Spring 2022

Essays on Discrimination and Spatial Inequality

Luca Perdoni
Yale University Graduate School of Arts and Sciences, luca.perdoni@yale.edu

Abstract

Essays on Discrimination and Spatial Inequality

Luca Perdoni

2022

The dissertation investigates institutional sources of spatial inequality, such as government-sponsored "redlining" and family structure. Throughout my doctorate, I developed an interest in the geographic sources of socioeconomic gaps that characterize US society according to race, ethnicity and gender. My research has focused on institutional features of the social environment that shape "neighborhood effects," such as place-based government interventions, family arrangements, and peer influences. Methodologically, I developed empirical strategies to estimate causal effects in observational settings where a control group is not immediately apparent. To carry out my research plans, I built and analyzed complex spatial datasets using a variety of geographic software programs and machine learning algorithms that have proven useful in defining valid and innovative control groups.

The first chapter is "*The Effects of Federal 'Redlining' Maps: A Novel Estimation Strategy*," joint work with Disa M. Hynsjö. Redlining, the systematic denial of credit to residents of a community, is often cited by activists and policymakers as one cause of enduring urban inequality. It is widely understood that the federal government started redlining in the 1930s. Government maps, identifying disadvantaged neighborhoods with the color red, have become a symbol of institutional discrimination. However, historians have disputed the ultimate influence of such maps on access to credit, and evidence of any causal economic impacts is scarce due to a lack of data and estimation challenges.

This paper investigates the causal effects of the Home Owners' Loan Corporation (HOLC) maps and the neighborhood grades they assigned to summarize lending risk in the second half of the 1930s. In particular, we estimate the effects of different grades on homeownership rates, property values and shares of African-Americans between 1940 and 2010. In

their time, the HOLC maps were a data analytics tool at the forefront of real estate appraisal techniques that soon became influential in the housing market at large. Our study illustrates how institutional practices can coordinate individual choices and amplify their discriminatory consequences.

To measure the short and long-term effects of the HOLC mapping intervention, we propose a new estimation strategy. Spatial discontinuity designs, often used in the literature on this topic, suffer from endogeneity concerns: multiple authors documented socioeconomic differences on opposite sides of boundaries traced by the agency, indicating that the HOLC did not assign border locations and grades randomly. Instead, we exploit an exogenous population threshold that determined which cities were mapped and a machine learning algorithm drawing HOLC maps in control cities. Using the grades predicted by the machine learning model, we apply a grouped difference-in-differences design to measure the causal effects of the HOLC intervention. The causal effects are identified by differences between neighborhoods in treated cities and areas in control cities that would have received the same grade but were not mapped. This empirical strategy is possible thanks to a new spatial dataset we constructed geocoding full-count Census records between 1910 and 1940. In addition, geographic coordinates let us join tract-level Census data for 1960-2010 and CoreLogic real property data to measure long-term outcomes.

We find a substantial reduction in property prices and a 2.4 percentage points decrease in homeownership rates in the lowest grade (red) areas in the short term. For this same grade, the HOLC maps caused a 1.6 percentage points increase in the local share of African-American residents in 1940. We also find a sizable house value reduction in the second to last grade (yellow) areas, showing that the causal impacts were not confined to red areas. Such negative effects for property prices persisted until the early 1980s, shortly after the federal government introduced legislative measures to counteract redlining.

The second chapter is titled "*The Long-Term Effects of Exposure to Non-Traditional Family Structures*." Single-mother households have become common in the US over the past fifty years. Economists, sociologists, and psychologists have documented that children from single-headed families have lower intergenerational mobility because of a lack of resources and the type of parenting they receive. However, little is known about the effects of children from single-mother families on their school peers. Taking advantage of the Add Health panel data structure, I estimate the effect of this feature of the adolescents' social environment on educational achievement and long-run labor market outcomes. My identification strategy is based on cohort-to-cohort variation in the percentage of children without a father figure within a school. The preliminary estimates indicate that exposure to peers with a higher rate of father absence does not have much of an effect on education, employment, or wages.

Essays on Discrimination and Spatial Inequality

A Dissertation
Presented to the Faculty of the Graduate School
of
Yale University
in Candidacy for the Degree of
Doctor of Philosophy

by
Luca Perdoni

Dissertation Director:
Joseph Altonji

May 2022

For Disa

Contents

List of Tables

List of Figures

Acknowledgments

This dissertation is the result of years of guidance, advice and mentoring from the Labor-Public Group of the Department of Economics at Yale University. I feel lucky I learned how to be a researcher in a group that combined academic rigor with kindness and openness. Such a precious environment is due to the leadership of Joe Altonji, who provides a constant example of scientific excellence and relentless goodwill. John Eric Humphries gave shape to my research thoughts from the very beginning, and he has been a fundamental source of reassurance ever since. Cormac O'Dea believed in my research often more than I did, providing an essential sense of legitimacy. His help through the most challenging moments has been vital.

Teresa and Clara made me feel at home every day in New Haven with unconditional acceptance and love. Francesca, Sarah, Alice, Cecilia, Alessandro, Giacomo and Silvia have always been ready to cheer me up from different corners of the world. Diletta, Pietro and Sara have been a safe haven to come back to in my hometown. Jaehee, Paula and Sun kept asking me how I was doing through ups and downs. A special thanks to Pam and Anne, who always had a solution ready.

Mia mamma e mia sorella hanno avuto fiducia in me e nelle mie scelte, anche quando tutto sembrava più difficile del previsto.

Julian has always been present, constantly reminding me I was not alone. Together with Disa, we made it through every milestone of the Ph.D. together. Celebrating the end of this journey without Disa feels wrong. The time spent working with Disa shaped my research identity and taught me courage. We miss you.

Chapter 1

The Effects of Federal "Redlining" Maps: A Novel Estimation Strategy. Joint with Disa M. Hynsjo

1.1 Introduction

Discrimination in the housing market can take many forms. Racial covenants, restrictive zoning and barriers to credit access are only a few examples. While developers, real estate agents and bank executives have often been blamed for their exclusionary norms, public institutions took the lead in shaping discriminatory practices in some cases. Between 1935 and 1940, a federal agency – the Home Owners' Loan Corporation (HOLC) – created *Residential Security Maps* for more than 200 US cities to summarize the financial risk of granting loans in different neighborhoods. Color-coded maps assigned each neighborhood one of four security grades, from A (green) to D (red).[1] Standardized forms attached to the maps (*Area Descriptions*) invariably described the presence of African Americans, Jews, and European immigrants as detrimental to a neighborhood's grade. In the late 1970s, urban historian K. T. Jackson rediscovered the maps at the National Archives and

[1] As an example, a scan of the HOLC map for New Haven, CT is available in Figure 1.1.

proposed them as an example of structural discrimination (Jackson, 1980). Since then, the view that the HOLC maps were a source of residential redlining, the systematic denial of mortgages to residents of a community, has steadily gained popularity (Rothstein, 2017; Coates, 2014). Nevertheless, some historians have emphasized the program's timing and confidentiality to raise doubts about whether a federal data collection program could have strongly influenced the housing market (Hillier, 2003, 2005; Fishback et al., 2021). Quantitative evidence supporting either claim about the maps' influence is scarce due to the lack of detailed and comprehensive datasets and the non-random assignment of different grades. In this paper, we propose a new strategy to measure the causal effects of a federal government initiative that has been proposed as a symbol of structural discrimination by journalists, activists, academics and presidential candidates.[2]

This paper estimates the short and long-term causal effects of the Home Owners' Loan Corporation maps using a new empirical strategy. Our approach exploits an exogenous population threshold: only cities above 40,000 residents were mapped. We use a machine learning classification algorithm to draw residential security maps in control cities with a population below the 40,000-resident threshold.[3] Using the grades predicted by the classification model, we apply a grouped difference-in-differences design to measure the causal effects of the HOLC maps. Our outcomes of interest are homeownership rates, property values, rent prices and shares of African American residents in 1940. We also analyze the evolution of these outcomes between 1960 and 2010. The estimated causal effects are identified by the differences between neighborhoods in treated cities and neighborhoods that would have received the same security grade if their city had been mapped.

The effects we find in red neighborhoods support the view that HOLC maps reduced access to credit and led to urban disinvestment. In 1940, shortly after the introduction of the maps, we find a sizable reduction in property prices and homeownership rates in

[2]As an example, see Rothstein (2017) and Coates (2014). Historical government support of redlining practices has been proposed by President J. Biden and Senator E. Warren as a motivation for their housing plans.

[3]Defining the control group with a machine learning algorithm is an alternative to synthetic control methods. In our case, control units are actual observations grouped by a predictive model replicating an observed classification mechanism.

HOLC Residential Security Map of New Haven, CT

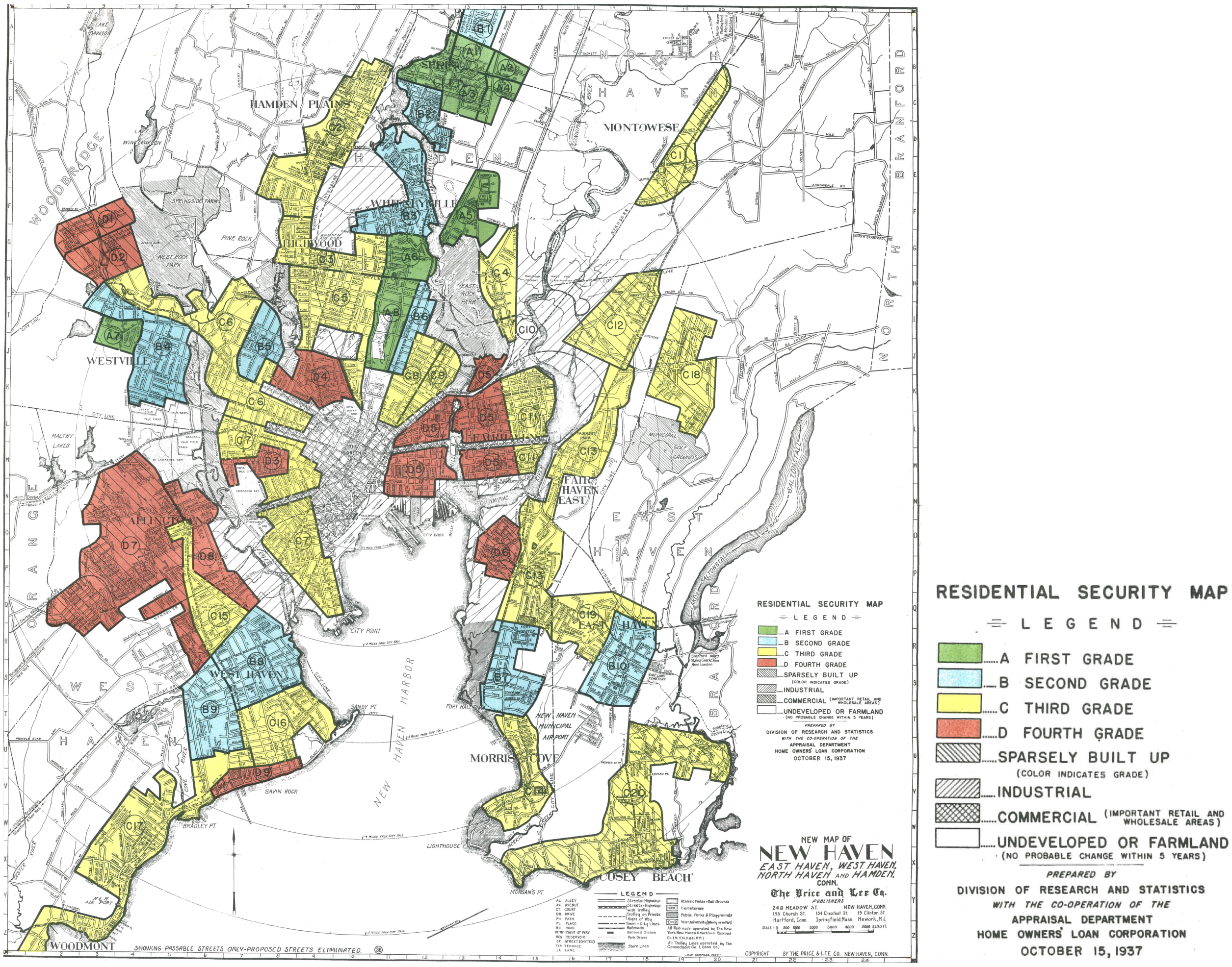


Figure 1.1 — The scan of the original Residential Security Map of New Haven, CT has been provided by *Mapping Inequality* (Nelson et al., 2021).

D (red) areas, along with an increase in the percentage of African Americans living in those neighborhoods. Property value reductions are also detected in C (yellow) areas.[4] The negative effects on property prices in yellow and red areas persisted until the early 1980s, shortly after the introduction of legislative measures meant to improve access to residential credit.[5]

The credibility of our approach relies on the performance of the machine learning clas-

[4] This result is consistent with a finding in Aaronson et al. (2021b) labelled by the authors as "yellow-lining".

[5] The Equal Credit Opportunity Act (1974), the Home Mortgage Disclosure Act (1975) and the Community Reinvestment Act (1977) had the common goal of increasing access to mortgages in neighborhoods previously ignored by financial institutions.

sification model. To assess its precision, we build a test dataset randomly excluding 25% of neighborhoods from the algorithm's training procedure. We then compare observed and predicted grades in the test dataset. Our trained random forest algorithm assigns the correct grade to more than 90% of test neighborhoods, and its predicted maps are convincing replicas of those made by HOLC. Even if the model is trained on the complete set of municipalities surveyed by HOLC, including all American metropolises, its performance is robust in cities close to our threshold of interest as well. When we restrict the test dataset to neighborhoods from cities with fewer than 50,000 residents, the overall precision is still above 90%.

Our empirical strategy is possible thanks to a spatial dataset we constructed using the 1910-1940 full-count census records (Ruggles et al., 2020), National Historical Geographic Information System (NHGIS) information (Manson et al., 2021) and CoreLogic property deeds. We clean and impute household addresses for each census decade following best practices from the urban history literature (Logan and Zhang, 2018). Detailed geographic coordinates are assigned to census observations using a state-of-the-art locator. Georeferenced data allow us to match census records with digitized HOLC maps and alternative sources of information to expand our dataset to the years beyond 1940. We include sociodemographic information from the National Historical Geographic Information System (NHGIS) (Manson et al., 2021) along with disaggregated property transaction prices from the CoreLogic deed database to estimate the long-term causal effects of the maps. The resulting dataset covers major US urban areas between 1910 and 2010.

Our paper contributes to a growing literature in economics studying the consequences of HOLC policies. This paper is most closely related to the work of Aaronson, Hartley and Mazumder (2021b), who use a border regression discontinuity design to measure the local effects of lower grades in cities surveyed by HOLC. Unlike Aaronson et al. (2021b), our estimation method compares similar neighborhoods in mapped and unmapped cities. In particular, we avoid spatial discontinuity designs because of endogeneity concerns due to differential pre-trends in socioeconomic variables on different sides of the borders traced by the HOLC.[6] Moreover, our empirical approach measures a different type of effect. While the existing literature has focused on the local effects of receiving a lower HOLC grade, we capture the global effects of the four HOLC grades. In our case, the counterfactual is made of similar neighborhoods not mapped by HOLC rather than nearby areas with a higher evaluation. With respect to Fishback, LaVoice, Shertzer and Walsh (2020), who investigates whether HOLC grades were racially-biased, we study a different question: the effects of HOLC grades on property prices and demographic characteristics. While there is a growing body of research on HOLC mapping,[7] we are the first to propose a predictive model replicating the HOLC maps and employing an exogenous population threshold for estimating the effects of different grades.

Focusing on the effects of institutionalizing a set of exclusionary attitudes is an important complement to existing research in the economics of discrimination. Economists have mainly focused on individuals who discriminate based on their taste (Becker et al., 1971), because of imperfect information[8] or implicit bias (Bertrand et al., 2005; Bertrand and Duflo, 2017). These different mechanisms originate from individual choices and cannot be readily applied to settings where something other than an individual discriminate.[9] This paper provides an empirical analysis focused on *institutional* discrimination, revealing new evidence of an overlooked source of socioeconomic inequality. The results capture the impact of institutional assessment practices developed by a governmental organization and adopted in the real estate market at large.

[6] The differential pre-trends, due to non-random location of borders and non-random assignment of grades, have been documented by Fishback et al. (2020) and Aaronson et al. (2021b) themselves. The authors of the latter paper employ propensity scores and a subset of idiosyncratic borders to address endogeneity concerns.

[7] All the recent papers on this topic either focus on different questions or employ different empirical approaches and datasets. In addition to Aaronson et al. (2021b) and Fishback et al. (2020), see Fishback et al. (2011), Faber (2020), Aaronson et al. (2021a) and Fishback et al. (2021). There is also a number of contemporaneous working papers on this topic using spatial regression discontinuities designs: see Anders (2019), Appel and Nickerson (2016) and Krimmel (2018).

[8] See Fang and Moro (2011) for a review of research on statistical discrimination originated by Phelps (1972) and Arrow (1973).

[9] See Small and Pager (2020) and Lang and Kahn-Lang Spitzer (2020) for a comparison of perspectives on discrimination between Sociology and Economics. The sociological literature has focused more on institutional sources of discrimination, if compared with economics.

Another relevant dimension of the HOLC initiative was its technological content. The agency undertook an unprecedented data collection effort,[10] creating a data analytics tool at the forefront of real estate appraisal techniques of the time. HOLC maps can be interpreted as an innovation in statistical technology that led to increased automation in the processing of mortgage applications. Today, concerns about algorithmic bias (Rambachan et al., 2020; Ludwig and Mullainathan, 2021) and distributional impacts of statistical technology are widespread (Fuster et al., 2021). Our results characterize the effects of a federal initiative that provided a powerful and practical tool to evaluate local housing market conditions.[11] They also offer a cautionary tale of how institutional practices can coordinate individual biases and amplify their consequences.

The paper also contributes to the literature investigating the causes of segregation and urban inequality.[12] As outlined in Boustan (2013), residential segregation can result from individual choices by white households,[13] Black self-segregation,[14] or collective action.[15] Our findings give an example of the last of these causes since the federal agency's practices had the effect of reinforcing residential exclusion. In terms of methods, we contribute to a relatively recent body of literature using machine learning algorithms to build control groups for causal inference in observational studies.[16]

The paper proceeds as follows. Section 1.2 provides additional details about HOLC activities and the circulation of its maps, while Section 1.3 contains a description of our novel dataset. We outline our empirical strategy in Section 1.4, along with results about the

[10]See Michney (2021) for a description, based on HOLC staff correspondence, of how the maps were developed.

[11]In particular, we provide an empirical analysis of the consequences of a collection of federal maps. See Nagaraj and Stern (2020) for a review of recent work about the Economics of maps.

[12]See Glaeser and Vigdor (2012), Cutler et al. (1999) and Logan and Parman (2017) for an overview of trends for different urban segregation measures.

[13]The mechanism is often referred to as "white flight". See Boustan (2010) and Boustan (2016).

[14]This possible source of segregation does not find strong empirical support. See Krysan and Farley (2002) and Ihlanfeldt and Scafidi (2002).

[15]Collective action to induce segregation can take many forms. Some examples are racial covenants (Jones-Correa, 2000; Sood et al., 2019), urban renewal programs (Collins and Shester, 2013) and public housing programs (Chyn, 2018; Tach and Emory, 2017).

[16]An example is Liberman et al. (2018). See Mullainathan and Spiess (2017) for a review of machine learning algorithms within the econometric toolbox.

performance of our classification algorithm and an array of validity checks. The estimated effects of HOLC maps can be found in Section 1.5. Section 1.6 concludes.

1.2 Historical Background

In the aftermath of the Great Depression, the Roosevelt Administration developed several programs to tackle a mortgage crisis characterized by soaring default rates and falling property values. The Home Owners' Loan Corporation (HOLC) was created in 1933 to aid homeowners "in hard straits largely through no fault of their own" (Federal Home Loan Bank Board, 1937). Under the direction of the Federal Home Loan Bank Board (FHLBB), the first task of the HOLC was to refinance mortgages in distress with longer terms, lower interest rates and higher loan-to-value ratios. In particular, the HOLC granted fully amortized loans with 15-year minimum terms at 5% interest rate, financing up to 80% of the property value.[17] The agency concluded its $3 billion lending effort in 1936 after refinancing over one million loans and holding approximately 10% of US non-farm mortgages (Jackson, 1980).

As a consequence of their lending program, HOLC gained considerable exposure to the housing market. Government officials believed that a healthier lending industry was necessary to safeguard the value of federal real estate investments (Hillier, 2005). In particular, they considered the standardization of appraisal techniques critical to achieving price stability. For this reason, the FHLBB directed HOLC to develop a systematic evaluation process for US neighborhoods, following a growing interest in ecological models across the real estate industry.[18] In 1936, HOLC started the *City Surveys* program, producing maps (*Residential Security Maps*) and standardized forms (*Area Descriptions*) for 239 major U.S. cities. The initiative was completed by 1940.

Field agents drew HOLC maps based on published reports, public records, federal

[17]These terms were much more convenient to homeowners than the 5-year interest-only loans, with interest up to 7%, that were prevalent in the market up to that time. See Fishback et al. (2011).

[18] See Jackson (1980) and Light (2010) for a discussion about how ecological models, newly-developed by the Chicago School of Sociology, became an influential theory for real estate appraisal.

maps and detailed surveys of local financial institutions (Michney, 2021). The availability of these sources varied between cities, and HOLC agents relied on their networks in the real estate community to supply any missing information. The result was meant to be "a composite opinion of competent realtors engaged in residential brokerage, good mortgage lenders and the HOLC appraisal staff."[19] HOLC agents traced boundaries to divide residential areas into homogeneous neighborhoods. They then assigned a grade on a four-level scale meant to summarize the financial security of real estate investment in each zone.[20] Areas colored in green (grade A) were the first-tier neighborhoods, while blue neighborhoods (grade B) were deemed still good. The color yellow (grade C) highlighted neighborhoods becoming obsolete or at risk of "infiltration of a lower grade population" (Hillier, 2005). Red neighborhoods (grade D) were considered "hazardous" (Hillier, 2003) for investment. The agency also produced detailed *Areas Descriptions* for each neighborhood. In these forms, they described housing conditions, local amenities and the area's demographic composition.[21] The presence of African Americans, Jews and certain European immigrants was inevitably characterized as a "detrimental influence" that "infiltrated" the American social fabric with fatal effects on local housing markets (Jackson, 1980). While the inclusion of racial and ethnic hierarchies in real estate appraisal was pervasive at that time, HOLC practices implemented these notions at an unprecedented scale with the coordinated effort of more than 20,000 employees distributed across more than 200 local offices, and the stamp of federal approval.

HOLC could not have used the results of the *City Surveys* in its lending decisions since the maps were created after the agency completed its refinancing effort. Therefore, the economic impact of the maps depends on how widely these documents circulated among other federal agencies and private financial institutions. The literature offers diverging views on this topic. Hillier (2003) reports that the FHLBB intended to restrict access to

[19] Corwin A. Fergus to T. L. Williamson, October 2, 1936, Roll 431, Home Owners Loan Corporation, microfilm copies of General Administrative Correspondence 1933-36, National Archives II (College Park, MD). As cited in Michney (2021).

[20] A scan of the HOLC map drawn in 1937 for New Haven, CT can be found in Figure 1.1.

[21] As an example, a scan of the HOLC Area Description for New Haven D-4 neighborhood is available in Figure 1.2.

"agencies within the FHLB" and "such government agencies having interests allied with those of the Board" while no copies were granted to "private interests". However, the author concedes that the maps were in strong demand among the public and that local consultants employed by the HOLC had access to these documents. An opposite stance, first proposed by Jackson (1980) and more popular today, argues that HOLC's findings were widely distributed and quickly became a benchmark for real estate appraisal both in government agencies and the private sector.

Even if we lack definitive evidence about the circulation of *City Surveys*, there is proof that another federal agency – the Federal Housing Administration (FHA) – received multiple copies of HOLC maps. The FHA evaluated applications to its mortgage insurance program with manuals that described the presence of "undesirable racial or nationality groups" in a neighborhood as detrimental. Moreover, the FHA employed a collection of maps that categorized neighborhoods on a four-level scale according to their financial security (Aaronson et al., 2021b). Today, a systematic comparison between the HOLC security maps and those of the FHA is impossible, but historical research provides evidence that the two collections were often similar.[22] If so, HOLC maps can be considered the best available proxy for FHA standards of neighborhood appraisal.[23] While the HOLC ceased its activities in 1951, the FHA continued its operations in the following decades.[24]

The explicit inclusion of racial or ethnic criteria in real estate financing became illegal in 1968 with the introduction of the Fair Housing Act. A further series of federal laws enacted in the 1970s addressed concerns about the lasting effects of financial exclusion. In particular, the Equal Credit Opportunity Act (1974), the Home Mortgage Disclosure Act

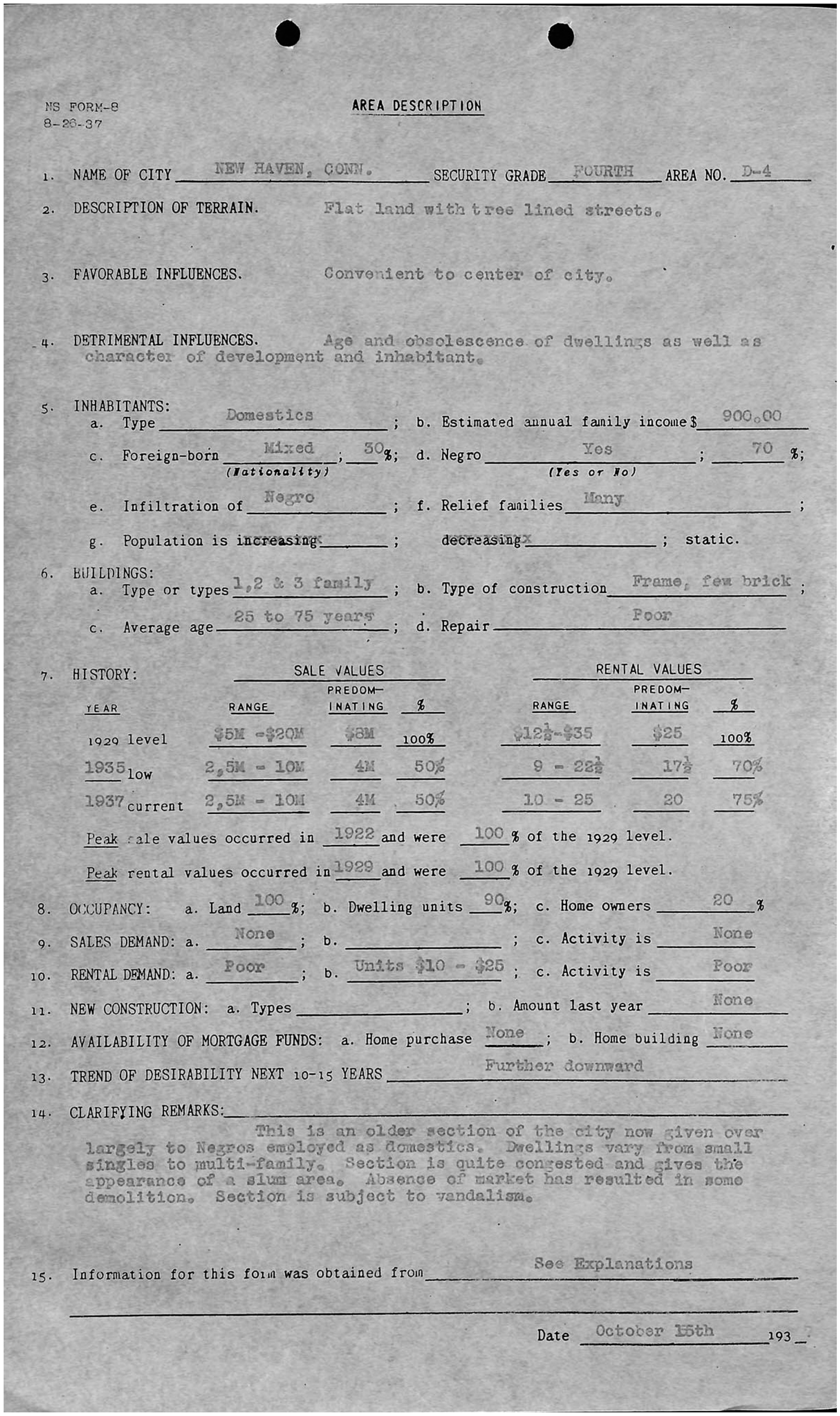
[22]Nearly all FHA maps are missing. A limited comparison is possible thanks to a reproduction of the FHA map of Chicago, IL (Light, 2010).

[23]Fishback et al. (2021) study FHA-backed mortgages in three US cities between 1935 and 1940. The vast majority of loans were granted in areas rated A or B by HOLC maps. However, the authors argue that FHA exclusionary patterns were established before HOLC maps were drawn, and they did not change throughout their study period.

[24]There is no historical evidence of different FHA appraisal practices according to the 40,000 resident threshold, or any other population threshold. Moreover, there is no evidence about how the FHA used HOLC maps. We assume homogeneous FHA practices in treatment and control cities, except for the availability of the HOLC maps.

(1975) and the Community Reinvestment Act (1977) were meant to counteract redlining by reinforcing anti-discrimination legislation, introducing mortgage disclosure requirements and supervising credit supply at the local level.

HOLC Area Description, Neighborhood D-4, New Haven, CT



NS FORM-8
8-20-37

AREA DESCRIPTION

1. NAME OF CITY NEW HAVEN, CONN. SECURITY GRADE FOURTH AREA NO. D-4
2. DESCRIPTION OF TERRAIN. Flat land with tree lined streets.
3. FAVORABLE INFLUENCES. Convenient to center of city.
4. DETRIMENTAL INFLUENCES. Age and obsolescence of dwellings as well as character of development and inhabitant.
5. INHABITANTS:
   a. Type Domestics ; b. Estimated annual family income $ 900.00
   c. Foreign-born Mixed (Nationality) ; 30 %; d. Negro Yes (Yes or No) ; 70 %;
   e. Infiltration of Negro ; f. Relief families Many ;
   g. Population is increasing ; decreasing x ; static.
6. BUILDINGS:
   a. Type or types 1,2 & 3 family ; b. Type of construction Frame, few brick ;
   c. Average age 25 to 75 years ; d. Repair Poor
7. HISTORY:

YEAR	SALE VALUES RANGE	PREDOMINATING	%	RENTAL VALUES RANGE	PREDOMINATING	%
1929 level	$5M -$20M	$8M	100%	$12½-$35	$25	100%
1935 low	2,5M - 10M	4M	50%	9 - 22½	17½	70%
1937 current	2,5M - 10M	4M	50%	10 - 25	20	75%

Peak sale values occurred in 1922 and were 100 % of the 1929 level.
Peak rental values occurred in 1929 and were 100 % of the 1929 level.

8. OCCUPANCY: a. Land 100 %; b. Dwelling units 90 %; c. Home owners 20 %
9. SALES DEMAND: a. None ; b. ; c. Activity is None
10. RENTAL DEMAND: a. Poor ; b. Units $10 - $25 ; c. Activity is Poor
11. NEW CONSTRUCTION: a. Types ; b. Amount last year None
12. AVAILABILITY OF MORTGAGE FUNDS: a. Home purchase None ; b. Home building None
13. TREND OF DESIRABILITY NEXT 10-15 YEARS Further downward
14. CLARIFYING REMARKS:
    This is an older section of the city now given over largely to Negros employed as domestics. Dwellings vary from small singles to multi-family. Section is quite congested and gives the appearance of a slum area. Absence of market has resulted in some demolition. Section is subject to vandalism.
15. Information for this form was obtained from See Explanations

Date October 15th 193_

Figure 1.2 — The scan of the original Area Description for neighborhood D-4 of New Haven,CT has been provided by *Mapping Inequality* (Nelson et al., 2021).

1.3 Data

We construct a new dataset drawing from three sources: digitized HOLC maps, census data and CoreLogic deeds records. Our classification algorithm is trained on 1930 census data merged with HOLC maps. The short-term effects of the HOLC maps are measured with 1930 and 1940 census data, while the long-term effects are estimated by combining full-count census data with CoreLogic real property data and tract-level census data for the decades between 1960 and 2010.[25]

We plan to extend the analysis to 1950 when the full-count census data for that decade becomes available in April 2022. In the following sections, we provide additional details about each data source.

1.3.1 HOLC Residential Security Maps

We incorporate HOLC grades in our project using the digitized maps provided by the Digital Scholarship Lab at the University of Richmond (Nelson et al., 2021). The files contain maps for 202 cities in 38 states. We convert neighborhood shapes originally traced by HOLC into a regular grid of hexagons. Hexagons are our fundamental spatial unit of observation, and their use simplifies the construction of HOLC maps in control cities.[26] The area of one hexagon approximates the typical size of a block in US grid plan cities such as New York City and Chicago.[27] We assign a grade to a hexagon if one color occupies at least 75% of its surface.[28] This spatial transformation has a negligible impact on the overall distribution of the grades. Appendix Table 1.11 shows the proportions of each grade according to different spatial definitions, while Figure 1.3 compares the digitized version

[25] Validity checks of the empirical strategy employ census data between 1910 and 1930, the pre-treatment decades.

[26] More details on why we choose to use a grid of hexagons can be found in Section 1.4.1.

[27] The grid is made of regular hexagons with an area of 0.025 squared kilometers (7.3 acres) and a side of approximately 100 meters (328 feet).

[28] The results are robust to modest variations in the 75% threshold. Given the small dimension of each spatial unit, the vast majority of hexagons (81.2%) contains only one grade. 7.5% of hexagons do not meet the 75% threshold and have a missing grade.

of the HOLC map of New Haven, Connecticut, with its hexagon-level counterpart. The percentage reductions for A and B grades are due to the smaller average size of HOLC neighborhoods in these classes. Furthermore, while our hexagons have a fixed area, the HOLC neighborhoods do not, which explains the minor discrepancies in the shares for C and D grades.

Comparison of HOLC Digitized Map and its Hexagon Version

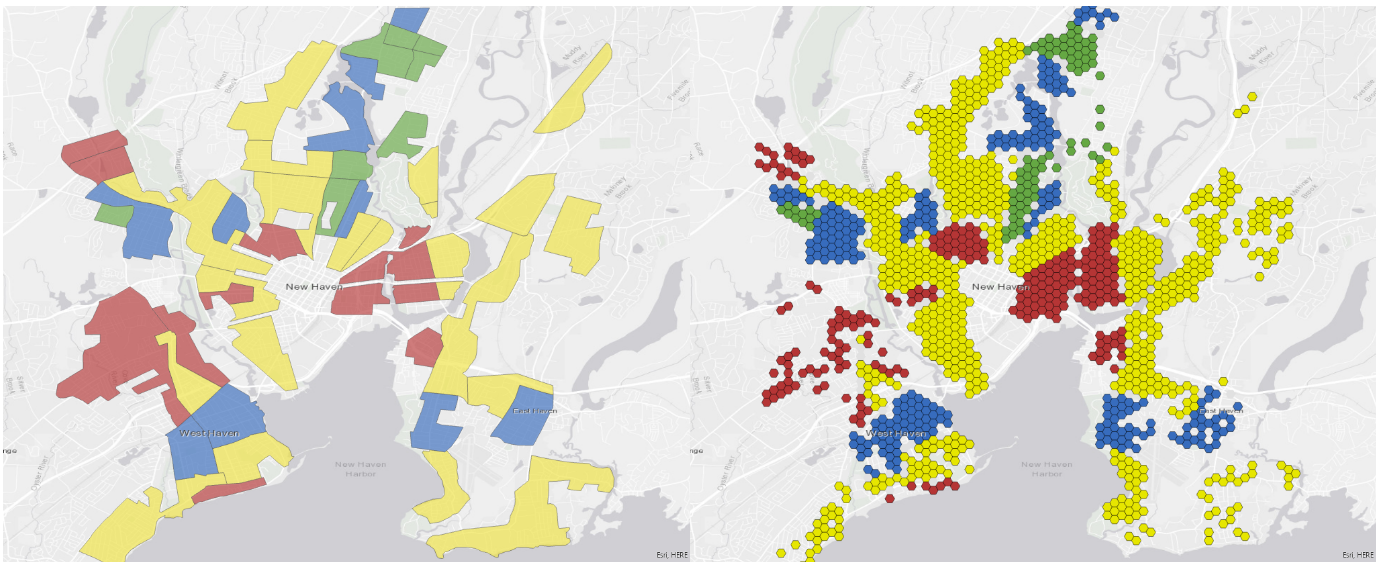


Figure 1.3 — The digitized version of the Residential Security Map of New Haven, CT, shown in the left panel, has been provided by *Mapping Inequality* (Nelson et al., 2021). Details about the definition of the hexagon grid can be found in Section 1.3.1. The right panel shows our hexagon-level replica of the original HOLC map. All the maps are north-oriented.

1.3.2 Census Data

Full-Count Census Data

We rely on full-count census records for data between 1910 and 1940. We geocode the heads of household by taking advantage of the addresses available in the proprietary version supplied by *ancestry.com* and IPUMS (Ruggles et al., 2020). Census addresses are cleaned following best practices found in the spatial history literature (Logan and Zhang, 2018), and geographic coordinates are assigned by a state-of-the-art locator (ESRI

StreetMap Premium 2019) that combines parcel centroids and street locations.[29] Detailed geographic coordinates allow us to construct neighborhood-level averages by combining individual observations with our graded hexagon grid. Population distributions according to the grades can be found in Table 1.1. Yellow and red areas include 77.8% of the general population in our 1930 sample, but contain 95.6% of African American respondents. Table 1.2 reports descriptive statistics of the 1930 census according to the HOLC grades. Even before the agency's intervention, African Americans were concentrated mainly in red neighborhoods. Homeownership rates, property values, and rent prices are all positively correlated with the HOLC scale.

		Population Shares			
	N	A	B	C	D
General Population	30,945,584	3.9%	18.3%	41.7%	36.1%
By Race					
White	28,801,136	4.1%	19.4%	43.9%	32.6%
Black	2,094,493	0.9%	3.1%	12.0%	83.6%

Notes: The sample includes 1930 Census individuals with a valid geocode in neighborhoods with a digitized HOLC map.

Table 1.1 — 1930 Population Distribution According to HOLC Grades

NHGIS Data

Starting in 1950, we must rely on publicly available census data. We obtain tract-level data for homeownership rates, property values, rent prices, and the shares of African Americans between 1950 and 2010 from the National Historical Geographic Information System (NHGIS) at IPUMS (Manson et al., 2021). We focus on census tracts since they are the smallest geographical units identifiable between 1950 and 2010.[30] Census tracts became available in smaller cities only in later decades. Hence, this source does not provide full

[29] The overall proportions of matched addresses for 1910,1920,1930 and 1940 are 60.5%, 65.4%, 76.1% and 73.5% respectively.

[30] The median population of a census tract in our dataset is 231, while it is 68 for hexagons. The hexagon area is constant while the one of the census tract is not. A census tract is always bigger than the hexagons we defined in surface terms. The median census tract contains 34 hexagons in surface terms.

	Grade			
	A	B	C	D
Black	0.01 (0.09)	0.01 (0.085)	0.02 (0.11)	0.17 (0.32)
Home Owner	0.76 (0.28)	0.66 (0.27)	0.56 (0.27)	0.43 (0.28)
House Value	12,938 (7,013)	8,805 (5,134)	6,638 (3,967)	5,038 (3,648)
Rent	143 (352)	73 (174)	51 (120)	42 (112)
Income Score	7.30 (0.26)	7.23 (0.22)	7.12 (0.21)	6.931 (0.30)
First Gen Immigrant	0.14 (0.17)	0.16 (0.17)	0.22 (0.20)	0.23 (0.24)
Unemployed, Men	0.04 (0.11)	0.06 (0.11)	0.10 (0.13)	0.13 (0.14)
Owns a Radio	0.76 (0.27)	0.69 (0.26)	0.58 (0.27)	0.38 (0.27)

Notes: The sample includes 1930 census individuals with a valid geocode in neighborhoods with a digitized HOLC map. See Appendix Section 1.7.2 for definitions of Census variables in our dataset.

Table 1.2 — 1930 Descriptive Statistics According to HOLC Grades

coverage of our sample of interest until 1980.[31] This is the best available nationwide source of harmonized data for demographic characteristics and homeownership rates in the second half of the twentieth century.

Geographical coordinates allow us to harmonize information from HOLC maps, 1910-1940 full count census data and 1960-2010 NHGIS data. We can use our composite dataset to describe the socioeconomic evolution of US neighborhoods throughout the twentieth century according to the grades assigned by HOLC in the late 1930s. Figure 1.4 contains trends for our four outcomes of interest between 1910 and 2010, showing that the HOLC ranking in terms of homeownership rates, property values and rent prices was stable during the last century. D (red) neighborhoods were, and still are, the most likely residence for African Americans. The percentage of Black American residents increased in A, B and C

[31] As mentioned in Section 1.4, we focus on cities with population between 30,000 and 50,000. Appendix Table 1.12 reports rates of coverage of NHGIS data for our sample of interest.

areas after World War Two. In particular, C (yellow) neighborhoods reached a 10% share of African Americans in 1970, while B (blue) neighborhoods met the same threshold in 1990. A (green) neighborhood did not attain the same level in 2010 yet.

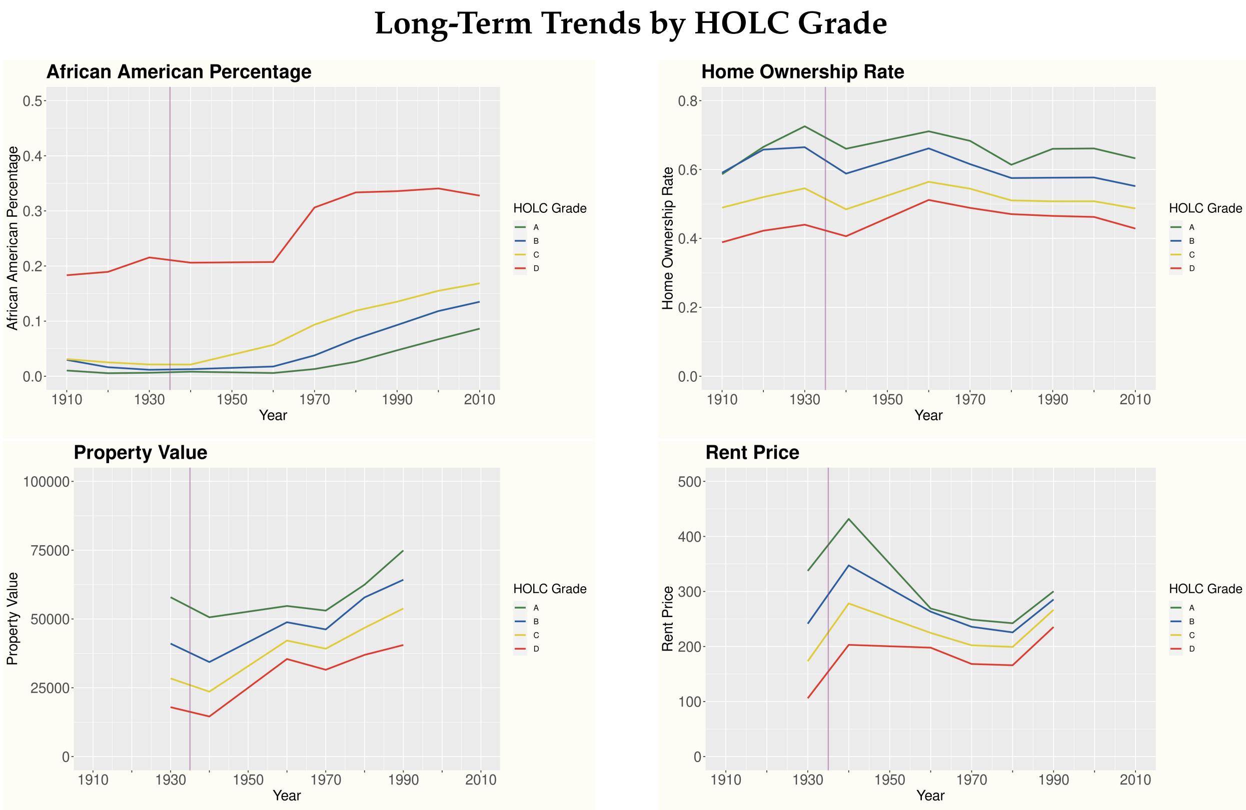


Figure 1.4 — The sample includes neighborhoods located within a digitized HOLC map. See Appendix Section 1.7.2 for definitions of Census variables in our dataset. The data sources are US full count census for 1910-1940 and NHGIS for 1960-2010.

1.3.3 CoreLogic Deeds Records

We supplement NHGIS tract-level data with sale records obtained from CoreLogic which contains transaction data collected from county assessors and deed registries, including information about sale prices, dates of sale and the geographic coordinates of the buildings. In our dataset, transactions are binned into 5-year windows according to the sale year and month. As expected, the number of sales recorded in the dataset is much higher in recent years.[32] The nature of CoreLogic data is different from data sources we have described so far. They are administrative records of realized sales, while census property values are the results of extensive surveys based on self-reports.

1.4 A Novel Estimation Strategy

We propose a new strategy to measure the short and long-term effects of the HOLC maps. Our approach does not rely on border discontinuities designs, which have been prevalent in the literature on the topic.[33] Instead, we exploit an exogenous population threshold: HOLC staff focused on cities with a population of at least 40,000 residents. Accordingly, we define cities with a population between 40,000 and 50,000 as the treated cities, while the municipalities between 30,000 and 40,000 residents are included in the control group. Figure 1.5 illustrates this threshold and highlights our definitions of treated cities in purple and control cities in orange.[34] However, a city-level analysis cannot measure grade-specific effects. Moreover, ignoring the heterogeneous effect of the four different grades would lead to empirical results that would mischaracterize the legacy of this federal intervention. Appendix Table 1.13 shows the results we obtain when we apply a simple difference-in-difference design to estimate the effect of HOLC mapping imposing homogenous effects for the four different grades. We do not detect any significant effect on homeownership rates or African American percentages, while we find weak evidence of a reduction in property values. To measure the consequences of HOLC maps, we need an empirical approach to estimate the impacts of four different interventions that share the same treatment assignment and timing.

[32] Additional details about CoreLogic's coverage of our cities of interest can be found in Appendix Table 1.12.

[33] The previous literature has mainly focused on the local effect of a lower HOLC grade using spatial regression discontinuity designs. This approach suffers from endogeneity concerns due to the non-random location of borders and non-random assignment of grades. Both Aaronson et al. (2021b) and Fishback et al. (2020) document how locations on opposite sides of HOLC borders showed differential trends in a variety of observables before the introduction of the maps. Aaronson et al. (2021b) employ propensity scores and a subset of idiosyncratic borders to address this issue.

[34] See Appendix 1.7.4 for a list of cities.

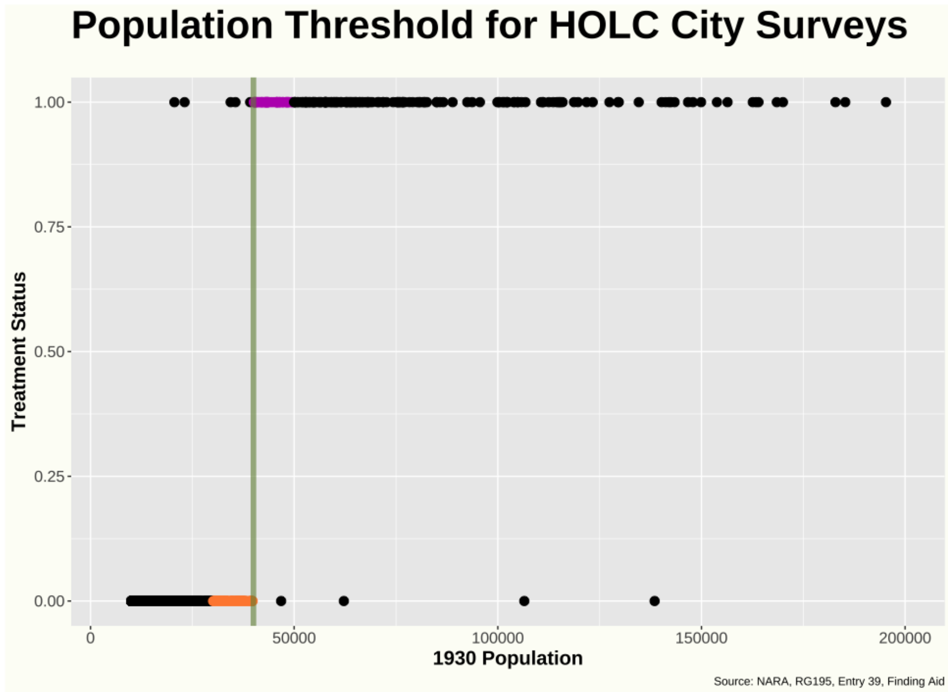


Figure 1.5 — The graph shows the treatment status of US cities according to their 1930 population. The vertical line highlights the 40,000 people threshold. Orange points identify cities in the control group, while the color purple highlights treated cities.

We develop a strategy to compare areas with a given grade in treated cities with neighborhoods that would have received the same grade if their city had been mapped. A machine learning classification algorithm assigns grades to neighborhoods in control cities, replicating HOLC assessment standards. The algorithm is trained to link observed grades from the whole set of HOLC maps to 1930 census observables aggregated at the neighborhood (hexagon) level. Using the predicted grades, we then apply a grouped difference-in-differences design to measure the causal effects of the maps' four different grades. To provide an intuition, let us restrict the analysis to two cities: a treated city, Phoenix, Arizona, and a control city, Raleigh, North Carolina. To estimate the effect of a D (red) grade, we compare the outcomes for observations geocoded in Phoenix D areas with Raleigh observations in neighborhoods (hexagons) that the HOLC would have rated D if its agents had surveyed the area.

Building the control group with a machine learning model is a matching technique alternative to synthetic control methods (Abadie and Gardeazabal, 2003; Abadie et al., 2010).

Instead of estimating a complete set of weights so that control areas could mirror treated areas, we classify neighborhoods by replicating HOLC standardized evaluations. In the same spirit of synthetic controls, we do not use any post-treatment data when designing the control group classes and the contribution of each observation to the counterfactual is explicit.[35] Moreover, the "donor pool" for each class can be easily visualized on a map. Unlike the synthetic control method, in our procedure control group units are never used in the training procedure that determines the counterfactual composition. The resulting control group will have to meet validity checks, such as the parallel trends assumption, that were not targeted during its design.[36] Hence, our approach reduces even more the possibility of manipulation in developing a synthetic counterfactual. Harnessing an institutional feature of our research setting, we defined control groups that, while being synthetic, are particularly plausible.

1.4.1 A Classification Algorithm

The success of our strategy relies on convincingly replicating HOLC evaluations using 1930 census data. In particular, we are interested in recovering a function that can credibly predict y, the HOLC grade, based on X, a set of neighborhood observables.

Since HOLC appraisers traced area boundaries and assigned grades simultaneously, our classification algorithm should replicate both outcomes. We tackle these goals by classifying a regular grid of hexagons into the four different grades.[37] This approach imitates HOLC methods and tackles the complex task of drawing grade borders in control

[35]Our approach can be thought of as a special case of the synthetic control method where weights are assigned by a classification model. In particular, for treated observations with grade j we are building a control group assigning weights with only two values, either 0 or 1. Let $\hat{g} \in \{A, B, C, D\}$ be the grade predicted by the classification model, we are assigning the 0 weight to all control observations such that $\hat{g} \neq j$ and a weight equal to 1 if $\hat{g} = j$. The resulting weighted mean will be rescaled by $n_{\hat{g}}$, the number of observations with a predicted class $\hat{g}$.

[36]The machine learning training dataset does not contain information about pre-treatment trends. It only includes 1930 census information.

[37] Hexagons, rather than triangles or squares, are well suited for our goals because they are the most circular-shaped polygon that can generate a regular grid. In particular: they reduce sampling bias, capture curved patterns more easily, reduce the projection distortion due to earth curvature and provide a better definition of neighbors because of their centroid properties. For more details see Birch et al. (2007).

cities. While the hexagon grid is useful to imitate the original borders, 1930 census data are the best nationwide data source to replicate HOLC surveys. Our dataset provides sufficient detail about the sociodemographic composition of neighborhoods and housing prices. However, we lack information about mortgages, defaults and interest rates that the agency regularly collected surveying local financial institutions.

We implement a random forest algorithm (Breiman, 2001) to classify hexagons into one of four HOLC grades. This machine learning method proves effective in dealing with the class imbalance of our classification problem[38] and outperformes other popular classification algorithms.[39] In short, the random forest is a nonparametric and nonlinear model based on decision trees. A tree is a hierarchical series of splitting rules for covariates X. In practice, the goal is to find the best binning structure of covariates X, together with the hierarchy of these splits, to predict class y. Since random forests are widely employed in recent economics literature, we will highlight only a few relevant features of the algorithm.[40]

A decision tree provides flexible binning of multiple covariates to maximize the predictive power for the outcome class y. The definition of bins is entirely data-driven, and the process flexibly takes into account interactions between covariates. The resulting bins define a link between covariates X and the predicted class $\hat{y}$ as a nonlinear multivariate function. This approach usually returns a good in-sample fit, but it often suffers from poor out-of-sample predictions due to overfitting. Bootstrapped aggregation (bagging) techniques offer a remedy. The solution is to fit several trees on bootstrapped samples of the data, thus growing a forest. Moreover, each split is determined only by m randomly selected covariates. These steps reduce the correlation between the predictions of each tree, characterizing the forest as "random" and providing reliable out-of-sample predictions. Once the algorithm is trained, the predicted class is the one most voted on by all

[38] The minority class share (Grade A) is 7.8% while the maximum one (Grade C) is 42.1%. More details can be found in Table 1.11.

[39] More details about the performance of an ordered logit model in this setting can be found in section 1.4.2.

[40] See Fuster et al. (2021) for a more detailed explanation of the random forest algorithm.

the trees in the forest. The number of variables to use at each split (m), and the fraction of observations to sample for each tree, are parameters that need tuning.[41]

1.4.2 Classification Model Results

We train the random forest algorithm with a hexagon-level dataset containing all cities mapped by the HOLC with a population below 3,000,000.[42] The dataset includes 48 different 1930 census variables; some are included at different geographical levels, bringing the total number of training variables to 163.[43] The total number of neighborhoods in our dataset is 192,016.

We assess the performance of our classification model on a test set of spatial units (hexagons) randomly drawn from the original dataset.[44] These observations were excluded from the random forest training procedure and represent an out-of-sample validation of the model performance. Table 1.3 presents a matrix comparing the observed and predicted grades in the test set (*Confusion Matrix*). The probability of correctly classifying a neighborhood (*Accuracy*) is 91.55%, while the probabilities of correct predictions conditional on observed grades (Class-specific *Sensitivities*) are above 90% for B, C, and D classes. Comparing the predicted grade distribution (*Detection Prevalence*) with the observed class frequencies (*Prevalence*) shows that our model does not alter the overall distribution of HOLC grades. Given that the identification strategy focuses on neighborhoods in cities between 30,000 and 50,000 residents, we are interested in the performance of our model in smaller cities. Appendix Table 1.14 shows the results we obtain if we restrict our test set to cities with a population below 50,000. Accuracy is still above 90% percent and performance metrics are similar overall.

[41] More details on tuning of our random forest can be found in Appendix section 1.7.3.

[42] The results in this section are robust to variations in the population threshold determining which cities are included in the training procedure. The range of variation for this threshold is between 50,000 and 7,000,000 residents.

[43] The complete list of variables is available in Appendix 1.7.3.

[44] The test dataset represents 25% of the original dataset, while 75% of the observations were used in training the model. The random sampling was stratified according to HOLC grade and city population. The results are robust to changes in the sampling procedure.

		Data			
		D	C	B	A
Prediction	D	12940	668	62	3
	C	927	18771	827	74
	B	93	792	9939	462
	A	7	36	153	2792
	Accuracy	91.55%			
	Class Sensitivity	92.65	92.62	90.51	83.82
	Prevalence	28.77	41.75	22.62	6.86
	Detection Prevalence	28.17	42.43	23.25	6.15

Notes: The matrix compares the observed and the predicted grades for a test set of observations excluded from the training procedure. The test set is a 25% random subsample of the original dataset selected with stratified sampling according to city population and HOLC grade. The level of observation is a neighborhood (hexagon). See Section 1.3.1 for details about the hexagon definition. The sample includes every hexagon in a mapped city with a 1930 population below 3,000,000 and containing at least 20 residents in 1930. The results are robust to different sample definitions in terms of population cutoffs. A predicted grade is the class predicted by the trained random forest algorithm. See Section 1.4.1 and Appendix Section 1.7.3 for details about the Random Forest training procedure. Overall Accuracy is the percentage of hexagons whose predicted grades correspond to observed ones. Class sensitivity for a grade j is the proportion of correctly predicted hexagons among the spatial units with grade j. Prevalence reports the share of each observed grade in the test set, while detection prevalence shows the distribution of predicted grades.

Table 1.3 — Random Forest Performance, Confusion Matrix

It is worth comparing our machine learning procedure to classification models traditionally used in the economics literature. Accuracy levels above 90% are a substantial performance improvement compared to what we would obtain with an ordered logit. Appendix Table 1.15 shows the performance of an ordered logit estimated on the same dataset used by the random forest. Overall accuracy reaches only 64.35% and the predicted grades severely overestimate the prevalence of C neighborhoods, while underestimating the presence of D and A neighborhoods. It should be noted that a logit-type model is more transparent than a random forest in characterizing the contribution of each variable to determine the probability of a grade. However, the improvement in predictive accuracy offered by the latter is so significant that it compensates for the loss in interpretability.

Prediction accuracy characterizes the model's precision, but it does not provide any insight into the spatial patterns of our predicted maps. In particular, the challenge is to obtain graded areas with a sufficient degree of compactness to mirror the HOLC maps. A comparison between the original HOLC map of Baltimore, Maryland, and our predicted map can be found in Figure 1.6. In general, the predicted neighborhoods are not dissimilar from the original neighborhoods in terms of shape.[45] When the classification model disagrees with HOLC evaluations, it tends to assign a different grade to whole clusters rather than to single hexagons. Examples of this behavior can be found in downtown Baltimore where the predicted grade is C (yellow) versus an original D (red), or in the northwest suburbs of the city, where an area with a B (blue) grade from HOLC is classified as A (green) by the model. The final goal of replicating HOLC grades is to draw "redlining" maps in cities between 30,000 and 40,000 people. Examples of predicted maps for control cities can be found in Figure 1.7. The model identifies areas for all four grades, returning neighborhood shapes similar to the ones observed in the original maps in larger cities.

The random forest algorithm we employ does not have any spatial constraint that would guarantee an output visually similar to HOLC maps. The results rely on a training dataset that includes several census observables at different levels of geographical definition. Figure 1.8 shows how the predicted map for New Haven, Connecticut changes when we train the classification models with datasets including different geographical levels of aggregation. The top left map is the output we obtain when each hexagon only includes information about its area, while the top right map adds city and county-level information to the dataset. These maps, while generally accurate, suffer from spatial noise and the resulting neighborhoods cannot be easily encircled into a compact shape.[46] The plausibility of the predicted maps increases when we include information about the area surrounding each hexagon. In particular, we construct averages of surrounding census observables using 500-meter and 1,000-meter radii (0.31 miles and 0.62 miles, respectively) for each hexagon. With the addition of these local averages, the classification algorithm returns

[45] The surface covered by our hexagon-level maps is slightly smaller than the area originally covered by HOLC. This is because the federal agency mapped even scarcely populated areas, while our strategy focuses on hexagons with at least 20 residents.

[46] The overall prediction accuracy of the random forest considering only neighborhood level data is 66.75%, while it rises at 75.45% when including city and county level covariates.

predicted maps with compact neighborhoods, as shown in panels (c) and (d) of Figure 1.8.[47]

1.4.3 Measuring the Effects of HOLC Maps

Our goal is to estimate the grade-specific causal effects of introducing HOLC maps, an innovative information tool for neighborhood appraisal, in the real estate market. In the early stages of the 20th century information revolution, the HOLC maps could act as a coordination device providing practical area evaluations to local financial institutions. To estimate these effects, we classify neighborhoods into four classes according to their predicted grades using our trained random forest. Then, we apply a difference-in-differences design comparing neighborhoods in treated cities, between 40,000 and 50,000 residents, with those in control cities, between 30,000 and 40,000 residents, separately for each grade. Our pre-treatment period is 1930, and 1940 is our first post-treatment period. If the empirical design assumptions are deemed credible, the estimated coefficients will capture the global effects of each grade assigned by HOLC.

In the short run, our specification is:

$$Y^{\hat{g}}_{i,h,c,t} = \alpha^{\hat{g}} D_c + \gamma^{\hat{g}} P_t + \beta^{\hat{g}} D_c P_t + \delta^{\hat{g}} X_{i,h,c} + \varepsilon^{\hat{g}}_{i,h,c,t} \tag{1.1}$$

In the equation, $Y^{\hat{g}}_{i,h,c,t}$ is the outcome for individual i, living in a neighborhood h with grade $\hat{g}$, in city c, at time t. The term D_c is a treatment dummy and P_t is a post-treatment indicator. $X_{i,h,c}$ includes neighborhood observables and information about the surrounding areas. Equation (1.1) will be estimated by group according to the grade $\hat{g}$ assigned by the trained random forest. The coefficients of interest are $\{\beta^A, \beta^B, \beta^C, \beta^D\}$. In section 1.5.1 we provide results for different specifications of equation (1.1), replacing the treatment term $\alpha^{\hat{g}} D_c$ with city fixed effects or neighborhood (hexagons) ones.

[47] A random forest trained on a dataset including neighborhoods and information about their surroundings achieves a 89.58% accuracy. If we add city and county level variables to the former dataset, accuracy increases to 91.55%.

We can extend equation (1.1) to measure medium and long-term effects. In particular, we estimate the following equation at the neighborhood level:[48]

$$Y^{\hat{g}}_{h,c,t} = \alpha^{\hat{g}} D_c + \Gamma^{\hat{g}} \bar{P}_t + \sum_{t \in T} \beta^{\hat{g}}_t D_c P_t + \delta^{\hat{g}} X_{h,c} + \varepsilon^{\hat{g}}_{h,c,t} \tag{1.2}$$

where $T = \{1930, 1940, 1960, 1965, 1970, 1975, \cdots, 2010\}$. The only new terms compared to equation (1.1) are $\bar{P}_t = (P_{1940}, P_{1960}, \cdots P_{2010})$ a vector of year dummies for all elements of T (except 1930) and its corresponding vector of coefficients $\Gamma^{\hat{g}}$. In this case, $X_{h,c}$ includes time-invariant geographic controls. Section 1.5.2 contains the results for different specifications of (1.2) where we replace the treatment term $\alpha^{\hat{g}} D_c$ with city fixed effects.

The validity of this empirical framework relies on two main assumptions. First, the maps did not affect dependent variables in control cities. Second, outcomes would have evolved in parallel in the absence of the policy. We examine the validity of these assumptions in the following section.

A limitation of our analysis is related to external validity. Our effects are estimated for cities with a population between 30,000 and 50,000 and might not be appropriate to describe the effects of the maps in American metropolises.[49] Another limitation is that our estimates rely on a first-stage classification model. The prediction errors of the random forest might attenuate, or inflate, the difference-in-differences estimates and affect their precision.[50] Ultimately, the high prediction accuracy of our trained machine learning algorithm and the soundness of the parallel trends assumption reassure us about the overall credibility of this strategy.

Previous research on this topic has focused on estimating the local effects of a lower

[48]We switch to a neighborhood level regression to estimate the long-term effects of HOLC maps. While CoreLogic data allow an individual level analysis with disaggregated deeds, NHGIS data do not. To ensure comparability between the two sources of post-1940, data we will focus on neighborhood level results. The long-term results obtained with the CoreLogic dataset are robust to using an individual-level specification.

[49] In Section 1.5.1 we show results when we expand the treatment group to include cities up to 60,000 inhabitants.

[50]Our results are robust to substituting the predicted grades with the observed ones in treated cities. Moreover, we propose additional checks of biases introduced by our classification exercise in section 1.4.4 to mitigate these concerns.

grade, such as D, compared to a higher one, e.g. C, with border regression discontinuity techniques. Such results characterize within-city local impacts, but it is not immediate to translate them in an aggregate measure of HOLC maps' effects. Instead, our approach returns the treatment effects of the four HOLC grades providing a direct description of the global effects of HOLC maps on US neighborhoods.

1.4.4 Validation of the Empirical Strategy

As noted in section 1.4.3, our difference-in-differences framework relies on the no-treatment-spillover assumption and parallel pre-trends. In our context, the no-spillover assumption means that control cities would not have been affected by the HOLC intervention because of their geographic location. Figure 1.9 shows that control cities, in blue, are scattered throughout the country, and they are not suburbs of treated cities, in red. To strengthen the assumption, we include in our analysis only control cities with a distance of at least 50km (31mi) to the nearest treated municipality.[51] While it is safe to assume that HOLC mapping did not directly affect control cities, it is harder to argue that the assignment of grade g in a certain area does not affect surrounding neighborhoods.[52] If we are worried about the spillovers of surrounding graded areas in treated cities, the coefficient β^g will combine the effect of grade g and the correlations with other local grades. In Table 1.4 we provide the grade composition of neighborhood surroundings according to their own grade and treatment status. D grades are surrounded, on average, by 62.8% red and 32.3% yellow neighborhoods. For all grades, the majority neighborhood share corresponds to the same class: for example, 76.7% of C neighbors belong to grade C. In section 1.5.1, we show that the results are robust when we include information about the local grade composition as an additional control.

[51] The results are robust to variations in this threshold between 30km (18.6mi) and 70km (43.5mi). The median distance between a control city and the closest mapped municipality is 144.7Km (89.9mi)

[52] This challenging problem is similar to estimating the effects of an exogenous shock when dealing with non-random exposure, as described by Borusyak and Hull (2020). In our setting, even if it is credible to characterize treatment assignment as random, we might think that neighborhood location will lead to non-random exposure to different grades from the surrounding areas.

	A	B	C	D
Share A	50.2%	4.9%	0.5%	0.3%
Share B	38.9%	64.0%	11.3%	4.6%
Share C	9.3%	27.4%	76.7%	32.3%
Share D	1.6%	3.6%	11.5%	62.8%

Notes: The Table reports average shares of surrounding grades according to neighborhoods grades and treatment status. Neighborhood surroundings are defined with a 1000mt. radius (0.63 miles). The sample includes neighborhoods with at least 20 residents in 1930 in cities with a population between 30,000 and 50,000. See Appendix Section 1.7.4 for a list of cities.

Table 1.4 — Shares of Local Grades, by Grade and Treatment Status

The soundness of the difference-in-differences framework hinges on the similar evolution of socioeconomic characteristics between treatment and control cities prior to the HOLC intervention. We graphically investigate the soundness of the parallel trends assumption in Figure 1.10 using data between 1910 and 1930, the decades before the HOLC intervention. The trends for African American percentage and homeownership rate evolved in parallel in C (yellow) and D (red) areas. The same is true for grade B, as can be seen in Appendix Figure 1.13. The assumption appears less valid for A (green) areas, and the results for this class, which represents approximately 3% of the sample, should be interpreted with caution.

This validity check cannot be completed for property values, one of our outcomes of interest, because the census started to record this variable only in 1930. As a partial remedy, we can investigate the trajectory of alternative socioeconomic variables. The bottom panels of Figure 1.10 compare trends for the imputed income score we built based on 1940 census information.[53] Given its definition, this variable can be interpreted as an index of socioeconomic status. The observed trends support our research designs. Additional plots investigating the trends for population density, number of children and percentage of first-generation migrants can be found in Appendix Figure 1.14. Table 1.5 contains the

[53]The Census did not record income before 1940. We impute an income score for wage-employed men aged 25-55 between 1910 and 1930 using income measures from 1940. More details can be found in Appendix section 1.7.2.

results for an analytical check of the parallel trends assumption. There are no substantial differences between treated and control cities in the evolution of demographic and economic variables between 1930 and 1920 in B, C, and D areas. These results are confirmed in Table 1.17 for changes between 1920 and 1910.

Testing 1930-1920 Trends By Treatment Status

Dependent Variable	A	B	C	D
Black	0.008 (0.006)	0.001 (0.002)	-0.001 (0.003)	-0.003 (0.010)
Home Owner	-0.080** (0.034)	-0.017 (0.019)	-0.006 (0.015)	-0.021 (0.014)
Income Score	-0.023 (0.026)	-0.019 (0.012)	-0.006 (0.008)	0.021* (0.012)
Education Score	-2.055 (1.413)	-0.478 (0.554)	0.173 (0.259)	0.074 (0.289)
First Gen. Immigrant	0.018 (0.016)	0.004 (0.007)	0.014** (0.007)	-0.012 (0.012)
Number of Children	0.014 (0.021)	0.004 (0.011)	0.011 (0.008)	0.001 (0.010)

Notes: The Table reports the coefficients from a set of regressions where the dependent variable is the 1930-1920 change in the variable reported in the left column, and the independent variable is an indicator for treatment status. See Appendix Section 1.7.2 for definitions of Census variables in our dataset. The level of observation is a neighborhood (hexagon). The sample includes every hexagon in cities with a 1930 population between 30,000 and 50,000 and at least 20 residents in 1930. See Appendix Section 1.7.4 for a list of cities. Standard errors, in parentheses, are clustered at the city-year level. Significance: * 0.10 ** 0.05 *** 0.01

Table 1.5 — Testing Differences in 1930-1920 Trends by Treatment Status

The classification algorithm we employ to replicate HOLC grades returns predicted maps that we can compare with the original ones in terms of socioeconomic characteristics. Table 1.6 compares averages according to observed and predicted grades, showing that the predicted maps do not alter the original socioeconomic composition of C and D neighborhoods. Moreover, we might be worried about the type of bias introduced in the analysis by spatial units receiving a "wrong" grade.[54] The averages in Appendix Table 1.16 show that even when the model assigns a neighborhood to the wrong class, we are not introducing significant sources of bias.

[54]Spatial units with different observed and predicted grades represent approximately 12% of our sample in treated cities.

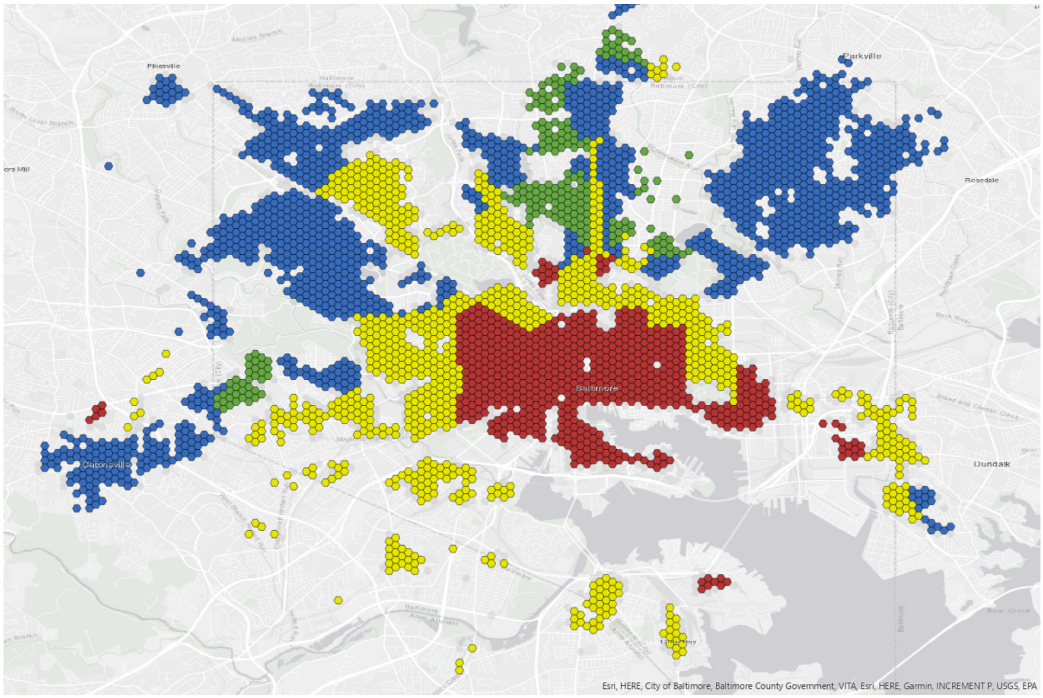
	Grade			
	C		D	
	HOLC	Predicted	HOLC	Predicted
Black	0.02	0.02	0.17	0.21
	(0.10)	(0.08)	(0.31)	(0.33)
Home Owner	0.53	0.52	0.41	0.40
	(0.23)	(0.23)	(0.24)	(0.23)
Property Value	6,843	6,869	5,253	4,729
	(3,841)	(3,861)	(3,632)	(3,267)
Rent	52	54	42	39
	(112)	(116)	(98)	(97)
Income Score	7.13	7.13	6.91	6.86
	(0.18)	(0.17)	(0.28)	(0.28)
First Gen Immigrant	0.22	0.23	0.24	0.23
	(0.17)	(0.18)	(0.22)	(0.23)
Unemployed, Men	0.10	0.10	0.13	0.13
	(0.09)	(0.08)	(0.10)	(0.10)
Owns a Radio	0.58	0.58	0.38	0.33
	(0.22)	(0.21)	(0.23)	(0.21)

Notes: The Table reports averages of 1930 census variables according to different classifications. The first two columns compare means between hexagons classified as *C* by the HOLC with those classified as C by our random forest algorithm. The third and fourth columns do the same for grade *D*. The level of observation is a neighborhood (hexagon). The sample includes all the hexagons intersecting a HOLC neighborhood digitized by Nelson et al. (2021) in 202 maps. Standard deviations are reported in parentheses.

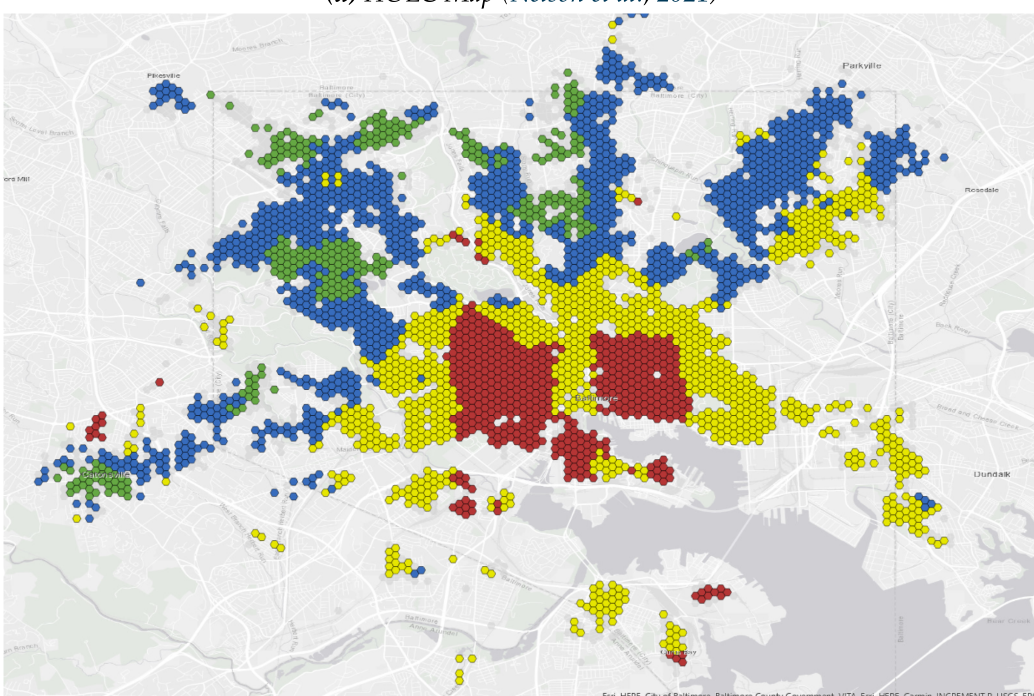
Table 1.6 — 1930 Descriptive Statistics According to HOLC and Predicted Grades

Another assumption implicit in our empirical approach is that HOLC practices did not change between different cities. In particular, a predictive model trained with US metropolises might not accurately replicate HOLC grades in smaller cities, the ultimate goal of our prediction exercise. Appendix Figure 1.15 shows that the accuracy level of our random forest algorithm is robust to different training datasets according to the population of cities included in the training set. When we restrict our attention to predicting grades for neighborhoods in cities below 50,000 residents, overall accuracy is still above 90%. We interpret these results as evidence of the high degree of standardization of HOLC grading procedures, making our predictive model a reliable source for HOLC evaluations in the smaller control cities.

Comparison of HOLC and Predicted Maps for Baltimore, MD



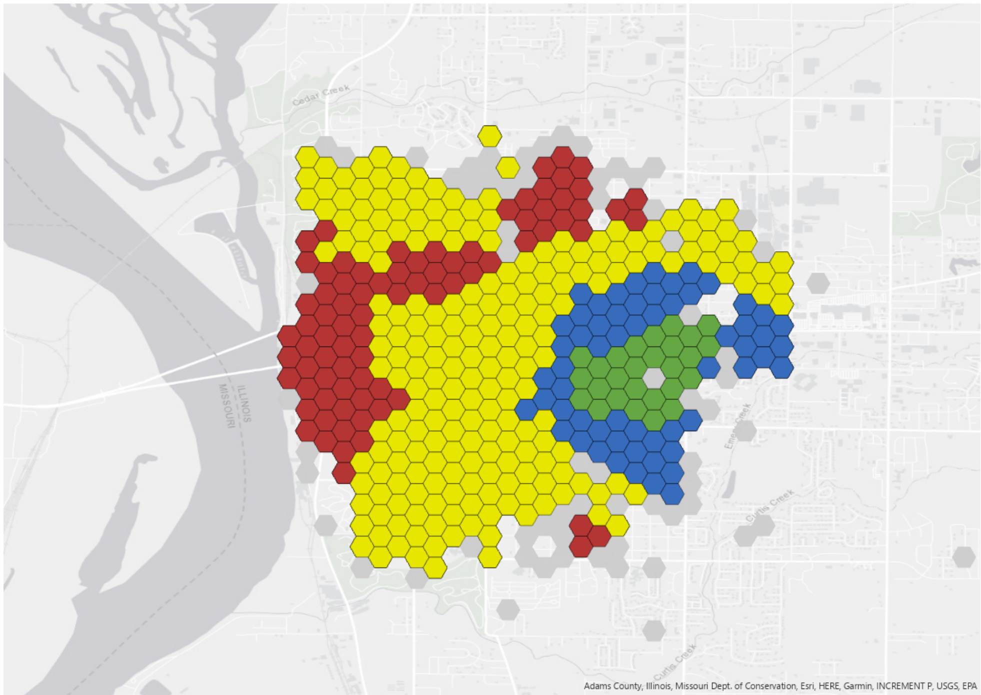
***(a)** HOLC Map (Nelson et al., 2021)*



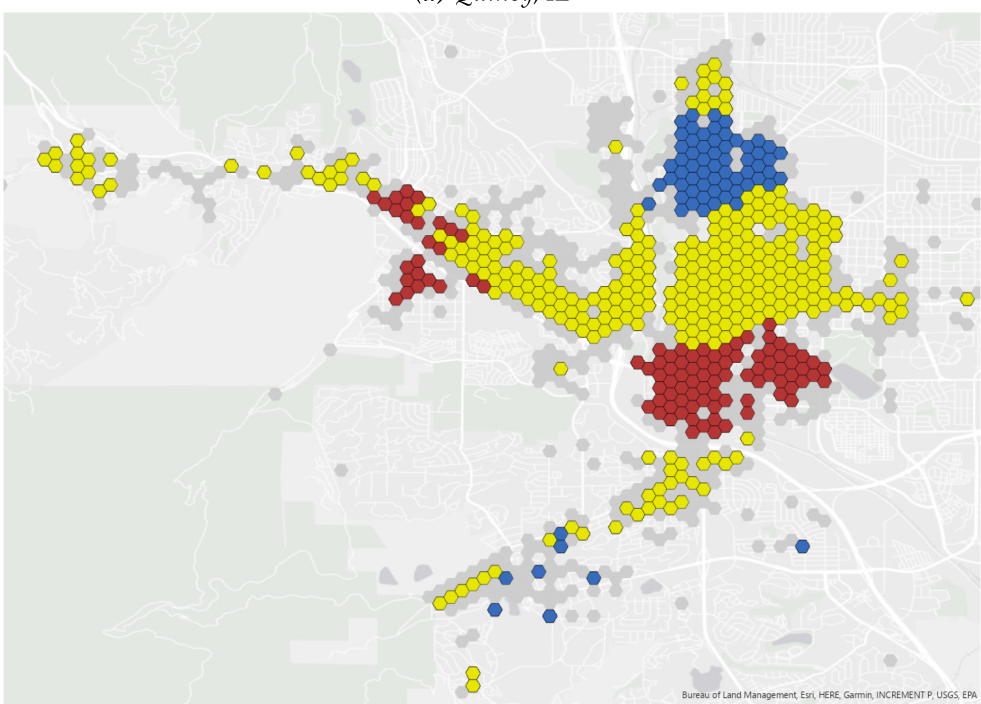
***(b)** Predicted Map, Random Forest Algorithm*

Figure 1.6 — The Figure compares the digitized version of the HOLC maps for Baltimore, MD (Nelson et al., 2021) with the hexagon-level map we predict with the trained random forest algorithm. The correspondence between colors and grades is: Green=A, Blue= B, Yellow=C, Red=D. Grey hexagons have less than 20 residents in 1930 and are excluded from the prediction exercise. All the maps are north-oriented.

Predicted Maps in Control Cities



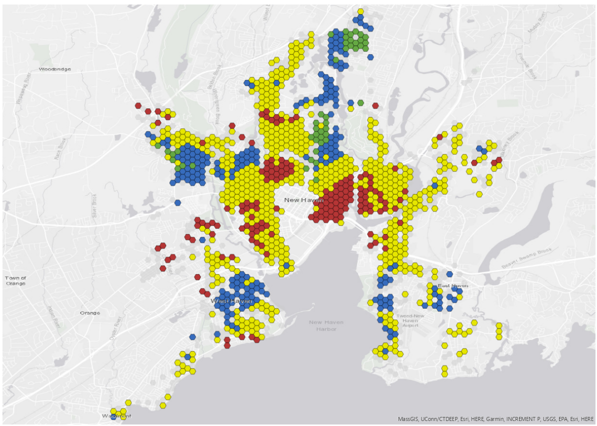
(a) *Quincy, IL*



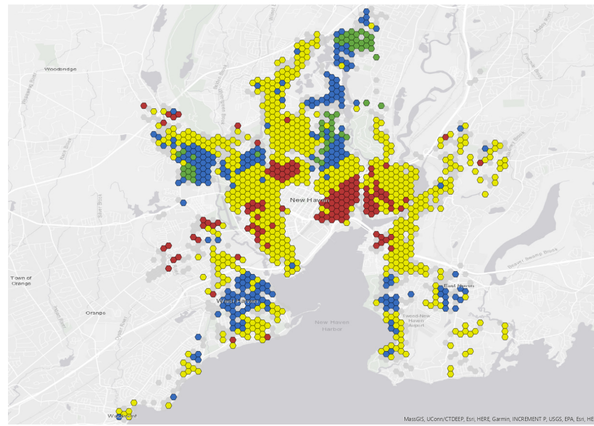
(b) *Colorado Springs, CO*

Figure 1.7 — The Figure compares the hexagon-level maps predicted with the trained random forest algorithm for Quincy, IL and Colorado Springs, CO. The correspondence between colors and grades is: Green=A, Blue= B, Yellow=C, Red=D. Grey hexagons have less than 20 residents in 1930 and are excluded from the prediction exercise. All the maps are north-oriented.

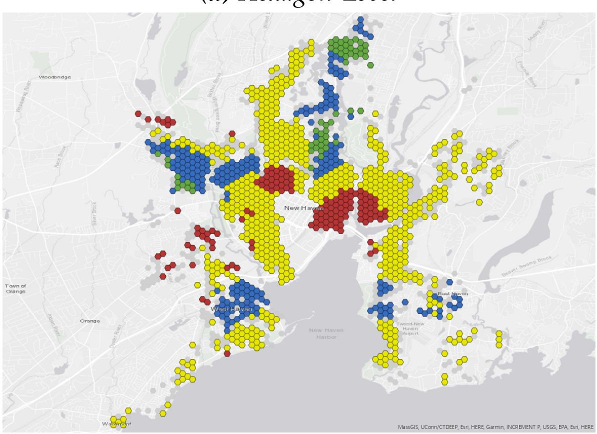
Comparison of Predicted Maps with Different Training Datasets



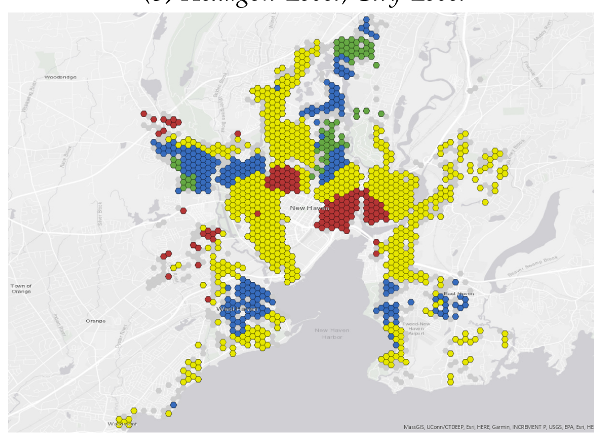
***(a)** Hexagon-Level*



***(b)** Hexagon-Level, City-Level*



***(c)** Hexagon-Level and Local Level*



***(d)** Hexagon-Level, Local-Level, City-Level*

Figure 1.8 — The maps show neighborhoods (hexagons) for New Haven, CT. The colors represent the grade predicted by the random forest algorithm. The correspondence between colors and grades is: Green=A, Blue= B, Yellow=C, Red=D. Grey hexagons have less than 20 residents in 1930 and are excluded from the prediction exercise. Different panels show predicted grades for random forests trained on four different sets of variables. The training datasets differ in terms of their levels of geographical aggregation, but not because of the variables included. The top-left panel shows predicted grades when only hexagon-level variables are included. The top-right panel adds city-level variables. The bottom-left panel replaces city-level variables with local-level information about the surrounding area. The surrounding area includes any hexagon whose centroids is within a 500mt. or 1000mt. radius. The bottom-right panel shows the predicted grades when we include all the previously mentioned variables. See Section 1.4.1 for details about the training of our random forest algorithm. All the maps are north-oriented.

Locations of Treatment and Control Cities

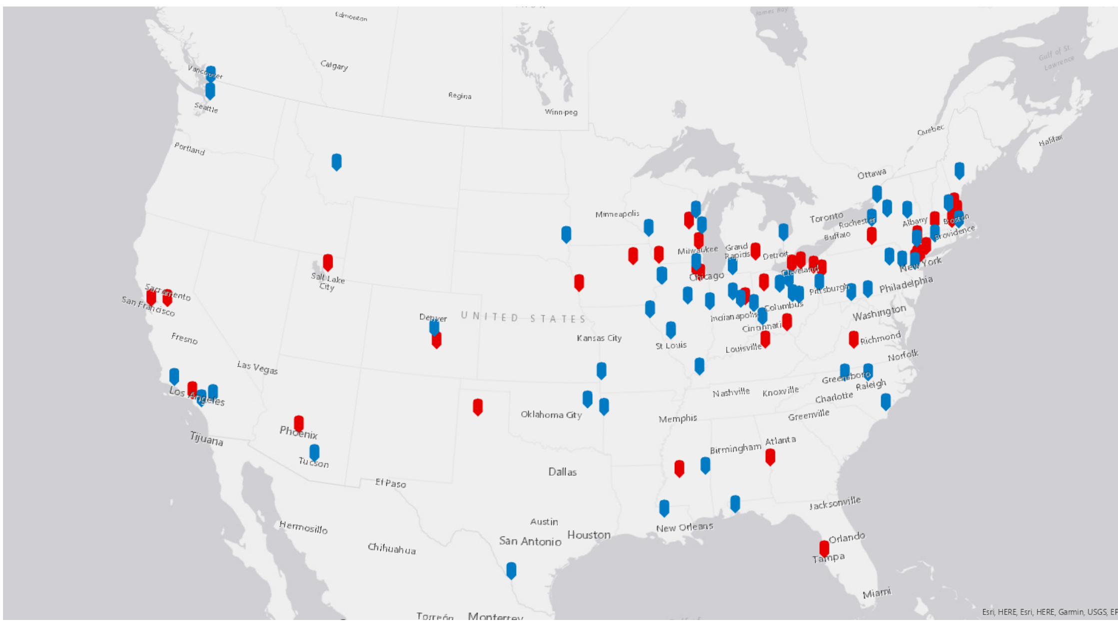


Figure 1.9 — The Figure shows the location of cities contained in the control and treatment groups. Control group cities are labeled in blue, while red pins are used for treatment group cities.

Observable Pre-Trends, Grades C and D

Figure 1.10 — The Figure shows pre-trends for selected variables for C and D grades. The point estimates are averages of hexagon-level observations. The bars show the respective standard errors of each mean. The sample includes hexagons in cities with a 1930 population between 30,000 and 50,000, with at least 20 residents in 1930. The vertical line highlights 1930, the last pre-treatment decade. See Appendix Section 1.7.2 for definitions of census variables in our dataset.

1.5 The Effects of HOLC Maps

1.5.1 Short-Term Results

We start by estimating equation (1.1) separately for each grade. The coefficients of interest reported in the tables of this section are $\left\{\beta^A, \beta^B, \beta^C, \beta^D\right\}$. The dataset includes individual-level observations from 1930, the pre-intervention period, and 1940, the post-intervention one. Standard errors are clustered at the city-year level.[55]

Table 1.7 reports the results for local homeownership rates. The results from a simple difference-in-differences design are reported in column (1). Replacing the treatment indicator with a city fixed effect does not substantially alter the coefficients but considerably increases the estimates' precision. Instead, we do not gain additional precision if we substitute a city-level fixed effect with a neighborhood (hexagon) fixed effect as in column (3). The last column reports our preferred specification where we combine a city fixed effect with neighborhood level sociodemographic controls. We find a 2.4 percentage points decrease in the percentage of homeowners in D (red) zones in 1940, shortly after the introduction of the maps. A weaker effect can be detected in C (yellow) areas, while we find no effects in B (blue) neighborhoods. On the contrary, we find a 4.5 percentage point increase in the best-rated areas (A, green), but the caveats we mentioned in section 1.4.4 apply in this case.

In terms of African American percentage, we find a 1.8 percentage point increase in the lowest-rated areas (D, red) a 9.4% increase with respect to the baseline period, as reported in Table 1.8. We do not find any other effect of this policy in other areas, given the near absence of Black Americans in A, B and C neighborhoods. The results for the full set

[55] In our research design, the treatment is assigned at the city level. Since in our data different periods are separated by 10-year gaps, we do not allow for within city serial correlation of standard errors. This clustering choice does not address unobserved, within-city, serially correlated shocks over a ten-year time span. One more threat is that hexagon-specific error components could be serially correlated across decades. Note that our results are robust to the inclusion of city fixed effects, which will absorb constant city-level error components, or neighborhood (hexagon) fixed effects. We report results with standard errors clustered at the city level, our most conservative option, in Appendix Table 1.25. The median ratio between city-year clustered standard errors and city clustered standard errors is 0.702, the average one is 0.782.

	Dependent variable: Homeownership Rates			
	(1)	(2)	(3)	(4)
DiD_A	0.032	0.047***	0.031**	0.045***
$\bar{Y}^A = 0.61$	(0.093)	(0.016)	(0.014)	(0.015)
DiD_B	0.002	0.003	0.007	-0.002
$\bar{Y}^B = 0.62$	(0.027)	(0.009)	(0.009)	(0.009)
DiD_C	-0.010	-0.012	-0.009	-0.017**
$\bar{Y}^C = 0.49$	(0.031)	(0.007)	(0.007)	(0.007)
DiD_D	-0.022	-0.017**	-0.015**	-0.024***
$\bar{Y}^D = 0.39$	(0.044)	(0.008)	(0.008)	(0.009)
City Fixed. Eff.		X		X
Hexagon Fixed Eff.			X	
Hexagon Controls				X

Notes: The Table reports difference-in-differences coefficients obtained estimating equation (1.1) by grade. Each row contains the DiD coefficients for a given grade. $\bar{Y}^J$ is the average for the outcome of interest, homeownership rates, in the pre-treatment period (1930) for grade J. The first column report the DiD coefficients resulting from a simple DiD framework. The second one replaces the indicator for treatment, which is assigned at the city level, with a city fixed effect, while the third replaces it with a neighborhood (hexagon) fixed effect. Column 4 reports the DiD coefficients when we add geographic and demographic controls at the hexagon level to the specification from the second column. The list of controls includes geographic coordinates, a scaled measure of distance from the city center, spatial unit's population density, imputed income score and family size. The regressions are estimated with individual-level observations. The sample includes individuals with valid geocodes in cities with a 1930 population between 30,000 and 50,000, living in hexagons with at least 20 residents in 1930. See Appendix Section 1.7.3 for a list of cities. The number of observations according to each grade, N^J, are: $N^A = 137,144, N^B = 979,145, N^C = 3,116,521, N^D = 1,195,213$. Standard errors, in parentheses, are clustered at the city-year level. Significance: * 0.10 ** 0.05 *** 0.01

Table 1.7 — Short-Term Difference-in-Differences Results, by Grade

of specifications for this outcome can be found in Appendix Table 1.18. Table 1.8 also shows the estimated coefficients for property values. The assignment of C (yellow) and D (red) grades caused sizable reductions in property prices. While the reduction in property prices in red areas confirms the popular narrative for "redlining", the negative effect for C neighborhoods is more surprising, and it was first documented in Aaronson et al. (2021b). Our empirical design does not find any significant effect in B areas.[56] Appendix Table

[56]The results for the full set of specifications for property values can be found in Appendix Table 1.19. We are hesitant in interpreting the estimated coefficient β^A as the causal effect of grade A. The significantly smaller sample size, paired with weak evidence of parallel pre-trend for this class, invite caution when considering the results for this grade.

1.20 confirms the negative effects in C and D areas when we apply a log transformation to property values, although the estimates are less precise. The discrepancy between the linear and logarithmic specifications suggests the presence of heterogenous effects across the property value distribution. Since the logarithmic form reduces the contribution of higher prices, it seems that HOLC maps in 1940 had a stronger impact on more expensive houses in C and D areas relatively to lower-value ones. The patterns we find for property values do not translate to rent prices. The last column of Table 1.8 shows that the HOLC maps did not affect rental prices in 1940 for our cities of interest.

Robustness

All the short-term results are confirmed if we replicate our estimates using a neighborhood (hexagon) level dataset instead of an individual one. The coefficients and their standard errors can be found in Tables 1.21 to 1.24. These alternative specifications are estimated on a two-period panel of neighborhoods in our cities of interests. The stability of the results across the individual and the neighborhood levels datasets mitigates the concern that the short-term results might be driven by strong changes in cohort composition between 1930 and 1940.

In the results we discussed so far, observations were grouped according to the grade predicted by the random forest algorithm. In Table 1.26 we show that the results are robust when we replace predicted grades with observed HOLC classes for treated observations. Another robustness check is presented in Table 1.27 where we confirm that the estimated coefficients do not change if we extend the treatment group to cities up to 60,000 residents according to the 1930 census.[57]

Our strategy might be capturing structural differences in the evolution of smaller versus bigger cities between 1930 and 1940. Table 1.28 shows that we do not find meaningful effects if we focus on placebo outcomes such as female percentage, number of children or male unemployment rate. As an additional check, we replicate our procedure with a

[57]Similar results can be obtained by changing the treatment-group population limit to 70,000, 80,000 and 100,000.

	Dependent Variables		
	African American Percentage	Property Values	Rent Prices
DiD_A	-0.005	1,046***	39.3***
	(0.006)	(378)	(12.0)
DiD_B	0.006*	-106	8.8
	(0.004)	(170)	(12.1)
DiD_C	0.001	-502***	5.9
	(0.001)	(403)	(9.6)
DiD_D	0.018***	-302**	3.1
	(0.004)	(153)	(8.1)
$\bar{Y}^A$	0.008	9,305	107.4
$\bar{Y}^B$	0.011	6,836	57.0
$\bar{Y}^C$	0.020	5,274	43.4
$\bar{Y}^D$	0.190	3,500	29.0

Notes: The Table reports difference-in-differences coefficients obtained estimating equation (1.1) by grade for three different outcomes. Each row contains the DiD coefficients for a given grade. $\bar{Y}^J$ is the average for the outcome of interest in the pre-treatment period (1930) for grade J. The Table shows the DiD coefficients resulting from a DiD framework with a city fixed effect and geographic and demographic controls at the hexagon level. The regression specification is analogous to the one in column (4) of Table 1.7. The list of controls includes geographic coordinates, a scaled measure of distance from the city center, spatial unit's population density, imputed income score and family size. The regressions are estimated with individual-level observations. The sample includes individuals with valid geocodes in cities with a 1930 population between 30,000 and 50,000, living in hexagons with at least 20 residents in 1930. See Appendix Section 1.7.4 for a list of cities. Standard errors, in parentheses, are clustered at the city-year level. Significance: * 0.10 ** 0.05 *** 0.01

Table 1.8 — Short-Term Difference-in-Differences Results, by Grade

placebo population threshold set at 60,000 people. This new threshold defines a new treatment group (cities between 60,000 and 70,000 residents) and a new control group (cities between 50,000 and 60,000 residents). Our research design should not replicate our main results with the new thresholds since HOLC practices did not differ between these two new sets of cities. Table 1.29 contains the results of this exercise. We do not find any relevant and significant effect for homeownership rates, African American percentage, and property values.[58]

In section 1.4.4 we discussed how treatment spillovers from surrounding areas could be a threat to the validity of our analysis. As a first step to address this concern, Table 1.30 shows that our results are robust when we control for local grade composition. In particular, we include the prevalence of the four grades in a 1000mt (0.63miles) radius.

1.5.2 Long-Term Results

At the time of writing, full individual census data are not available starting in 1950. To investigate the effects of HOLC maps in the second half of the twentieth century, we must employ alternative data sources. As mentioned in section 1.3.2, we use tract-level NHGIS data between 1960 and 2010. Unfortunately, this data source does not provide sufficient coverage for our cities of interest in 1950, so we drop this decade in the analysis.[59] We estimate the model described in equation (1.2) separately for each grade with neighborhoods (hexagons) as the unit of observations. Standard errors are clustered at the city-year level. Since census tracts are always bigger than hexagon neighborhoods in our cities of interest, the geographical variation underlying these estimates is much smaller than the variation we exploited in previous estimates.

Figure 1.11 shows the estimated DiD coefficients for homeownership rates and African American shares in C and D areas. We find reductions in homeownership rates between 4.4 and 5.6 percentage points in D (red) neighborhoods, while no statistically significant effects can be found in C (yellow) ones. The increase in the local shares of Black Americans in D (red) areas we found in 1940 is confirmed in later years, as depicted in Figure 1.11. We also detect similar significant increases in C (yellow) zones. While the estimated effects in red areas are coherent with the short term results, the ones in yellow areas are more surprising. They show an increase between 5.9 and 7.9 percentage points in the percentage

[58]The positive effect on property values in D areas has the opposite sign of what we find in our main results.

[59]We plan to extend the current analysis to 1950 when full-count Census data for that year becomes available in April 2022.

of African Americans in C areas starting in 1980. The complete results for these two outcomes can be found in Table 1.9. Because of their high level of geographical aggregation, NHGIS data do not provide enough information to characterize the long term effects of HOLC maps on property values. To make up for the lack of precision in these estimates, we include an additional data source in our long-term analysis of property values.

	Dependent variable:			
	Homeownership Rates		African American Percentage	
	C	D	C	D
DiD_{60}	-0.003	-0.045**	0.015	0.080***
	(0.015)	(0.019)	(0.017)	(0.027)
DiD_{70}	-0.019	-0.047**	0.036**	0.077***
	(0.014)	(0.019)	(0.015)	(0.023)
DiD_{80}	-0.015	-0.046**	0.059***	0.098***
	(0.014)	(0.020)	(0.014)	(0.025)
DiD_{90}	-0.010	-0.056***	0.072***	0.089***
	(0.013)	(0.018)	(0.016)	(0.024)
DiD_{00}	-0.020	-0.054***	0.079***	0.082***
	(0.012)	(0.017)	(0.016)	(0.025)
DiD_{10}	-0.026**	-0.044**	0.078***	0.068***
	(0.013)	(0.019)	(0.017)	(0.025)
N	104,887	41,328	104,899	41,332
R^2	0.225	0.376	0.491	0.580

Notes: The Table reports Difference-in-Differences coefficients obtained estimating equation (1.2) by grade. Each row contains the DiD coefficients for a given grade in the corresponding year. In the reported specification, we replace the indicator for treatment, which is assigned at the city level, with a city fixed effect. The list of controls includes geographic coordinates and their squares together with state-specific linear time trends. The regressions are estimated with neighborhood (hexagon) level observations. Observations are weighted by log-transformed 1930 neighborhood population. The sample includes neighborhoods in cities with a 1930 population between 30,000 and 50,000, with at least 20 residents in 1930. See Appendix Section 1.7.4 for a list of cities. The data source for post-1940 outcomes is NHGIS; see Section 1.3.2 for details. Standard errors, in parentheses, are clustered at the city-year level. Significance:* 0.10 ** 0.05 *** 0.01

Table 1.9 — Long-Term Difference-in-Differences Results, by Grade. Census Data

We turn to an alternative, more granular, source of information: the CoreLogic deeds. This additional dataset allows us to assess the impact of HOLC maps on real estate transactions between 1965 and 2005.[60] We estimate equation (1.2) with neighborhood level

[60] The choice of this time span is based on the coverage of CoreLogic deeds for our cities of interest.

observations grouped in time bins with a 5-year frequency, so that the results are directly comparable with estimates obtained with NHGIS data. Each individual transaction is assigned a neighborhood, and hence a HOLC grade, using their geographic coordinates, available in CoreLogic. As in previous specifications, standard errors are clustered at the city-year level. Figure 1.12 shows the resulting coefficients and 95% confidence intervals. We find negative causal effects of HOLC maps on house values between 1965 and 1980 in D (red) and C (yellow) areas. The results for D neighborhoods describe a somewhat steady convergence of property values between treated and control cities in neighborhoods classified as D by the random forest algorithm. Statistically significant effects cannot be detected starting in mid 1980s. Table 1.10 shows the results for the four different HOLC grades.[61] As we mentioned in section 1.2, between 1974 and 1977, three critical legislative measures were introduced with the primary goal of counteracting redlining in the mortgage market. Our results suggest that the combined effects of the Equal Credit Opportunity Act, the Home Mortgage Disclosure Act and the Community Reinvestment Act might have been sufficient to offset persistent effects of HOLC maps on property prices.

In particular, the dataset does not provide enough transactions to obtain reliable estimates prior to 1965. Additional details about CoreLogic's coverage of our cities of interest can be found in Appendix Table 1.12. We stop the analysis in 2005 to avoid including the effects of the subprime mortgage crisis of the early 2000s.

[61]The results are robust to a log-transformation of the outcome, as it is shown in Appendix Table 1.31 and Appendix Figure 1.18.

	Dependent variable: Property Values			
	Grade			
	A	B	C	D
DiD_{65}	.	292	-12,360***	-13,669***
	(.)	(3,781)	(2,689)	(1,976)
DiD_{70}	-10,514	-373	-16,607***	-17,601***
	(8,839)	(4,510)	(3,547)	(2,790)
DiD_{75}	-7,939	-1,259	-13,155***	-14,949***
	(8,669)	(4,449)	(3,087)	(2,804)
DiD_{80}	-8,876	-6,208	-21,296***	-11,320**
	(9,052)	(5,168)	(6,009)	(4,708)
DiD_{85}	18,516**	12,686	6,112	-6,907
	(9,435)	(10,943)	(11,426)	(7,885)
DiD_{90}	15,363	9,267	2,030	-5,321
	(10,427)	(7,779)	(8,101)	(8,057)
DiD_{95}	-1,621	-904	-7,065	-9,326*
	(11,155)	(6,663)	(5,933)	(4,928)
DiD_{00}	6,795	3,297	2,978	-3,095
	(14,086)	(11,277)	(9,409)	(7,059)
DiD_{05}	20,774	-5,353	5,457	-342
	(17,199)	(12,564)	(11,622)	(9,584)
N	5,399	40,267	97,015	28,421
R^2	0.150	0.132	0.169	0.266

Notes: The Table reports Difference-in-Differences coefficients obtained estimating equation (1.2) by grade. Each row contains the DiD coefficients for a given grade in the corresponding year. The regressions are estimated with neighborhood-level observations. The sample includes neighborhoods with at least 20 residents in 1930 in cities with a population between 30,000 and 50,000. See Appendix Section 1.7.4 for a list of cities. The data source for post-1940 outcomes is CoreLogic, see Section 1.3.3 for details. The outcome variable is adjusted with CPI to 1980 dollars. Standard errors, in parentheses, are clustered at the city-year level. Significance: * 0.10 ** 0.05 *** 0.01

Table 1.10 — Long-Term Difference-in-Differences Results, by Grade. CoreLogic

Long-Term Difference-in-Differences Results, C and D Grades. Census Data

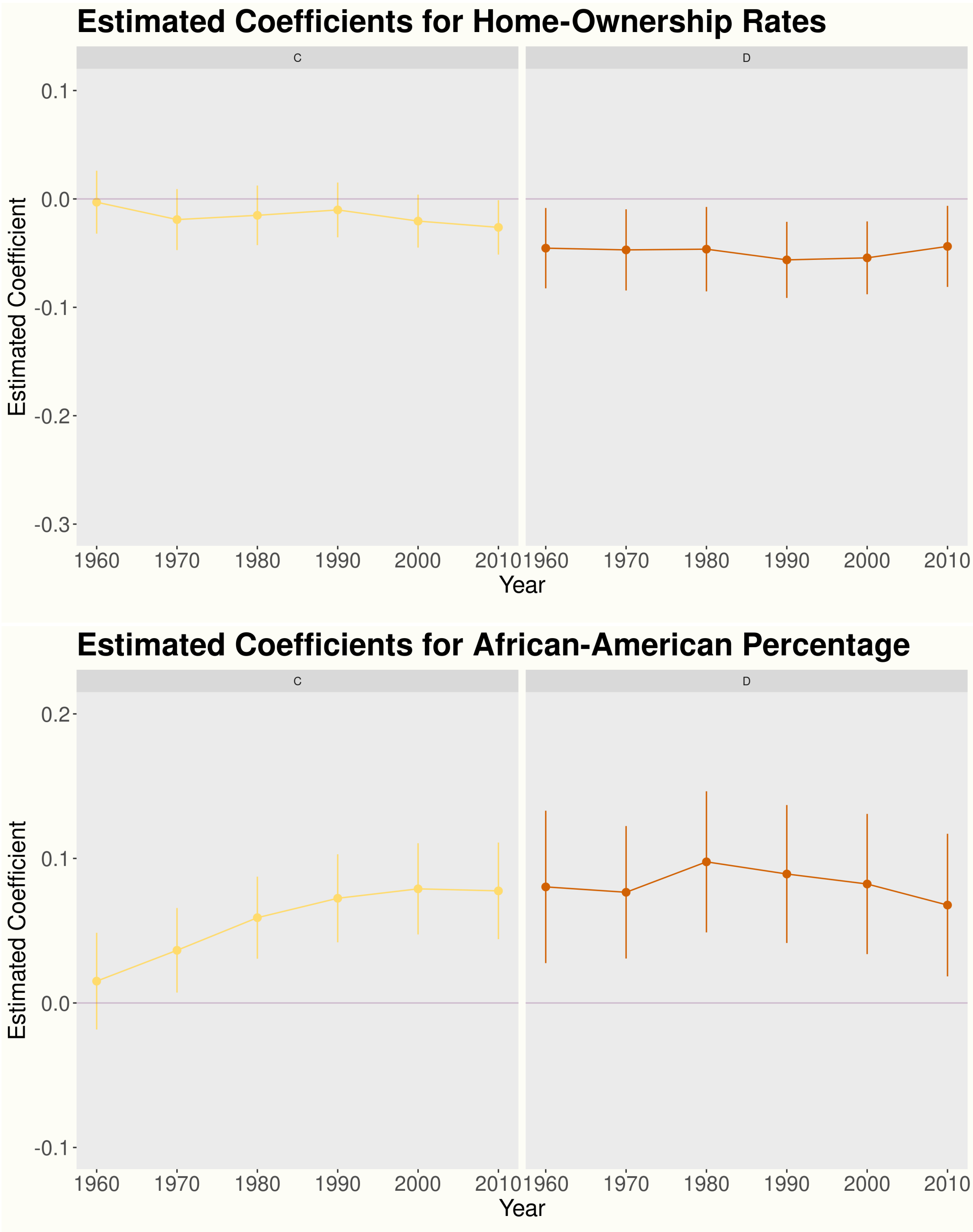


Figure 1.11 — The Figure shows the estimated coefficients for regression (1.2) and their 95% confidence intervals for homeownership rates and African American percentage. The Figure includes the results for grades C and D. The coefficients and standard errors are the ones reported in Table 1.9. See the Notes of Table 1.9 for estimation details.

Long-Term Difference-in-Differences Results, C and D Grades. CoreLogic Data

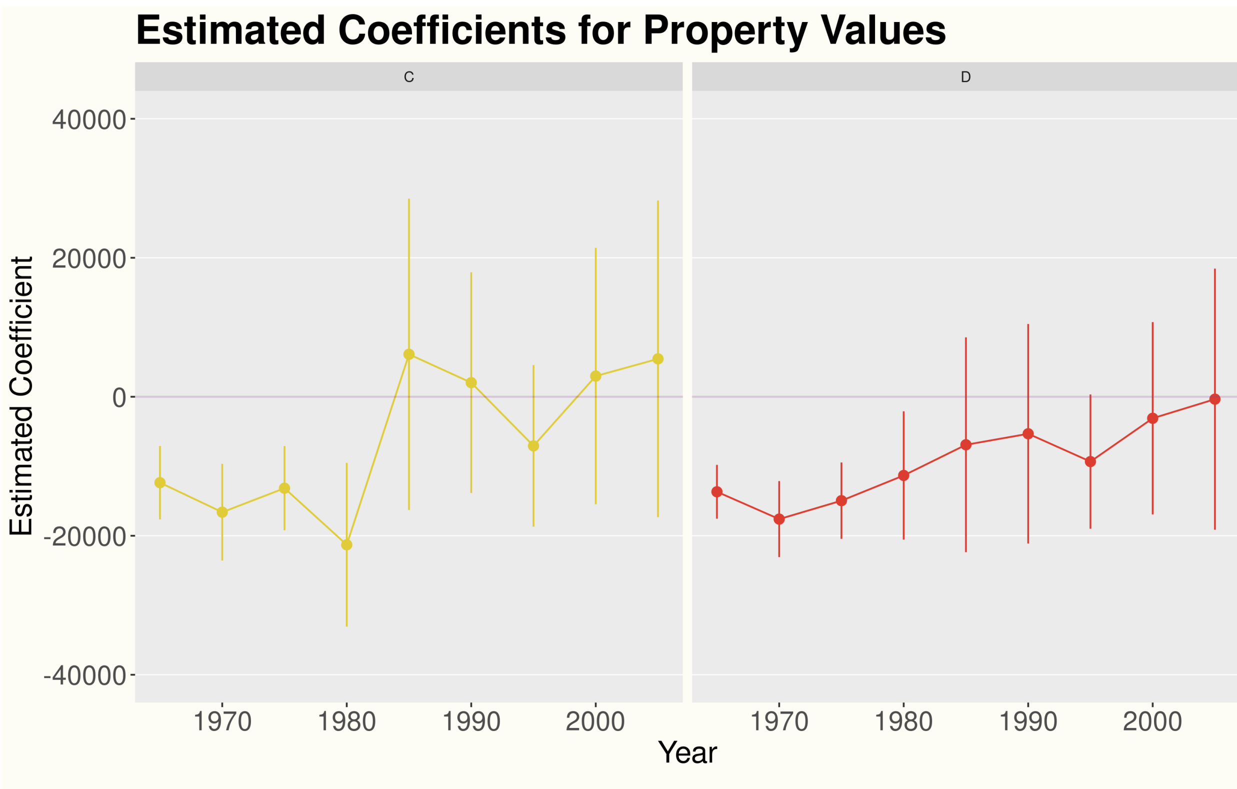


Figure 1.12 — The Figure shows the estimated coefficients for regression (1.2) and their 95% confidence intervals for property values. The Figure includes the results for grades C and D. The coefficients and standard errors are the ones reported in Table 1.10. See the Notes of Table 1.10 for estimation details.

1.6 Conclusion

In the second half of the 1930s, a federal agency undertook an unprecedented survey of the demographic and housing conditions of US neighborhoods in more than 200 cities. Its goal was to provide unified standards to assess real estate properties and stabilize a market that had just begun to recover from the Great Depression. The initiative was a data-driven effort based on the most advanced theories of urban development of the time, and its resulting maps were a data analytics tool in high demand among real estate professionals. Less than a hundred years later, the Home Owners' Loan Corporation maps have become a symbol of structural racism in the popular press and the political debate. Today's negative judgments of HOLC practices are based on non-discrimination principles that have guided US public institutions since the civil rights movement. Such condemnations are backed by historical evidence and are valid independent of quantitative estimates of causal effects. At the same time, measuring the consequences of the HOLC maps is an interesting exercise to understand the role of public institutions in coordinating and standardizing individual discriminatory behaviors.

The main challenge in estimating the causal effects of different HOLC grades is that the agency's personnel traced neighborhood borders and assigned evaluations with precision. Different HOLC grades within a city closely mirror socioeconomic trends we can observe in the census data. Instead of relying on spatial discontinuity designs, we take advantage of an exogenous threshold determining which cities the agency surveyed. Since we are interested in estimating the effects of different grades, we compare neighborhoods evaluated by HOLC with analogous neighborhoods in control cities. To classify neighborhoods in control cities, we train a random forest algorithm to replicate HOLC grades. Our spatial classification model has an out-of-sample accuracy of more than 90% and returns predicted maps that are credible replicas of HOLC maps. Using the predicted grades, we then estimate the HOLC grades' short- and long-term effects with a grouped difference-in-differences design.

We find that this government intervention had adverse effects in areas that received the lowest grade. In the short-term, we find a 2.4 percentage points reduction in homeownership rates, a sizable reduction in property prices, together with a 1.8 percentage points increase in the percentage of African American residents. No effects on rent prices is found. We also find a 9.5% decrease in property values in C neighborhoods compared to the 1930 values. This consequence of "yellow-lining" is rarely discussed and was first highlighted by Aaronson et al. (2021b).

We have evidence that the negative effects of D grades in terms of homeownership rates and percentages of African American residents have persisted until the present, but we sometimes lack precision since the data are available only at the census tract level. For property prices, we exploit a more granular data source – CoreLogic deeds – to estimate the long-term evolution of the causal effects. We find significant negative effects on property prices in C and D neighborhoods until the early 1980s. This result differs from that of Aaronson et al. (2021b), who find significant effects on property prices up until 2010, using a different set of cities and an alternative identification strategy. In our case, the effects of the maps can no longer be detected in the decades following the introduction of legislation targeting residential redlining.

We have analyzed a government policy that institutionalized discrimination by standardizing appraisal standards. Given that discriminatory practices in the housing market were widespread at the time, it is not obvious that a graphical representation of mortgage risk could have affected homeownership rates or property prices. Our results show that an organization's acceptance and reproduction of discriminatory practices can have an economic effect. The HOLC maps could have influenced discrimination in the housing market via at least two mechanisms that are not mutually exclusive: first, replacing heterogenous individual biases with one homogeneous set of biases; and second, solving an information asymmetry by circulating an information tool useful to discriminate between different neighborhoods. While our empirical strategy cannot differentiate between these two mechanisms, future research could attempt to disentangle the roles of bias standardization and information provision to provide a more nuanced understanding of the consequences of institutional discrimination.

1.7 Appendix

1.7.1 Geocoding Procedure

Address Cleaning

We clean addresses in decades between 1910 and 1940 following the procedure outlined in Logan and Zhang (2018). In particular:

- We cleaned street names. Names containing geographic indicators were removed, if they were not street names, and dummy variables were created for group quarters (hotels, apartments, convents, hospitals, group homes). House number information was extracted from street names.

- We cleaned house numbers. If the number found in the house number variable conflicted with the house number extracted from the street variable and the home was rented, the house number variable was interpreted to represent an apartment number.

- We interpolated missing street names and house numbers, conservatively. For observations on the same census page and within 6 house numbers from one another, missing streets were given the street name of the prior observation. For rented homes, missing house numbers were given the house number of the prior observation. For owned homes, if the street name was the same as the prior observation, missing house numbers were assigned a value equal to the house number of the prior observation plus two.

Geocoding

We geocoded the head of each household using ESRI Streetmap Premium 2019. These new-generation locator combines street addresses routing coordinates and parcel centroids databases to improve the number and the quality of the matches. Each address-coordinate match is assigned a 0-100 score by ESRI algorithm. We include in our analysis only matches with a score of at least 85. This choice is rather conservative and reduces measurement errors due to wrong locations of census households.

1.7.2 Census Variables Definitions

The following definitions are based on information provided by IPUMS.com documentation.

Census Variable Definitions	
Homeownership	Indicates whether the housing unit was owned, rather than rented, by its inhabitants.
African American	Based on census race variable. Prior to 1960, the census enumerator was responsible for categorizing persons and was not specifically instructed to ask the individual his or her race.
Property Values	For 1930 and 1940, enumerators consulted with the owners to estimate the sale value of the housing unit.
Rent Prices	Amount of the household's monthly contract rent payment.
First-Generation-Migrant	Whether an individual was foreign born. Based on the census variable *nativity*.
Unemployed, Men	Indicator defined accordind to census variable *empstat* for men between 18 and 65 years of age.
Radio Ownership	Whether any member of a family or housing unit owns a radio set.
Education Score	Census-built percentage of people in the respondent's occupational category who had completed one or more years of college
Number of Children	The number of own children residing with each individual.
Population Density	For any neighborhood the ratio between the area population and surface. Hexagon surface is fixed at $0.025km^2$.

Income Score Imputation

We estimate a log-wage regression on 1940 census data focusing on men aged 25-55 living in urban areas who were employed for wages. We regress self-reported wage income on a second degree polynomial in age, dummies indicating black, hispanic and immigration status, 3-digit occupation and state of residence indicators. Moreover, we include interactions between each of black, hispanic and immigration status with the age polynomial and interactions between each of the demographic variables with 1-digit occupations and state of residence. The inputed income score in decades 1930, 1920, 1910 is the prediction based on the resulting estimates for men aged 25-55 who are employed for wages in those years.

1.7.3 Random Forest Training Procedure

We train the random forest algorithm with a hexagon-level dataset ($N = 192,016$) containing all cities mapped by the HOLC with a population up to 3,000,000. The dataset includes 48 different 1930 census variables. The variables are included at different geographical levels, bringing the total number of training variables to 163. The geographical levels employed in the training procedure are: hexagon, hexagon surroundings (500mt and 1000mt), city, county. In particular the variables are:

Random Forest Training Variables

- •Share of African Americans
- •Share of Women
- •Share of Home-Owners
- •Share of Population Not Speaking English
- •Share Married
- •Share of families owning a Radio
- •Family Size
- •Number of Children
- •Age at First Marriage
- •House Values

- Rent Prices
- Imputed Income
- Earning Scores
- Educational Scores
- House Distance from City Center
- Neighborhood Population Density
- Labor Force Participation, by gender
- Unemployment Rates, by gender
- Self-Employed and waged employees
- Share of First Generation Immigrants
- Share of Second Generation Immigrants
- Domestic Migrants from the South
- Domestic Migrants from the Mid West
- Detailed Job Categories shares
- Detailed Country of Birth shares

Before starting the training model, we follow a standard pre-processing machine learning procedure: we impute missing values with the corresponding median values, and we standardize all our predictors. The random forest is trained on a 75% random sample of the original dataset selected with stratified sampling according to HOLC grades and city population. We set the parameter $m = 41$, which determines the number of variables randomly selected at each split, following the results of an automated tuning procedure employing model-based optimization (MBO) (Probst et al., 2018). The fraction of observations randomly sampled for each tree is grade-specific to counteract the class imbalance of HOLC grades. Hence, the less frequent class (A) has the highest sampling fraction (92%), while the most frequent classes (C and D) have lower fractions (63% and 70%). The results are robust to using a unique sample fraction. In particular the one suggested by the automated MBO procedure of Probst et al. (2018) is 0.89.

1.7.4 List of Cities

The difference-in-differences results are based on the definition of control and treatment group outlined in Section 1.4. As a reminder, treated cities are municipalities surveyed by the HOLC with a population between 40,000 and 50,000. The control group includes cities between 30,000 and 40,000 residents that had a distance of least 50km (31mi) from the nearest mapped city.

Treated Cities

- Amarillo, TX
- Aurora, IL
- Chelsea, MA
- Chicopee, MA
- Columbus, GA
- Council Bluffs, IA
- Dubuque, IA
- Elmira, NY
- Haverhill, MA
- Jackson, MS
- Joliet, IL
- Lexington, KY
- Lima, OH
- Lorain, OH
- Lynchburg, VA
- Muncie, IN
- Oshkosh, WI
- Phoenix, AZ
- Portsmouth, OH
- Poughkeepsie, NY
- Salem, MA
- S. Petersburg, FL
- Stamford, CT
- Stockton, CA
- Waterloo, IA
- Jackson, MS
- Woonsocket, RI
- Pueblo, CO
- Waltham, MA
- Warren, OH
- Ogden, UT
- Everett, MA

Control Cities

- Baton Rouge, LA
- Bellingham, WA
- Butte, MT
- Colorado Springs
- Fort Smith, AR
- Hagerstown, MD
- Joplin, MO
- La Crosse, WI
- Laredo, TX
- Lewiston, ME
- Mansfield, OH
- Meridian, MS
- Moline, IL
- Muskogee, OK
- Norwood, OH
- Paducah, KY
- Pensacola, FL
- Quincy, IL

- Rock Island, IL
- San Bernardino
- Santa Barbara
- Alton, IL
- Amsterdam, NY
- Auburn, NY
- Bloomington, IL
- Cumberland, MD
- Danville, IL
- Elkhart, IN
- Everett, WA
- Hazleton, PA
- High Point, NC
- Marion, OH
- Newark, OK
- Port Huron, MI
- Raleigh, NC
- Rome, NY
- Sheboygan, WI
- Steubenville, OH
- Kokomo, IN
- Meriden, CT
- Green Bay, WI
- Easton, PA
- Santa Ana, CA
- Richmond, IN
- Sioux Falls, SD
- Tucson, AZ
- Watertown, NY
- Wilmington, NC
- Zanesville, OH

Spatial Unit	N	Proportions A	B	C	D
HOLC Neighborhoods	8872	11.7%	26.3%	38.1%	23.9%
Hexagons	278066	7.8%	22.0%	42.1%	28.0%

Notes: The sample of HOLC neighborhoods includes all the shapes digitized by Nelson et al. (2021) for 202 cities. We obtain the sample of hexagons overlaying a regular grid of hexagons with an area of 0.025 km^2 and a side of approximately 100 $mt.$ over the digitized HOLC shapes. The grade of a hexagon is the one occupying the majority of its area. We keep only hexagons whose area is occupied by a single grade for at least 75%.

Table 1.11 — Neighborhood Distribution according to HOLC Grades

	Share of Coverage				
	Census Data			CoreLogic Deeds	
Decade	Neighborhood	City	Year	Neighborhood	City
1910	45.6%	98.8%			
1920	56.9%	100.0%			
1930	84.5%	100.0%			
1940	96.6%	100.0%			
1950	5.2%	14.8%			
1960	54.4%	51.9%			
			1965	3.4%	35.21%
1970	83.5%	80.2%	1970	5.6%	38.0%
			1975	8.6%	50.7%
1980	87.4%	93.8%	1980	11.4%	57.7%
			1985	21.1%	76.0%
1990	99.4%	100.0%	1990	35.8%	84.5%
			1995	52.7%	88.7%
2000	100.0%	100.0%	2000	64.4%	91.5%
			2005	76.0%	94.4%
2010	100.0%	100.0%	2010	77.9%	94.4%

Notes: The Table reports the percentages of coverage for neighborhoods and cities of interest. The sample includes every hexagon in cities with a 1930 population between 30,000 and 50,000 and at least 20 residents in 1930. CoreLogic deeds are binned in 5-year time periods according to their sale year and month.

Table 1.12 — Census and CoreLogic coverage of Neighborhoods and Cities.

	Dependent variables:			
	Homeownership Rates	African American Percentage	Property Values	Rent Prices
DiD	-0.006 (0.007)	0.002 (0.002)	-254* (132)	6.3 (8.1)
$\bar{Y}$	0.49	0.06	5439	42.6

Notes: The Table reports difference-in-differences coefficients obtained estimating equation (1.1) by grade. The regressions are estimated with individual-level observations. The sample includes individuals with valid geocodes in cities with a 1930 population between 30,000 and 50,000, living in hexagons with at least 20 residents in 1930. See Appendix Section 1.7.4 for a list of cities. Standard errors, in parentheses, are clustered at the city-year level. See the Notes of Table 1.7 for additional estimation details.
Significance: * 0.10 ** 0.05 *** 0.01

Table 1.13 — Short-term Diff-in-Diff Results. No Grade Heterogeneity

		Data			
		D	C	B	A
Prediction	D	2139	117	12	0
	C	233	4435	219	18
	B	36	214	2647	153
	A	3	18	45	866
	Accuracy	90.43%			
	Class Sensitivity	88.72	92.70	90.56	83.51
	Prevalence	21.61	42.89	26.20	9.30
	Detection Prevalence	20.33	43.97	27.34	8.35

Notes: The matrix compares the observed and the predicted grades for a test set of observations excluded from the training procedure. The test set is a 25% random subsample of the original dataset selected with stratified sampling according to city population and HOLC grade. In this case, the test set is restricted to include only hexagons in cities with a population below 50,000. For other details, see the notes of Table 1.3, which contains the predicted grades for the complete test set.

Table 1.14 — Random Forest Performance, Confusion Matrix, Restricted Test Set

		Data			
		D	C	B	A
Prediction	D	8732	2101	118	12
	C	5076	15338	4215	379
	B	158	2801	6225	1998
	A	1	27	423	942
	Accuracy	64.35%			
	Class Sensitivity	62.52	75.68	56.69	28.20
	Prevalence	28.77	41.75	22.62	6.86
	Detection Prevalence	22.58	51.51	23.03	2.86

Notes: The matrix compares the observed and the predicted grades for a test set of observations excluded from the training procedure of a logit model. The test set is a 25% random subsample of the original dataset selected with stratified sampling according to city population and HOLC grade. The logit model is estimated with the same estimation steps of the random forest. For other details, see the notes of Table 1.3, which contains the predicted grades for a random forest algorithm.

Table 1.15 — Logit Performance, Confusion Matrix

	Predicted Grade			
	C		D	
	Correct	Wrong	Correct	Wrong
Panel A: 1930 Levels				
Black	0.02	0.02	0.22	0.12
	(0.08)	(0.08)	(0.34)	(0.26)
Home Owner	0.53	0.49	0.39	0.43
	(0.22)	(0.24)	(0.23)	(0.21)
Income Score	7.13	7.13	6.85	6.92
	(0.17)	(0.16)	(0.28)	(0.24)
First Gen Immigrant	0.22	0.26	0.23	0.23
	(0.17)	(0.18)	(0.23)	(0.21)
Panel B: 1930-1920 Trends				
Black	-0.005	-0.001	0.03	0.01
	(0.07)	(0.07)	(0.16)	(0.19)
Home Owner	-0.01	0.03	-0.02	-0.03
	(0.24)	(0.23)	(0.25)	(0.25)
Income Score	0.01	0.03	-0.02	-0.04
	(0.18)	(0.19)	(0.22)	(0.25)
First Gen Immigrant	-0.02	-0.03	-0.05	-0.03
	(0.14)	(0.15)	(0.17)	(0.14)

Notes: The Table reports averages of 1930 Census variables according to different classifications. The first two columns compare means between hexagons classified as C by our classification model. The first columns reports averages for hexagons whose observed grade is C, while the second refers to neighborhoods with a HOLC grade other than C. The third and fourth columns do the same for grade D. The level of observation is a neighborhood (hexagon). The sample includes all the hexagons intersecting a HOLC neighborhood digitized by Nelson et al. (2021) in 202 maps. Standard deviations are reported in parentheses.

Table 1.16 — 1930 Descriptive Statistics According to Predicted Grades

	A	B	C	D
Black	0.00002	0.008	0.007	0.011
	(0.004)	(0.006)	(0.005)	(0.014)
Home Owner	0.026	0.009	0.015	0.037*
	(0.036)	(0.023)	(0.010)	(0.020)
Income Score	0.017	-0.008	-0.003	-0.002
	(0.046)	(0.014)	(0.007)	(0.013)
Education Score	2.501	-0.548	0.013	0.235
	(2.569)	(0.528)	(0.226)	(0.318)
First Gen Immigrant	-0.025	-0.010	0.013	-0.014
	(0.023)	(0.008)	(0.008)	(0.011)
Number of Children	-0.015	0.009	0.009	0.008
	(0.028)	(0.012)	(0.008)	(0.014)

Notes: The Table reports the coefficients from a set of regressions where the dependent variable is the 1920-1910 change in the variable reported in the left column, and the independent variable is an indicator for the treatment status. See Appendix Section 1.7.2 for definitions of Census variables in our dataset. The level of observation is a spatial unit (hexagon). The sample includes every hexagon in cities with a 1930 population between 30,000 and 50,000 and at least 20 residents in 1930. See Appendix Section 1.7.4 for a list of cities. Standard errors, in parentheses, are clustered at the city level. Significance: * 0.10 ** 0.05 *** 0.01

Table 1.17 — Testing Differences in 1920-1910 Trends by Treatment Status

	Dependent variable: African American Percentage				
	(1)	(2)	(3)	(4)	(5)
DiD_A	-0.003	-0.006	-0.004	-0.005	-0.005
$\bar{Y}^A = 0.008$	(0.014)	(0.007)	(0.006)	(0.006)	(0.006)
DiD_B	0.009	0.006*	0.004	0.006*	0.006*
$\bar{Y}^B = 0.01$	(0.012)	(0.003)	(0.003)	(0.004)	(0.004)
DiD_C	0.004	0.002	0.001	0.001	0.001
$\bar{Y}^C = 0.02$	(0.009)	(0.002)	(0.001)	(0.002)	(0.001)
DiD_D	0.012	0.010***	0.014***	0.016***	0.018***
$\bar{Y}^D = 0.19$	(0.096)	(0.003)	(0.004)	(0.004)	(0.004)
City Fixed. Eff.		X		X	X
Spatial Unit Fixed Eff.			X		
Spatial Unit Controls				X	X
Local Area Controls					X

Notes: The Table reports difference-in-differences coefficients obtained estimating equation (1.1) by grade. Each row contains the DiD coefficients for a given grade. The regressions are estimated with individual-level observations. The sample includes individuals with valid geocodes in cities with a 1930 population between 30,000 and 50,000, living in hexagons with at least 20 residents in 1930. See Appendix Section 1.7.4 for a list of cities. Standard errors, in parentheses, are clustered at the city-year level. The table structure is analogous to Table 1.7. See the Notes of Table 1.7 for additional estimation details. Significance: * 0.10 ** 0.05 *** 0.01

Table 1.18 — Short-term Difference-in-Differences Results, by Grade

	Dependent variable: Property Values				
	(1)	(2)	(3)	(4)	(5)
DiD_A	946	1,219***	1,142***	1,239***	1,046***
$\bar{Y}^A = 22,907$	(1,742)	(435)	(378)	(403)	(378)
DiD_B	-178	-147	-34	-106	-105
$\bar{Y}^B = 16,828$	(604)	(173)	(171)	(178)	(170)
DiD_C	-510	-493***	-487***	-468***	-502***
$\bar{Y}^C = 12,983$	(383)	(144)	(150)	(148)	(148)
DiD_D	-319	-284*	-301*	-251	-302**
$\bar{Y}^D = 8,615$	(432)	(156)	(157)	(155)	(153)
City Fixed. Eff.		X		X	X
Spatial Unit Fixed Eff.			X		
Spatial Unit Controls				X	X
Local Area Controls					X

Notes: The Table reports difference-in-differences coefficients obtained estimating equation (1.1) by grade. Each row contains the DiD coefficients for a given grade. The regressions are estimated with individual-level observations. The sample includes individuals with valid geocodes in cities with a 1930 population between 30,000 and 50,000, living in hexagons with at least 20 residents in 1930. See Appendix Section 1.7.4 for a list of cities. Standard errors, in parentheses, are clustered at the city-year level. The table structure is analogous to Table 1.7. See the Notes of Table 1.7 for additional estimation details. Significance: * 0.10 ** 0.05 *** 0.01

Table 1.19 — Short-term Difference-in-Differences Results, by Grade

	Dependent variable: Property Values (logs)				
	(1)	(2)	(3)	(4)	(5)
DiD_A	0.135	0.168***	0.152***	0.168***	0.156***
$\bar{Y}^A = 9.82$	(0.244)	(0.027)	(0.022)	(0.026)	(0.025)
DiD_B	-0.057	-0.048*	-0.035	-0.040	-0.038
$\bar{Y}^B = 9.52$	(0.123)	(0.028)	(0.028)	(0.028)	(0.028)
DiD_C	-0.052	-0.048*	-0.049*	-0.043	-0.049*
$\bar{Y}^C = 9.20$	(0.095)	(0.027)	(0.028)	(0.028)	(0.028)
DiD_D	-0.006	-0.011	-0.018	-0.005	-0.022
$\bar{Y}^D = 8.66$	(0.190)	(0.045)	(0.043)	(0.044)	(0.042)
City Fixed. Eff.		X		X	X
Spatial Unit Fixed Eff.			X		
Spatial Unit Controls				X	X
Local Area Controls					X

Notes: The Table reports difference-in-differences coefficients obtained estimating equation (1.1) by grade. Each row contains the DiD coefficients for a given grade. The regressions are estimated with individual-level observations. The sample includes individuals with valid geocodes in cities with a 1930 population between 30,000 and 50,000, living in hexagons with at least 20 residents in 1930. See Appendix Section 1.7.4 for a list of cities. Standard errors, in parentheses, are clustered at the city-year level. The table structure is analogous to Table 1.7. See the Notes of Table 1.7 for additional estimation details. Significance: * 0.10 ** 0.05 *** 0.01

Table 1.20 — Short-term Difference-in-Differences Results, by Grade

	Dependent variable: Home-Ownership Rates			
	(1)	(2)	(3)	(4)
DiD_A	0.034	0.041**	0.037**	0.037**
$\bar{Y}^A = 0.71$	(0.128)	(0.017)	(0.017)	(0.016)
DiD_B	0.0002	-0.0001	-0.008	-0.003
$\bar{Y}^B = 0.65$	(0.023)	(0.009)	(0.009)	(0.009)
DiD_C	-0.014	-0.015	-0.019**	-0.014
$\bar{Y}^C = 0.53$	(0.029)	(0.010)	(0.010)	(0.009)
DiD_D	-0.023	-0.019**	-0.027***	-0.020**
$\bar{Y}^D = 0.43$	(0.046)	(0.009)	(0.010)	(0.009)
City Fixed. Eff.		X	X	X
Spatial Unit Controls			X	X
Local Area Controls				X

Notes: The Table reports difference-in-differences coefficients obtained estimating equation (1.1) by grade. Each row contains the DiD coefficients for a given grade. The regressions are estimated with neighborhood-level observations. The sample includes neighborhoods with at least 20 residents in 1930 in cities with a population between 30,000 and 50,000. See Appendix Section 1.7.4 for a list of cities. Standard errors, in parentheses, are clustered at the city-year level. Significance: * 0.10 ** 0.05 *** 0.01

Table 1.21 — Short-term Difference-in-Differences Results. Neighborhood Level

	Dependent variable: African American Percentage			
	(1)	(2)	(3)	(4)
DiD_A	0.005	0.003	0.004	0.004
$\bar{Y}^A = 0.009$	(0.016)	(0.006)	(0.005)	(0.005)
DiD_B	0.002	0.001	0.002	0.002
$\bar{Y}^B = 0.01$	(0.010)	(0.003)	(0.003)	(0.003)
DiD_C	0.002	0.001	0.0003	0.0005
$\bar{Y}^C = 0.02$	(0.009)	(0.001)	(0.001)	(0.001)
DiD_D	0.022	0.015**	0.018***	0.018***
$\bar{Y}^D = 0.22$	(0.090)	(0.006)	(0.006)	(0.006)
City Fixed. Eff.		X	X	X
Spatial Unit Controls			X	X
Local Area Controls				X

Notes: The Table reports difference-in-differences coefficients obtained estimating equation (1.1) by grade. Each row contains the DiD coefficients for a given grade. The regressions are estimated with neighborhood-level observations. See the Notes of Appendix Table 1.21 for additional details. Significance: * 0.10 ** 0.05 *** 0.01

Table 1.22 — Short-term Difference-in-Differences Results. Neighborhood Level

	Dependent variable: Property Values			
	(1)	(2)	(3)	(4)
DiD_A	629	831	849	747
$\bar{Y}^A = 25,162$	(1,931)	(378)	(354)	(350)
DiD_B	-124	-144	-149	-163
$\bar{Y}^B = 17,371$	(735)	(188)	(189)	(173)
DiD_C	-394	-397***	-348***	-369***
$\bar{Y}^C = 12,727$	(434)	(126)	(129)	(131)
DiD_D	-548**	-523***	-462***	-510***
$\bar{Y}^D = 7,923$	(374)	(143)	(144)	(146)
City Fixed. Eff.		X	X	X
Spatial Unit Controls			X	X
Local Area Controls				X

Notes: The Table reports difference-in-differences coefficients obtained estimating equation 1.1 by grade. Each row contains the DiD coefficients for a given grade. The regressions are estimated with neighborhood-level observations. See the Notes of Appendix Table 1.21 for additional details. Significance: * 0.10 ** 0.05 *** 0.01

Table 1.23 — Short-term Difference-in-Differences Results. Neighborhood Level

	Dependent variable: Rent Prices			
	(1)	(2)	(3)	(4)
DiD_A	60.0	68.8***	66.2***	64.7***
$\bar{Y}^A = 562.20$	(37.4)	(25.2)	(25.0)	(24.8)
DiD_B	18.3	17.6	18.3	18.2
$\bar{Y}^B = 342.76$	(19.4)	(13.3)	(12.6)	(12.5)
DiD_C	22.6	22.3**	21.4**	21.6**
$\bar{Y}^C = 241.72$	(15.5)	(10.4)	(10.3)	(10.3)
DiD_D	6.9	7.5	8.5	7.6
$\bar{Y}^D = 172.63$	(11.5)	(7.5)	(7.5)	(7.5)
City Fixed. Eff.		X	X	X
Spatial Unit Controls			X	X
Local Area Controls				X

Notes: The Table reports difference-in-differences coefficients obtained estimating equation 1.1 by grade. Each row contains the DiD coefficients for a given grade. The regressions are estimated with neighborhood-level observations. See the Notes of Appendix Table 1.21 for additional details. Significance: * 0.10 ** 0.05 *** 0.01

Table 1.24 — Short-term Difference-in-Differences Results. Neighborhood Level

	Dependent Variables			
	Homeownership Rates	African American Percentage	Property Values	Rent Prices
DiD_A	0.045**	-0.005	1,046**	39.3***
	(0.021)	(0.009)	(538)	(16.3)
DiD_B	-0.002	0.006*	-106	8.8
	(0.013)	(0.005)	(248)	(17.0)
DiD_C	-0.017	0.001	-502**	5.9
	(0.011)	(0.002)	(212)	(13.3)
DiD_D	-0.024*	0.018***	-302 **	3.1
	(0.013)	(0.006)	(218)	(11.4)

Notes: The Table reports difference-in-differences coefficients obtained estimating equation (1.1) by grade for four different outcomes. Each row contains the DiD coefficients for a given grade. The Table shows the DiD coefficients resulting from a DiD framework with a city fixed effect and geographic and demographic controls at the hexagon level. The regression specification is analogous to the one in column (4) of Table 1.7. The list of controls includes geographic coordinates, a scaled measure of distance from the city center, spatial unit's population density, imputed income score and family size. The regressions are estimated with individual-level observations. The sample includes individuals with valid geocodes in cities with a 1930 population between 30,000 and 50,000, living in hexagons with at least 20 residents in 1930. See Appendix Section 1.7.4 for a list of cities. Standard errors, in parentheses, are clustered at the city level. Significance: * 0.10 ** 0.05 *** 0.01

Table 1.25 — Short-term Diff-in-Diff by Grade. City Level S.E. Clustering

	Dependent Variables		
	Homeownership Rates	African American Percentage	Property Values
DiD_A	0.034** (0.015)	-0.004 (0.004)	1,253*** (360)
DiD_B	0.004 (0.008)	0.006* (0.003)	-171 (187)
DiD_C	-0.010 (0.007)	0.002 (0.002)	-407*** (144)
DiD_D	-0.018* (0.008)	0.018*** (0.004)	-405 ** (167)

Notes: The Table reports difference-in-differences coefficients obtained estimating equation (1.1) by grade for three different outcomes. Each row contains the DiD coefficients for a given grade. Differently from the other results, we group observations in treated cities according to their observed HOLC grade, rather than the predicted one. See the Notes of Table 1.25 for estimation details. Standard errors, in parentheses, are clustered at the city-year level. Significance: * 0.10 ** 0.05 *** 0.01

Table 1.26 — Short-term Diff-in-Diff. Grouping with Observed Grades

	Dependent Variables		
	Homeownership Rates	African American Percentage	Property Values
DiD_A	0.027 (0.019)	0.008** (0.004)	1,657** (835)
DiD_B	0.001 (0.008)	0.001 (0.002)	-766*** (255)
DiD_C	-0.017** (0.007)	0.0002 (0.001)	-578*** (130)
DiD_D	-0.024*** (0.008)	0.019*** (0.006)	-855 *** (192)

Notes: The Table reports difference-in-differences coefficients obtained estimating equation (1.1) at the neighborhood level, by grade, for three different outcomes. Each row contains the DiD coefficients for a given grade. Differently from the other results, we extend the treatmet group to include cities between 40,000 and 60,000 residents in the 1930 census. See the Notes of Table 1.25 for estimation details. Standard errors, in parentheses, are clustered at the city-year level. Significance: * 0.10 ** 0.05 *** 0.01

Table 1.27 — Short-term Diff-in-Diff. Extended Treatment Group

	Dependent Variables		
	Female Percentage	Number of Children	Unemployment Rate, Men
DiD_A	0.004	0.019	-0.045***
	(0.002)	(0.013)	(0.008)
DiD_B	0.005***	0.004	0.011**
	(0.002)	(0.002)	(0.005)
DiD_C	-0.001	-0.006	-0.003
	(0.001)	(0.005)	(0.007)
DiD_D	0.002	0.022***	0.012
	(0.002)	(0.006)	(0.013)
$\bar{Y}^A$	0.523	0.824	0.070
$\bar{Y}^B$	0.527	0.744	0.059
$\bar{Y}^C$	0.509	0.778	0.099
$\bar{Y}^D$	0.495	0.821	0.136

*Notes:*The Table reports difference-in-differences coefficients obtained estimating equation (1.1) by grade for three different outcomes. Each row contains the DiD coefficients for a given grade. See the Notes of Table 1.25 for estimation details. Standard errors, in parentheses, are clustered at the city-year level. Significance: * 0.10 ** 0.05 *** 0.01

Table 1.28 — Short-term Diff-in-Diff. Placebo Outcomes

	Dependent Variables		
	Homeownership Rates	African American Percentage	Property Values
DiD_A	0.016 (0.026)	0.004 (0.004)	-1,999 (1,498)
DiD_B	-0.015 (0.014)	-0.003* (0.002)	717 (492)
DiD_C	-0.008 (0.010)	0.001 (0.002)	53 (301)
DiD_D	0.004 (0.008)	0.009 (0.006)	970*** (324)

Notes: The Table reports difference-in-differences coefficients obtained estimating equation (1.1) at the neighborhood level, by grade, for three different outcomes. Each row contains the DiD coefficients for a given grade. Differently from the other results, we set a placebo population threshold at 60,000. Accordingly, we define neighborhoods in cities between 60,000 and 70,000 residents as treated, while areas in cities between 50,000 and 60,000 are included in the control group. See the Notes of Table 1.25 for estimation details. Standard errors, in parentheses, are clustered at the city-year level. Significance: * 0.10 ** 0.05 *** 0.01

Table 1.29 — Short-term Diff-in-Diff. Placebo Population Threshold

	Dependent Variables		
	Homeownership Rates	African American Percentage	Property Values
DiD_A	0.042*** (0.015)	-0.005 (0.006)	1,223*** (398)
DiD_B	-0.002 (0.009)	0.006* (0.004)	-106 (177)
DiD_C	-0.017** (0.007)	0.001 (0.002)	-472*** (147)
DiD_D	-0.022** (0.009)	0.015*** (0.004)	-274* (155)

Notes: The Table reports difference-in-differences coefficients obtained estimating equation (1.1) at the neighborhood level, by grade, for three different outcomes. Each row contains the DiD coefficients for a given grade. Differently from the other results, we include the grade composition of the surrounding neighborhoods as control. The surrounding neighborhoods are defined with a 1000mt radius (0.63 miles). See the Notes of Table 1.25 for estimation details. Standard errors, in parentheses, are clustered at the city-year level. Significance: * 0.10 ** 0.05 *** 0.01

Table 1.30 — Short-term Diff-in-Diff. Neighborhood Grade Index.

	Dependent variable: Property Values (logs)			
	Grade			
	A	B	C	D
DiD_{65}	.	0.05	-0.52***	-0.95***
	(.)	(0.10)	(0.09)	(0.13)
DiD_{70}	-0.24	0.025	-0.68***	-1.05***
	(0.17)	(0.121)	(0.12)	(0.14)
DiD_{75}	-0.15	-0.010	-0.53***	-0.94***
	(0.17)	(0.113)	(0.10)	(0.15)
DiD_{80}	-0.17	-0.11	-0.77**	-0.70***
	(0.17)	(0.12)	(0.19)	(0.22)
DiD_{85}	0.18	0.19	-0.09	-0.51**
	(0.23)	(0.18)	(0.22)	(0.25)
DiD_{90}	0.29	0.23*	-0.06	-0.40
	(0.19)	(0.18)	(0.18)	(0.249)
DiD_{95}	-0.02	0.02	-0.28*	-0.55***
	(0.21)	(0.16)	(0.15)	(0.19)
DiD_{00}	0.07	0.06	-0.14	-0.39**
	(0.23)	(0.16)	(0.18)	(0.19)
DiD_{05}	0.21	-0.06	-0.13	-0.35
	(0.25)	(0.19)	(0.19)	(0.21)

Notes: The Table reports Difference-in-Differences coefficients obtained estimating equation (1.2) by grade. Each row contains the DiD coefficients for a given grade in the corresponding year. The regressions are estimated with neighborhood-level observations. The sample includes neighborhoods with at least 20 residents in 1930 in cities with a population between 30,000 and 50,000. See Appendix Section 1.7.4 for a list of cities. The data source for post-1940 outcomes is CoreLogic, see Section 1.3.3 for details. The outcome variable is adjusted with CPI to 1980 dollars. Standard errors, in parentheses, are clustered at the city-year level. Significance: * 0.10 ** 0.05 *** 0.01

Table 1.31 — Long-Term Difference-in-Differences Results, by Grade. CoreLogic

Observable Pre-Trends, Grades A and B

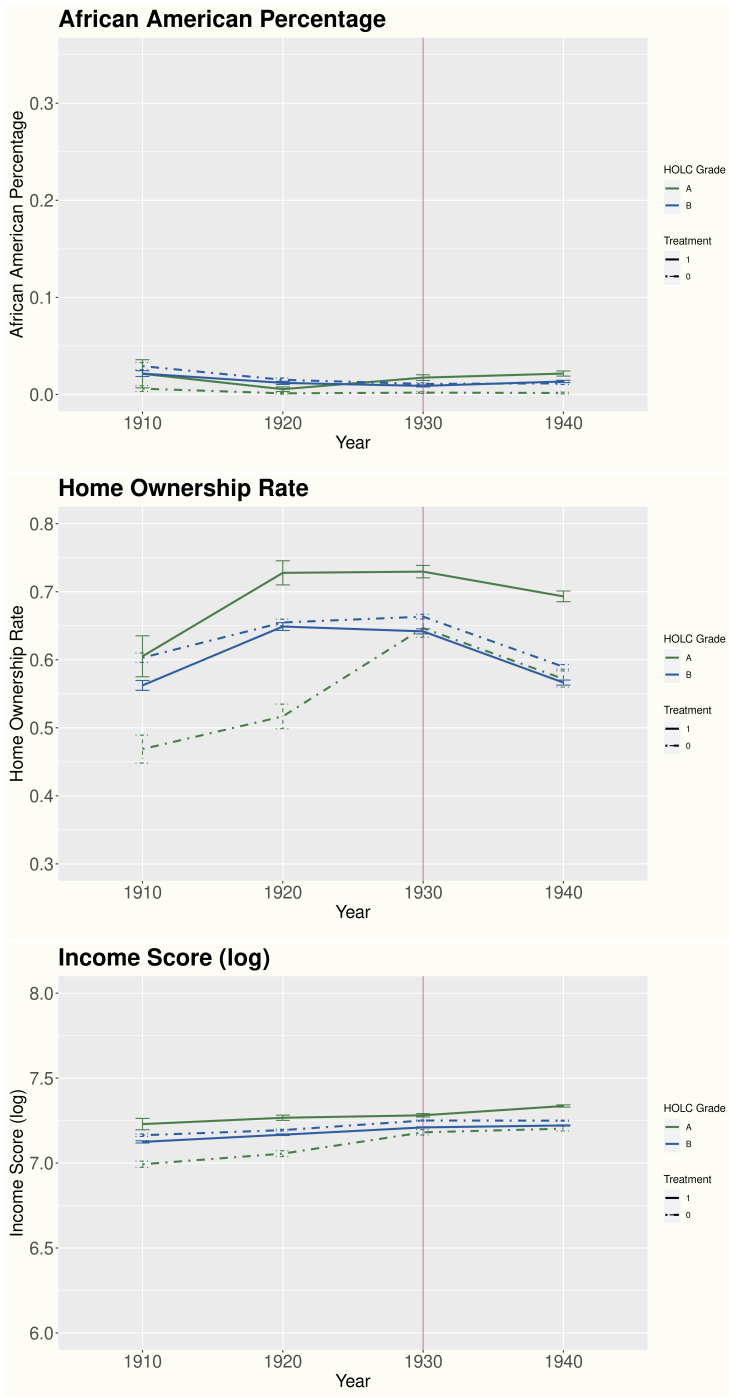


Figure 1.13 — The Figure shows pre-trends for selected variables for A and B grades. The point estimates are averages of hexagon-level observations. The bars show the respective standard errors of each mean. The sample includes neighborhoods in cities with a 1930 population between 30,000 and 50,000, with at least 20 residents in 1930. The vertical line highlights 1930, the last pre-treatment decade. See Appendix Section 1.7.2 for definitions of Census variables in our dataset.

Observable Pre-Trends, Additional Variables

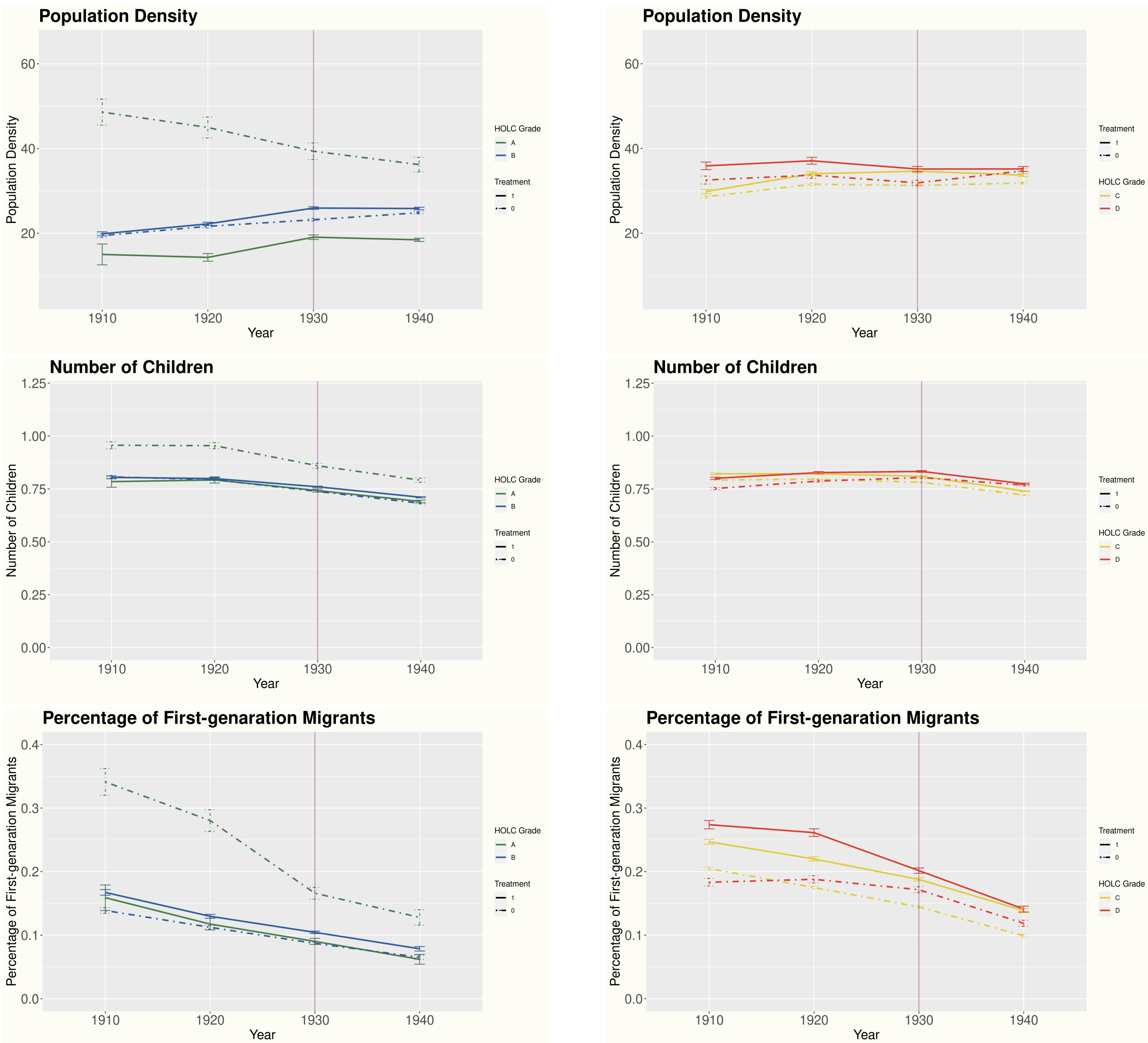


Figure 1.14 — The Figure shows pre-trends for selected variables according to their predicted grade. The point estimates are averages of hexagon-level observations. The bars show the respective standard errors of each mean. The sample includes hexagons in cities with a 1930 population between 30,000 and 50,000, with at least 20 residents in 1930. The vertical line highlights 1930, the last pre-treatment decade. See Appendix Section 1.7.2 for definitions of Census variables in our dataset.

Accuracy Levels according to Dataset Definition

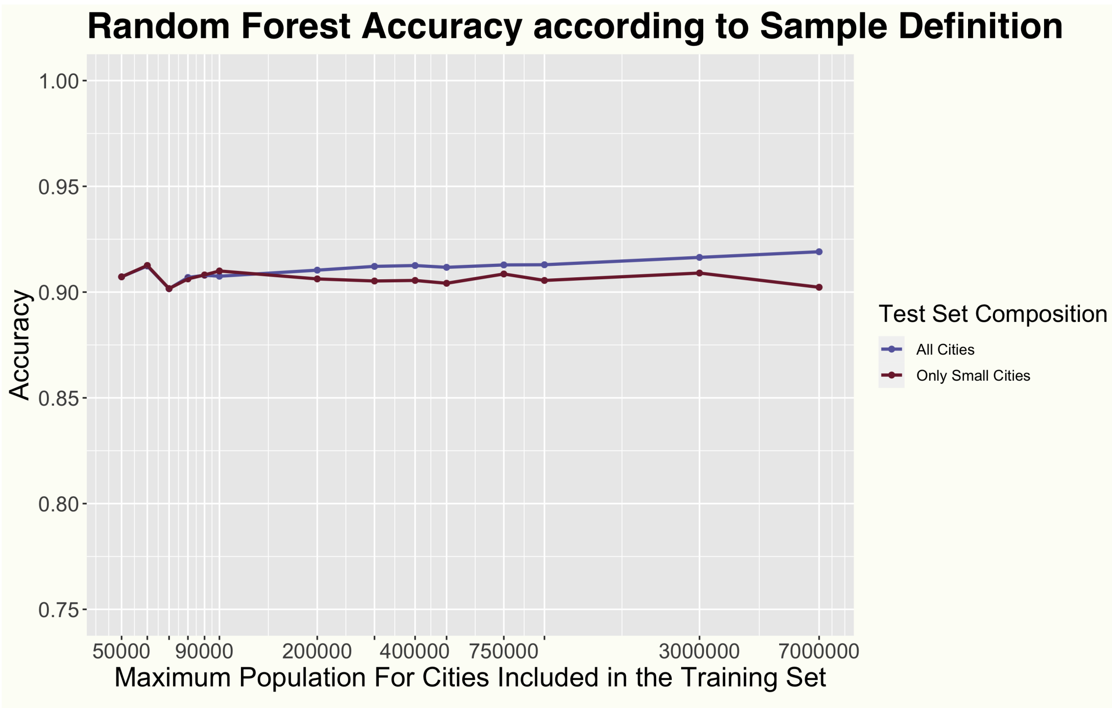


Figure 1.15 — The Figure shows the accuracy level we obtain when we train the random forest classification algorithms with different datasets according to the size of cities we include. The purple line plots the Accuracy obtained with a test set defined as a 25% random subsample of the original dataset, selected with stratified sampling according to city population and HOLC grade. The brown line shows Accuracy levels when we restrict the test set to cities with at most 50,000 residents. The level of observation is a neighborhood (hexagon). See Section 1.3.1 for details about the hexagon definition. The complete dataset includes every hexagon in a mapped city containing at least 20 residents in 1930. Overall Accuracy is the percentage of hexagons whose predicted grades correspond to observed ones. A predicted grade is the class predicted by the trained random forest algorithm. See Section 1.4.1 and Appendix Section 1.7.3 for details about the Random Forest training procedure.

Long-Term Difference-in-Differences Results, A and B Grades. Census Data

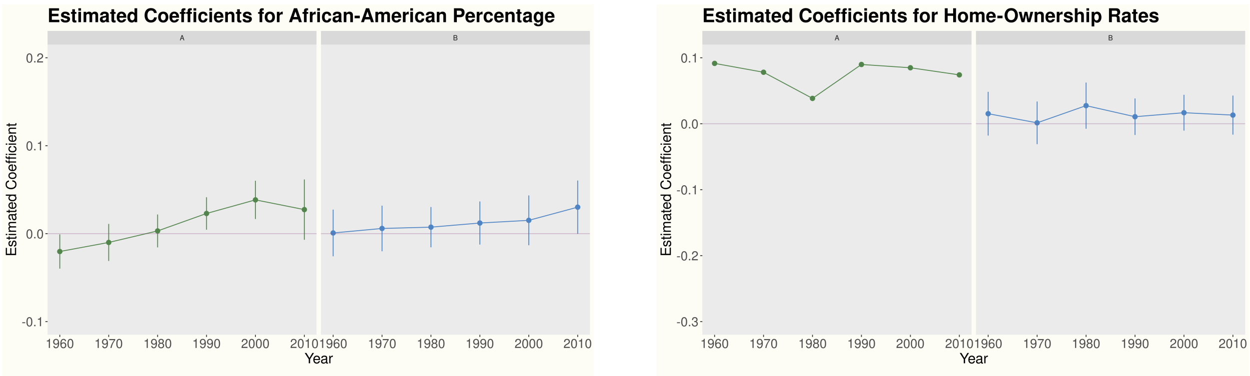


Figure 1.16 — The Figure shows the estimated coefficients for regression (1.2) and their 95% confidence intervals for grades A and B. The coefficients and standard errors are analogous to the ones reported in Table 1.9, but focus on different grades. See the Notes of Table 1.9 for estimation details.

Long-Term Difference-in-Differences Results, A and B Grades. CoreLogic Data

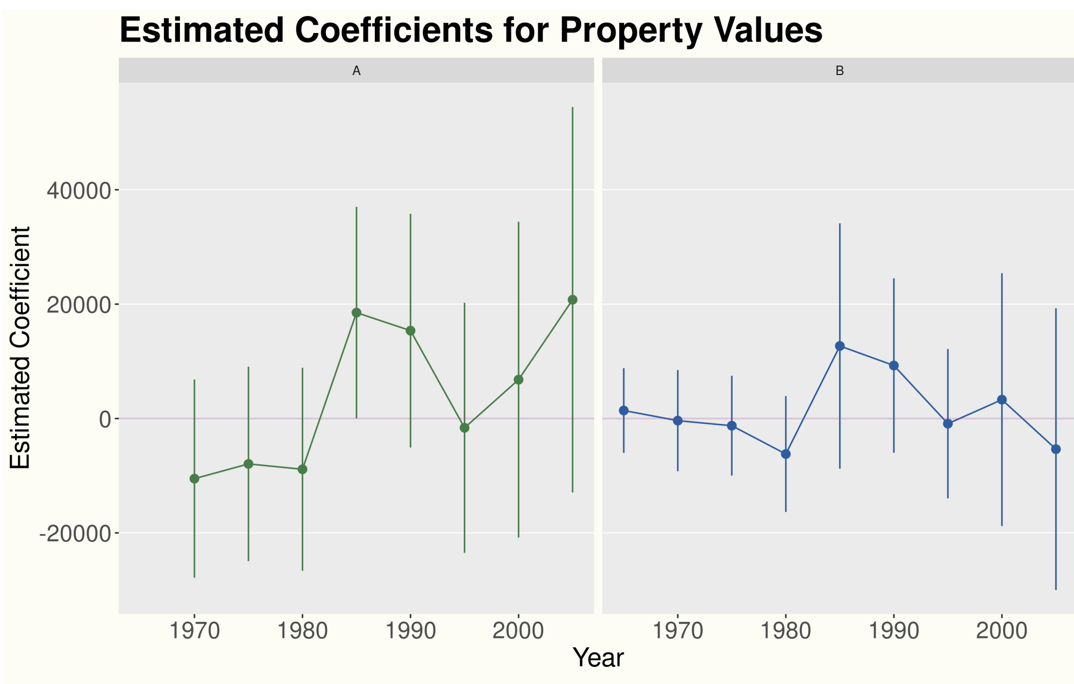


Figure 1.17 — The Figure shows the estimated coefficients for regression (1.2) and their 95% confidence intervals for property values. The Figure includes the results for grades C and D. The coefficients and standard errors are the ones reported in Table 1.10. See the Notes of Table 1.10 for estimation details.

Long-Term Difference-in-Differences Results, by Grade. CoreLogic Data

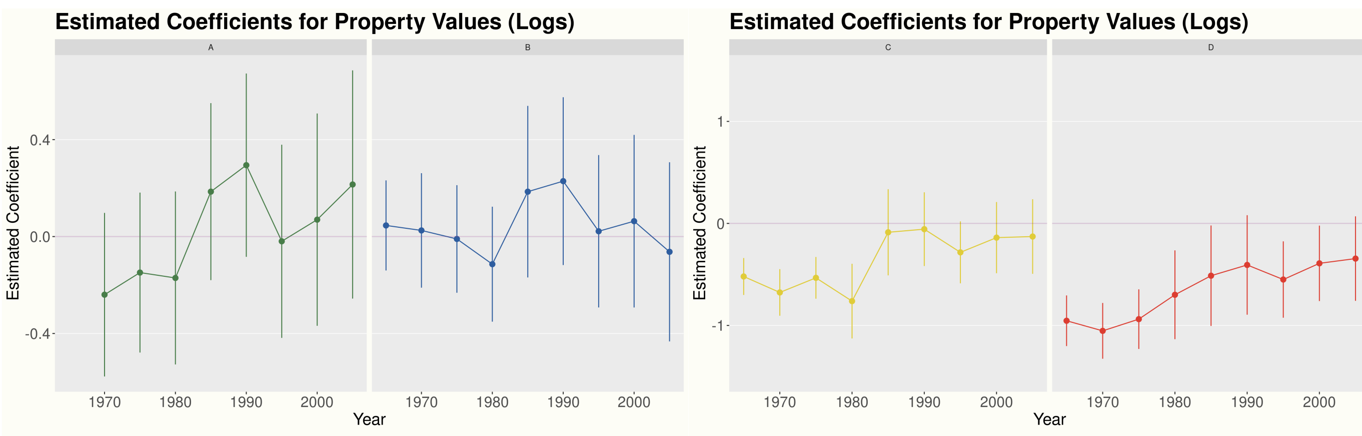


Figure 1.18 — The Figure shows the estimated coefficients for regression (1.2) and their 95% confidence intervals for property values. The coefficients and standard errors are analogous to the ones reported in Table 1.10, but the outcome is a log-transformation. See the Notes of Table 1.31 for estimation details.

Chapter 2

The Long-Term Effects of Exposure to Non-Traditional Family Structures

2.1 Introduction

Economists have extensively investigated how socioeconomic status is transmitted across generations. Families are a fundamental institution in shaping children's life trajectories, and different domestic arrangements can influence the intergenerational persistence of inequality. The start of the contemporary debate on the role of family structure in the reproduction of poverty can be traced to Moynihan's 1965 report[1] which argued that African American children growing up without a father faced reduced chances of economic success. Since then, American families have dramatically changed. In the post-war period, most couples got married early, and out-of-marriage births were rare. Over the past sixty years, though, an educational gradient in family arrangements has emerged. College graduates are much more likely to be married than high school graduates today. Moreover, by 2013, birth to an unmarried mother was the typical outcome for those without a college degree.[2]

[1] Moynihan (1965).

[2] According to Lundberg et al. (2016), the rate of non-marital childbearing for high school graduates is 58% versus 11% for college graduates.

Not surprisingly, social scientists started to investigate the implications of this phenomenon. Sociologists have produced an enormous literature linking single-headed, unstable families to worse outcomes for their children.[3] Economists, meanwhile, have contributed much less to research in this area due to methodological concerns. In particular, the lack of reliable sources of exogenous variation has prevented an extensive investigation of this topic. Another aspect that is often overlooked is how the increasing prevalence of missing-father families might create social spillovers. In fact, little is known about the effects of children from single-mother families on their school peers. Switching to a social environment perspective can characterize an additional consequence of non-traditional families. Such investigation is feasible since it is easier to find exogenous variation in children's social environment rather than inside one's family.

A piece of motivating evidence comes from Chetty et al. (2019). The authors show that Black father presence, at the census tract level, is associated with higher employment rates and income at age 30 or later for Black men who grew up in that neighborhood. The surprising strong correlation outpaces any other tract-level characteristics, and it is robust to controlling for the marital status of one's parents. The same effect is not detected for Whites of both genders and Black women. Why should family structure at the neighborhood level be relevant only for Black boys? A partial answer can be found in Sampson (1987), which emphasizes the role of formal and informal social controls provided by families who rely on more than one working parent. Specifically, father figures provide mentoring and shape the social environment in which their children will grow up. Such guidance is especially important for those children who start their life in a position of socioeconomic disadvantage.

This paper investigates the long-term causal effects of differential exposure to peers living without a residential father. The underlying idea is that adolescents who faced reduced access to local father figures received less mentoring, guidance, and support at a critical age of their development. Following the terminology from Manski (1993), I am

[3] McLanahan and Percheski (2008).

focusing on a contextual effect: the impact of peers' demographic outcomes on long-term educational achievements, labor market trajectories, income and mental health. The empirical strategy takes advantage of within-school, cohort-to-cohort variation in shares of missing fathers following an approach first proposed by Hoxby (2000). Given that parents choose their children's school based on average school characteristics, one cannot employ any variation across schools to identify the peer effects of interest. Instead, once I control for students' endogenous sorting, it is credible to characterize the residual variation in cohort composition as exogenous. This allows me to interpret the resulting regression coefficients as causal effects.

In order to implement the empirical strategy, I require a dataset containing information on multiple student cohorts for several different schools. Moreover, it is critical to observe a wide range of adult outcomes. The Add Health survey meets all these needs with a longitudinal survey of approximately 15,000 students regularly interviewed from adolescence to adulthood. Importantly, this survey started with an in-school census that precisely measured cohort composition in terms of demographic and socioeconomic characteristics. I provide evidence that the variation in the percentage of missing fathers can be safely considered exogenous with respect to unobservable determinants of students' outcomes. At the same time, the available amount of residual variation is similar to that seen in several other papers that successfully identified contextual peer effects with the same dataset.

The empirical strategy does not find any causal effect of exposure to peers from single-mother households on long-term educational achievements, labor market outcomes, total income or mental health. Even if the Add Health data confirm the association between higher rates of father absence at the local level and long-term socioeconomic disadvantage, such a relationship is due to endogenous sorting across schools and it should not have a causal interpretation.

The paper proceeds as follows: Sections 2.2 and 2.3 provide a literature review and a dataset description, respectively. The details of the empirical strategy can be found in Section 2.4 , and Section 2.5 provides checks about the validity of the econometric approach. The results are presented in Section 2.6.

2.2 Literature Review

This paper is related to two strands of social science research: the first examines the effect of one's family structure on later life outcomes, while the second is the extensive literature on peer effects.

The majority of papers investigating the association between family structure and child development comes from the field of sociology.[4] Economists, for their part, have focused on marriage choices[5] and how marriage patterns are shaped by economic and institutional shocks.[6] A separate literature deals with child development of cognitive and non-cognitive skill[7] focusing on stable two-parent families. More recently, economists have started to focus on family disadvantage to explain the reverse gender gap in education. Autor et al. (2019) show that boys' educational outcomes are more vulnerable to economic disadvantage while Bertrand and Pan (2013) suggest that non-traditional families are particularly detrimental for boys' non-cognitive development, increasing the prevalence of externalizing behaviors.

The empirical strategy exploits idiosyncratic variation in peer family structure across different cohorts within the same school. First proposed by Hoxby (2000), this approach has been widely used to investigate peer effects in education.[8] Several papers have exploited the Add Health school census to investigate different sources of peer effects. Bifulco et al. (2011) and Cools et al. (2019) use cohort composition variation in terms of parental education, such as maternal college graduation and parental post-college degrees, to detect an effect on post-secondary outcomes. Similarly, Olivetti et al. (2018) ex-

[4]McLanahan (1985); McLanahan and Percheski (2008); Lee and McLanahan (2015).
[5]Becker (1973).
[6]Charles and Luoh (2010); Autor et al. (2018).
[7]Cunha and Heckman (2007); Agostinelli and Wiswall (2016); Del Boca et al. (2013).
[8]Lavy and Schlosser (2011); Sacerdote (2014).

ploit cohort-to-cohort variation in the percentage of working mothers, finding a positive effect on daughters' labor force participation. Marriage patterns are analyzed by Merlino et al. (2019), who measure the effect of one's cohort racial composition on the number of interracial marriages. Khan and Anand (2018) analyze how the timing and outcome of pregnancies in the peer group affect fertility choices of Add Health respondents. Lastly, Fruehwirth et al. (2019) use cohort-to-cohort variation in peers' religiosity as an instrument to deal with selection into religiosity, finding a positive effect on depression.

To the best of my knowledge, the only papers investigating family structure using Add Health data have been published in the sociology literature. Cavanagh and Fomby (2011) investigate family instability as a moderating factor between one's family structure and academic outcomes, while Gaydosh and Harris (2018) test the link between family instability in childhood and biological markers of health. Both works, however, do not address endogeneity concerns with their empirical strategy.

2.3 Data

Peer effects studies often rely on school administrative data precisely measuring students' cohort composition and educational outcomes.[9] However, investigating the long-term effects of peer characteristics requires a longitudinal dataset including both information about adolescent peer groups and an adequate time span to measure long-term outcomes. The National Longitudinal Study of Adolescent to Adult Health (Add Health)[10] dataset is one of the few existing surveys satisfying such data requirements, and it has been extensively used to investigate the long-term effects of schools' social environment.[11]

This paper uses the restricted version of the Add Health, which started with an in-school census survey administered in the 1994–1995 academic year. Participating US schools were sampled to be nationally representative in terms of size, racial and ethnic composi-

[9] Hoxby (2000); Lavy and Schlosser (2011).

[10] Harris et al. (2019).

[11] Bifulco et al. (2011); Calvo-Armengol et al. (2009); Olivetti et al. (2018); Cools et al. (2019); Fruehwirth et al. (2019); Merlino et al. (2019).

tion, school type and geographical location. A further in-home interview was designed for a subsample of students randomly selected within schools with a clustered sampling design.[12] The in-home interview collected extensive demographic, socioeconomic, behavioral, cognitive, and health information about students and their families. The core-sample students were followed with later surveys in 1996 (Wave 2), 2002 (Wave 3), 2009 (Wave 4) and 2018 (Wave 5).

The results in this paper are obtained by combining Waves 1 and 4 of the Add Health data. In particular, I employ Wave 1 data to measure the main variables of interest: the percentage of missing fathers in a cohort within a school.[13] Wave 1 variables are also used as additional controls. Specifically, we can control for several individual characteristics, such as race-ethnicity, family structure and ability,[14] together with their socioeconomic status, proxied by parents' levels of education. Moreover, we can control for additional peer composition variables such as cohort racial and ethnic shares and the percentage of peers' parents with graduate degrees. One key control in this analysis is students' individual experiences of father absence. The detailed questions about family structure of Add Health in Waves 1,2, and 3 enable me to construct a measure of father absence throughout adolescence. All the regressions in this paper control for the number of years spent without a resident father until 16 years of age.[15] The long-term outcomes are obtained from Add Health Wave 4, which was completed when respondents were between 24 and 32 years old. This later survey provides information about completed levels of education, labor market trajectories, and economic and financial conditions together with mental health assessments.

[12] According to (Bifulco et al., 2011): "The median cohort within a school consists of 275 students, and the median number selected from a high school cohort for the longitudinal Wave 1 sample is 42 students."

[13] This variable is constructed using in-school census survey information, which provides the most precise information about students' cohort composition.

[14] Add Health provides a measure of ability with a Picture Vocabulary Test (PVT) score, a shorter version of the Peabody Picture Vocabulary Test (PPVT). Literature in psychology has shown that the PPVT is a reliable proxy of individual ability (Cools et al., 2019).

[15] The average years of father's absence in the Add Health sample is 3.63 years. The share of respondents that never experienced father absence while adolescent is 58.2%. Among those that grew up without a residential father, the median absence spell is 8 years.

Although approximately 20,000 students were included in the core sample of Wave 1, only 15,701 respondents were followed until Wave 4. I drop roughly 3,700 observations with missing information on the peer variable of interest, the long-term outcomes or the baseline set of control variables. Moreover, I exclude from the sample 609 students in schools with only two cohorts or that were single-sex institutions. As a last step, I drop 145 students belonging to cohorts with less than 20 students, corresponding to 1.47% of the sample.

This sample selection procedure returns a dataset of 9,697 observations sampled across 94 different schools. Table 2.1 provides descriptive statistics. The outcomes of interest are measured in Wave 4, when the respondents are adults. "*Years of Education*" is a continuous measure of educational achievement built using the highest self-reported academic degree, while "*College Graduation*" is a 0–1 indicator. The variable "*Employed*" is another indicator equal to one if the respondent is currently working any hours for pay. The continuous income measure, in logs, and the homeownership indicator are based on self-reports. Add Health also measures depression symptoms following the Center for Epidemiological Studies Depression (CES-D) scale. The variable is constructed by Add Health staff based on answers to a questionnaire included in the in-home survey. On average, individuals included in the sample experienced 2.67 years of father absence when they were younger than 16 years of age. African Americans make up 17% of the selected sample, together with 14% self-identified Hispanic individuals and 8% Asians. The remaining share comprises white respondents.

Growing up without a resident father is likely correlated with several socioeconomic variables. Table 1.2 provides suggestive evidence for this hypothesis splitting the sample between those with and without a resident father during 1994–1995, the time of the school census survey. The sample of respondents with missing fathers is disproportionately African American; moreover, these adolescents attended schools with higher Black presence and higher percentages of missing fathers. Wave 4 outcomes show relative socioeconomic disadvantage for respondents with missing fathers during adolescence. The

	Whole Sample	Resident Father	Non-Resident Father
Var. Measured when Adult,(W.4)			
Years of Education	14.30 (2.31)	14.42 (2.32)	13.93 (2.23)
College Graduation	0.34 (0.47)	0.36 (0.48)	0.27 (0.44)
Employed	0.84 (0.37)	0.85 (0.36)	0.81 (0.39)
Log. Total Income	10.28 (0.99)	10.33 (0.96)	10.13 (1.09)
Home Ownership	0.45 (0.50)	0.47 (0.50)	0.38 (0.49)
Depression Scale	2.50 (2.51)	2.43 (2.47)	2.75 (2.61)
Years of Father Absence	2.67 (4.83)	1.52 (3.89)	6.57 (5.63)
Var. Measured when Adolescent,(W.1)			
African American	0.17 (0.37)	0.12 (0.32)	0.34 (0.47)
Asian	0.08 (0.27)	0.08 (0.28)	0.05 (0.22)
Hispanic	0.14 (0.34)	0.15 (0.35)	0.11 (0.31)
Male	0.51 (0.50)	0.52 (0.50)	0.47 (0.50)
% Peers with Missing Fathers	0.23 (0.1)	0.23 (0.09)	0.26 (0.11)
% Peers African American	0.17 (0.22)	0.15 (0.2)	0.23 (0.26)
% Peers with Employed Father	0.95 (0.04)	0.95 (0.04)	0.95 (0.04)
No. Observations	9,697	6,910	2,787
No. Schools	94	94	94

Notes: The sample includes Add Health Wave 4 respondents. See section 2.3 for the sample selection procedure. The second and third columns split the sample according to whether respondents reported living with a residential father when adolescent (Wave 1 survey). Standard deviations are reported in parentheses.

Table 2.1 — Descriptive Statistics

sample of individuals living with their biological father during the 1994–1995 academic year exhibits higher levels of education, lower unemployment percentages, higher income levels, and higher rates of homeownership. They also report fewer depression symptoms according to the CES-D scale.

2.4 Empirical Strategy

The main challenge in estimating the long-term effects of exposure to non-traditional families is caused by the correlation between unobserved student characteristics and peer composition. It is reasonable to assume that residential sorting and school choice will create a strong association between students' individual features and their school cohort makeup. If students from single-mother families on average attend schools in less affluent neighborhoods, any detrimental effect of higher rates of missing fathers might be due to other dimensions of economic disadvantage. I exploit a source of exogenous variation in the percentage of non-resident fathers in a student's cohort following an established approach in the peer effects literature. In particular, the long-term effects are identified by variation in average peer characteristics across cohorts within schools. The assumption underlying this empirical strategy is that while parents select schools based on average school attributes, they cannot forecast cohort-to-cohort idiosyncratic variation in their child's grade. The estimated effects capture the association between cohort-to-cohort changes in long-term outcomes and cohort-to-cohort changes in missing fathers' shares within the same school environment. Intuitively, I am comparing the long-term outcomes of students from adjacent cohorts with similar attributes and identical school environments, except that in some cohorts missing-father families were more prevalent. The results will not capture any effect of school-wide student composition on students' long-term outcomes. It is also important to note that our peer variable of interest is pre-determined. Namely, long-term student outcomes cannot affect peers' family residential arrangements preventing the typical reflection problem (Manski, 1993) often harming identification in peer effects studies.

This identification strategy can be implemented by regressing long-term outcomes on the peer variable of interest, a school fixed effect, a cohort fixed effect together with additional controls. More specifically, the coefficients of interest are estimated using the following regression:

$$y_{i,g,s,t+1} = \alpha_s + \gamma_g + \beta MF_{i,g,s,t} + \theta X_{i,g,s,t} + \phi Z_{g,s,t} + \varepsilon_{i,g,s,t+1} \tag{2.1}$$

where $y_{i,g,s,t+1}$ is the outcome when adult for individual i in school s and cohort g. It includes α_s a school fixed effect, and γ_g a grade fixed effect. $MF_{i,g,s,t}$, the main variable of interest, represents the percentage of peers that do not live with their biological father. The variable is individual specific because it is computed as the average for the leave-one-out distribution, following Townsend (1994), Cools et al. (2019), Olivetti et al. (2018). The model includes controls for individual students' attributes $X_{i,g,s,t}$, together with a vector of peer characteristics $Z_{g,s,t}$. The main assumption required to interpret β's estimates as the causal effects of exposure to missing-father families is that the unobserved determinants of long-term outcomes are uncorrelated with $MF_{i,g,s,t}$. Namely, we assume that parents do not choose schools based on their child's cohort idiosyncratic characteristics. In order to allow for the possibility that parents can forecast trends in school composition, regression Equation 2.1 can be extended with a school-specific linear trend. Section 2.6 will also include the results of the following equation:

$$y_{i,g,s,t+1} = \alpha_s + \gamma_g + \delta_s g + \beta MF_{i,g,s,t} + \theta X_{i,g,s,t} + \phi Z_{g,s,t} + \varepsilon_{i,g,s,t+1} \tag{2.2}$$

where $\delta_s g$ is the additional school-specific linear trend. Equation (2) measures the causal effect of interest by exploiting school-specific deviations in the percentage of missing fathers from a long-term linear trend. Section 2.5 provides evidence that the variation in the peer variable of interest can be considered exogenous, and it is appropriate to identify the presence of peer effects. However, this empirical strategy will not provide a correct characterization of the effects of one's own father's absence. Growing up in a single-mother household is likely correlated with unobservable determinants of later-life outcomes. Because it is particularly challenging to obtain a source of exogenous variation in own father's absence, I will not focus on the results for own father's absence in the results section.[16] The goal of the empirical strategy is instead to credibly measure the spillover effects from adolescent peers' family structure on adult outcomes.[17]

What does the β, the reduced-form coefficient of interest in Equations 2.1 and 2.2, measure? In order to clarify this, we can follow Bifulco et al. (2011) and consider the following toy model for $y_{i,g}$, the long-term outcome of an individual i belonging to group g:

$$\begin{cases} y_{i,g} = \delta_y z_{i,g} + \beta_y X_i + \phi_y \bar{X}_{i,g} + \varphi_y \bar{z}_{i,g} + \mu_i & \text{(2.3a)} \\ z_{i,g} = \beta_z X_i + \phi_z \bar{X}_{i,g} + \varphi_z \bar{z}_{i,g} + \varepsilon_i & \text{(2.3b)} \end{cases}$$

where $\bar{X}_{i,g}$ and $\bar{z}_{i,g}$ are leave-one-out means of peer attributes. Individual i long-term outcome depends on individual characteristics X_i, an intermediate outcome $z_{i,g}$, an average of peer observable attributes $\bar{X}_{i,g}$, and an average of peer short-term outcomes $\bar{z}_{i,g}$. The intermediate outcome $z_{i,g}$ in turn depends on individual traits X_i, peer characteristics $\bar{X}_{i,g}$ and peer contemporaneous choices $\bar{z}_{i,g}$. In principle, the goal of the empirical strategy in a peer-effect investigation is to identify ϕ_y. Starting from these two definitions, I can obtain reduced form equations for $z_{i,g}$ and $\bar{z}_{i,g}$ that depend only on individual and peers' observables characteristics. The analytical steps can be found in Appendix Section 2.8. Replacing the reduced-form definitions of $z_{i,g}$ and $\bar{z}_{i,g}$ in (2.3a) it is possible to express β in structural terms according to the parameters of the toy model. In particular, for sufficiently large school cohorts, the following equation between the reduced-form coefficient and the structural parameters holds:

$$\beta = \phi_y + \frac{\phi_z(\delta_y + \varphi_y) + \beta_z(\delta_y \varphi_z + \varphi_y)}{1 - \varphi_z} \tag{2.4}$$

[16] All the specifications in Section 2.6 include a control for the time spent without a father figure throughout adolescence.

[17] Balance checks in Section 2.5 shows that own father's absence and peers' share of missing fathers are not correlated conditional on the fixed effects outlined in Equations 2.1 and 2.2.

The parameter of interest β will include the desired structural parameter ϕ_y, the direct effect of peers' observables on individual long-term outcomes, together with an exposure term. The latter term is a complex combination of peers' exogenous and endogenous effects on intermediate outcomes, ϕ_z and φ_z respectively, and the effect of peers' choices on the adult outcome, φ_y. These effects are further mediated by the relationship between the intermediate and long term outcome δ_y. If we are willing to assume that the long term and intermediate outcomes $y_{i,g}$ and $z_{i,g}$ are exogenous with respect to average peers' outcomes $\bar{z}_i g$, the intuition regarding the components captured by β can be sharper. Thus, we obtain the following:

$$\beta = \phi_y + \phi_z \delta_y \tag{2.5}$$

Thanks to the additional assumptions, the exposure term can be reduced to $\phi_z \delta_y$, the interaction between short-term peer effects and the persistence of intermediate choices on later life outcomes δ_y. In principle, we would be interested in measuring ϕ_y separately, but the proposed strategy does not allow that. The estimated coefficient β will represent the effect of being exposed to peers with non-traditional family arrangements rather than the effect of peer family structure. Such limitation is widespread among peer effect studies and would not be solved even with random assignment. Even if the empirical strategy cannot directly identify the structural parameter ϕ_y, the estimated effect is still relevant to understanding the role of non-traditional families in the reproduction of inequality.

2.5 Validation of Empirical Strategy

The empirical strategy proposed to measure the long-term effects of exposure to single-mother families is credible if two conditions are met.[18] First, there should be enough residual variation in the percentage of missing fathers within schools to provide sufficient statistical power. Second, the causal interpretation of our results requires that school specific

[18]These validation exercises follow Lavy and Schlosser (2011) and Bifulco et al. (2011)

cohort-to-cohort changes in the share of missing fathers are uncorrelated with unobservable determinants of long-term socioeconomic outcomes. Since it is impossible to test this exogeneity requirement directly, I will provide evidence that my primary variable of interest is orthogonal to a wide array of background characteristics when controlling for the fixed effect required by the empirical framework. Table 2.2 reports averages and standard deviations for the main variable of interest: the cohort percentage of peers living without a resident father. When I remove variation due to school sorting and time fixed effects, there is a 70% reduction in the variable's standard deviation; adding a school-specific linear time trend further reduces the standard deviation to roughly 20% of the original raw variation. While the reduction in the amount of available variation is particularly steep, Table 2.2 shows that similar papers in the peer effects literature have successfully estimated long-term effects of alternative peer attributes with analogous amounts of residual variation.

	Mean	S.D.	Min	Max
% Peers without a Resident Father				
Raw Cohort Variable	0.24	0.10	0.005	0.57
Residuals net of cohort and school fixed effects	0.00	0.03	-0.11	0.17
Residuals net of cohort, school fixed effects and school trends	0.00	0.02	-0.08	0.09
% Peers Black or Hispanic *Bifulco et al. (2011)*				
Raw Cohort Variable	0.37	0.31	0.00	1.00
Residuals net of cohort and school fixed effects	0.00	0.03	-0.14	0.18
Residuals net of cohort, school fixed effects and school trends	0.00	0.02	-0.14	0.13
% Peers with Working Mothers *Olivetti et al. (2018)*				
Raw Cohort Variable	0.81	0.07	0.51	0.97
Residuals net of cohort and school fixed effects	0.00	0.03	-0.18	0.14
Residuals net of cohort, school fixed effects and school trends	0.00	0.02	-0.11	0.10

Notes: The table reports descriptive statistics for cohort variables of interest. Residuals are obtained regressing the variable of interest on cohort fixed effects, school fixed effects and school specific school trends. See section 2.3 for the sample selection procedure. The regressions include longitudinal sampling weights.

Table 2.2 — Residual Variation in Peer Measures

The "balancing tests" are designed to show that the variation in the percentage of missing fathers is not correlated with alternative student characteristics. Evidence supporting the orthogonality of the regressor of interest mitigates the concerns about endogenous selection on unobservable determinants.[19] The results of these "balancing tests" (reported in Table 2.3) are obtained by regressing each variable on the cohort share of missing fathers. Columns (2) and (3) add cohort-grade fixed effects, school fixed effects and school-specific linear time trends. Column (1) of Table 2.3 shows that the share of missing fathers in a student's cohort is unconditionally correlated with several different background characteristics, both at the individual and at the cohort level. However, these associations mostly disappear when the fixed effects required by the empirical strategy are introduced. Notably, none of the coefficients in column (3) are statically different from zero, providing evidence that the source of variation of this empirical strategy can be considered exogenous.

[19] Altonji et al. (2005)

	% Peers without Resident Father		
	(1)	(2)	(3)
Individual Variables			
Male	0.002***	-0.001	-0.002
	(0.001)	(0.001)	(0.002)
Years of Father Absence	0.035***	-0.017	-0.010
	(0.012)	(0.021)	(0.020)
Black	0.017***	0.003***	0.002
	(0.002)	(0.001)	(0.001)
Hispanic	0.005	-0.0001	0.0005
	(0.003)	(0.001)	(0.002)
Parent with a Post-Graduate Degree	-0.006***	-0.001	0.01
	(0.002)	(0.002)	(0.002)
Cohort Variables			
% Peers Black	0.014***	0.001**	0.001
	(0.001)	(0.001)	(0.001)
% Peers Hispanic	0.004**	-0.00002	-0.0003
	(0.002)	(0.0004)	(0.0004)
% Peers Parent with a Post-Graduate Degree	-0.002***	-0.001	-0.001
	(0.001)	(0.001)	(0.001)
Grade Fixed Effects	No	Yes	Yes
School Fixed Effects	No	Yes	Yes
School Trends	No	No	Yes

Notes: The table reports correlations between the variables on the left and the percentage of peers without a resident father. The coefficients are obtained regressing the variable of interest on the peer percentage, race-ethnicity dummies, maternal education and family structure. The second and third columns report results when school fixed effects, cohort fixed effects and school specific linear trends are included. Standard errors are clustered at the school level. See section 2.3 for the sample selection procedure. The regressions include longitudinal sampling weights. Significance: * 0.10 ** 0.05 *** 0.01.

Table 2.3 — Balancing Tests for Cohort Composition Measures

2.6 Results

Table 2.4 reports the estimated effect of the cohort share of missing fathers on eight different long-term outcomes drawn from the fourth wave of Add Health. All the specifications include controls for individual race or ethnicity, adolescent verbal ability, one's mother's education, and census region. They also include controls for the cohort makeup in terms of race-ethnicity and socioeconomic status. All the standard errors are clustered at the school level.[20] The first column shows unconditional correlations between dependent variables and the peer share of missing fathers. The results in this column confirm the association between higher rates of father absence at the local level and socioeconomic disadvantage, even for long-term outcomes. Higher percentages of single-mother households are correlated with lower graduation rates from high school and college, lower income levels, lower homeownership rates and higher incidence of depression symptoms. In some cases, the magnitude of the estimated relationship is sizable. A one standard deviation increase in the percentage of missing fathers corresponds to a 3-percentage-point reduction in graduation rates from college, which is equivalent to an 8.8% drop with respect to the sample baseline. An analogous increase in the peer variable of interest would lead to a 3-percentage-point reduction in homeownership rates and a 0.16 increase on the CES-D depression scale, corresponding to a -6.6% and a +6.4% change with respect to sample averages.

Introducing school and grade fixed effects enables one to switch from a descriptive approach to a causal one. Most of the correlations reported in column (1) vanish when we move to column (2), which reports the results when I include the fixed effects required by the empirical strategy for a causal interpretation. Introducing school-specific trends in column (3) does not substantially change the results of column (2). The empirical strategy does not find any causal effect of exposure to peers from single-mother households on long-term educational achievements, labor market outcomes, total income, or mental

[20] The clustering level is chosen according to the school-based sampling structure of the Add Health dataset.

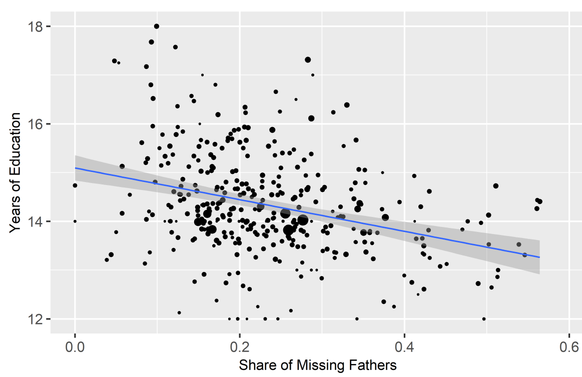
Dependent Variable	% Peers without Resident Father		
	(1)	(2)	(3)
Years of Education	-0.016**	0.005	0.004
$\bar{Y} = 14.30$	(0.007)	(0.010)	(0.010)
High School Graduate	-0.002*	-0.001	0.0004
$\bar{Y} = 0.94$	(0.001)	(0.001)	(0.001)
College Graduate	-0.003**	0.003	0.003
$\bar{Y} = 0.34$	(0.001)	(0.002)	(0.002)
Employed	-0.001	0.002	0.001
$\bar{Y} = 0.84$	(0.001)	(0.002)	(0.002)
Total Income, Logs	-0.009***	0.002	0.002
$\bar{Y} = 10.28$	(0.003)	(0.004)	(0.005)
Home Owner	-0.003*	-0.002	-0.004
$\bar{Y} = 0.45$	(0.002)	(0.002)	(0.003)
Depression Scale	0.016**	0.021**	0.007
$\bar{Y} = 2.5$	(0.008)	(0.008)	(0.012)
Grade Fixed Effects	No	Yes	Yes
School Fixed Effects	No	Yes	Yes
School Trends	No	No	Yes

Notes: The table reports the estimated effects of the percentage of peers without a resident father on the outcomes of interest. Outcomes are measured with Wave 4 of the Add Health. The second and third columns report results when school fixed effects, cohort fixed effects and school specific linear trends are included. Sample averages for each outcome are reported below the variable names. Standard errors are clustered at the school level. See section 2.3 for the sample selection procedure. The regressions include longitudinal sampling weights. Significance: * 0.10 ** 0.05 *** 0.01.

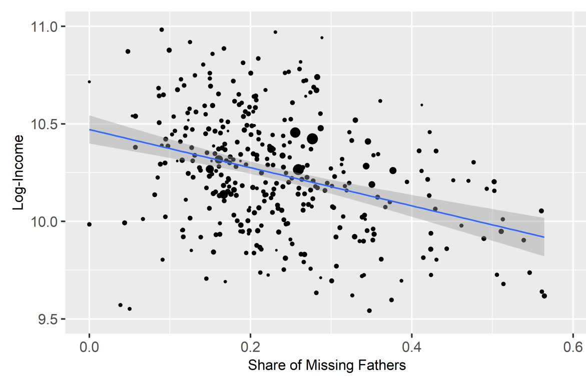
Table 2.4 — Effects of Exposure to Missing-Father Peers

health. The discrepancies between correlational and causal perspectives in characterizing the role of missing fathers in shaping adult outcomes can be confirmed graphically. Figure 2.1 shows clear correlations between the share of missing fathers in a school cohort and years of education, total income, and the CES-D depression scale. Such associations disappear in Figure 2.2, which contains correlation plots between the share of missing fathers and the dependent variables' residuals when I control for school and cohort fixed effects. The comparison between the two figures confirms once more that, according to the Add Health data, the relationship between father absence and socioeconomic disadvantage is a result of students' sorting across schools, and it should not have a causal interpretation.

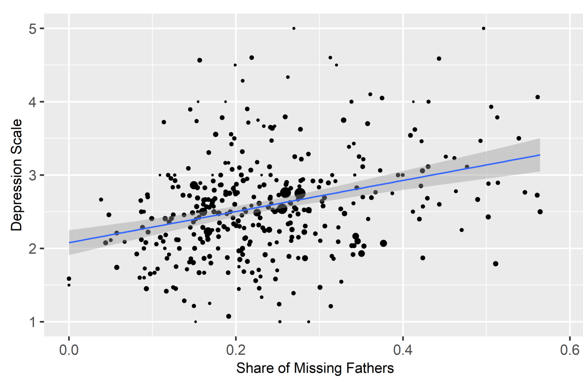
Raw Dependent Variables and Share of Missing Fathers



(a) *Years of Education*



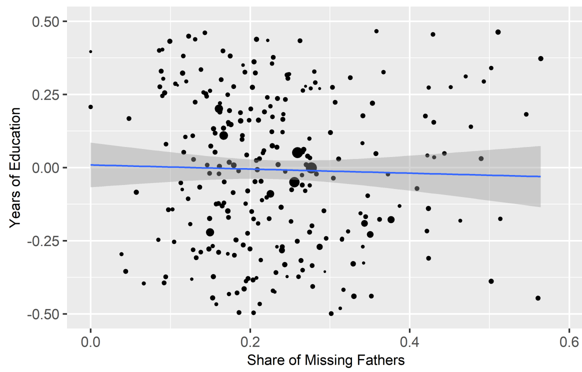
(b) *Total Income, Logs*



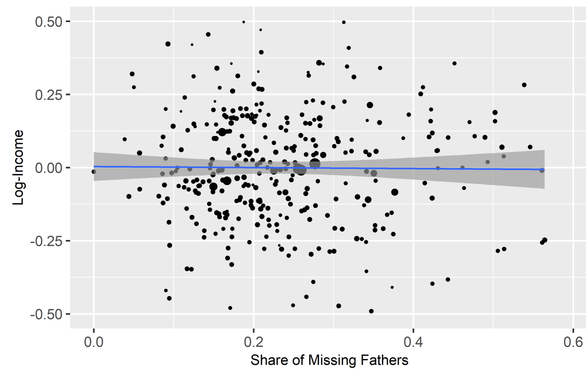
(c) *Depression Scale*

Figure 2.1 — The Figure shows correlations between three outcomes of interest and the share of missing fathers. Every point is a cohort within a school, and the cohort's student size gives the size of each point. The blue line is a linear model fitted to minimize the mean squared errors, and the shaded area represents its 95% confidence interval. Outcomes are measured with Wave 4 of the Add Health. See section 2.3 for the sample selection procedure.

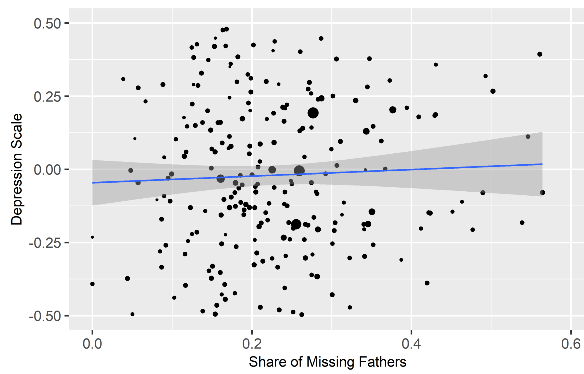
Residualized Dependent Variables and Share of Missing Fathers



(a) *Years of Education*



(b) *Total Income, Logs*



(c) *Depression Scale*

Figure 2.2 — The Figure shows correlations between residualized outcomes of interest and the share of missing fathers. The residualized outcomes are obtained by regressing the variable on a school fixed effect, cohort fixed effect and school-specific time trend. Every point is a cohort within a school, and the cohort's student size gives the radius of each point. The blue line is a linear model fitted to minimize mean squared errors, and the shaded area represents its 95% confidence interval. Outcomes are measured with Wave 4 of the Add Health. See section 2.3 for the sample selection procedure.

The mute results in the general Add Health sample might mask heterogeneous results according to demographic characteristics. Chetty et al. (2020) provide a suggestive example that uses IRS administrative data to show how father presence at the census tract level is a strong predictor of higher adult income for African American men but not for Black women. Table 2.5 replicates the last specification of Table 2.4, but splits the sample by gender. While the majority of the null results are confirmed by the heterogeneous effects of Table 2.5, I find statistically significant and sizable impacts of the share of missing fathers on two outcomes. A one standard deviation increase in the peer percentage of missing fathers leads to a 7-percentage-point reduction in homeownership rates for men in the sample, which is a -16% change with respect to the baseline. Moreover, the same increase in the share of missing fathers causes a 0.35 increase in the CES-D depression scale (a 15% increase). Splitting the sample according to race and ethnicity, or even creating subsamples combining these two demographic attributes, does not highlight any further heterogeneous effect of local father absence.[21]

A further hypothesis is that missing fathers in one's social group might have a different effect according to the peer's gender. For example Autor et al. (2019) and Bertrand and Pan (2013) provide evidence that single-mother households are particularly harmful for boys' early development. It could be the case that the percentage of boys from single-mother households has a different impact than the share of girls without a residential father. Table 2.6 tests this hypothesis by replicating Equation 2.2. However, in this case we regress outcomes on separate percentages of missing fathers for boys and girls. The results do not suggest differential mechanisms according to peers' gender, confirming the null baseline results presented in Table 2.4.

[21] Appendix Table 2.7 replicates Table 2.4 for these subsamples. The only additional insight from this heterogeneity exercise is that the detrimental effect of fathers' absence on men's depression symptoms is driven by African-American respondents rather than white men.

Dependent Variable	% Peers without Resident Father	
	Male	Female
Years of Education	0.005 (0.014) [14.05]	0.001 (0.014) [14.57]
High School Graduate	-0.001 (0.002) [0.93]	0.002 (0.001) [0.95]
College Graduate	0.006** (0.003) [0.30]	-0.0005 (0.003) [0.38]
Employed	0.003 (0.003) [0.89]	-0.001 (0.003) [0.79]
Total Income, Logs	0.002 (0.006) [10.44]	-0.001 (0.008) [10.10]
Home Owner	-0.007** (0.003) [0.45]	-0.001 (0.004) [0.46]
Depression Scale	0.035** (0.017) [2.25]	-0.023 (0.016) [2.76]
Number of Observations	4,946	5,198

Notes: The table reports the estimated effects of the percentage of peers without a resident father on outcomes of interest. The first column reports results for the male sample, while the second column restricts the sample to women. The regression includes school fixed effects, cohort fixed effects and school specific linear time trends. Outcomes are measured with Wave 4 of the Add Health. Standard errors, in parentheses, are clustered at the school level. Baseline averages are reported in square brackets. See section 2.3 for the sample selection procedure. The regressions include longitudinal sampling weights. Significance: * 0.10 ** 0.05 *** 0.01.

Table 2.5 — Effects of Exposure to Missing-Father Peers, By Gender

	Dependent Variable			
	Years of Educ.	High School Graduate	College Graduate	Employed
% Male Peers without Res. Father	0.006 (0.008)	0.0003 (0.001)	0.001 (0.001)	0.0001 (0.001)
% Female Peers without Res. Father	-0.010 (0.008)	-0.001 (0.001)	0.001 (0.001)	0.0003 (0.002)
	Dependent Variable			
	Total Income, Logs	Home Owner	Depression Scale	
% Male Peers without Res. Father	0.004 (0.003)	-0.005** (0.002)	0.009 (0.010)	
% Female Peers without Res. Father	-0.0004 (0.003)	-0.001 (0.002)	0.002 (0.009)	

Notes: The table reports the estimated effects of the percentages of peers without a resident father on the outcomes of interest. The peer variables are gender-specific. They report the percentage of boys and girls without a resident father separately. The regression includes school fixed effects, cohort fixed effects and school specific linear time trends. Outcomes are measured with Wave 4 of the Add Health. Standard errors are clustered at the school level. See section 2.3 for the sample selection procedure. The regressions include longitudinal sampling weights. Significance: * 0.10 ** 0.05 *** 0.01.

Table 2.6 — Effects of Exposure to Missing-Father Boys and Girls

2.7 Conclusion

This paper investigates the long-term effects of exposure to peers living in single-mother households. The empirical strategy credibly isolates an exogenous source of variation in the cohort share of missing fathers, thus providing sufficient statistical power and allowing for a causal interpretation of the coefficients. The results are obtained using Add Health data, a long-term survey that followed more than 15,000 respondents from adolescence to adulthood. This survey precisely recorded student cohort composition when in school and a wide range of adult socioeconomic outcomes. We find extensive evidence of a correlation between higher rates of father absence at the local level and long-term socioeconomic disadvantage. However, the correlational evidence is not confirmed with a causal framework. Indeed, my estimates do not detect causal effects of exposure to peers from single-mother households on long-term educational achievements, labor market outcomes, total income or mental health in the Add Health general sample. There is suggestive evidence of detrimental effects of local father absence on homeownership rates and depression symptoms for men in the sample, but no impact is found for women.

2.8 Appendix

The appendix provides the analytical steps required to establish a correspondence between the reduced form parameter β ant the toy model's structural parameters. This allows to have a better understanding of how to interpret the future results of the analysis. We deal with the model proposed in section 2.4.

$$\begin{cases} y_{i,g} = \beta_y X_i + \delta_y z_{i,g} + \phi_y \bar{X}_{i,g} + \varphi_y \bar{z}_{i,g} + \mu_i & \text{(2.6a)} \\ z_{i,g} = \beta_z X_i + \phi_z \bar{X}_{i,g} + \varphi_z \bar{z}_{i,g} + \varepsilon_i & \text{(2.6b)} \end{cases}$$

Given that equation 2.6b is valid for all the peers of i we can expand the leave-one out mean in the following way

$$\bar{z}_{i,g} = \frac{1}{n-1}\sum_{j\neq i} z_{j,g} = \frac{\beta_z}{n-1}\sum_{j\neq i} X_{j,g} + \frac{\phi_z}{(n-1)^2}\sum_{j\neq i}\sum_{l\neq j} X_{l,g} + \frac{\varphi_z}{(n-1)^2}\sum_{j\neq i}\sum_{l\neq j} z_{l,g} + \frac{1}{n-1}\sum_{j\neq i}\varepsilon_{j,g} \tag{2.7}$$

The fundamental algebraic step is to break the double sum in two components according to

$$\sum_{j\neq i}\sum_{l\neq j} y_l = (n-1)y_i + (n-2)\sum_{j\neq i} y_j \tag{2.8}$$

where n is the size of i' s peer group.

Once we use this summation property, we can write $\bar{z}_{i,g}$ in terms of X_i, z_i and $\bar{X}_{i,g}$

$$\bar{z}_{i,g} = \left(\frac{1}{(n-1)-(n-2)\varphi_z}\right)\left[\phi_z X_{i,g} + \varphi_z z_{i,g} + \left[(n-1)\beta_z + (n-2)\phi_z\right]\bar{X}_{i,g} + \sum_{j\neq i}\varepsilon_{j,g}\right]$$

Substituting this into 2.6b we finally obtain a closed form for Λ and Γ

$$z_{i,g} = \underbrace{\left(\frac{(n-1-(n-2)\varphi_z)\beta_z + \phi_z \varphi_z}{n-1-(n-2)\varphi_z - \varphi_z^2} \right)}_{\Lambda} X_{i,g} + \underbrace{\left(\frac{(n-1)\left[\phi_z + \beta_z \varphi_z\right]}{n-1-(n-2)\varphi_z - \varphi_z^2} \right)}_{\Gamma} \bar{X}_{i,g} + \tilde{\varepsilon}_{i,g} \quad (2.9)$$

In order to gain some intuition it is useful to consider the case where n is big enough so that we can approximate $n-2 \approx n-1 \approx n$ and we can disregard coefficients that are not multiplied by n. We obtain the following approximation

$$\Lambda \approx \beta_z \quad (2.10)$$

$$\Gamma \approx \frac{\phi_z + \beta_z \varphi_z}{1 - \varphi_z} \quad (2.11)$$

If we have an identification strategy allowing us to estimate the reduced form equation 2.9 free of bias, for cohort size n big enough, then Γ would capture the exogenous peer effect of interest φ_z only if the role of peer outcomes were negligible ($\varphi_z \approx 0$).

Dependent Variable	% Peers without Resident Father					
	A.A.	Whites	A.A.Men	A.A. Women	White Men	White Women
Years of Education	0.009	0.001	0.030	-0.031	0.002	-0.012
	(0.022)	(0.013)	(0.033)	(0.0.026)	(0.019)	(0.017)
High School Graduate	0.002	0.0002	0.002	0.0005	-0.002	0.001
	(0.004)	(0.002)	(0.006)	(0.0.004)	(0.002)	(0.002)
College Graduate	0.002	0.002	0.009	-0.007	0.006	-0.002
	(0.001)	(0.004)	(0.007)	(0.005)	(0.005)	(0.004)
Employed	-0.0003	0.002	0.0004	-0.004	0.003	0.001
	(0.005)	(0.003)	(0.007)	(0.006)	(0.004)	(0.004)
Total Income, Logs	0.005	-0.004	-0.016	0.016	0.003	-0.011
	(0.014)	(0.006)	(0.020)	(0.017)	(0.007)	(0.008)
Home Owner	0.009	-0.009	0.008	0.013	-0.013**	-0.003
	(0.005)	(0.004)	(0.006)	(0.007)	(0.004)	(0.005)
Depression Scale	0.051**	-0.010	0.082**	-0.001	0.016	-0.042
	(0.025)	(0.017)	(0.036)	(0.041)	(0.021)	(0.024)
Number of Observations	2,138	5,091	879	1,259	2,416	2,675

Notes: The table reports the estimated effects of the percentage of peers without a resident father on outcomes of interest for several different samples. The table's header indicates the definition of the sample. The regression includes school fixed effects, cohort fixed effects and school specific linear time trends. Outcomes are measured with Wave 4 of the Add Health. Standard errors are clustered at the school level. See section 2.3 for the sample selection procedure. The regressions include longitudinal sampling weights. Significance: * 0.10 ** 0.05 *** 0.01.

Table 2.7 — Effects of Exposure to Missing-Father Peers, By Gender

Bibliography

Aaronson, Daniel, Jacob Faber, Daniel Hartley, Bhashkar Mazumder, and Patrick Sharkey (2021a) "The long-run effects of the 1930s HOLC "redlining" maps on place-based measures of economic opportunity and socioeconomic success," *Regional Science and Urban Economics*, Vol. 86, p. 103622, URL: https://www.sciencedirect.com/science/article/pii/S0166046220303070, DOI: http://dx.doi.org/https://doi.org/10.1016/j.regsciurbeco.2020.103622.

Aaronson, Daniel, Daniel Hartley, and Bhashkar Mazumder (2021b) "The Effects of the 1930s HOLC "Redlining" Maps," *American Economic Journal: Economic Policy*, Vol. 13, pp. 355–92, URL: https://www.aeaweb.org/articles?id=10.1257/pol.20190414, DOI: http://dx.doi.org/10.1257/pol.20190414.

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*, Vol. 105, pp. 493–505.

Abadie, Alberto and Javier Gardeazabal (2003) "The economic costs of conflict: A case study of the Basque Country," *American economic review*, Vol. 93, pp. 113–132.

Agostinelli, Francesco and Matthew Wiswall (2016) "Estimating the Technology of Children's Skill Formation," Working Paper 22442, National Bureau of Economic Research.

Altonji, Joseph G, Todd E Elder, and Christopher R Taber (2005) "An evaluation of instrumental variable strategies for estimating the effects of catholic schooling," *Journal of Human resources*, Vol. 40, pp. 791–821.

Anders, John Paul (2019) "The Long Run Effects of Transformational Federal Policies: Redlining, the Affordable Care Act and Head Start."

Appel, Ian and Jordan Nickerson (2016) "Pockets of poverty: The long-term effects of redlining," *Available at SSRN 2852856*.

Arrow, Kenneth J. (1973) *The Theory of Discrimination*, pp. 3–33: Princeton University Press, URL: http://www.jstor.org/stable/j.ctt13x10hs.5.

Autor, David, David Dorn, and Gordon Hanson (2018) "When work disappears: Manufacturing decline and the falling marriage-market value of young men," *American Economic Review: Insights*.

Autor, David, David Figlio, Krzysztof Karbownik, Jeffrey Roth, and Melanie Wasserman (2019) "Family Disadvantage and the Gender Gap in Behavioral and Educational Outcomes," *American Economic Journal: Applied Economics*, Vol. 11, pp. 338–381.

Becker, Gary S et al. (1971) "The Economics of Discrimination," *University of Chicago Press Economics Books*.

Becker, Gary S (1973) "A theory of marriage: Part I," *Journal of Political economy*, Vol. 81, pp. 813–846.

Bertrand, M. and J. Pan (2013) "The Trouble with Boys: Social Influences and the Gender Gap in Disruptive Behavior," *American Economic Journal-Applied Economics*, Vol. 5, pp. 32–64.

Bertrand, Marianne, Dolly Chugh, and Sendhil Mullainathan (2005) "Implicit Discrimination," *American Economic Review*, Vol. 95, pp. 94–98, URL: https://www.aeaweb.org/articles?id=10.1257/000282805774670365, DOI: http://dx.doi.org/10.1257/000282805774670365.

Bertrand, Marianne and Esther Duflo (2017) "Field experiments on discrimination," *Handbook of economic field experiments*, Vol. 1, pp. 309–393.

Bifulco, R., J. M. Fletcher, and S. L. Ross (2011) "The Effect of Classmate Characteristics on Post-Secondary Outcomes: Evidence from the Add Health," *American Economic Journal-Economic Policy*, Vol. 3, pp. 25–53.

Birch, Colin P.D., Sander P. Oom, and Jonathan A. Beecham (2007) "Rectangular and hexagonal grids used for observation, experiment and simulation in ecology," *Ecological Modelling*, Vol. 206, pp. 347–359, URL: https://www.sciencedirect.com/science/article/pii/S0304380007001949, DOI: http://dx.doi.org/https://doi.org/10.1016/j.ecolmodel.2007.03.041.

Borusyak, Kirill and Peter Hull (2020) "Non-Random Exposure to Exogenous Shocks: Theory and Applications," Working Paper 27845, National Bureau of Economic Research.

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

——— (2013) "Racial Residential Segregation in American Cities," Working Paper 19045, National Bureau of Economic Research.

——— (2016) *Competition in the Promised Land*: Princeton University Press.

Breiman, Leo (2001) "Random forests," *Machine learning*, Vol. 45, pp. 5–32.

Calvo-Armengol, Antoni, Eleonora Patacchini, and Yves Zenou (2009) "Peer Effects and Social Networks in Education," *Review of Economic Studies*, Vol. 76, pp. 1239–1267.

Cavanagh, Shannon E. and Paula Fomby (2011) "Family Instability, School Context, and the Academic Careers of Adolescents," *Sociology of Education*, Vol. 85, pp. 81–97.

Charles, Kerwin Kofi and Ming Ching Luoh (2010) "Male incarceration, the marriage market, and female outcomes," *The Review of Economics and Statistics*, Vol. 92, pp. 614–627.

Chetty, Raj, Nathaniel Hendren, Maggie R Jones, and Sonya R Porter (2019) "Race and economic opportunity in the United States: An intergenerational perspective," *NBER Working Paper No. 24441*.

——— (2020) "Race and economic opportunity in the United States: An intergenerational perspective," *The Quarterly Journal of Economics*, Vol. 135, pp. 711–783.

Chyn, Eric (2018) "Moved to opportunity: The long-run effects of public housing demolition on children," *American Economic Review*, Vol. 108, pp. 3028–56.

Coates, Ta-Nehisi (2014) "The Case for Reparations," *Atlantic*.

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

Cools, Angela, Raquel FernÃ¡ndez, and Eleonora Patacchini (2019) "Girls, Boys, and High Achievers," *NBER Working Paper No. 25763*.

Cunha, Flavio and James Heckman (2007) "The technology of skill formation," *American Economic Review*, Vol. 97, pp. 31–47.

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

Del Boca, Daniela, Christopher Flinn, and Matthew Wiswall (2013) "Household choices and child development," *Review of Economic Studies*, Vol. 81, pp. 137–185.

Faber, Jacob W. (2020) "We Built This: Consequences of New Deal Era Intervention in America' s Racial Geography," *American Sociological Review*, Vol. 85, pp. 739–775, URL: https://doi.org/10.1177/0003122420948464, DOI: http://dx.doi.org/10.1177/0003122420948464.

Fang, Hanming and Andrea Moro (2011) "Theories of statistical discrimination and affirmative action: A survey," *Handbook of social economics*, Vol. 1, pp. 133–200.

Federal Home Loan Bank Board (1937) "Fifth Annual Report of the Federal Home Loan Bank Board, For the Period July 1, 1936 through June 30, 1937."

Fishback, Price V, Alfonso Flores-Lagunes, William C. Horrace, Shawn Kantor, and Jaret Treber (2011) "The influence of the Home Owners' loan corporation on housing markets during the 1930s," *The Review of Financial Studies*, Vol. 24, pp. 1782–1813.

Fishback, Price V, Jessica LaVoice, Allison Shertzer, and Randall Walsh (2020) "The HOLC Maps: How Race and Poverty Influenced Real Estate Professionals' Evaluation of Lending Risk in the 1930s," Working Paper 28146, National Bureau of Economic Research.

Fishback, Price V, Jonathan Rose, Kenneth A Snowden, and Thomas Storrs (2021) "New Evidence on Redlining by Federal Housing Programs in the 1930s,"Technical report, National Bureau of Economic Research.

Fruehwirth, Jane Cooley, Sriya Iyer, and Anwen Zhang (2019) "Religion and depression in adolescence," *Journal of Political Economy*, Vol. 127, pp. 1178–1209.

Fuster, Andreas, Paul Goldsmith-Pinkham, Tarun Ramadorai, and Ansgar Walther (2021) "Predictably Unequal? The Effects of Machine Learning on Credit Markets," *Journal of Finance, Forthcoming*.

Gaydosh, Lauren and Kathleen Mullan Harris (2018) "Childhood family instability and young adult health," *Journal of health and social behavior*, Vol. 59, pp. 371–390.

Glaeser, Edward and Jacob Vigdor (2012) "The end of the segregated century," *Manhattan Institute for Policy Research, January*, pp. 23–26.

Harris, Kathleen Mullan, Carolyn Tucker Halpern, Eric A Whitsel, Jon M Hussey, Ley A Killeya-Jones, Joyce Tabor, and Sarah C Dean (2019) "Cohort Profile: The National Longitudinal Study of Adolescent to Adult Health (Add Health)," *International Journal of Epidemiology*, Vol. 48, pp. 1415–1415k, URL: `https://doi.org/10.1093/ije/dyz115`, DOI: `http://dx.doi.org/10.1093/ije/dyz115`.

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

——— (2005) "Residential security maps and neighborhood appraisals: the Home Owners' Loan Corporation and the case of Philadelphia," *Social Science History*, Vol. 29, pp. 207–233.

Hoxby, Caroline (2000) "Peer effects in the classroom: Learning from gender and race variation,"Technical report, National Bureau of Economic Research.

Ihlanfeldt, Keith R and Benjamin Scafidi (2002) "Black self-segregation as a cause of housing segregation: Evidence from the multi-city study of urban inequality," *Journal of Urban Economics*, Vol. 51, pp. 366–390.

Jackson, Kenneth T (1980) "Race, ethnicity, and real estate appraisal: The Home Owners Loan Corporation and the federal housing administration," *Journal of Urban History*, Vol. 6, pp. 419–452.

Jones-Correa, Michael (2000) "The origins and diffusion of racial restrictive covenants," *Political Science Quarterly*, Vol. 115, pp. 541–568.

Khan, Lisa B. and Priyanka Anand (2018) "Peer Effects of Teen Pregnancy on Sexual Behavior," *Working Paper*.

Krimmel, Jacob (2018) "Persistence of prejudice: Estimating the long term effects of redlining."

Krysan, Maria and Reynolds Farley (2002) "The residential preferences of blacks: Do they explain persistent segregation?" *Social Forces*, Vol. 80, pp. 937–980.

Lang, Kevin and Ariella Kahn-Lang Spitzer (2020) "Race Discrimination: An Economic Perspective," *Journal of Economic Perspectives*, Vol. 34, pp. 68–89, URL: `https://www.aeaweb.org/articles?id=10.1257/jep.34.2.68`, DOI: `http://dx.doi.org/10.1257/jep.34.2.68`.

Lavy, Victor and AnalÃa Schlosser (2011) "Mechanisms and Impacts of Gender Peer Effects at School," *American Economic Journal: Applied Economics*, Vol. 3, pp. 1–33.

Lee, D. and S. McLanahan (2015) "Family Structure Transitions and Child Development: Instability, Selection, and Population Heterogeneity," *American Sociological Review*, Vol. 80, pp. 738–763.

Liberman, Andres, Christopher Neilson, Luis Opazo, and Seth Zimmerman (2018) "The Equilibrium Effects of Information Deletion: Evidence from Consumer Credit Markets," Working Paper 25097, National Bureau of Economic Research.

Light, Jennifer S (2010) "Nationality and neighborhood risk at the origins of FHA underwriting," *Journal of Urban History*, Vol. 36, pp. 634–671.

Logan, John R and Weiwei Zhang (2018) "Developing GIS Maps for US Cities in 1930 and 1940," in *The Routledge Companion to Spatial History*: Routledge, pp. 229–249.

Logan, Trevon D. and John M. Parman (2017) "The National Rise in Residential Segregation," *The Journal of Economic History*, Vol. 77, DOI: http://dx.doi.org/10.1017/S0022050717000079.

Ludwig, Jens and Sendhil Mullainathan (2021) "Fragile Algorithms and Fallible Decision-Makers: Lessons from the Justice System," Working Paper 29267, National Bureau of Economic Research.

Lundberg, S., R. A. Pollak, and J. Stearns (2016) "Family Inequality: Diverging Patterns in Marriage, Cohabitation, and Childbearing," *Journal of Economic Perspectives*, Vol. 30, pp. 79–102.

Manski, C. F. (1993) "Identification of Endogenous Social Effects - the Reflection Problem," *Review of Economic Studies*, Vol. 60, pp. 531–542.

Manson, Steven, Jonathan Schroeder, David Van Riper, Tracy Kugler, and Steven Ruggles (2021) "IPUMS National Historical Geographic Information System: Version 16.0," DOI: http://dx.doi.org/http://doi.org/10.18128/D050.V16.0.

McLanahan, S. and C. Percheski (2008) "Family structure and the reproduction of inequalities," *Annual Review of Sociology*, Vol. 34, pp. 257–276.

McLanahan, Sara (1985) "Family structure and the reproduction of poverty," *American journal of Sociology*, Vol. 90, pp. 873–901.

Merlino, Luca Paolo, Max Friedrich Steinhardt, and Liam Wren-Lewis (2019) “More than just friends? School peers and adult interracial relationships,” *Journal of Labor Economics*, Vol. 37, pp. 663–713.

Michney, Todd M. (2021) “How the City Survey’ s Redlining Maps Were Made: A Closer Look at HOLC’s Mortgagee Rehabilitation Division,” *Journal of Planning History*, Vol. 0, p. 15385132211013361, URL: https://doi.org/10.1177/15385132211013361, DOI: http://dx.doi.org/10.1177/15385132211013361.

Moynihan, Daniel Patrick (1965) *The Negro family: The case for national action*, No. 3: US Government Printing Office.

Mullainathan, Sendhil and Jann Spiess (2017) “Machine learning: an applied econometric approach,” *Journal of Economic Perspectives*, Vol. 31, pp. 87–106.

Nagaraj, Abhishek and Scott Stern (2020) “The Economics of Maps,” *Journal of Economic Perspectives*, Vol. 34, pp. 196–221, URL: https://www.aeaweb.org/articles?id=10.1257/jep.34.1.196, DOI: http://dx.doi.org/10.1257/jep.34.1.196.

Nelson, Robert K., LaDale Winling, Richard Marciano, and Nathan Connolly (2021) “Mapping Inequality.”

Olivetti, Claudia, Eleonora Patacchini, and Yves Zenou (2018) “Mothers, Peers, and Gender-Role Identity,” *Journal of the European Economic Association*.

Phelps, Edmund S (1972) “The statistical theory of racism and sexism,” *The American Economic Review*, Vol. 62, pp. 659–661.

Probst, Philipp, Marvin Wright, and Anne-Laure Boulesteix (2018) “Hyperparameters and Tuning Strategies for Random Forest,” *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, DOI: http://dx.doi.org/10.1002/widm.1301.

Rambachan, Ashesh, Jon Kleinberg, Jens Ludwig, and Sendhil Mullainathan (2020) “An economic perspective on algorithmic fairness,” in *AEA Papers and Proceedings*, Vol. 110, pp. 91–95.

Rothstein, Richard (2017) *The color of law: A forgotten history of how our government segregated America*: Liveright Publishing.

Ruggles, Steven, Catherine A. Fitch, Ronald Goeken, Josiah Grover, J. David Hacker, Matt Nelson, Jose Pacas, Evan Roberts, and Matthew Sobek (2020) "IPUMS Restricted Complete Count Data: Version 2.0."

Sacerdote, Bruce (2014) "Experimental and quasi-experimental analysis of peer effects: two steps forward?" *Annu. Rev. Econ.*, Vol. 6, pp. 253–272.

Sampson, R. J. (1987) "Urban Black Violence - the Effect of Male Joblessness and Family Disruption," *American Journal of Sociology*, Vol. 93, pp. 348–382.

Small, Mario L. and Devah Pager (2020) "Sociological Perspectives on Racial Discrimination," *Journal of Economic Perspectives*, Vol. 34, pp. 49–67, URL: https://www.aeaweb.org/articles?id=10.1257/jep.34.2.49, DOI: http://dx.doi.org/10.1257/jep.34.2.49.

Sood, Aradhya, William Speagle, and Kevin Ehrman-Solberg (2019) "Long Shadow of Racial Discrimination: Evidence from Housing Covenants of Minneapolis."

Tach, Laura and Allison Dwyer Emory (2017) "Public housing redevelopment, neighborhood change, and the restructuring of urban inequality," *American journal of sociology*, Vol. 123, pp. 686–739.

Townsend, Robert M (1994) "Risk and insurance in village India," *Econometrica: journal of the Econometric Society*, pp. 539–591.