chicago" (400). More broadly, Greer (2012) claims that thousands of real estate professionals played a role in the creation of the maps and many remained involved in the industry through the post-War era.[14]

We will likely never know the degree to which lenders used the actual maps or received the underlying information in the maps.[15] It is clear, however, that the FHLBB fostered the practice of using maps to classify the credit worthiness of neighborhoods. If, in fact, the maps developed by lenders differed from the original HOLC maps such that boundaries were drawn along slightly different streets, it suggests that our estimates are, if anything, likely to *understate the overall effects of the general practice of redlining* even if they capture the effects of the HOLC maps.

[12] Examples included (i) the creation of a Joint Committee on Appraisal and Mortgage Analysis in 1937 that included three private agencies whose purpose was "to share appraisal data throughout all segments of the national lending industry" (1038) and (ii) the dissemination of a monthly FHLBB journal entitled the *Federal Home Loan Bank Review* (circulation of 6,000) with articles "that provided painstaking detail regarding the influence of neighborhood demographics on mortgage finance" (1038). The list of subscribers "was so extensive that it reached a representative cross section of the national urban housing industry" (1040). Woods (2012) also cites a 1935 *Federal Home Loan Bank Review* article: "It is inevitable, therefore, that the HOLC's appraisals should exert a major influence in setting values on urban-home properties throughout the country. The magnitude of the operation insures that this influence shall be more than temporary, and that the Corporation's appraisals will affect all property values for many years" (1041).

[13] The FHLBB required that lending practices take into account neighborhood demographics. Woods (2012) specifically argues that "there existed a relationship between the HOLC security maps and FHLBB lending policies" (1043). In particular, as a matter of policy, the balance sheets of lending institutions had to include a "security map of the institution's lending area" (1043) and that institutions were instructed that "the best method of grading residential neighborhoods as lending areas is to make a scientific analysis of the entire community and of each neighborhood within it" (1043). Woods further notes that "The FHLBB widely distributed the instructions necessary for creating this critical appraisal material throughout the national lending industry. The Mortgage Rehabilitation Division of the FHLBB 'has prepared simple instructions for making the security maps of residential neighborhoods' available 'to any experienced mortgage lender.' The Rehabilitation Division of the FHLBB 'recognize[d] four broad categories of lending areas, ranging from most desirable to least desirable. Each category was represented by a different color, so that the map could be read at a glance.' These four categories were identical to those created by the HOLC" (1043).

[14] To take one publicly available example, 8 of the 14 reviewers of the Cuyahoga County (Cleveland) HOLC map were from local lending institutions or appraisers. See https://library.osu.edu/documents/redlining-maps-ohio/area-descriptions/CuyahogaCounty_Explanation_and_A1-A31_Area_Description.pdf.

[15] In contrast to some researchers who take the stated policies of federal agencies such as the FHLBB and FHA at face value based on documents available at the National Archives, we think researchers should be open to the possibility that these policies may not have reflected actual practices of federal employees given the conflicting evidence and the prevailing racial attitudes at the time.

FHA Manuals and Maps.—The FHA created a parallel set of maps that likewise rated neighborhoods on a color-coded A to D scale and were based on a systematic appraisal process that took demographic characteristics of neighborhoods into account. Indeed, the 1930s and 1940s FHA manuals explicitly emphasize "undesirable racial or nationality groups" as one of the underwriting standards; their use was ultimately outlawed by the 1968 Federal Housing Act and the 1977 CRA.[16] The enormous influence of the FHA is highlighted by the fact that by 1949, the mortgages on one-third of newly constructed homes were insured by them (Woods 2012). Therefore, perhaps at least as important as whether lenders had direct access to the HOLC maps is whether the HOLC maps were shared with the FHA and thus influenced the provision of housing credit through the FHA's decisions regarding whether to insure loans in low-graded neighborhoods. On this issue, there is agreement among historians that the HOLC influenced the FHA maps.[17] Unfortunately, the limited availability of FHA maps today makes a broad comparison with the HOLC maps infeasible (Light 2010).[18]

Related Literature.—Several recent papers examine the long-run effects of redlining. A contemporaneously written paper by Appel and Nickerson (2016) finds that the HOLC maps affected home prices in 1990. They use a border-based regression discontinuity approach on a smaller set of cities, making stronger assumptions than we do on the exogeneity of the borders.[19] Similarly, Krimmel (2017) uses a difference-in-difference (DD) approach along borders and finds an effect of redlining on housing supply and population density between 1940 and 1970. Anders (2019) uses the same 40,000-population cutoff design as ours and finds that cities with redlining maps experienced higher rates of crime. Faber (2020) compares cities that were appraised by HOLC to cities that were not and finds that the former set of cities experienced higher levels of segregation than the latter group. Earlier studies, such

[16] See Jackson (1980) and Light (2010) for discussions of how FHA risk maps and underwriter instructions were created. The 1934 FHA manual includes race as one of the underwriting standards to be applied to new loans: "The more important among the adverse influential factors (of a neighborhood's character) are the ingress of undesirable racial or nationality groups. … All mortgages on properties in neighborhoods definitely protected in any way against the occurrence of unfavorable influences obtain a higher rating. The possibility of occurrence of such influences within the life of the mortgage would cause a lower rating or disqualification" (FHA 1934).

[17] Light (2010, 671) highlights "ample evidence" to support the influence of the HOLC appraisal methods and maps on the FHA's practices. Footnote 85: "FHA records indicate the agency kept the HOLC security maps on file in connection with the construction of its Economic Data System … and comments from Federal Home Loan Bank Board general counsel Horace Russell on how the FHA 'was fortunate in being able to avail itself of much of the [t]raining and experience in appraisal and the development of appraisal data by Home Owners Loan Corporation' underscores the two agencies' close ties" (671). Woods (2012) cites a 1938 FHA underwriting manual that provided examples taken directly from HOLC appraisals. Hillier (2003) also states that the HOLC maps were shared with the FHA as well as other government agencies but minimizes the link between the FHA and HOLC by noting that the FHA had their own independent sources of information for developing maps.

[18] The FHA map for Chicago is available and shows a strong resemblance to the HOLC map. We find that roughly 82 percent of population-weighted Chicago has the same grade on both maps, including 86 percent of D-graded areas. However, we cannot speak to the similarity of other cities. Therefore, we argue that our estimates capture the sum of any HOLC and FHA effects where the boundaries align and only the HOLC effect where the boundaries differ.

[19] Appel and Nickerson (2016) assume that there were no preexisting discontinuities along HOLC borders, which we show does not hold in our data. They also combine all border types, only examine outcomes in 1990, and use 1940 as the pretreatment period.

as Hillier's (2005) seminal study on Philadelphia and Fishback's (2014) on New York, focus on individual cities.

II. Data and Descriptive Facts

A. *HOLC Maps and Area Description Files*

We obtained geocoded renderings of the original HOLC maps for 149 cities from the Digital Scholarship Lab at the University of Richmond.[20] Figure 1 shows that the geographic coverage is extensive. The 149 cities comprise 89 percent of residents of the 100 largest cities in 1930 and 1940, including 9 of the largest 10 and 17 of the largest 20 cities as well as 30 of the 42 cities with a population above 200,000.[21] The maps for three prominent cities—Chicago, New York, and San Francisco—are displayed in Figure 2. The large set of boundaries separating neighborhood types, especially evident in New York and San Francisco, illustrate our main identification strategy that takes advantage of households living in a narrow band on each side of an HOLC border.

To identify HOLC boundaries, we begin with outlines of cities from the census 2000 place boundary shapefile. An ID is assigned to each line segment of an HOLC boundary that is at least a quarter-mile long.[22] We then draw rectangles that extend a quarter-mile on each side of a boundary. These areas are referred to interchangeably as boundary buffer zones, buffer zones, or buffers. Each boundary has two buffers: the lower-graded side and higher-graded side. We also refer to boundaries between C and D neighborhoods as "D–C" and those separating B and C areas as "C–B."[23]

B. *1910 to 2010 Censuses*

We match the geocoded maps to the 1910 to 2010 censuses. For 1910 to 1940, we use the 100 percent count files and match between 60 and 80 percent of household heads with nonmissing street addresses to modern street locations. Ultimately, roughly 50 to 80 percent of respondents are assigned HOLC neighborhoods.[24] We aggregate our measures to the boundary buffer level by taking means of all observations that fall

[20] See online Appendix Table A1 for the list of cities. Accompanying the maps are a set of ADFs that provide additional quantitative and qualitative detail on the neighborhoods. An example of an ADF for a Tacoma, Washington, neighborhood is provided in online Appendix Figure A1.

[21] Of the 20 most populous cities, we are missing Los Angeles (#5), Washington DC (#11), and Cincinnati (#17). Our 149 cities contain over a quarter of the total US population.

[22] We experimented with different segment lengths, and it made little difference to our results. For both D–C and C–B borders, the median boundary length is 0.37 miles, with about 75 percent less than 0.5 miles.

[23] There are too few "A" areas to study B–A boundaries. In the spirit of analyzing similar neighbors, we exclude boundaries separated by more than one grade (e.g., D–B). See online Appendix Figure A2 for a depiction of New York City boundary buffers as an example.

[24] See online Appendix Table A2. As might be expected, characteristics such as race and home ownership predict the probability of being geocoded. However, our empirical strategy (described below) of comparing changes over time in boundary differences to changes over time in comparison boundary differences (a triple difference) should be robust to any sample selection concerns around geocoding. Regardless, our results are robust to focusing on cities with high geocoding rates (see Section IV). Some additional detail about data consistency with regard to housing measures is provided in the online Appendix.

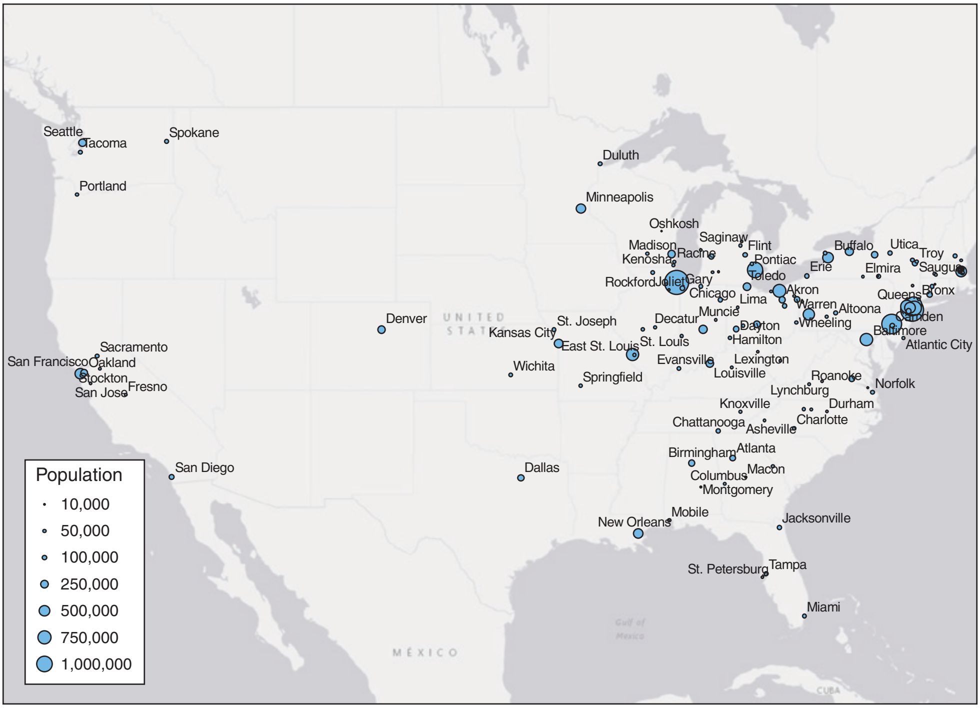

Figure 1. Geographic Coverage of Digitized HOLC Maps

inside of a buffer zone. We restrict our sample to boundary buffers that contain at least three households in the buffer zone of each side of the boundary in 1930.

After 1940, we must use publicly available aggregate data. The smallest geography currently available for 1950 to 1980 is the census tract (IPUMS, US NHGIS-B 2020). Since tracts change over time, we overlay tract boundaries from each census with our boundary buffer shapes and calculate weighted means of any tract for which at least 15 percent of the area of the tract lies within the boundary buffer.[25] Starting in 1990, the census provides smaller geographic tabulations called blocks, which contain on average roughly 100 people.[26] Since blocks are much smaller than tracts, we use weighted means of blocks for which their area is more than 50 percent within the boundary buffer. Combined, this procedure produces a balanced panel of boundary buffer means from 1910 to 2010. In Section IV, we show that our results are not driven by changing the underlying geography from addresses to tracts to blocks over time or from selection into the sample.

[25] The choice of the 15 percent threshold balances a tradeoff between sample size and measurement precision. Our results are robust to alternative census tract inclusion thresholds such as 10, 20, or 25 percent.

[26] Some variables, notably house value, rent, house age, and foreign-born population, are only reported at the block group level, which are aggregates of blocks and typically contain between 600 and 3,000 people. For these variables, we assign the block the values of the block group it is in. In 2000 (2010), there were over 8 (11) million blocks, 208,790 (217,740) block groups, and 65,443 (73,057) census tracts.

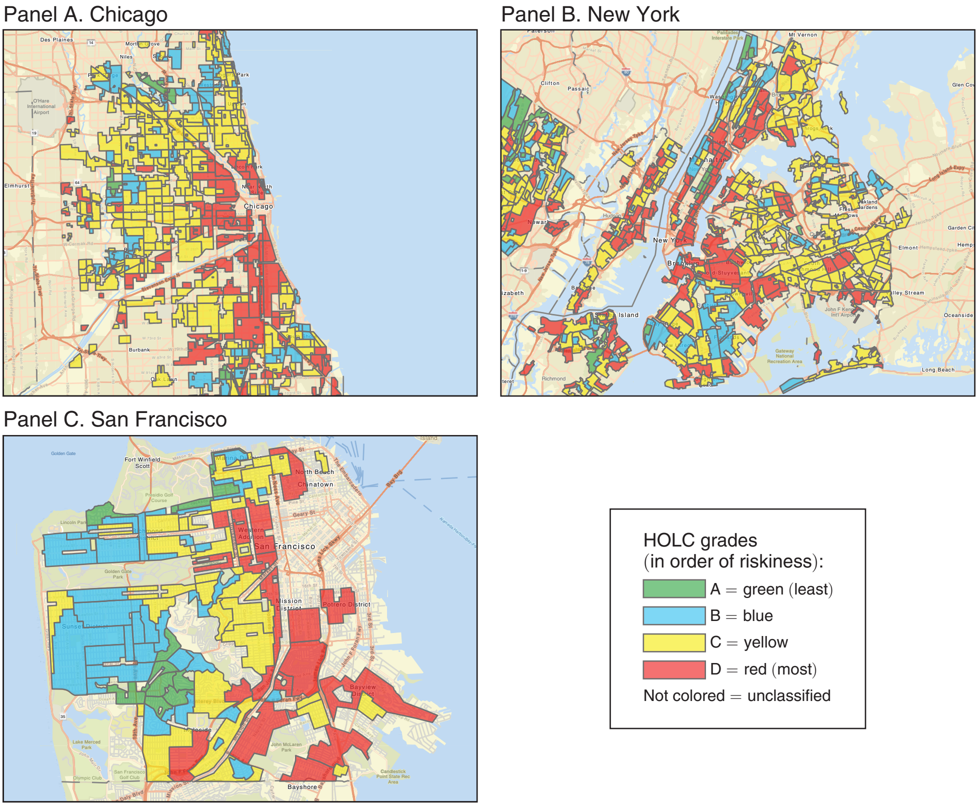

Figure 2. HOLC Maps for Chicago, New York, and San Francisco

Note: Maps digitized by the Digital Scholarship Lab at the University of Richmond.

C. *Summary Statistics*

In 1930, before the maps were drawn, African Americans comprised 14.6 percent of residents living in D neighborhoods but only 1.5 percent of those living in C neighborhoods. By 1980, the share of African American residents grew to 46.2 and 30.7 percent in D and C neighborhoods. These rates converged to 35.7 and 29.0 percent, respectively, by 2010. The complete time patterns by neighborhood grade are shown graphically in panel A of Figure 3. Table 1 shows summary statistics by neighborhood grade along the boundaries. Panel A reports the share of African Americans over time. Columns 1 to 4 show statistics for those living in a buffer zone on each side of the boundaries that meet our criteria. The C–B and D–C boundary differences or gaps are reported in columns 5 and 6.

The gap in the share of African Americans is always smaller along the D–C boundary buffer zones than between the full D and C neighborhoods. For example, in 1930, the gap along the D–C boundary buffers was 6.2 percentage points (column 6), compared to 13.1 percentage points across all D–C residents (not shown). The racial gaps within the D–C boundary peaked at between 14 and 15 percentage points between 1950 and 1970 before declining sharply thereafter. By 2010, the

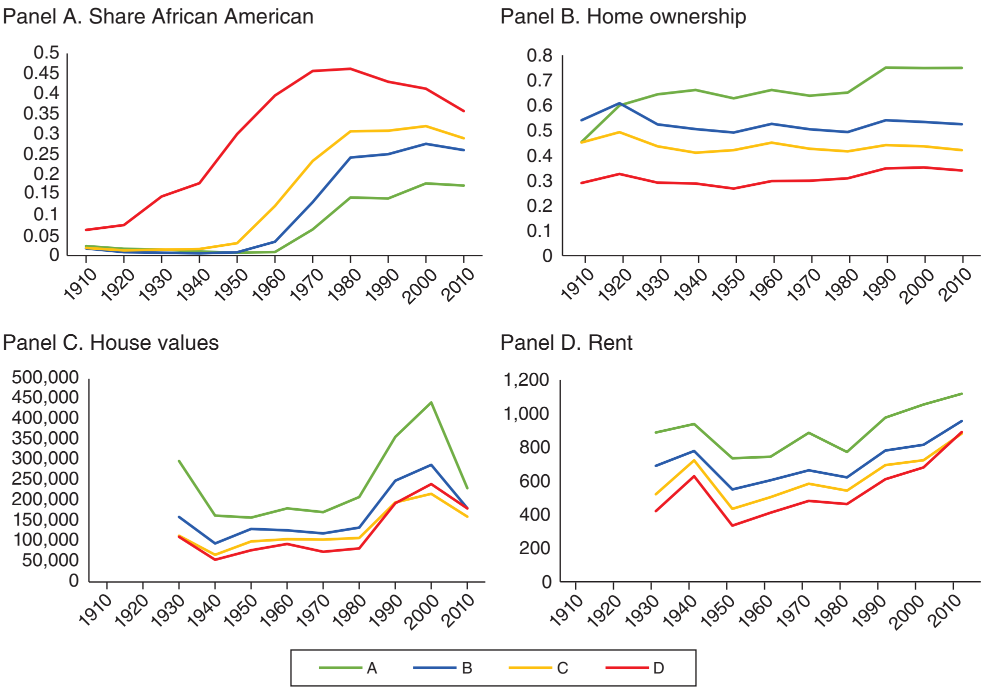

FIGURE 3. MEAN OUTCOMES, BY HOLC NEIGHBORHOOD GRADE AND TIME

Notes: Panels A to D plot summary statistics by HOLC grade over the period 1910–2010. Colors represent those used in the HOLC maps. Summary statistics are weighted by neighborhood population. Data are drawn from the full-count US census (1910–1940), census tracts (1950–1980), and census blocks and block groups (1990–2010). House values and rents are in 2010 US dollars. See text for more detail.

gap stood at just 3.0 percentage points. This secular, hump-shaped pattern in the racial gap also characterizes the C–B boundaries. There was a relatively meager 0.4 percentage point gap in 1930 that grows to 5.2 percentage points by 1970 and then subsequently declines.

Panel B of Table 1 and Figure 3 show corresponding patterns for home ownership. In 1930, the D–C and C-B home ownership boundary gaps were 4.4 and 4.7 percentage points, respectively. These gaps increased to 5.7 and 6.1 percentage points by 1960. Thereafter, the patterns diverge by border type. As of 2010, the home ownership gap declined to just 2.3 percentage points along the D–C boundaries but remained elevated at 6.3 percentage points for C–B borders. Panels C and D of Figure 3 plot the patterns for house values and rents.[27]

[27] House values and rents become available in 1930 and are expressed in 2010 dollars. House values typically refer to the owner-occupied portion of single-family nonfarm houses and land. Table 1 also shows secular patterns in share immigrant and credit scores. The online Appendix reports statistical models of the determinants of HOLC grades. Like Hillier (2005) and Fishback (2014), who examine single cities, we find a clear monotonic relationship between grades and nearly all the key census economic and housing measures, including those not reported in Table 1, whether considered individually or simultaneously.

Table 1—Summary Statistics

Sample type		Boundary buffer zones					
		C–B borders		D–C borders		Buffer gaps	
Grade		B	C	C	D	C–B	D–C
Observations		1,965	1,965	2,111	2,111	1,965	2,111
	Year	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Share African American							
	1910	0.018	0.014	0.025	0.050	−0.004	0.026
	1920	0.008	0.009	0.019	0.051	0.001	0.033
	1930	0.008	0.012	0.024	0.086	0.004	0.062
	1940	0.006	0.010	0.029	0.108	0.003	0.079
	1950	0.006	0.038	0.052	0.206	0.032	0.153
	1960	0.080	0.111	0.213	0.352	0.031	0.139
	1970	0.169	0.221	0.311	0.457	0.052	0.146
	1980	0.315	0.305	0.370	0.482	−0.010	0.112
	1990	0.341	0.348	0.395	0.438	0.007	0.043
	2000	0.364	0.363	0.400	0.428	−0.001	0.028
	2010	0.323	0.330	0.354	0.385	0.007	0.030
Panel B. Home ownership rate							
	1910	0.546	0.472	0.422	0.361	−0.073	−0.061
	1920	0.582	0.521	0.449	0.385	−0.062	−0.064
	1930	0.462	0.415	0.379	0.334	−0.047	−0.044
	1940	0.435	0.391	0.355	0.309	−0.044	−0.046
	1950	0.361	0.306	0.358	0.296	−0.055	−0.062
	1960	0.395	0.334	0.361	0.304	−0.061	−0.057
	1970	0.344	0.295	0.337	0.287	−0.049	−0.050
	1980	0.340	0.294	0.348	0.293	−0.046	−0.056
	1990	0.430	0.361	0.365	0.337	−0.068	−0.028
	2000	0.425	0.360	0.359	0.337	−0.065	−0.022
	2010	0.411	0.348	0.339	0.316	−0.063	−0.023
Panel C. House value (1,000s)							
	1930	166	133	115	105	−33	−10
	1940	91	76	62	55	−15	−7
	1950	128	116	92	76	−12	−16
	1960	127	121	102	88	−7	−14
	1970	125	115	95	78	−10	−18
	1980	133	115	104	77	−18	−27
	1990	241	231	188	179	−11	−9
	2000	263	245	212	192	−19	−19
	2010	182	174	164	173	−8	8

(*continued*)

III. Identification and Methodology

Our strategy is guided by the historical narrative that the creators of the HOLC maps explicitly considered neighborhood characteristics and their trends when drawing borders. This narrative is confirmed by the HOLC's ADFs that accompanied the maps and provided explanations for the grades.[28] Therefore, we use multiple approaches to try to overcome this obstacle to identification.

[28] See the online appendix for examples in the ADFs of how race was sometimes critical in determining grades.

Table 1—Summary Statistics (*continued*)

Sample type		Boundary buffer zones					
		C–B borders		D–C borders		Buffer gaps	
Grade		B	C	C	D	C–B	D–C
Observations		1,965	1,965	2,111	2,111	1,965	2,111
	Year	(1)	(2)	(3)	(4)	(5)	(6)
Panel D. Rent							
	1930	680	586	488	449	−95	−39
	1940	745	861	661	674	116	13
	1950	507	463	406	342	−44	−63
	1960	571	536	493	438	−35	−55
	1970	618	591	534	483	−28	−51
	1980	583	558	506	456	−25	−50
	1990	707	691	631	600	−16	−31
	2000	742	724	678	646	−18	−32
	2010	894	883	862	844	−12	−18
Panel E. Share foreign born							
	1910	0.176	0.183	0.198	0.214	0.008	0.016
	1920	0.143	0.158	0.180	0.200	0.015	0.019
	1930	0.181	0.197	0.209	0.224	0.017	0.015
	1940	0.147	0.160	0.164	0.167	0.013	0.003
	1950	0.171	0.177	0.155	0.146	0.006	−0.009
	1960	0.155	0.151	0.127	0.107	−0.004	−0.021
	1970	0.145	0.143	0.110	0.088	−0.002	−0.022
	1980	0.175	0.184	0.122	0.116	0.009	−0.006
	1990	0.283	0.304	0.200	0.197	0.021	−0.003
	2000	0.212	0.237	0.191	0.203	0.026	0.011
	2010	0.221	0.244	0.205	0.218	0.023	0.014

Notes: Data for panels A to E are drawn from the full counts of the 1910 to 1940 censuses, the census tract aggregation of the 1950 to 1980 censuses, and the census block and block group aggregations of the 1990 to 2010 censuses. Columns 1 to 4 report averages for each side of the C–B and D–C buffer zones. Columns 5 and 6 report the simple difference or gap between each side of a border type (e.g., column 5 = column 2 − column 1).

A. *Differencing*

We begin by considering a naive DD strategy. DD compares changes over time in neighborhood-level outcomes, pre– and post–construction of the HOLC maps in places that are spatially proximate but on different sides of an HOLC boundary, similar in spirit to a border regression discontinuity design (RD) used extensively elsewhere (e.g., Holmes 1998; Black 1999; Bayer, Ferreira, and McMillan 2007; Dube, Lester, and Reich 2010; Dhar and Ross 2012; and many others). Along the line segments that make up these boundaries, we compare nearby neighbors that live within buffer zones a quarter-mile (1,320 feet) from the boundary. This allows us to remove potentially important, but typically hard to measure, confounding factors that influence residents on both sides of a border, such as access to labor markets, public transportation, retail stores, and the like.

The statistical model underlying the DD estimator is

$$y_{gbt} = \sum_{t=1910}^{2010} \beta_t \mathbf{1}[lgs]\,\gamma_t + \beta_{lgs}\mathbf{1}[lgs] + \gamma_t + \alpha_b + \epsilon_{gbt}, \tag{1}$$

where y_{gbt} is an outcome in geographic unit g (e.g., quarter-mile boundary buffer) on boundary b in census year t, $\mathbf{1}[lgs]$ is an indicator that the geographic unit is on the lower-graded side of the HOLC boundary, γ_t are year dummies, and α_b are boundary fixed effects. Differencing across the boundary is captured by the α_b terms. Our coefficients of interest, the β_t terms, capture the change in the mean outcome in year t relative to 1930 (the census year before the maps were drawn, which we omit). The gap in the mean outcome in year t is therefore $\beta_t + \beta_{lgs}$ for years other than 1930 and β_{lgs} for 1930.

B. *Parallel Trends Assumption Likely Violated*

The DD strategy relies on the strong assumption that in the absence of the policy change, trends in characteristics would be parallel for both the treatment and comparison groups. The plausibility of this assumption is typically gauged by examining trends in the pretreatment period. In our case, we have good reason to expect that pretrends are not parallel. We know from the ADFs that the choice of the placement of borders was based in part on demographic and housing characteristics, which were already diverging along these boundaries. Indeed, this divergence can be seen clearly in columns 9 and 10 of Table 1. Ideally, the 1910 to 1930 D–C and C–B outcome gaps are negligible and constant. However, as early as 1910, there was a 3 percentage point African American gap between the D and C sides, which grew to 7 percentage points by 1930. Similarly, there is no evidence that gaps were stable prior to 1940 in home ownership (panel B), house values (panel C), or rents (panel D). While the racial gap along C–B boundaries is virtually nonexistent before the maps were drawn, that is not the case for home ownership, house values, and rents.

Moreover, an RD design will likely not satisfy the assumption of continuity along the borders. We show examples of several distance plots in online Appendix Figure A3 where each dot represents the mean characteristic (regression adjusted for border fixed effects) in bins of 0.01 miles (roughly 50 feet) of distance in each direction from the D–C or C–B border. For several of our outcomes, even limiting our sample to observations that are just a city block away from the border would lead to meaningful discontinuities and render an RD design invalid. Bayer, Ferreira, and McMillen (2007) demonstrate that similar issues arise across modern school attendance zone boundaries, and Dhar and Ross (2012) show even larger differences when examining school district boundaries.

We propose two strategies to address the failure of parallel trends to be a plausible assumption along the HOLC borders.

C. *Comparison Boundaries*

The first strategy creates a set of comparison boundaries with similar characteristics and trends to the HOLC treatment boundaries before the maps were drawn. We motivate this approach by what we refer to as "missing" HOLC borders. It may have been difficult to construct polygons that reflected completely homogeneous neighborhoods if there were small areas within neighborhoods that were fundamentally different. A stylized illustration is depicted in the top panel of online Appendix

Figure A4, where there is a small island of C type streets within a larger ocean of D. The Chicago HOLC map (Figure 2) also shows that this is plausible. Among the large swath of D (red) in the heart of Chicago, there are surely pockets that might be appropriately labeled C or higher.

We identify these potential comparison boundaries by first drawing half-mile by half-mile grids over each city. We then create quarter-mile boundary buffers around any grid line segment that does not overlap with HOLC treatment boundaries. This set of boundaries is referred to as our "grid" comparison group.[29] Propensity scoring methods are applied to construct weights for the grid comparison group. A byproduct of weighting the comparison group is that the pretreatment differences in outcomes and covariates become very small.[30] We use the logic that if pretreatment differences are eliminated using these weights, then it may be valid to interpret any posttreatment difference between treatment and comparison boundaries as an estimate of the causal effect of the HOLC grade.

Since each set of treated boundaries has a side deemed riskier by the HOLC (such as the D side of a D–C boundary), an analogous construct is needed for the comparison boundaries. In most cases, the characteristics of the "more risky" side of these comparison boundaries will be very similar to the "less risky" side. However, in some cases, the comparison boundaries will turn out to separate areas with quite different characteristics. We would anticipate that the "risky" side of the pseudo-border will often have less favorable characteristics than the safe side, like actual HOLC borders. The propensity score method essentially uses this special set of comparison boundaries as a control group for the HOLC boundaries. Both types of boundaries separate areas with different characteristics, but only the HOLC borders formally limited credit to the risky side. To better preserve the statistical properties of the comparison boundaries, we randomly assign one side of each comparison boundary to be the riskier or lower-graded side.[31] In parallel to the treatment boundaries, we then construct the difference or gap between the mean of our outcome on the "higher-graded" and "lower-graded" side.

To construct the propensity score, we pool the treatment and grid comparison boundaries, where each boundary is an observation. For each grade-type difference (D–C or C–B), only comparison boundaries from the same HOLC graded areas are used. That is, when we estimate the effects of the D–C borders, we only include C–C

[29] See online Appendix Figure A5 for an example of a grid placed over New York City. We also considered an alternative comparison border group that uses "same-grade" (e.g., B–B, C–C, or D–D) borders. The HOLC often drew borders separating two "unique" neighborhoods with the same grade. We are not sure why this was done, but we speculate that cities were first broken into neighborhoods, and then each neighborhood was evaluated. Using same-grade borders as a comparison group yields similar, albeit less precise, estimates (online Appendix Figures A7 and A8) than what we find using our grid-based comparison group where we can create much larger samples. We prefer using grids because the same-grade borders may induce some treatment effect due to having a border associated with it.

[30] We also tried the synthetic control method (SCM) of Abadie, Diamond, and Hainmueller (2010) and found similar results. We prefer the propensity score method for our application, as SCM is more difficult to implement without a balanced panel of geographic units (in our case, address-tract-block).

[31] Random assignment ensures that the distribution of the within-boundary differences in our comparison group is representative of all comparison boundaries and is not skewed toward either tail of the distribution. Note that reweighting of comparison boundaries occurs after the randomization.

or D–D boundaries and not A–A or B–B boundaries. We then estimate the following probit separately for D–C and C–B boundaries:

$$(2)\quad \mathbf{1}\{Treated\}_{b,c} = \alpha_c + \sum_{k=1}^{K} \beta_{1910}^{k} z_{b,c}^{k,1910} + \beta_{1920}^{k} z_{b,c}^{k,1920} + \beta_{1930}^{k} z_{b,c}^{k,1930} + \epsilon_{b,c},$$

where $\mathbf{1}\{Treated\}_{b,c}$ is an indicator variable for whether boundary b in city c is a "treated" boundary, α_c is a city fixed effect, and $z_{b,c}^{k,t} = x_{lgs,b,c}^{k,t} - x_{hgs,b,c}^{k,t}$ is the gap between an explanatory variable k on the lower-graded side (lgs) and the higher-graded side (hgs) at time $t =$ 1910, 1920, and 1930. The variables indexed by k include the gaps in the share African American, African American population density, White population density, the share foreign born, the home ownership rate, the share of homeowner households that have a mortgage, log house value, and log rent.[32]

We also experimented with including three measures of housing quality—the average age of houses, the share of houses in a "poor state of repair," and the share of houses that are brick versus wood frame or stucco—transcribed from the ADF report cards that accompanied the maps. These report cards are only available for a small sample of neighborhoods (see online Appendix). Nevertheless, including them did not improve the predictive power of our model and resulted in similar albeit somewhat noisier estimates.

Our estimate of the propensity score ($pscore$) is equal to the predicted probability of treatment from equation (2). Inverse probability weights (IPW) are then formed for the comparison boundaries as $w = pscore/(1 - pscore)$ and for the "treated" boundaries as $w = 1$. This has the effect of "up-weighting" comparison boundaries that are most similar to treated boundaries and "down-weighting" those that are less similar.

This procedure produces considerable overlap in the distributions of the propensity scores for the treated and comparison groups (online Appendix Figure A6, panels A and B). The sample is trimmed to exclude treated borders with a propensity score above that of the maximum comparison border and comparison borders with a propensity score below that of the minimum treated border. Consequently, the reweighted comparison borders look more comparable to the treated borders than the unweighted comparison borders. Panels C and D of online Appendix Figure A6 demonstrates this critical implication for the home ownership rate.

We then estimate triple difference regressions using inverse probability weights for the comparison borders. The specification we use is as follows:

$$(3)\qquad y_{gbt} = \sum_{t=1910}^{2010} \beta_t \mathbf{1}[lgs]\mathbf{1}[treated]\gamma_t + \beta_{lgs\times treated}\mathbf{1}[lgs]\mathbf{1}[treated] + \beta_{lgs}\mathbf{1}[lgs] + \gamma_t + \alpha_b + \epsilon_{gbt},$$

[32] The model is run using a balanced panel in which at least one of the following three variables contains no missing values (on either boundary side) from 1910 through 1930: share African American, the home ownership rate, or share foreign born. House values and rents are only available in 1930. Whether the household has a mortgage is only available in 1910 and 1920. Measures that should be available but are missing are recoded to a constant value, and a missing indicator variable is turned on. The probit models are weighted by the log total population of the buffers on both sides of the boundary. We also experimented with nearest-neighbor matching but found our samples were too thin once we limit neighbors to the same city.

where $\mathbf{1}[treated]$ is a dummy for being in the group of "treated" borders versus the group of weighted comparison borders. As in equation (1), 1930 is omitted from the summation of t from 1910 to 2010.

D. *Exploiting Idiosyncratic Borders*

A second simpler strategy to eliminating confounding factors takes advantage of the possibility that some HOLC boundaries might have been more idiosyncratic in nature and were drawn simply to close a polygon. Consider the hypothetical example of a "misaligned" border where the northern part of the neighborhood contains largely red blocks and the southern area contains largely yellow blocks. It may not have been entirely clear where exactly to draw the southern border, and the HOLC agents may have just chosen a major street several blocks from the actual red-yellow demarcation to define the neighborhood. Such "treated" boundaries may not reflect a discontinuous change in creditworthiness and would be much less likely to exhibit pre-trends in outcomes.[33] Once we focus on these unusual boundaries, it is possible a control group is no longer needed.[34]

Using idiosyncratic borders also addresses a possible concern with using city-grid-derived comparison boundaries that may not have been realistic choices for the HOLC. It may have been the case that the HOLC would have primarily drawn boundaries at major streets or other sharp demarcations of neighborhoods that the grid approach is not well suited to capture.[35]

To operationalize this idea, we identify our idiosyncratic boundaries by selecting borders whose propensity score—or predicted probability of being treated—is below the median. This approach is akin to the method of subclassification, also known as blocking or stratification, discussed in Imbens (2015) and Imbens and Rubin (2015). Their idea is to partition the sample into subclasses based on the value of the estimated propensity score so that within a subclass, differences in the covariate distribution are small. Causal effects can be inferred within a subclass as if assignment was random. Our application focuses on the low-propensity-score subclass, as these boundaries are most likely to be idiosyncratic since they have covariate distributions similar to our comparison group of randomly drawn grid boundaries. As we show later, the low-propensity-score subsample of treated borders exhibits virtually no pretrends. While this strategy is more straightforward and does not rely on the comparison group except when estimating the propensity scores, it reduces power and may not generalize to all borders if there are heterogeneous effects.

[33] A visual example is provided in the bottom panel of online Appendix Figure A4. One common example of such a situation is found in the ADF for Chicago's neighborhood D98, where the notes mention that "[t]he eastern portion of the area is not quite so heavily populated with foreign element." Therefore, the particular street used to demarcate the eastern boundary may have been idiosyncratic. We considered trying to directly capture the phenomenon of "closing the polygon" by looking only at neighborhoods that had multiple different grade treated boundaries and then using only the boundary that had the lowest propensity score. The logic is that the lowest-propensity-score border within a polygon is most likely drawn to close the shape. But, in practice, the sample of such boundaries is too small.

[34] We thank a referee for this insight.

[35] Another approach to addressing this concern is to use the "same-grade" boundaries discussed in footnote 29.

IV. Main Boundary Results

We start by describing our baseline results—a contrast of HOLC boundaries to weighted comparison boundaries—separately for D–C and C–B. We then turn to a second set of results based on low-propensity-score treated boundaries that we argue were most likely idiosyncratically chosen. Finally, we consider an array of robustness checks.

A. *Baseline Results along the D–C Boundary*

To show how we arrive at our baseline specification, Table 2 walks through a detailed accounting for one particular outcome, the share African American, along one type of boundary, D–C.[36] We start with the African American share because it is well suited for describing the potential problems with implementing a naïve DD approach. Column 1 begins by comparing entire neighborhoods (D versus C) rather than just the narrow buffer zones around HOLC borders. Specifically, we estimate a version of equation (1) where the geographic unit g indexes HOLC neighborhoods and boundary fixed effects α_b are excluded. Consistent with Table 1, the D–C gap in the share African American is large in 1930 at 13.5 (1.4) percentage points, rises to 25 (2.1) percentage points in 1960, and then falls to 8.1 (1.6) percentage points by 2010.[37] Adding city fixed effects (column 2) has little impact.

The consequence of using buffer zones becomes apparent when we move to column 3, which limits the analysis to households living within a quarter-mile of a boundary. Now, the D–C gap starts at just 6.4 (1.0) percentage points in 1930, rises to 13.8 (2.7) percentage points by 1970, and thereafter falls to 3.7 (0.8) percentage points by 2010. These estimates are modestly lower when we include boundary fixed effects (column 4). However, although the variation is now restricted to comparing residents living, at most, a quarter-mile from the same boundary, there are still significant pretrends from 1910 to 1930.

To further address the preexisting differences along our boundary buffers, column 5 shows estimates obtained from using our weighted comparison borders based on the propensity score analysis. The comparison borders successfully mimic the pretrends in the treated boundaries. For example, they show a D–C gap in African American share of 2.3 (0.6) percentage points in 1920 rising to 5.3 (1.2) percentage points in 1930. This 3.0 percentage point increase is essentially equivalent to the 3.3 percentage point increase in the treated boundaries.

However, *after the maps were drawn*, the treated and comparison estimates diverge sharply. These patterns are illustrated in panel A of Figure 4, which plots the estimates and standard error bands for both the treated and comparison groups. We find that the gap in the share African American in the treated group continues to rise in subsequent decades and peaks as high as 11.3 (2.3) percentage points by 1970 before declining. In contrast, the analogous gap in the comparison group drops

[36] Analogous tables for the other outcomes along the D–C borders are in online Appendix Tables A3 to A5.
[37] City-clustered standard errors are in parentheses. Bootstrapped standard errors, stratified by city, are similar in magnitude.

Table 2—Effects of D versus C Grade, Share African American

Sample Type	HOLC Neighorhoods		1/4 mile D–C boundaries				
					Grid	Triple	Low PS
Year	D–C	D–C	D–C	D–C	C.F.'s	Diff	D–C
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1910	0.061	0.053	0.026	0.025	0.019	−0.003	0.005
	(0.011)	(0.01)	(0.005)	(0.006)	(0.009)	(0.008)	(0.004)
1920	0.069	0.063	0.031	0.029	0.023	−0.003	0.003
	(0.009)	(0.008)	(0.006)	(0.006)	(0.006)	(0.009)	(0.004)
1930	0.135	0.133	0.064	0.062	0.053	—	0.007
	(0.014)	(0.013)	(0.01)	(0.011)	(0.012)	—	(0.004)
1940	0.150	0.147	0.076	0.074	0.039	0.026	0.020
	(0.015)	(0.013)	(0.012)	(0.012)	(0.009)	(0.006)	(0.007)
1950	0.224	0.214	0.119	0.101	0.009	0.082	0.041
	(0.02)	(0.019)	(0.026)	(0.024)	(0.011)	(0.025)	(0.018)
1960	0.250	0.234	0.121	0.094	0.001	0.085	0.044
	(0.021)	(0.018)	(0.031)	(0.026)	(0.016)	(0.03)	(0.015)
1970	0.216	0.203	0.138	0.113	−0.007	0.111	0.091
	(0.024)	(0.02)	(0.027)	(0.023)	(0.016)	(0.031)	(0.021)
1980	0.172	0.159	0.107	0.087	0.003	0.075	0.062
	(0.028)	(0.023)	(0.028)	(0.021)	(0.02)	(0.025)	(0.022)
1990	0.130	0.126	0.059	0.056	0.016	0.031	0.032
	(0.018)	(0.014)	(0.011)	(0.011)	(0.007)	(0.013)	(0.01)
2000	0.106	0.103	0.042	0.038	0.009	0.019	0.018
	(0.017)	(0.013)	(0.01)	(0.01)	(0.005)	(0.011)	(0.008)
2010	0.081	0.079	0.037	0.034	0.006	0.019	0.016
	(0.016)	(0.012)	(0.008)	(0.007)	(0.005)	(0.012)	(0.008)
Cities	148	148	115	115	115	115	97
Neighborhoods	3,534	3,555	—	—	—	—	—
Boundaries	—	—	—	1,134	4,217	5,351	567
Observations	27,814	27,814	16,690	16,690	61,467	78,157	8,525
R^2	0.215	0.383	0.426	0.644	0.685	0.676	0.646
Fixed effects	None	City	City	Boundary	Boundary	Boundary	Boundary

Notes: Table entries are from regressions that estimate the gaps between D- and C-rated neighborhoods in the share African American. Columns 1 and 2 use entire neighborhoods. Columns 3 to 7 use quarter-mile boundary buffer zones. Columns 3 and 4 use actual HOLC "treated" boundaries. Column 5 shows effects on counterfactual boundaries weighted by propensity scores to be similar to treated boundaries. Column 6 shows the difference in the gap between treated and comparison boundaries relative to 1930. Column 7 uses only those treated boundaries with below-median propensity scores.

slightly to 3.9 (0.9) percentage points in 1940 before reverting to roughly 0 by 1960. By 2010, the estimates are 3.4 (0.7) percentage points in the treated group and 0.6 (0.5) percentage points in the comparison group.

A set of "triple difference" estimates that differences the treatment and comparison group effects relative to 1930 are reported in column 6 of Table 2 (and plotted in online Appendix Figure A7). A racial gap emerges in 1940 and continues to rise, peaking at 11.1 (3.1) percentage points in 1970 before beginning to converge. Nevertheless, there remains an economically relevant 2 to 3 percentage point racial gap during 1990 to 2010, more than a half-century after the maps were drawn.

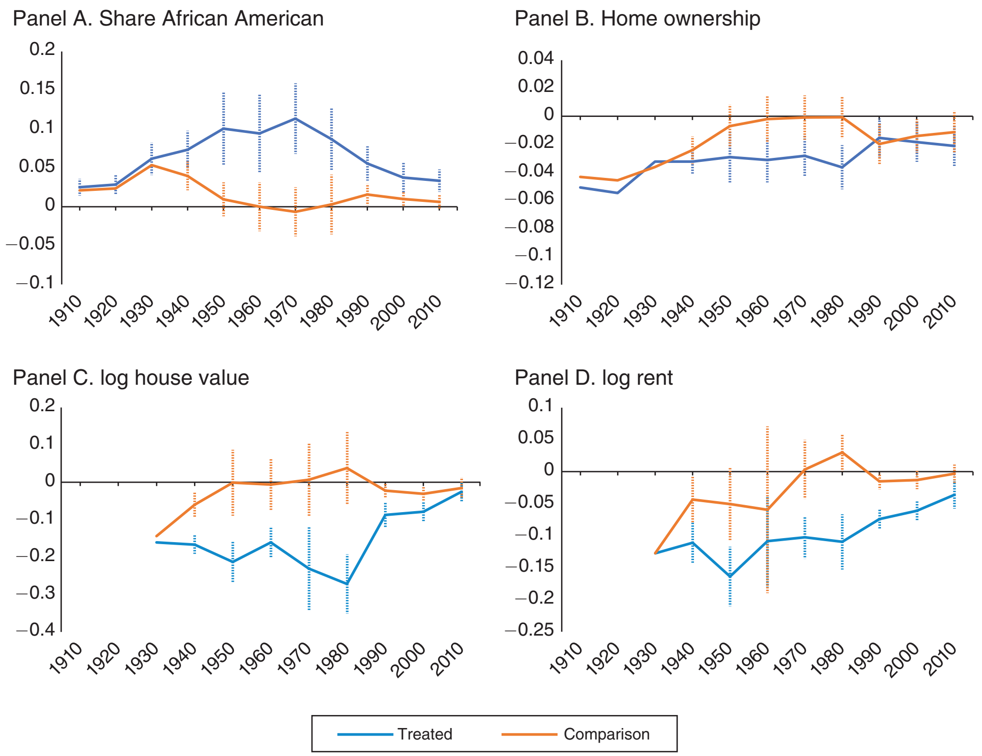

Figure 4. Main Effects along D–C Boundaries, by Outcome

Notes: The treatment estimates (blue lines) are derived from a quarter-mile buffer zone around the D–C boundaries. The comparison boundaries are based on a quarter-mile buffer zone drawn around grids over each city and weighted by propensity scores to mirror pre-map trends (see text for more detail). Vertical bands denote 95 percent confidence intervals.

A parallel analysis for our three housing outcomes—home ownership, house values, and rents—is displayed in the remaining panels of Figure 4 (detailed estimates are shown in online Appendix Tables A3 through A5). In all three cases, we again document preexisting gaps along the HOLC boundaries, which we can successfully reproduce using propensity score weighting of comparison boundaries. We also find that meaningful differences emerge between the treated and comparison boundaries starting in 1940, generally grow larger in subsequent decades, and persist to varying degrees through 2010.

Specifically, after the HOLC maps were drawn, the home ownership gap was relatively constant at around -3 to -4 percentage points through 1980 before falling to -1 to -2 percentage points by the 1990 to 2010 period. By contrast, the home ownership gap in the comparison boundaries closed relatively quickly and remained roughly 0 (with a standard error of around 1 percentage point) through 2010. The relative widening of the home ownership gap between the treatment and comparison boundaries, especially through 1980, was accompanied by parallel gaps in house values and rents. Among treated boundaries, the D–C house value gap starts at around -16 (1.2) percentage points in 1930, gradually climbs to around -27 (4.0) percentage points by 1980, before falling to around −8 (1.3) percentage points by 2000 and −2.6 (1.3) percentage points in 2010. The house value gap in the comparison boundaries quickly reverts from a similar level to the treated boundaries in 1930 to statistically indistinguishable from zero by 1950 through 1980. The peak in the rent gap occurs earlier than the other outcomes and steadily declines after 1950. But as of 2010, it also remains statistically and economically meaningful.

One concern with our approach is multiple hypothesis testing given that we have estimated triple-difference coefficients for seven posttreatment decades (1940–2010) and four different outcomes. To address this in a conservative manner, we ran F-tests on the joint significance of all seven coefficients in each regression and then implemented a Bonferroni correction by multiplying the p-values by four to account for our four different outcomes.[38] For the African American share, the Bonferroni-corrected p-value on the joint significance of the 1940 to 2010 coefficients is 9.1×10^{-5}. The corresponding p-values for home ownership, house values, and rents are 0.077, 9.2×10^{-5}, and 0.0071, respectively.

B. *Baseline Results along the C–B Boundary*

Figure 5 presents results along the C–B boundaries. As noted earlier, the African American population was sparse in B and C neighborhoods in 1930, so, as expected, pretrends for racial gaps are not an issue (panel A). After the maps were drawn, however, a meaningful gap of about 4 percentage points opens up by 1950 and continues to rise to a peak of over 8 percentage points by 1970 before gradually reverting to about 2 percentage points by 2010. In contrast, we estimate a virtually flat line around 0 for the comparison boundaries. These results suggest that restricted access to credit in yellow areas ("yellow-lining") was also a meaningful phenomenon.[39]

We find consistent evidence of C–B gaps opening up in housing market measures as well (Figure 5, panels B to D). The C–B home ownership gap was roughly 5.5 percentage points by 1950 and peaked at around 7 to 8 percentage points from 1990 through 2010. By comparison, the D–C home ownership gap topped out around 4 percentage points and is about 2 percentage points as of 2010. Likewise, as of 2010, the C–B gap in house values stood at 7.5 percentage points, 3 times larger than our estimated D–C house value gap.[40] The p-values conservatively adjusted for multiple hypothesis testing are 0.088 for African American share, 0.0001 for home ownership, 0.0094 for house values, and 0.097 for rent. In Section VI, we consider explanations for a larger impact along the C–B borders.

C. *Estimates from Low-Propensity-Score Borders*

Our second strategy attempts to isolate borders that may have been more idiosyncratic in nature by honing in on the sample of low-propensity-score borders that our

[38] A nice alternative approach to multiple hypothesis testing that is less conservative is provided by Bifulco, Fletcher, and Ross (2011).

[39] Triple difference estimates are reported in online Appendix Figure A8.

[40] Appel and Nickerson (2016) report a 4 percentage point gap in 1990 house values across all boundaries.

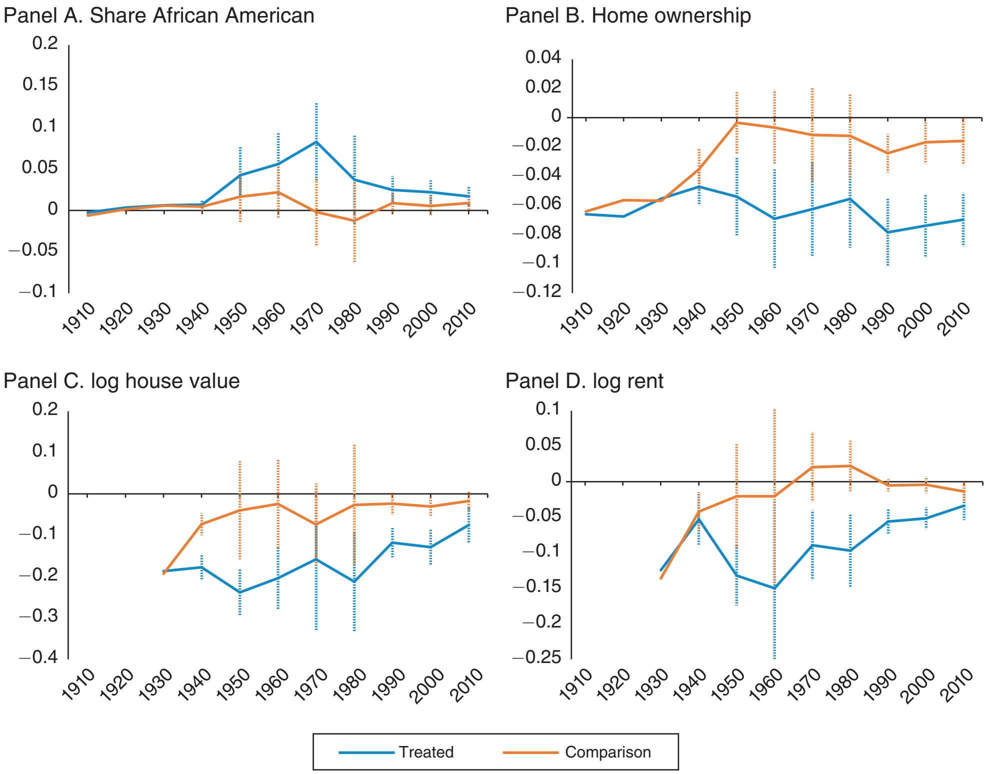

Figure 5. Main Effects along C–B Boundaries, by Outcome

statistical model predicts were least likely to have been drawn.[41] The low-propensity point estimates for the share African American gap are shown in column 7 of Table 2 and plotted in the blue line of Figure 6, panel A. For ease of reference, the gray line reproduces estimates for all treated borders (column 4 of Table 2).

Perhaps what is most compelling about this strategy is that there is no longer a pretrend for the low-propensity D–C borders—the gap in 1910, 1920, and 1930 is essentially zero. If pretrends are broadly nonexistent with this method, the low-propensity method might be particularly useful in cases where we do not have pre-1940 data. Moreover, this research design also produces a smooth continuous function with no abrupt change near the border in 1930 (see panel A of online Appendix Figure A9). On the other hand, there is the possibility that the results from these borders may not generalize to the full population. If there are heterogeneous effects, we suspect that the low-propensity-score method could lead us to understate the average effect. This would occur if low-propensity-score borders had other positive features that protected the neighborhood housing stock and therefore understate the typical effect of receiving a low grade from the HOLC. We acknowledge, however, that this

[41] Low-propensity-score boundaries are somewhat more prevalent in the Northeast.

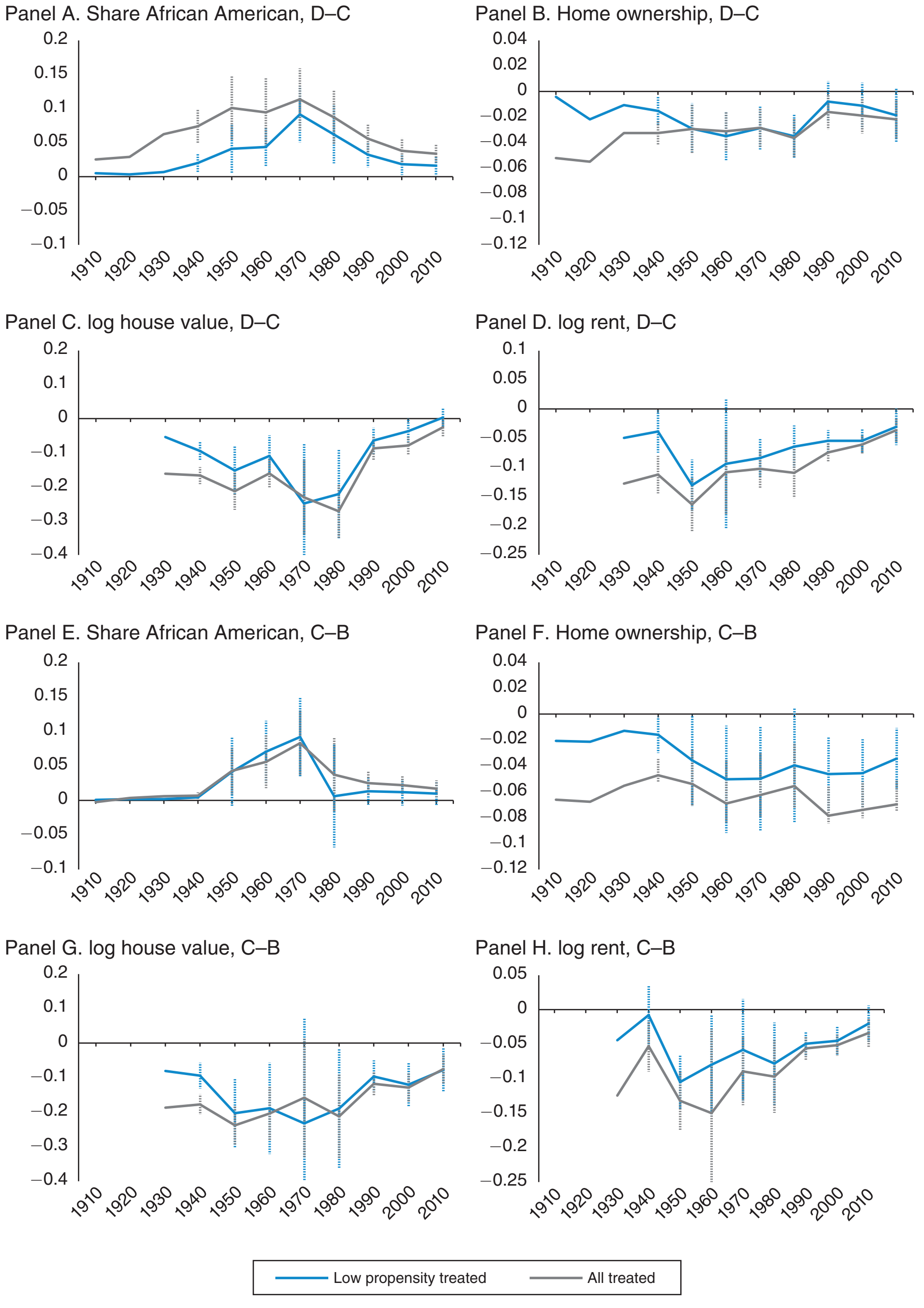

Figure 6. Effects on D–C and C–B Gaps, Using Low Propensity for Treatment Boundaries

Notes: The "low-propensity treated" sample is restricted to boundaries with below-median propensity scores. See notes to Figure 4. Error bars in panels C, G, and H may extend below the lower bound of the *y*-axis.

is just a conjecture and there may be valid reasons why effects could be larger along these borders.

Using the low-propensity-score borders, we find that there is a meaningful rise in the D–C gap in share African American after the maps are drawn that peaks at a 9.1 (2.1) percentage point difference in 1970 before falling to 1.6 (0.8) percentage points by 2010. Notably, the timing and magnitude of these gaps is similar to the triple difference estimates that use the full sample of D–C borders.

The remaining panels of Figure 6 plot the D–C estimates for the three housing outcomes and for the corresponding race and housing outcomes along C–B boundaries. In every case, the use of the low-propensity boundaries largely eliminates gaps and trends in the premap period. Further, for most of the outcomes, the low-propensity estimates after the maps were drawn are nearly identical to the triple difference estimates using all treated boundaries. An exception is home ownership along the C–B borders; here, the gap did not grow nearly as large in the postmap period using the low-propensity borders. As of 2010, the C–B home ownership gap based on low-propensity treated boundaries was roughly half as large (3.5 percentage points) as estimates using all treated boundaries (7.0 percentage points).

Overall, we are agnostic as to method. Regardless, the maps appear to have had economically significant negative effects on the lower-graded side of the HOLC border.

D. *Robustness Checks*

In this section, we briefly discuss the robustness of the baseline results to several alternative methodological choices.

One important concern is the varying use of address, census tract, and block-level data over the 1910 to 2010 time period. This is potentially a critical issue because we often find a sharp increase in outcome gaps between 1950 and 1980, which is precisely when we must use the most highly aggregated census tract data. Therefore, we reran our models but aggregated to a consistent level of geography—census tracts—in all years. These results are shown in Figure 7. We continue to find that the 1950 to 1980 period remains well above the preperiod and convergence begins post-1980.

As an added safeguard, we also constructed a "geography-consistent" time series that adjusts the baseline 1950 to 1980 point estimates by an estimate of the potential bias from using tracts in these years. Specifically, we construct a block-to-tract adjustment ratio $\beta^{block}_{1990-2010,c}/\beta^{tract}_{1990-2010,c}$ based on block-level and census tract-level estimates derived from the 1990 to 2010 censuses. This adjustment, shown in the dashed lines in each panel, typically, but not always, lowers our 1950 to 1980 estimates but has little impact on the general contours of our results.[42]

We considered three other robustness checks. First, we use a narrower one-eighth-mile-wide cutoff on each side of the boundary to construct our buffer zone. There is a tradeoff in using a narrower buffer between having a more comparable

[42] The one exception where this adjustment fails is the share African American along the C–B boundary (panel B). This is because the denominator is close to zero.

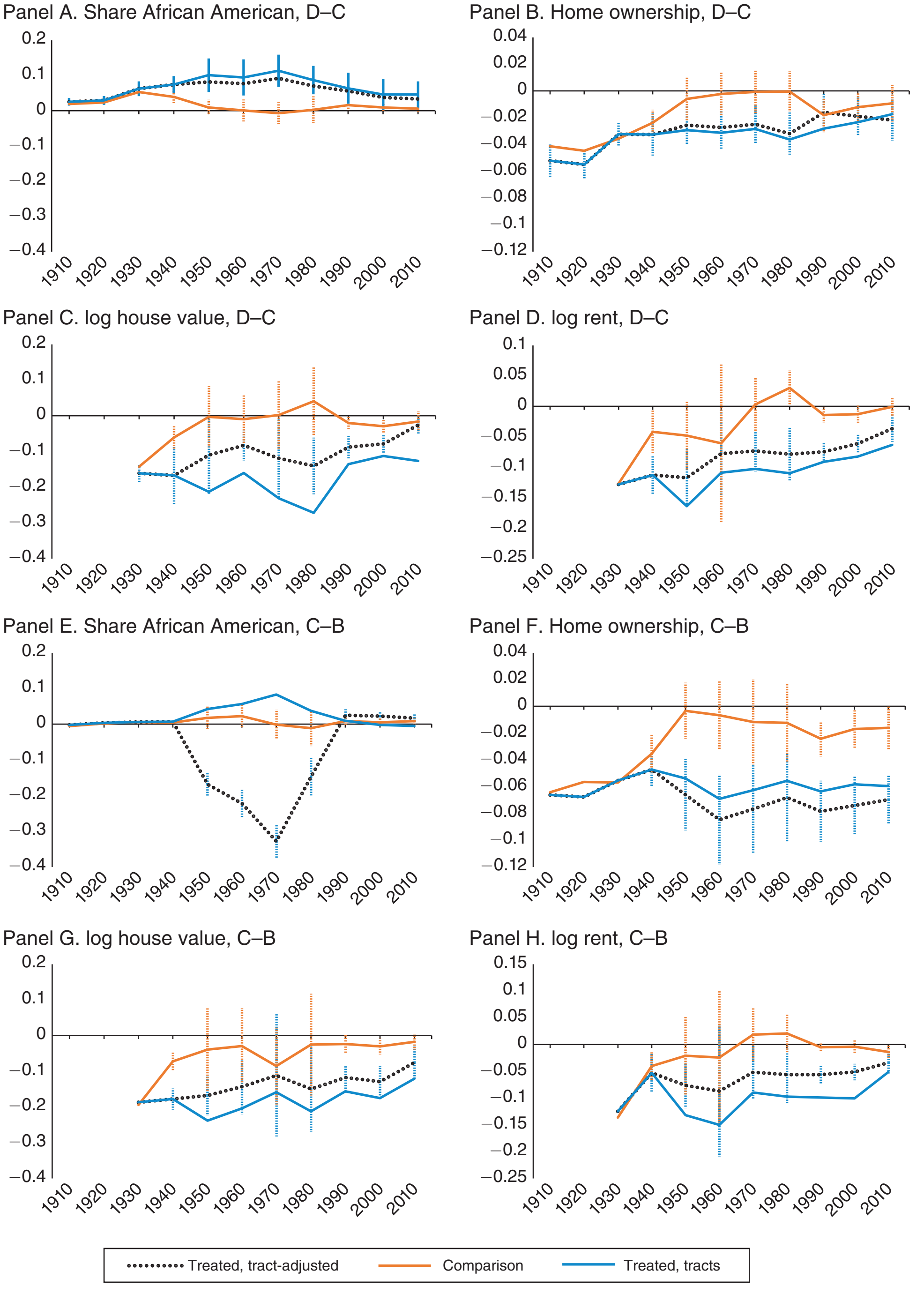

Figure 7. Robustness of Main Outcomes to Using Census Tracts in All Years

Notes: We multiply our main tract results in 1950 to 1980 by the mean block-to-tract ratio in 1990–2010 to generate "tract-adjusted" estimates for these years and combine these with our main block-based estimates from 1990 to 2010 (dotted line). See text for more detail.

across-boundary group on the one hand and having smaller samples and a greater share of the sample that is potentially contaminated by across-border spillovers. Second, we restrict our sample to cities with a high (above-the-median) rate of geocoding in 1920. Third, we exclude all borders with a significant overlap with rivers or railroads under the assumption that these borders may be most prone to preexisting trends. In all of these exercises, which are organized by outcome and border type in online Appendix Figures A11 to A14, we find that the results are broadly comparable to our benchmark estimates in Figures 4 and 5.[43]

V. Identification of Aggregate City-Level Effects: The 40,000-Population Cutoff

One limitation of our analysis thus far is that our estimates ignore possible effects of the HOLC maps in other parts of cities. It is possible that we could be underestimating the overall effects by only focusing on residences within a few blocks of a boundary and that lower-graded areas located farther from the borders were more negatively (or positively) impacted. Or, there could have been negative spillovers on the entire city from having concentrated areas of disinvestment and poverty. On the other hand, the maps could have simply led to a reallocation in the city of where investment would take place and where poor residents would live, and in the aggregate, the maps might have had little overall effect.

To address these citywide issues, we exploit a discontinuity in the appraisal program (see also Anders 2019). The FHLBB chose to draw maps only in places with a population of 40,000 or more. This cutoff enables a comparison of the outcomes of cities with a premap population just above 40,000 to cities just below. Our working assumption is that cities on either side of 40,000 would not be systematically different with respect to the outcomes we consider. Therefore, any relative difference that emerges over subsequent decades could be attributed to the HOLC maps. We have not encountered any evidence that the population cutoff was a strategic choice, but we assess this assumption below by analyzing placebo population cutoffs.

A nice feature of this identification strategy is that it overcomes a possible criticism of the border strategy, that there may have been unobservable factors, unaccounted for in our propensity score models, that allowed the real estate professionals who drew the maps to correctly predict the trajectory of US neighborhoods over subsequent decades. Although this second method is cleaner and one can make a stronger case for exogenous treatment, we have much smaller samples and therefore less statistical power. Moreover, we may not be able to generalize the effects from small to large cities.[44]

Consequently, we consider the city analysis a complement to our boundary estimates. An economically significant impact at the city level could suggest that our localized estimates may not be offset in the aggregate due to other countervailing

[43] We also tried trimming our sample further to exclude boundaries with propensity scores above the ninety-fifth percentile and below the fifth percentile. These results are slightly noisier but otherwise the same.

[44] Estimates derived from only the smallest HOLC cities might not be externally valid if there are heterogeneous effects by city size. In the online Appendix, we show that there is no compelling evidence of differences by city size. However, that analysis excluded cities with very few borders, including virtually all cities with a population under 50,000 in 1930 that we use in this exercise.

forces. On the other hand, the absence of effects at the aggregate level could still obscure important effects at the local level uncovered by the boundary analysis.

We compare 27 redlined treatment cities with a 1930 population between 40,000 and 50,000 to a comparison group of 26 nonredlined cities with a population of 30,000 to 40,000.[45] A list of the cities along with their 1930 population and their mean outcomes in 1930 and 1980 can be found in online Appendix Table A6. Our control sample of cities (panel A) appears to be, if anything, more negatively selected on housing characteristics and has a higher share of African Americans than the treated cities (panel B). However, the differences are relatively small. For example, the 1930 mean home ownership rate is 46 percent in the comparison cities and 48 percent in the treated cities. By 1980, however, the comparison cities have a higher home ownership rate at 58 percent compared to 55 percent in the treated group of redlined cities. This shift in the relative gap in home ownership of 5 percentage points happens to be of a similar magnitude to the 4 and 4.5 percentage points in the 1980 D–C and C–B home ownership gap using our boundary triple difference estimates.

Figure 8 plots the results. From 1910 to 1940, housing outcomes were largely similar in both treated and untreated cities. But in subsequent decades, home ownership rates, house values, and rents grew faster in untreated cities, only reverting somewhat back to pre-1940 norms in recent decades. The racial gap also begins to diverge after 1940 but surprisingly has yet to show evidence of retreat as of 2010. Of course, to be clear, the racial composition gap we are measuring here is between entire cities rather than across borders, so it is not directly comparable to the previous analysis. Nevertheless, the figures are striking and, we believe, lend additional credence to the economic implications of our baseline boundary findings.[46]

Moreover, there is little evidence of similar trends at nearby population levels. For example, Figure 9 plots the difference in outcomes between cities just above and below a 25,000-population threshold (and for easy reference, the same calculation for the HOLC 40,000 cutoff).[47] Again, assuming that cities on either side of 25,000 would not be systematically different with respect to the outcomes we consider, we expect

[45] We exclude any nonredlined city within 50 miles of a redlined city to avoid the possibility that it might have effectively been treated. Our redlining cities were drawn from the 149 cities that were digitized by the University of Richmond's Digital Scholarship Lab and a list of additional HOLC mapped cities from Price Fishback. The additional list resulted in the inclusion of Jamestown, New York, and Perth Amboy, New Jersey. City population size was based on published volumes of the 1930 census.

[46] The decline in aggregate home ownership may seem surprising given the program's intent to improve the functioning of lending markets. However, in smaller redlined cities, almost 90 percent of census tracts are graded either C or D, where lending was potentially restricted by the maps. Ideally, we would separately compare B, C, and D areas of these cities. Of course, by definition, grades are not available for nonmapped areas. Moreover, many of these small cities were not tracted until as late as 1990, making 1950 to 1980 especially difficult to infer. Instead, we use an ordered probit estimated on the full sample of HOLC cities to predict grades for each 1990 census tract based on 1920–1930 trends and 1930 levels of tract characteristics. We then compare the evolution of demographic characteristics in 1940 and 1990 by the predicted grade of the tract. Relative to nonredlined cities, we find that share African American grew more in the predicted D and C neighborhoods and less in the predicted B neighborhoods in cities that were redlined, consistent with our other evidence. Unfortunately, results on housing outcomes are too imprecise to draw inferences. That imprecision may be, in part, because the housing effect had dissipated by 1990.

[47] As of 1930, there were 103 cities with a population between 25,000 and 35,000 and 257 cities with a population between 15,000 and 25,000. As far as we have been able to ascertain, none of the 360 small cities used for our placebo exercise were redlined. The 25,000 line in Figure 9 plots the mean outcome of the 25,000–35,000 cities less the mean outcome of the 15,000–25,000 cities.

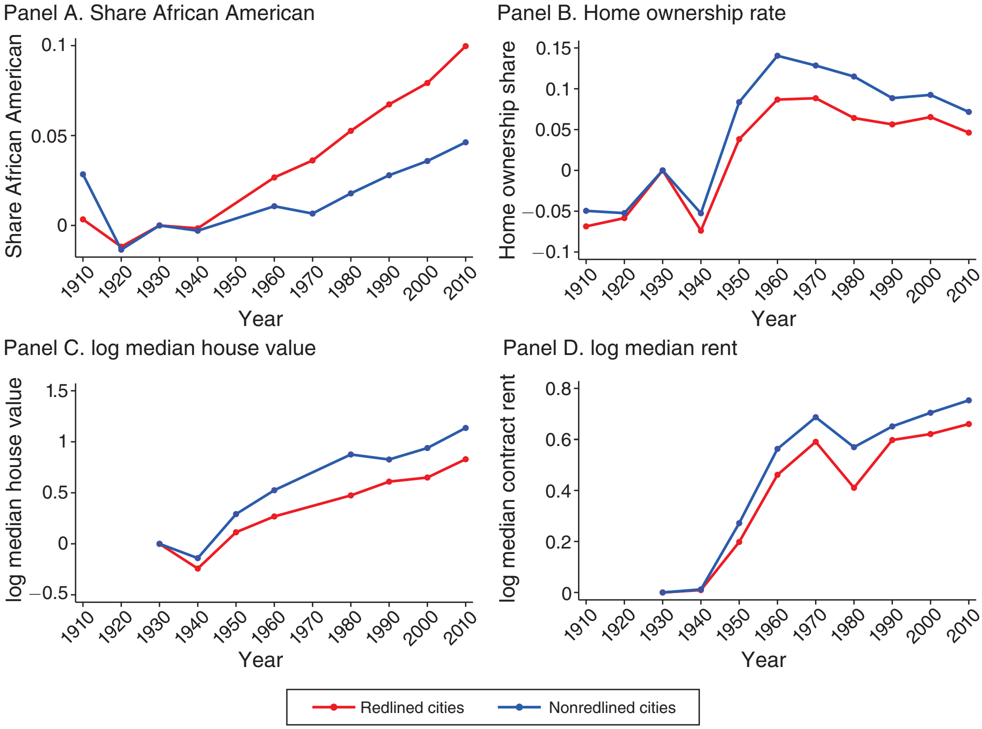

Figure 8. Comparison of Redlined versus Nonredlined Cities Using 40,000-Population Cutoff

Notes: The red line shows trends in outcomes for 27 redlined cities with populations between 40,000 and 50,000 in 1930. The blue line shows trends in outcomes for 26 nonredlined cities with a population between 30,000 and 40,000 in 1930 that were located at least 50 miles away from the closest redlined city. The sample of nonredlined cities was constructed using published volumes of the 1930 US census. All estimates are normalized to equal zero in 1930, and house values and rents are in 2010 US dollars.

this exercise to produce roughly a flat line. That is the case with 3 of 4 outcomes and lies in sharp contrast to the patterns observed at the 40,000 cutoff. Home ownership rates increase faster for the 25,000–35,000 cities relative to the 15,000–25,000 cities between 1950 and 1970, but this pattern moves in the opposite direction to the larger (redlined) cities relative to the smaller (nonredlined) cities around the 40,000-population threshold.

VI. Discussion

A. *Mechanisms Leading to Urban Disinvestment*

The most straightforward explanation of the significant and lasting impact of the HOLC maps is reduced access to credit and therefore higher borrowing costs. Typically, a decrease in household credit conditions impacts financial asset values like housing via a number of channels. But perhaps most pertinent here are a) lower housing demand, which can then cause household consumption and investment to fall as a result of negative wealth effects, and b) deteriorating leverage, possibly leading

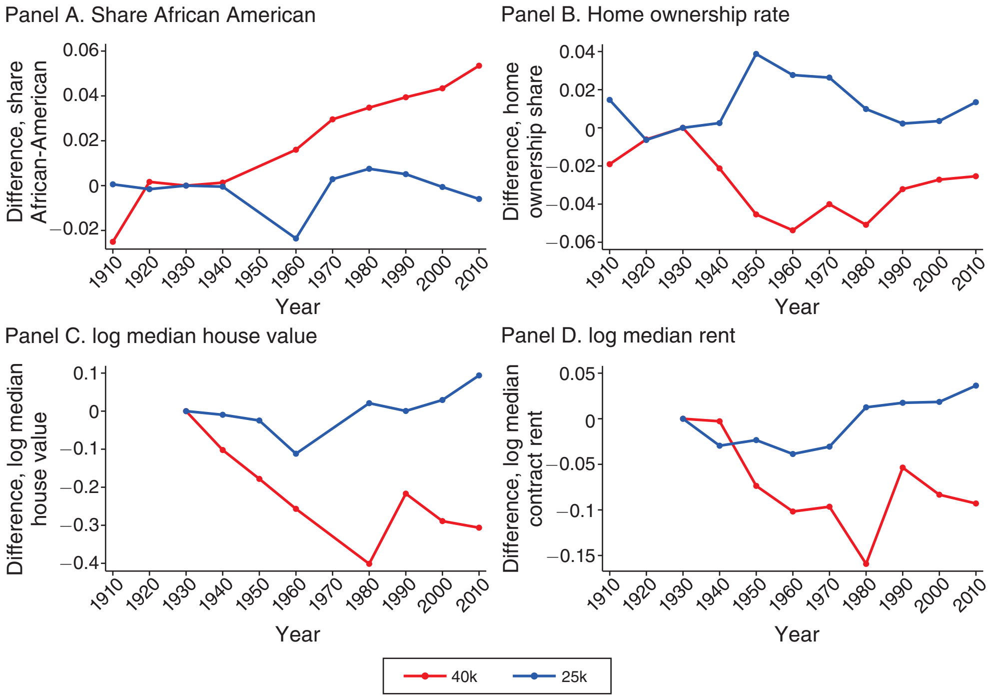

Figure 9. City Outcomes Using a Placebo 25,000-Population Cutoff

Notes: The red lines show the difference in trends between 27 redlined cities with populations between 40,000 and 50,000 in 1930 and 26 nonredlined cities with populations between 30,000 and 40,000 in 1930 (the differences between the lines in Figure 8). The blue lines show the difference in trends between 103 cities with populations between 25,000 and 35,000 in 1930 and 257 cities with populations between 15,000 and 25,000 in 1930. The trends are normalized to zero in 1930, and house values and rents are in 2010 US dollars. See the text for additional details. Population data were drawn from published volumes of the 1930 US census.

to problems associated with debt overhang (e.g., Mian, Rao, and Sufi 2013).[48] Such a credit-induced decline in house prices raises the likelihood that property owners with mortgages could be left owing more than the market value of their property (Glaeser and Gyourko 2005). Homes with a market value below replacement cost are in turn much less likely to be maintained and improved (Gyourko and Saiz 2004; Haughwout, Sutherland, and Tracy 2013; Melzer 2017). Consequently, disinvestment in housing can occur when investment in maintenance does not keep pace with depreciation. Another related explanation is that rental housing, in general, tends to depreciate faster than owner-occupied housing (Shilling, Sermans, and Dombrow 1991; Rosenthal 2014). Therefore, stricter credit conditions that reduce home ownership can lead to more rapid depreciation.

In our setting, a number of institutional features may have exacerbated disinvestment in some communities. One example, specific to D-graded, heavily African

[48] Empirically, several recent papers have been able to convincingly show a causal link between credit access and the housing boom and bust of the 1990s and 2000s using the deregulation of bank branching (Favara and Imbs 2015) and exposure to damaged financial institutions (Mondragon 2020).

American areas, is the common pre–World War II practice of contract sales (Satter 2009). Individuals who could not obtain mortgages through the formal lending sector, in some cases because of low HOLC grades, may have instead purchased homes by entering into long-term loans known as contract sales. Under these contracts, ownership did not transfer until the final payment was made, and failure to meet the terms of the loan at any point could lead to residents losing all equity in the home. Furthermore, contract sales typically had higher implicit interest rates than were available in the formal lending sector. More generally, housing and other community disinvestment likely altered household composition in lower-graded neighborhoods toward groups with less political representation or social capital. That change may have exacerbated the adverse impacts of post–World War II federal policies such as public housing siting (Hunt 2009), interstate construction (Brinkman and Lin 2019), and urban renewal (Rast 2019) that further reinforced disinvestment in these neighborhoods and gave the maps a long-lasting imprint.

But perhaps the strongest evidence of HOLC-related disinvestment is the decline in home ownership, housing values, and rents documented thus far. Other direct measures of disinvestment—such as the housing vacancy rate, total housing units, and ratings of housing quality—tend to have flaws for our purpose. Nevertheless, they also are consistent with long-run housing disinvestment in low-graded neighborhoods. In particular, we took the 50 redlined cities in which there are census tract housing vacancy data available beginning in 1940. We assume our low-propensity-score approach, in which the parallel trends assumption is fully satisfied for other housing variables, would take care of the problem here as well. If so, we find that the D–C gap in vacancy rates increased from 0.1 (0.3) percentage points in 1940 to 0.6 (0.3) percentage points by 1990. Over the same period, the C–B gap increased from 0.2 (0.4) percentage points to 0.7 (0.4) percentage points. A 0.5 percentage point increase explains 8 to 16 percent of the change in vacancy rates in our buffer zones between 1940 and 1990.[49] Further corroborative support of housing disinvestment is in Krimmel (2017), who finds a 20 percent relative reduction in housing units between 1940 and 1970 on the D-graded side of HOLC borders compared to the C-graded side. Finally, the 1960 census asked directly about housing quality. Conditioning on a rich set of census income, education, and occupation variables, we estimate that the rate of deteriorating or dilapidated housing is 6.2 (1.1) percentage points higher in D neighborhoods than in C and B neighborhoods.[50] Together, we view these suggestive patterns as providing additional empirical support that declining investment caused long-term harm to lower-graded neighborhoods after the HOLC maps were drawn.

[49] The results are similar using the full sample of borders in the same cities. In 1990, the treated estimates are 0.9 (0.3) and 1.2 (0.3) for the D–C and C–B boundaries and 0.1 (0.3) and 0.5 (0.3) for the D–C and C–B grid-based controls. The mean 1940 vacancy rate is 3.6 and 4.0 percent in D–C and C–B buffer zones. The mean 1990 vacancy rate is 9.8 and 7.1 percent in D–C and C–B buffer zones.

[50] The mean rate of deteriorating or dilapidated housing in 1960 is 31, 12, and 5 percent in D, C, and B neighborhoods. Of course, we acknowledge that it is difficult to interpret this result without a preperiod.

B. *Possible Explanations for Differences by Border Type*

Our housing results, particularly for home ownership, often uncovered larger and more persistent negative effects among C–B borders. Strikingly, some estimated effects reverse course along D–C borders after 1970 or 1980, but not along C–B borders. We can think of at least three possible reasons.

One explanation could be that policies enacted later in the twentieth century, such as the Fair Housing Act of 1968 and the CRA of 1977, that were designed to address discriminatory housing practices may have successfully targeted D- but not C-graded areas. The CRA instituted a process whereby regulators examine whether banks were providing adequate levels of loans to low- and moderate-income individuals in the areas they serve. Since low- and moderate-income individuals are more likely to be in D-graded neighborhoods than C-graded neighborhoods, lending by banks to satisfy CRA compliance could have led to a reduction in home ownership and housing value gaps between and D and C areas, but less so between C and B areas. However, because the CRA was instituted federally and not locally, we cannot convincingly show that its passage causally led to the reversal in racial and housing gaps that began sometime in the 1970s.

A second hypothesis is that the effects of the HOLC grades may have had significantly more "bite" in C-graded neighborhoods than D-graded neighborhoods. If lending tended to be more restrictive in D areas than C ones in the premap period, the marginal effect of the maps might have been most pronounced in C areas, leading to a larger initial impact on the C–B boundary that also takes longer to dissipate. Relatedly, it may have been the case that the maps revealed more information concerning the long-term prospects of C neighborhoods. This is consistent with the fact that the preexisting gaps between B and C areas were less pronounced than the gaps between C and D areas.

A third plausible reason is that D areas were quicker to redevelop, causing D–C gaps to fade faster than C–B gaps. This pattern might arise if D areas are closer to the central business district (CBD), which can lead to earlier gentrification (Brueckner and Rosenthal 2009, Baum-Snow and Hartley 2017). That said, we found mixed evidence that proximity to the CBD played a critical role in the long run.[51] Alternatively, the building stock in D areas may have depreciated more rapidly and been more suitable and less costly to redevelop; this seems consistent with significantly higher levels of older, vacant, and deteriorating and dilapidated properties in D neighborhoods by midcentury.

C. *Population Dynamics by Race*

We find strong evidence, both in our border design experiments and city discontinuity design, that the HOLC maps affected the rise and eventual decline of urban

[51] We divided the sample into terciles by distance to the CBD of the city. We then compared the first tercile to a sample combining the second and third terciles. We find suggestive evidence that the effects on share African American from 1950 to 1980 were larger along borders that were closer to the CBD. However, that difference disappears by 1990 to 2010, suggesting any gentrification effect related to proximity to the CBD happened later in the century.

racial segregation during the post–World War II period. There are several factors potentially driving this phenomenon. One possibility is that receiving a low grade could have made a neighborhood less desirable for *every* household, but if Black households have fewer outside options, they end up predominantly moving to (and staying in) low-graded areas. In this case, older housing units would "filter down" to African Americans (Rosenthal 2014). Alternatively, it could be that a lack of credit access is less relevant for Black households than for White households. This distinction could arise from wealth disparities that blocked many Black households from purchasing a home or perhaps because Black households faced other barriers to credit, even if they lived in higher-graded neighborhoods, and therefore did not have a differential loss from moving to a D neighborhood. Both explanations, which we cannot separate, would suggest that a pickup in Black inflows was driving the maps' impact on racial segregation. On the other hand, the maps may have also acted as a coordination mechanism for the outward expansion of African American neighborhoods by lowering house values in primarily White neighborhoods that were near African American neighborhoods. This may have amplified the well-known phenomenon of White flight.[52]

In Figure 10, we decompose the extent to which the maps' impact on rising segregation was driven by White outflow or Black inflow. We find that the population flows responsible for the increase in share African American vary by border type.[53] A combination of increased White outflow and Black inflow boosted the share African American along D–C borders (panels B and C). In total, while Black inflow initially increased the *overall* population density along the D side of the D–C boundaries (panel A), ultimately population density reverted in subsequent decades as White flight commenced and Black inflow slowed. Panel D shows that the population dynamics of immigrants followed a similar pattern as that of Whites.

Along the C–B borders, the rising share of African Americans is driven entirely by increased and persistent inflows of African American residents (panel F). If anything, there is evidence of a relative inflow of White population along the C side compared to the B side in 1950 (panel G), although that reverts by 1960, when African American population density begins to increase on the C side. Therefore, White flight appears to be associated with the redlined boundaries but not the yellow-lined ones.

We also examined whether there were especially large changes in racial gaps that might occur if there were tipping points, as in Card, Mas, and Rothstein (2008). We found that to be the case, whether measured by indicators of the African American share being at least 50, 75, or 90 percent, along both D–C and C–B boundaries. This result suggests that our mean racial gap estimates conceal an even larger impact on the upper tail of the racial gap distribution.

[52] There is a vast literature that discusses the importance of urban White flight on racial segregation. Recent studies include Card, Mas, and Rothstein (2008); Boustan (2010); and Shertzer and Walsh (2019).

[53] We measure the across-border differences in density rather than population levels to account for the different units of geography available in each census.

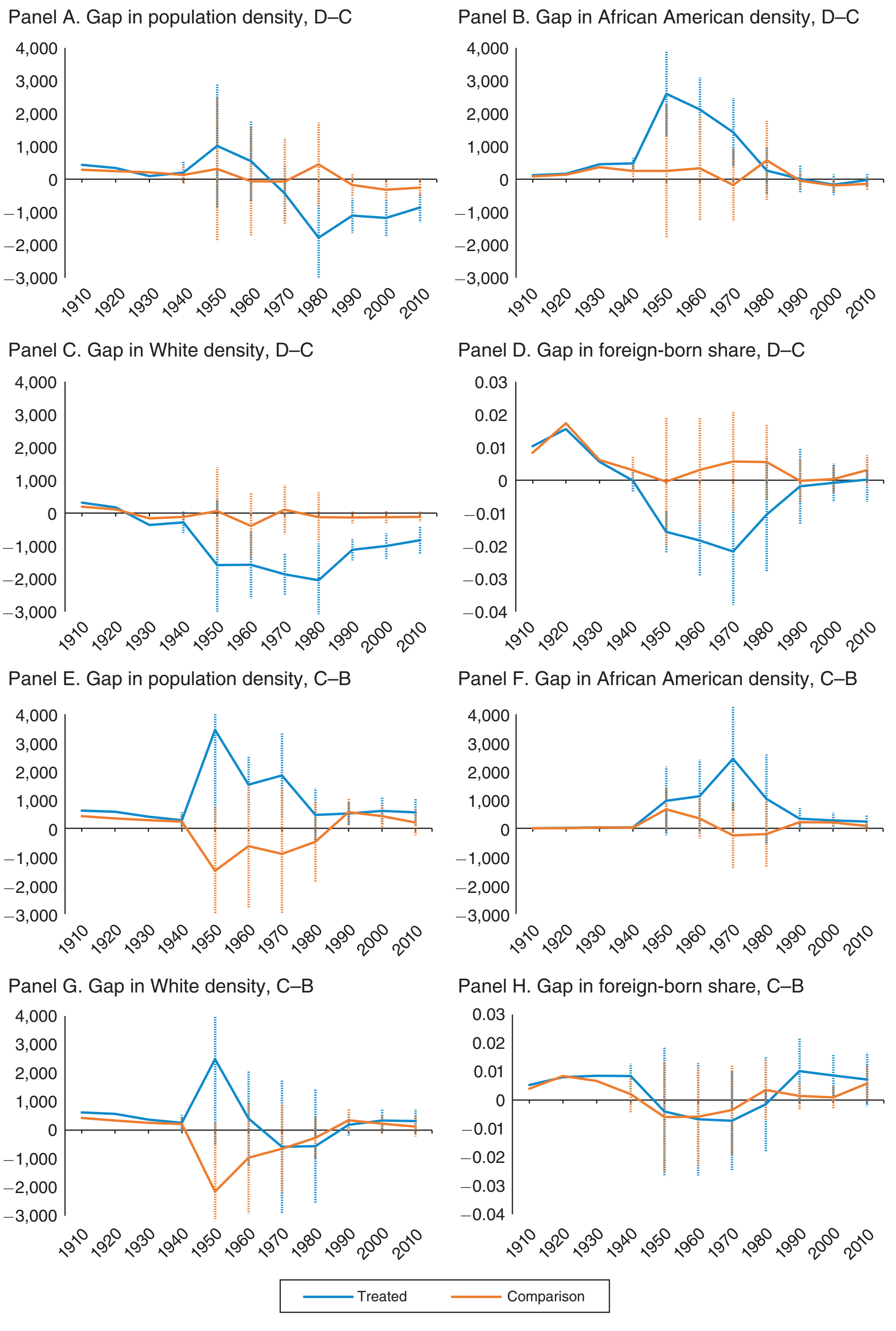

Figure 10. Population Dynamics along the D–C and C–B Boundaries

Note: See notes to Figure 5. See notes to Figure 4. Error bars in panels A, C, E, and G may extend above or below the bounds of the *y*-axis.

D. *City Heterogeneity*

We find substantial heterogeneity across cities and, in the online Appendix, make a preliminary effort to try to associate this heterogeneity with preexisting differences in bank competition, the relative coarseness of boundaries (fewer borders might make the maps less effective), city size, the level of inflows of Blacks due to the Great Migration, and World War II rent control measures. We find no compelling evidence that the first three factors played an economically important role in mediating the HOLC effects and mixed evidence that the Great Migration mattered.[54]

We uncover some evidence that the expansion of rent control during the War, which, as Fetter (2016) documents, led to the conversion of rental housing to owner-occupied units in the 1940s, may have attenuated the impact of redlining practices on rental prices immediately after the War. In particular, we split the 40 cities available in both our data and Fetter's into two equal-sized groups based on Fetter's (2016) median city-level rent control severity measure.[55] In cities that instituted tighter rent control policies, we find little difference in rental price gaps between HOLC treatment and control borders. We take that as evidence that rents were unable to adjust differentially to HOLC grades in these cities. By contrast, there are economically and statistically significant differences in rental prices in cities with less rent control (see online Appendix Figure A15). That said, we also expected to see the effect of rent control filter into the owner-occupied housing market, as in Fetter (2016), but based on our limited data, we are unable to find compelling evidence that that happened.

E. *Back-of-the-Envelope Economic Impact*

Finally, to provide a sense of the relative economic importance of the maps, we calculated how much our estimates could account for the overall gaps between different HOLC neighborhood grades (not just narrowly across boundary buffer zones). In particular, we divided our low-propensity boundary estimates by the full neighborhood estimates for share African American, home ownership, and house values for each border type and by two time periods (1950 to 1980 and 1990 to 2010).[56] Overall, we conclude that the maps account for between 15 and 30 percent of the overall gap in share African American and home ownership over the 1950 to 1980 period and 40 percent of the gap in house values. If we focus on just the C versus B neighborhoods over the 1950–1980 period, the maps account for roughly half of the home ownership and house value gaps.[57] After 1980, our estimates decline

[54] See the online Appendix for more detail.

[55] The 11 cities in his database that we do not have are Bridgeport, Connecticut; Cincinnati; Des Moines; Fall River, Massachusetts; Houston; Los Angeles; Memphis; Newark; Omaha; Providence; and Wilmington, Delaware. We use his rent control measure: pre-control max rent – freeze rent/max rent, which ranges from 0 to 0.09 with a median of 0.013.

[56] We concentrate on the low-propensity specifications because of the lack of a pretrend. Nevertheless, to be conservative, we still subtract out the 1930 estimate.

[57] There are very few African Americans in C–B neighborhoods until 1960, making it somewhat difficult to interpret this calculation for share African American, at least until later years.

in magnitude and therefore account for 0 to 20 percent of the D–C and C–B gaps in each of our outcomes.

VII. Conclusion

In response to the Great Depression, the federal government fundamentally reshaped the nature of housing finance to stabilize housing markets and support the lending industry. A slew of new federal agencies were created, including the FHLBB, and, under its auspices, the HOLC. Among their many initiatives, the FHLBB directed the HOLC to create a systematic and uniform scientific property appraisal process and produce residential security maps for all major cities. Some have argued that these initiatives had a profound and long-lasting influence on the real estate industry by initiating the so-called practice of "redlining." The residential security maps, which explicitly took into account demographic characteristics (e.g., race, ethnicity) of entire neighborhoods, were drawn for the purpose of influencing the property appraisal process. This in turn may have influenced lending as well as the provision of federal mortgage insurance.

We attempt to identify the causal effects of the HOLC maps on neighborhood development from 1940 through 2010. A major challenge for our analysis is that the maps were not exogenous and instead likely reflected existing neighborhood differences and trends. Therefore, there is a concern that the evolution of gaps in the postmap period reflects practices that would have occurred even in the absence of the maps. To address these challenges, we use a variety of empirical approaches, including the use of counterfactual boundaries that experienced the same preexisting trends but were where the HOLC did not ultimately draw borders. We also employ borders that may have been chosen for idiosyncratic reasons, where endogeneity is much less of a concern.

This border-based approach allows us to identify the local effects of the maps on areas proximate to borders. However, while there are benefits to this research design in terms of identification, it does not identify the effects in other parts of cities, much less capture citywide aggregate effects. Therefore, to examine aggregate effects at the city level, we also exploit a discontinuity in the HOLC's decision to only create maps for cities with a population above 40,000. The estimates from this last approach are fully immune to the criticism that the maps had no actual effects and that our estimates from the border design simply reflect the ability of real estate professionals to have somehow correctly predicted the future trajectory of US neighborhoods over many subsequent decades.

Using these approaches, we consistently find a significant and persistent causal effect of the HOLC maps on the evolution of housing markets in urban neighborhoods. We find that the maps had sizable effects on reducing home ownership rates, house values, and rents on the lower-graded sides of HOLC boundaries. These include 1) "D–C" boundaries, which demarcate borders between the lowest-graded neighborhoods ("D") marked in red and the next lowest-graded neighborhoods ("C") colored yellow, and 2) "C–B" boundaries, which are borders between the C-graded neighborhoods and the next highest-graded neighborhoods ("B") marked in blue. Intriguingly, the effects on home ownership—and, to a somewhat lesser extent, house values—dissipate over time along the D–C boundaries but remain highly persistent along the C–B boundaries. These patterns are consistent with the hypothesis that the maps led to reduced credit access and higher borrowing costs that, in turn, contributed to disinvestment in poor urban American neighborhoods, with long-run repercussions.

We also show that being on the lower-graded side of D–C boundaries led to a rising share of Black households from 1930 until about 1970 or 1980 before it started to decline thereafter. This same pattern appears along C–B borders, revealing for the first time that "yellow-lining" was also an important phenomenon. That the pattern begins to revert starting in the 1970s is at least suggestive that federal interventions like the Fair Housing Act of 1968, the ECOA of 1974, and the CRA of 1977 may have played a role in reversing the increase in segregation caused by the HOLC maps, though future research should consider a wide variety of alternative explanations for this reversal of trends. Nevertheless, racial segregation along both the C–B and D–C borders remains in 2010, almost three-quarters of a century later. We believe that our results highlight the key role that access to credit plays in the growth and long-running development of local communities.

REFERENCES

Aaronson, Daniel, Daniel Hartley, and Bhashkar Mazumder. 2021. "Replication data for: The Effects of the 1930s HOLC 'Redlining' Maps." American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor]. https://doi.org/10.38886/E127803V1.

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. 2010. "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program." *Journal of the American Statistical Association* 105 (490): 493–505.

Anders, John Paul. 2019. "The Long Run Effects of Transformational Federal Policies: Redlining, the Affordable Care Act and Head Start." PhD diss. Texas A&M University.

Appel, Ian, and Jordan Nickerson. 2016. "Pockets of Poverty: The Long-Term Effects of Redlining." https://www.idc.ac.il/en/schools/business/annual-conference/documents/pockets-of-poverty.pdf.

Baum-Snow, Nathaniel, and Daniel Hartley. 2017. "Accounting for Central Neighborhood Change, 1980–2000." Unpublished.

Bayer, Patrick, Fernando Ferreira, and Robert McMillan. 2007. "A Unified Framework for Measuring Preferences for Schools and Neighborhoods." *Journal of Political Economy* 115 (4): 588–638.

Bifulco, Robert, Jason M. Fletcher, and Stephen L. Ross. 2011. "The Effect of Classmate Characteristics on Post-secondary Outcomes: Evidence from the Add Health." *American Economic Journal: Economic Policy* 3 (1): 25–53.

Black, Sandra. 1999. "Do Better Schools Matter? Parental Valuation of Elementary Education." *Quarterly Journal of Economics* 114 (2): 577–99.

Black, Sandra, and Philip Strahan. 2002. "Entrepreneurship and Bank Credit Availability." *Journal of Finance* 57 (6): 2807–33.

Boustan, Leah Platt. 2010. "Was Postwar Suburbanization 'White Flight'? Evidence from the Black Migration." *Quarterly Journal of Economics* 125 (1): 417–43.

Boustan, Leah Platt. 2011. "Racial Residential Segregation in American Cities." In *Oxford Handbook of Urban Economics and Planning*, edited by Nancy Brooks, Kieran Donaghy, and Gerrit-Jan Knaap. Oxford, UK: Oxford University Press.

Breza, Emily, and Cynthia Kinnan. 2018. "Measuring the Equilibrium Impacts of Credit: Evidence from the Indian Microfinance Crisis." NBER Working Paper 24329.

Brinkman, Jeffrey, and Jeffrey Lin. 2019. "Freeway Revolts!" Federal Reserve Bank of Philadelphia Working Paper 19-29.

Brueckner, Jan, and Stuart Rosenthal. 2009. "Gentrification and Neighborhood Housing Cycles: Will America's Future Downtowns be Rich?" *Review of Economics and Statistics* 91 (4): 725–43.

Card, David, Alexandre Mas, and Jessie Rothstein. 2008. "Tipping and the Dynamics of Segregation." *Quarterly Journal of Economics* 123 (1): 177–218.

Carroll, Christopher. 2001. "A Theory of the Consumption Function, with and without Liquidity Constraints." *Journal of Economic Perspectives* 15 (3): 23–45.

Cameron, Stephen, and Christopher Taber. 2004. "Estimation of Educational Borrowing Constraints Using Returns to Schooling." *Journal of Political Economy* 112 (1): 132–82.

Chetty, Raj, Nathaniel Hendren, and Lawrence F. Katz. 2016. "The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment." *American Economic Review* 106 (4): 855–902.

Collins, William J., and Robert A. Margo. 2007. "The Economic Aftermath of the 1960s Riots in American Cities: Evidence from Property Values." *Journal of Economic History* 67 (4): 849–83.

Collins, William J., and Katherine L. Shester. 2013. "Slum Clearance and Urban Renewal in the United States." *American Economic Journal: Applied Economics* 5 (1): 239–73.

Dhar, Paramita, and Stephen L. Ross. 2012. "School District Quality and Property Values: Examining Differences Along School District Boundaries." *Journal of Urban Economics* 71 (1): 18–25.

Dube, Arindrajit, T. William Lester, and Michael Reich. 2010. "Minimum Wage Effects across State Borders: Estimates Using Contiguous Counties." *Review of Economics and Statistics* 92 (4): 945–64.

Faber, Jacob W. 2020. "We Built This: Consequences of New Deal Era Intervention in America's Racial Geography." *American Sociological Review* 85 (5): 739–75.

Favara, Giovanni, and Jean Imbs. 2015. "Credit Supply and the Price of Housing." *American Economic Review* 105 (3): 958–92.

Fetter, Daniel. 2016. "The Home Front: Rent Control and the Rapid Wartime Increase in Home Ownership." *Journal of Economic History* 76 (4): 1001–43.

Federal Housing Administration (FHA). 1934. *Underwriting Manual. Underwriting and Valuation Procedure Under Title II of the National Housing Act*. Excerpts in Richard C. Stearns, n.d., "Memorandum. Racial Content of FHA Underwriting Practices, 1934–1962." http://archives.ubalt.edu/aclu/pdf/Plex48.pdf.

Fishback, Price. 2014. "Panel Discussion on Saving the Neighborhood: Part III." https://arizonalawreview.org/pdf/syllabus/56ArizLRevSyl39.pdf.

Fishback, Price, Sebastian Fleitas, Jonathan Rose, and Kenneth Snowden. 2017. "Mortgage Foreclosure Overhangs and the Slow Recovery during the 1930s." Federal Reserve Bank of Chicago Working Paper 2017-12.

Fishback, Price, Alfonso Flores-Lagunes, William Horrace, Shawn Kantor, and Jaret Treber. 2011. "The Influence of the Home Owners' Loan Corporation on Housing Markets during the 1930s." *Review of Financial Studies* 24 (6): 1782–1813.

Ghent, Andra. 2011. "Securitization and Mortgage Renegotiation: Evidence from the Great Depression." *Review of Financial Studies* 24 (6): 1814–47.

Glaeser, Edward L., and Joseph Gyourko. 2005. "Urban Decline and Durable Housing." *Journal of Political Economy* 113 (2): 345–75.

Glaeser, Edward L., and Jacob Vigdor. 2012. *The End of the Segregated Century: Racial Separation in America's Neighborhoods, 1890–2010.* New York: Manhattan Institute.

Greer, James. 2012. "Race and Mortgage Redlining in the U.S." Unpublished.

Gyourko, Joseph, and Albert Saiz. 2004. "Reinvestment in the Housing Stock: The Role of Construction Costs and the Supply Side." *Journal of Urban Economics* 55 (2): 238–56.

Haughwout, Andrew, Sarah Sutherland, and Joseph Tracy. 2013. "Negative Equity and Housing Investment." Federal Reserve Bank of New York Staff Report 636.

Hillier, Amy. 2003. "Redlining and the Home Owners' Loan Corporation." *Journal of Urban History* 29 (4): 394–420.

Hillier, Amy. 2005. "Residential Security Maps and Neighborhood Appraisals: The Home Owners' Loan Corporation and the Case of Philadelphia." *Social Science History* 29 (2): 207–33.

Holmes, Thomas. 1998. "The Effect of State Policies on the Location of Manufacturing: Evidence from State Borders." *Journal of Political Economy* 106 (4): 667–705.

Hunt, D. Bradford. 2009. *Blueprint for Disaster: The Unraveling of Chicago Public Housing.* Chicago: University of Chicago Press.

Imbens, Guido. 2015. "Matching Methods in Practice: Three Examples." *Journal of Human Resources* 50 (2): 373–419.

Imbens, Guido, and Donald Rubin. 2015. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction.* Cambridge, UK: Cambridge University Press.

IPUMS, US NHGIS-A. 2020. IPUMS National Historical Geographic Information System: Version 15.0 [1950_tPH_Major, 1950_tPop_Misc, 1960_tPH, 1970_Cnt1, 1970_Cnt2, 1980_STF1, 1980_STF2a, 1990_STF1, 2000_SF1a, 2000_SF1b, 2010_SF1a, 2010_SF1b]. Minneapolis, MN: IPUMS. http://doi.org/10.18128/D050.V15.0 (accessed 2017–2020).

IPUMS, US NHGIS-B. 2020. Spatial Data Collection and Products Branch Geography Division, U.S. Census Bureau. IPUMS National Historical Geographic Information System: Version 15.0 [US Census's TIGER/line Shapefiles 2000 & 2010: Place, Track and Block-level]. Minneapolis, MN: IPUMS. http://doi.org/10.18128/D050.V15.0 (accessed 2016–2018).

Jackson, Kenneth. 1980. "Race, Ethnicity, and Real Estate Appraisal: The Home Owners Loan Corporation and the Federal Housing Administration." *Journal of Urban History* 6 (4): 419–52.

Krimmel, Jacob. 2017. "Persistence of Prejudice: Estimating the Long Term Effects of Redlining." https://faculty.wharton.upenn.edu/wp-content/uploads/2017/11/Krimmel-Redlining-DRAFT_Nov2017_v2.pdf.

Light, Jennifer. 2010. "Nationality and Neighborhood Risk at the Origins of FHA Underwriting." *Journal of Urban History* 36 (5): 634–71.

Melzer, Brian. 2017. "Mortgage Debt Overhang: Reduced Investment by Homeowners at Risk of Default." *Journal of Finance* 72 (2): 575–612.

Mian, Atif, Kamalesh Rao, and Amir Sufi. 2013. "Household Balance Sheets, Consumption, and the Economic Slump." *Quarterly Journal of Economics* 128 (4): 1687–1726.

Minnesota Population Center, and Ancestry.com. 2013. "IPUMS Restricted Complete Count Data: Version 1.0." IPUMS. Minneapolis, MN: University of Minnesota. https://ipums.org/projects/ipums-usa/d014.v1.0 (accessed 2016–2017).

Mondragon, John. 2020. "Household Credit and Employment in the Great Recession." https://www.dropbox.com/s/vjnnw9ox7zndkt9/Mondragon2020.pdf?dl=0.

Nelson, Robert K., LaDale Winling, Richard Marciano, Nathan Connolly, et al. 2021. "Mapping Inequality." American Panorama. https://dsl.richmond.edu/panorama/redlining/#loc=5/39.1/-94.58 (accessed December 15, 2016).

Nicholas, Tom, and Anna Scherbina. 2013. "Real Estate Prices During the Roaring Twenties and the Great Depression." *Real Estate Economics* 41 (2): 278–309.

Rast, Joel. 2019. *The Origins of the Dual City: Housing, Race and Redevelopment in Twentieth-Century Chicago.* Chicago: University of Chicago Press.

Rose, Jonathan. 2011. "The Incredible HOLC? Mortgage Relief during the Great Depression." *Journal of Money, Credit and Banking* 43 (6): 1073–1107.

Rosenthal, Stuart S. 2014. "Are Private Markets and Filtering a Viable Source of Low-Income Housing? Estimates from a 'Repeat Income' Model." *American Economic Review* 104 (2): 687–706.

Rothstein, Richard. 2017. *The Color of Law: A Forgotten History of How Our Government Segregated America*. New York: Liveright.

Satter, Beryl. 2009. *Family Properties: Race, Real Estate and the Exploitation of Black Urban America*. New York: Metropolitan Books.

Shilling, James D., C.F. Sirmans, and Jonathan F. Dombrow. 1991. "Measuring Depreciation in Single-Family Rental and Owner-Occupied Housing." *Journal of Housing Economics* 1 (4): 368–83.

Shertzer, Allison, Tate Twinam, and Randall Walsh. 2018. "Zoning and the Economic Geography of Cities." *Journal of Urban Economics* 105: 20–39.

Shertzer, Allison, and Randall Walsh. 2019. "Racial Sorting and the Emergence of Segregation in American Cities." *Review of Economics and Statistics* 101 (3): 415–27.

Sood, Aradhya, William Speagle, and Kevin Ehrman-Solberg. 2019. "Long Shadow of Racial Discrimination: Evidence from Housing Covenants of Minneapolis." https://www.frbatlanta.org/-/media/documents/news/conferences/2019/12/12/5th-biennial-real-estate-conference/sood_racial-discrimination-housing-covenants.pdf.

Wheelock, David. 2008. "The Federal Response to Home Mortgage Distress: Lessons from the Great Depression." *Federal Reserve Bank of St. Louis Review* 90 (31): 133–48.

White, Eugene N. 2014. "Lessons from the Great American Real Estate Boom and Bust of the 1920s." In *Housing and Mortgage Markets in Historical Perspective*, edited by Eugene White, Kenneth Snowden, and Price Fishback, 115–58. Chicago: University of Chicago Press.

Woods, Louis Lee. 2012. "The Federal Home Loan Bank Board, Redlining, and the National Proliferation of Racial Lending Discrimination." *Journal of Urban History* 38 (6): 1036–59.