

Beyond Fairness: Reparative Algorithms to Address Historical Injustices of Housing Discrimination in the US

Wonyoung So
Department of Urban Studies and
Planning, MIT
USA
wso@mit.edu

Pranay Lohia
Microsoft
India
pranaylohia@microsoft.com

Rakesh Pimplikar
IBM Research AI
India
rakesh.pimplikar@in.ibm.com

A.E. Hosoi
Institute for Data, Systems, and
Society, MIT
USA
peko@mit.edu

Catherine D'Ignazio
Department of Urban Studies and
Planning, MIT
USA
dignazio@mit.edu

ABSTRACT

Fairness in Machine Learning (ML) has mostly focused on interrogating the fairness of a particular decision point with assumptions made that the people represented in the data have been fairly treated throughout history. However, fairness cannot be ultimately achieved if such assumptions are not valid. This is the case for mortgage lending discrimination in the US, which should be critically understood as the result of historically accumulated injustices that were enacted through public policies and private practices including redlining, racial covenants, exclusionary zoning, and predatory inclusion, among others. With the erroneous assumptions of historical fairness in ML, Black borrowers with low income and low wealth are considered as a given condition in a lending algorithm, thus rejecting loans to them would be considered a “fair” decision even though Black borrowers were historically excluded from homeownership and wealth creation. To emphasize such issues, we introduce case studies using contemporary mortgage lending data as well as historical census data in the US. First, we show that historical housing discrimination has differentiated each racial group’s baseline wealth which is a critical input for algorithmically determining mortgage loans. The second case study estimates the cost of housing reparations in the algorithmic lending context to redress historical harms because of such discriminatory housing policies. Through these case studies, we envision what *reparative algorithms* would look like in the context of housing discrimination in the US. This work connects to emerging scholarship on how algorithmic systems can contribute to redressing past harms through engaging with reparations policies and programs.

CCS CONCEPTS

• **Theory of computation** → *Design and analysis of algorithms*; • **Applied computing** → *Sociology*.

KEYWORDS

fairness, mortgage lending, housing, racial wealth gap, reparations

ACM Reference Format:

Wonyoung So, Pranay Lohia, Rakesh Pimplikar, A.E. Hosoi, and Catherine D'Ignazio. 2022. Beyond Fairness: Reparative Algorithms to Address Historical Injustices of Housing Discrimination in the US. In *2022 ACM*

Conference on Fairness, Accountability, and Transparency (FAcCT '22), June 21–24, 2022, Seoul, Republic of Korea. ACM, New York, NY, USA, 18 pages. <https://doi.org/10.1145/3531146.3533160>

1 INTRODUCTION

Fairness in Machine Learning (ML) has tried to provide a set of meaningful criteria for different groups to be treated equally when developing an ML system [5, 16]. Because the system is developed using training data, it is natural to focus on assessing fairness within the particular moment of time that the data capture. The assumption of using the data is, in this case, that people of different groups represented in the data have been fairly treated. In other words, the group difference of credit scores in a loan application or a risk assessment score from a criminal justice algorithm is considered as fairly treated input data in that assumption. However, this assumption of historical equal treatment is erroneous in a society in which histories of disinvestment and discrimination produce unequal social hierarchies along the lines of race, gender, and class.

This is the case of mortgage lending discrimination, and broadly housing discrimination, in the US. Housing discrimination in the US is complicated with racism from private sectors, including racial covenants, and the history of state-sanctioned and institutionalized discriminatory policies, including redlining and exclusionary zoning, thus leading to residential segregation and targeted disinvestment in education and the built environment. These resulted in today’s significant racial wealth gap in the US—for instance, a study by Boston Federal Reserve Bank identifies that the estimated median net worth of White people was \$247,500 whereas that of Black people was \$8 in 2015 [54]. As a result, Black borrowers who apply for a mortgage loan have significantly fewer funds available for down payments and suffer from higher debt payments than White borrowers, as shown in Fig. 1.

While studies like this (and like our own) treat race as a quasi-natural category in order to measure disparities, it is important to keep in mind that racial categories such as Black, White, Asian, or Native American are socially and politically constructed. It is not so much that certain bodies “have” a race as an inherent property but rather that certain bodies are constantly undergoing racialization [58]. The racial stratification created because of such historical discrimination fundamentally challenges the assumptions of fairness

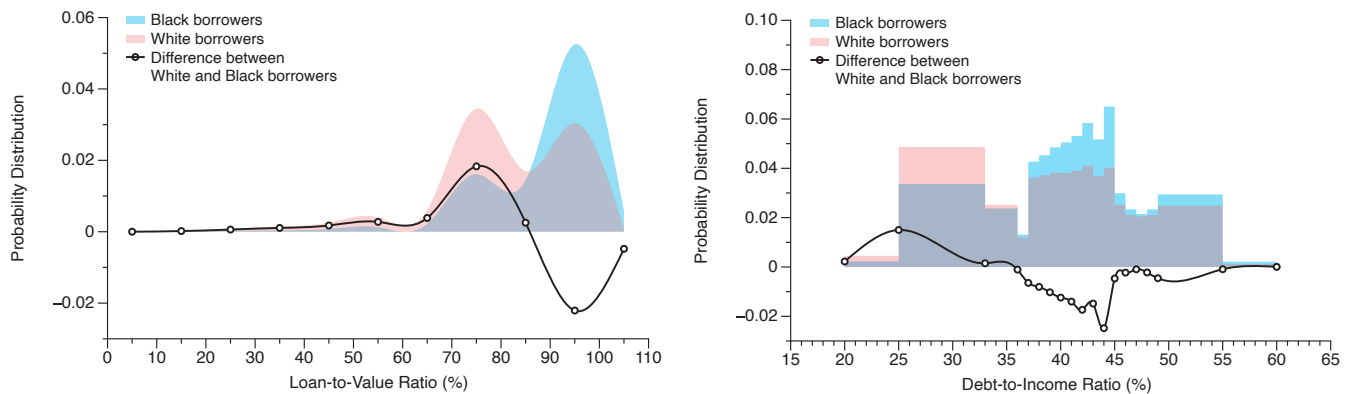


Figure 1: Distribution of the loan-to-value ratio and debt-to-income ratio by applicant's race. Source: Conventional first-lien mortgage applications of the 2020 Home Mortgage Disclosure Act (HMDA) data.

in ML by introducing historical bias into today's decision-making points [69]. Put simply, the base distribution of measuring creditworthiness by race, as shown in Fig. 1, has disparities by race and the reason can be attributed to historical discrimination which limited access to wealth accumulation through housing. Without acknowledging this, even if algorithmic lending systems achieve perfect accuracy, they will inevitably reproduce and amplify racial disparities. Thus, how can we develop a system and/or process that is to achieve the core objective of fairness in ML—in which different groups are treated *truly* equally?

We build on recent work on algorithmic reparations to argue that a *reparative approach* to developing an algorithmic process and/or system can contribute to redressing the historical harms that lead to deeply unequal base conditions in a decision-making or resource-allocation process [22]. A reparative approach urges us that there is an imminent need for deeply engaging with historical context to figure out why such population differences occur in the dataset. This should be the starting point of (1) analyzing what the counterfactual distribution would look like if there were no historical bias, and (2) intervening in algorithmic systems by suggesting a reparative scenario in such systems to converge the differences.

We specifically focus on reparations and housing related to African Americans for several reasons. Black people have been systemically harmed by unequal housing policy including redlining and residential segregation [32, 42, 45] and these policies are considered the biggest factor of widening the racial wealth gap [21, 52]. They are the group for whom reparations have been a public conversation since the end of the Civil War and the abolition of slavery [12, 20]. There is a growing literature on reparations for Black people in a variety of contexts [17, 37] and the practical experiments in reparations that have begun to take place in municipalities across the US [8, 73].

For navigating such a reparative approach, we conduct two case studies. The first case study shows how institutionalized redlining, which denied mortgage insurance to neighborhoods that Black people and immigrants most lived in from the 1940s to 1960s, affected the accumulation of wealth through housing. This can be understood as a first step in identifying historical bias in the system

and justifying the need for a reparative approach. Then, the second case study develops an ML system that shows the estimated cost of reparations if we wish to intervene in today's mortgage lending to address historical harms. It calculates the cost that is needed for Black borrowers who got rejected to get a prime loan to be accepted. This specific estimation, as well as the more generalized method, can contribute to the evaluation of housing reparations policies and programs that are trying to converge the racial wealth gap.

Through these case studies of housing discrimination in the US, we suggest an operationalizable reparative framework starting from recognizing and acknowledging past harms that affect today's data and developing an ML system that contributes to redressing the harms. Broadly, this study can exemplify ways in which an ML system can intervene in a society in which discrimination is historic and systemic. Algorithms alone cannot deliver or guarantee reparations; a fully reparative approach would also entail numerous political, social, and financial considerations. Nevertheless, given that today's mortgage loan decisions are increasingly made through automated decision-making systems [33, 39, 40], it is crucial to understand the underlying system and develop an ML process that is reflective of current mortgage lending algorithms to suggest a reparative intervention. Primarily, we propose that reparative algorithms may be mobilized to support policy analysis and thus contribute to the overall debates about interventions that can eliminate stratification between groups due to structural discrimination.

First, we review the literature on fairness in ML, the limitations of the concept of fairness, and discuss the potential ways that an ML system can address past injustices. Then, we give a brief history of housing discrimination in the US and discuss how such discrimination is related to today's racial wealth gap. Subsequently, we also review the concept of reparations and recent government-driven housing reparations programs in the US, as well as describe how algorithms and ML can contribute to such policies and practices. Then, we introduce two case studies using historical data and current mortgage lending data in the US with a discussion of their policy implications.

2 LITERATURE REVIEW

2.1 Limitations of Fairness in Machine Learning

In recent decades researchers have argued that computational systems and automated decision-making can bring potential harm to people, particularly politically and socially marginalized groups of people, including people of color, women, and low-income people. This is particularly because predictive systems are good at reifying existing structures [10] and due to *automation bias*: technological systems are considered as objective and neutral. The decisions of automated systems are accepted without critiques and, due to the proprietary nature of the systems, it is hard for people to challenge automated decisions. In relation to race, this can be especially exacerbated by systems that, instead of dealing with existing and real racial stratification, attempt to sweep it under the algorithmic carpet by trying to prevent racial identifiers from being considered by the system. This represents a kind of “colorblind” approach to fairness in automated systems [13, 14].

After the landmark investigation of a risk assessment algorithm named COMPAS [3], researchers concluded that algorithmic risk assessment can be biased and stressed the need to study how “fairness” of algorithms can be achieved [18]. It includes prohibiting protected attributes and proxies that correlated to the attributes [36], producing an algorithm to set similar accuracy between protected attributes including race and gender [66] or set similar error rates between those groups [2]. As a culmination of these studies, for instance, Narayan et al. [55] and Bellamy et al. [9] summarized that there are 21 different fairness criteria against which we can test our algorithms. However, these definitions of fairness can be incompatible with each other. Kleinberg et al. mathematically proved that it is impossible to meet both accuracy and demographic parity if the base rate between groups is unequal [49], and there are also unrecognizable errors because of the bias coming from training data sets.

More fundamentally, some researchers have questions about the fundamental limitations of the concept of fairness and its applicability in machine learning. In the context of risk assessment in the criminal justice system, Green asserted that algorithmic fairness relies on two false assumptions: one is that using an algorithm guarantees more objective decision making and the other is that risk assessment will introduce a more just criminal justice system [34]. In particular, he warned against the prevalent assumption that if we avoid the bias coming from human decision-making, then we could achieve fairness. This assumption ignores the fact that, particularly in the US context, there are base rate differences between groups—including race and gender—and these differences are “themselves the product of discrimination” [34]. Thus, even an algorithm that guarantees perfect accuracy would still discriminate against Black people and give undue advantages to White people [34]. Suresh and Gutttag argued that this “Historical Bias” can occur “even if data is perfectly measured and sampled if the world *as it is* or *as it was* leads to a model that produces harmful outcomes” [69]. Without acknowledging Historical Bias, fairness criteria “launders the product of historical discrimination into neutral and empirical fact” [34]. These issues show that training datasets can encode past discrimination. To overcome such issues, scholars have suggested

that a more fundamental shift in thinking is needed, such as pursuing “epistemic reform” which “challenges the discourses rather than the technical specifications of risk assessments” [34]. These perspectives are in line with what several law scholars propose as “algorithmic affirmative action,” which emphasizes the responsibility for historical discrimination that causes today’s inequalities [11, 15, 43]. Overall, these researchers ask how ML systems can be structured and mediate decision-making processes when there are accumulated injustices *before* the decision-making point [6], as well as how such machine learning systems can operate as “a mechanism to remedy past injustices” [5]. The critical question, then, is how current decision-making should be critically related to past histories, and how algorithms can be adjusted and adapted to intervene in domains with high levels of historical and structural discrimination.

2.2 A Brief History of Housing Discrimination and the Racial Wealth Gap in the US

Housing has been one of the crucial means of accumulating wealth in the US, thus it is important to understand housing discrimination, mortgage lending, and their role in producing the racial wealth gap. This section is not meant to paint a comprehensive picture of housing discrimination—Rather, it is to emphasize that systemic housing discrimination in the US is rooted in the transatlantic slave trade, the treatment of human beings as property with no rights, and post-Civil War efforts which continued (and continue) to deny people their rights through an array of public and private mechanisms.

First, after the European settlers arrived in the Americas, they constituted a form of property law consisting of two core characteristics: possession and a right to exclude [38]. However, in the process of executing the property rights, both were highly racialized. Obviously, the Americas were not empty and yet the property law did not guarantee the first possession or labor, and “[o]nly particular forms of possession—those that were characteristics of white settlement” were legitimized and supported by the law [38]. Furthermore, in the framework of the plantation economy, enslaved people were considered commodities. Therefore, Black people, particularly, had an “intimate relationship to property,” because Black people were the victim of the idea of property, and they were the property itself. Walcott argued that “thingification” and the modern conception of property should be centered in abolitionist politics because precisely such conception was created to police enslaved people [77]. This “crucial tension between property and humanity” indeed divided humans into an arbitrary fraction in the Representation Clause of the Constitution—The House of Representatives calculated a slave as $\frac{3}{5}$ of a person [38]. Slavery allowed to make Black people be tradable in the market, thus slaves could be even “transferred, assigned, inherited, or posted as collateral” [41]. Overall, through turning humans into property, White people have been able to accumulate their wealth through enslaved labor [38]. For instance, by enslaved people as collateral, the cotton plantation was able to get a mortgage and expand its business [24].

Even after slavery was abolished in 1865, *de jure* discrimination and segregation continued. Formerly enslaved people, for instance, were not granted any form of reparations, though they demanded

them. Rather, in contrast, the enacted plans ended up giving more land to mostly White people. For instance, the Homestead Act of 1862 granted over 240 million acres to homesteaders and thus provided ways of accumulating wealth through land. However, the Southern Homestead Act of 1866, which was initially designed to assist freedpeople, did not end up granting much land to them because not just as much of the land was unsuitable for farming, but many freedpeople could not afford the fee for the application and there was “continuing hostility and violence by whites against Black claimants” [26]. Eventually, the law was repealed earlier like many other Reconstruction programs [65].

Rather than granting enslaved people land, “separate but equal” policies dominated early in the 20th century. In the long period of the “Jim Crow era,” between the end of the Civil war in the 1870s and the beginning of the civil rights movements in the 1950s, even though African Americans became legal citizens, the social and economic mobility of Black people was limited by residential segregation. In the late 1930s, the establishment of the Home Owners’ Loan Corporation (HOLC) and the Federal Housing Administration (FHA) contributed to constructing institutionalized redlining [1]. The HOLC created a series of maps in the 1930s to measure and analyze the financial risk to banks for making loans in those neighborhoods. They graded neighborhoods in four groups from A (Best), B (Still Desirable), C (Definitely Declining) to D (Hazardous). The maps were the result of research coming from private and local assessors, lenders, and realtors. Although controversial on nuanced understanding, the white gaze towards racially and ethnically minoritized neighborhoods, all of whom were graded as ‘D,’ was endemic to both HOLC and FHA [80], and the FHA denied the insurance of the mortgage loan to those neighborhoods [30]. These maps are considered to clearly illustrate the long history of systemic, state-sanctioned racism in housing in the US, particularly “how decades of unequal treatment effectively limited where Black households lived in the 1930s” [31]. Institutionalized racism was exacerbated by racism in the private sphere like racial covenants and predatory loan terms like balloon payment or “on contract” terms, leading to further racial segregation [17].

Even after redlining became outlawed by the enactment of the Fair Housing Act of 1968, which prohibited discriminatory housing policies and practices on protected classes such as race and gender, *de facto* discrimination and segregation continued. Beginning in the 1970s, the FHA finally began insuring mortgages for Black homeowners. However, the loan conditions were not the same as those for White families who had benefited through the New Deal housing policy and suburbanization. Rather, Black homeowners were offered more predatory terms, and loans were still mostly granted in segregated communities of color where many of the homes available for purchase were in poor shape because of decades of neighborhood-level disinvestment. This “predatory inclusion,” combined with the advent of securitized mortgages, targeted and extracted higher interest rates from the segregated communities of color [71]. The harms of subprime lending culminated in the Great Recession of 2008, causing the current homeownership gap between Black people and White people to be wider than it was in the 1960s prior to the enactment of the Fair Housing Act [74].

The racial wealth gap in the US today, as a result, has tremendously widened because of these long and chained practices and

policies of housing discrimination. As of 2019, the Federal Reserve Bank of St. Louis measured that the median net wealth of White people was \$184,000, whereas the median net wealth of Black people was \$23,000 [48]. Such stark differences in the wealth gap are not just because of individual spending habits or that Black families spend more on education. Rather, it is mostly related to the structural concentration of wealth within certain groups of people. When controlling for socioeconomic status, Black families save slightly more money than the similar level of White families, and even with similar socioeconomic status, Black students are more likely to go to colleges than the similar socioeconomic level of White students. However, these college degrees are not well poised to accumulate wealth because they are more likely to borrow money to pay for college. This is because the ability to accumulate wealth to pay for college and buy a home comes from intergenerational wealth transition [21]. Put simply, “White Americans have had centuries of government assistance to accumulate wealth,” whereas Black Americans have not [37]. Derenoncourt et al. studied the racial wealth gap between the abolition of slavery and 2020 using the data from the historical census data, state tax records, and the Survey of Consumer Finances [23]. Their study showed that, first, the biggest determinant of the diverged wealth gap between White people and Black people is housing. Second, the wealth gap had converged 50 years after Emancipation, but it almost stopped converging by 1950 and was very steady as of 2020. Lastly, convergence under different scenarios seems difficult and has rather arrived at a “steady state” represented by today’s wealth gap. Even active interventions like lump-sum reparations may not resolve the racial wealth gap because the base rate for accumulating wealth is different by race. This indicates that more permanent policies that give minoritized groups the means of accumulating wealth, such as homeownership should be enacted to resolve the racial wealth gap [23].

2.3 Reparative approach: redressing historical harms through algorithmic reform

Reparations mean the full acknowledgment of past harms and figuring out the ways of redressing the past and ongoing harms [19, 79]. Recent successful reparations cases include lawsuits against Swiss banks and European insurance companies for harms of Nazi past and the monetary awards and official apology of the US government of illegal internment camps of Japanese Americans [53]. However, even though diverse sectors have endorsed reparations for slavery, there was no federal-level enactment of reparations for slavery in the US [20]. Still, as the monumental essay by Ta-Nehisi Coates noted, it is important to confront how America’s history has shaped the economy and political landscape and consider how to make an equitable society through the ideas of reparations [17].

After the abolition of slavery, there was a reparation plan called “40 acres and a mule,” which was supposed to give enslaved people a large land (40 acres) and a means of transportation (a mule) for reparations for southern enslaved people. However, President Andrew Johnson reversed the order and rather transferred back the land to former Confederate landowners. Even the legislative proposal to “study” the cost of reparations, called HR 40, could not pass [70]. However, the murder of George Floyd and the Black Lives Matter movement sparked the contemporary conversation

about reparations and culminated in the enactment of the first government-driven reparations program in Evanston, IL. The eligibility of the local reparations program is people who identify as Black who live or whose ancestors lived in Evanston between 1919 to 1968. They can receive \$25,000 to be used either by paying for a mortgage, applying for a new mortgage or remodeling their home [73]. Drawing on this first case, many other US cities, including Asheville, NC, and St. Paul, MN, started to either legislate the reparations bills or create a reparations commission to formulate a plan for addressing equity and wealth disparities in the city [63, 72]. These activities are in line with questions of how to operationalize reparations in the US by making a clearer connection to past harms in a local context. Municipal-level governments were able to specifically undertake these efforts in relation to housing which had such a direct impact on the racial wealth gap [46].

Given that the critical question of algorithmic fairness is that because today's data is the product of historical discrimination in an unequal society, we must think about how we can "develop new methods that recognize and account for the structural conditions of discrimination, oppression, and inequality" [34]. Relatedly, Davis et al. argued that "algorithmic reparation" could contribute to redressing the past harms of an unequal society using algorithms [22]. They suggested bringing the theories of intersectionality and a reparative approach to machine learning to critically assess and address systemic discrimination. However, because systemic discrimination is a combination of structural, organizational, and interpersonal policies and events, it is challenging to pinpoint the causal relationship between discriminatory policies and events. The question, then, is how we develop a reparative algorithm that may help identify histories of discrimination and provide some measures of redress. Overall, given that we are living in an unequal society built upon algorithms that reproduce inequalities, we need to think about what the "reparative algorithm" would be in the context of such practical problems out there. Barabas et al., for instance, argued that machine learning and statistical techniques should be used for identifying discrimination and formulating policies for intervention, rather than predicting risks [4].

3 CASE STUDIES: HOUSING DISCRIMINATION AND REPARATIVE APPROACH

We suggest a reparative framework in which an ML system can contribute to intervening in a reparative way in today's mortgage lending discrimination through two case studies using mortgage lending data in the US and the historical census data. Case study 1 establishes the need for a historical and reparative approach when addressing today's mortgage lending in the US. Then, case study 2 explores how ML can be applied not for risk assessment but for assessing the reparative potential of policy interventions. Before we introduce two case studies, we briefly review the literature on mortgage lending discrimination in the US, particularly in the context of FinTech which relies on automated lending algorithms.

Although there are numerous studies that have investigated mortgage lending discrimination in the US [27, 27, 28, 44, 61, 62], there are relatively few studies of more recent automated lending algorithms. Mortgage lending firms rapidly adopted machine

learning algorithms throughout their decision-making as FinTech emerged in the 2010s. Most of such automated lending systems make an instant decision after they have the input they request from the user, and traditional lenders have also started to integrate automated decision-making systems [40]. Interestingly, research shows that FinTech lending algorithms approve White applicants and non-White applicants at equal rates when controlling the effect of other critical characteristics of borrowers like a loan-to-value ratio and debt-to-income ratio. However, both studies also show that Black and Hispanic borrowers were charged higher subprime interest rates, compared to similar White borrowers [7, 39, 40].

Although these results may be related to the current racial lending landscape where lending institutions spatially target segregated communities of color [44, 68], the results do not fully account for the effects coming from the history of housing discrimination. In other words, these results may show how the "problem formulation" [59, 60] in FinTech mortgage lending, a process of translating traditional lending practices into data science problems, inadequately consider historical discrimination. In a lending environment that makes decisions with historical bias [69], these studies cannot account for disparities that occurred because of historical discrimination that limited the accumulation of wealth to specific racial groups. People with low income and low wealth are just understood as a given condition in the lending algorithm, thus rejecting loans to those people would be a "fair" decision even though the same people were historically excluded from housing ownership and wealth accumulation. The existence of such historical bias in mortgage lending data in the US challenges the applicability of fairness criteria in ML.

To examine this issue in mortgage lending discrimination, we conduct two studies. Each study is designed to identify the historical effects of discriminatory policies that cause the racial wealth gap and suggest an ML system that contributes to the analysis of the policies that seek to eliminate the racial wealth gap. The first case study examines how a neighborhood-level disinvestment policy called redlining caused a racialized pattern of accumulating wealth through housing. The second case estimates the amount of money for supporting today's Black borrowers to obtain prime mortgage loans, using the 2020 Home Mortgage Disclosure Act (HMDA) data.

3.1 Study 1: Identifying historical bias: accumulation of wealth through housing

Using causal inference methods, the first study illustrates the necessary step of identifying historical bias. We use residential security maps created by the Home Owners' Loan Corporation (HOLC) in the late 1930s and historical Census data to show how institutionalized redlining policies and racist white gazes on the redlined neighborhoods have shaped the ways of accumulating wealth through housing. Through this kind of study, governments can identify the wealth gap at the neighborhood level and thereby both characterize the historical bias as well as justify the need for a reparative approach rather than a fairness approach to deal with durable structural inequalities.

3.1.1 Data and Methods. We used two main sources for this study. The first data was a set of maps created by the HOLC in the late 1930s. The maps were digitized by the University of Richmond's

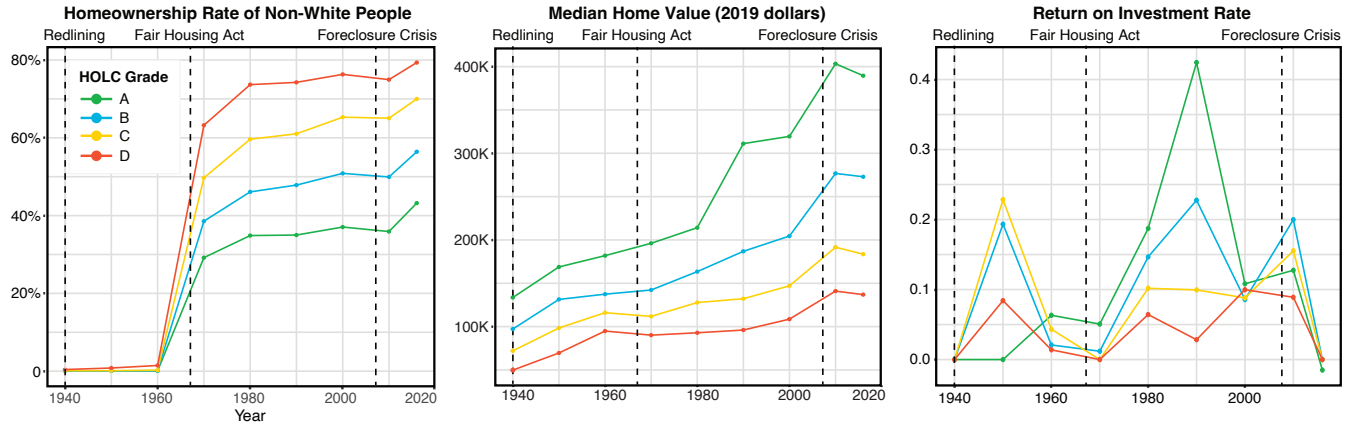


Figure 2: Homeownership rate of non-White people, median home value, return on investment rate by HOLC Grade.

Digital Scholarship Lab [57]. Each HOLC neighborhood has grades (A, B, C, D) and spatial boundaries. The second dataset comes from historical Census data. We used tract-level decennial Census data from 1940 to 2010 and 2016 American Community Survey 5-year estimates. We acquired the HOLC maps data and historical census data collected by The Legacy of Redlining: Residential Segregation in 147 American Cities [81]. Then we additionally collected median home value, owner-occupied households, and median rent from IPUMS's National Historical Geographic Information System (NHGIS) [50]. The data about median building age comes from HOLC's assessor document digitized and tabulated by Scott Markley [51]. Since HOLC maps' boundaries and census tract boundaries are not exactly matched, we used the proportional split method [64] to estimate the socioeconomic characteristics of census tracts including median rent, median home value, and homeownership rate. Fig 2. shows temporal patterns of homeownership rate of non-White people¹, median home value, and return on investment (ROI) rate. ROI refers to the expected return if one buys a home and holds the property in the neighborhood for 10 years, by HOLC grades. These plots illustrate how the pattern of accumulation of wealth is different by race. The homeownership of non-Whites was concentrated in D-grade neighborhoods, whereas the accumulation of wealth through housing was more concentrated in non-D-grade neighborhoods—in particular, the disparities of ROI observed in 1990 D-grade neighborhoods and non-D-grades are stark.

Our hypothesis for this study is that there are causal relationships in these three outcomes (homeownership rate of non-White people, median home value, and ROI) stemming from the grading of neighborhoods. Since the datasets are observational, we apply propensity score matching and weighting methods for inferring causal estimates. The main treatment variable of this study is a HOLC neighborhood being assigned D-grade. It is important to note that some researchers assert that the HOLC maps should be understood in a more nuanced way, as opposed to the popular understanding that HOLC maps directly caused banks to deny mortgage insurance to Black and other borrowers. Their reasoning

is based on the following facts. First, HOLC refinanced mortgages that were threatened by foreclosures because of the Great Depression, and they refinanced many Black borrowers. Second, The FHA (not the HOLC) is the institution that maintained racially discriminatory mortgage insurance policies, thus using HOLC maps to measure direct causal effects would have less explanatory power [82]. However, in our study, we justify the use of HOLC data because it is evident that the assessment of HOLC is ethically and racially discriminatory and there is some historical evidence that HOLC and FHA were in close relationship and their theory of land values was circulated in the same academic networks that were heavily affected by scientific racism and eugenics [80].

Therefore, in this study, we argue that the risk assessment of HOLC and the HOLC maps - and the treatment in this study - were indicators that show the *perceptions* of the 1940's real estate practices, both public and private sectors, and the trend towards racial stratification that was *already* in place. We use HOLC maps as proxies for how real estate professionals and governments perceive financial risks coming from neighborhoods in the late 1930s. We draw on the historical research that HOLC maps were influential to FHA practices through a close relationship with the FHA [80], as well as Nelson's argument that redlining maps demonstrate that the government failed to un-engineer the unequal patterns and rather directly engaged with numerous racially discriminatory housing policies and private investments after redlining [56]. In other words, *the period between the 1940s and the present can be understood as an opportunity period to undo disinvestment. In this way, we can track whether D-grade neighborhoods received an opportunity to emerge out from under a racist gaze over time, or whether the fact of receiving a D-grade induced more extraction and exploitation.*

We constructed matched control sets of HOLC neighborhoods using spatial adjacency. We chose the n nearest control observations. For instance, the matched set of control observations of a D-grade neighborhood in Boston is near non-D-grade neighborhoods in Cambridge and Brookline. We set the control units to be the same year as a treated unit and not farther than 15km (which is an average radius of US cities) from the treated unit. And we limited the maximum number of control sets to be 100. Using these

¹We used the rate of non-White people because the 1940 and 1950 Census did not collect race-specific owner-occupied households.

spatially adjacent matched control sets, we tested two different matching/weighting methods to estimate Average Treatment Effect for the Treated (ATT): propensity score matching and weighting with trimming the 5th and 95th percentile weights. The rest of the process is structured as follows. First, we checked the covariate balance of each method for $t = 1940$. The process of checking the balance of covariates before and after matching and weighting can be found in the appendices. After confirming that the covariates are well balanced after refining control sets using propensity score weights, we ran the matching and weighting estimators for three different outcomes using propensity score in 1940: homeownership rate of non-White people, median home value, and the ROI rate.

3.1.2 Results. The left plot of Fig 3. shows the effect of receiving D-grade neighborhood on the homeownership rate of non-White people. When redlining policies were enacted, homeownership was slightly more concentrated in D-grade neighborhoods, but in general, the homeownership rate of non-White people was only 1.1%. Around when the Fair Housing Act was enacted, non-White people finally could own housing through the FHA-insured mortgage. However, we can see evidence of “predatory inclusion” [71] in which the granting of mortgage loans to non-White people were more concentrated in segregated communities, as this plot shows that non-White people in the D-grade neighborhoods owned approximately 10% more homes than non-D-grades. As Taylor studied, the FHA’s mortgage insurance expanded to Black borrowers after the Fair Housing Act was enacted, but such investments were enacted *not* through active community integration but in segregated neighborhoods [71]. Additionally, the effect of the foreclosure crisis, as a result of predatory lending practices, hit more homeowners in D-grade neighborhoods than non-D-grades, as shown in the last part of the figure.

The center plot of Fig. 3 shows the difference in median home value between D-grade neighborhoods and non-D-grade neighborhoods. First of all, the median home value in D-grade neighborhoods is always lower than the value of homes in non-D-grade neighborhoods, suggesting how persistent the gap is. The home value gap between D-grade and non-D-grade neighborhoods was flattened a bit in the 1970s, potentially suggesting positive effects of the Fair Housing Act. However, the gap continued to diverge until the 2000s. But after the 2000s, we see evidence that the gap converged, which could be attributed to the result of subprime lending practices. In 2016, because of the effect of the foreclosure crisis, the gap diverged again dramatically.

The right plot of Fig. 3 shows the effect of receiving D-grade neighborhood on return on investment (ROI). The ROI gap in the 1940s was -0.16 when estimated by the matching estimator and -0.2 when the weighting estimator, meaning that when one buys a home in a D-grade neighborhood, they would lose 20% of the invested money for this home as compared to a counterfactual home in a non-D-grade neighborhood. But after the Fair Housing Act, the ROI gap converged to zero and the effect of ROI on D-grade neighborhoods was even flipped to positive until the 2000s with some fluctuation in the 1980s and the 2000s. Such a flip can be accounted for the fact that D-grade neighborhoods were slowly becoming gentrifiable for real estate developers [67] and ripe for predatory lending practices [27]. However, the ROI gap between

D-grade and non-D-grade neighborhoods went diverged again after the foreclosure crisis. The widened gap in 2016 was similar to the gap that existed before the Fair Housing Act.

These plots contribute to answering whether redlined neighborhoods have received an opportunity to emerge out of the discriminatory policies and practices or whether the neighborhoods were involved in more covert exploitation in the context of residential segregation. For instance, we observe that the homeownership of non-White people dramatically increased after the Fair Housing Act, but they could own home mostly in segregated neighborhoods. Their home was devalued whereas the home value in White-majority neighborhoods skyrocketed. The ROI of D-grade neighborhoods was the highest in the 2000s and the home value gap was suddenly decreased, but it implies that the neighborhoods became attractive to banks which could spatially target vulnerable neighborhoods to sell subprime loans [44, 68] (one of the main causes of the foreclosure crisis). These spatially targeted loan practices were possible because of the residential segregation that redlining practices made possible. In other words, redlining brought not just *direct* effects such as the denial of mortgage insurance to redlined neighborhoods, but redlining also brought *indirect* effects that produced vulnerabilities. These indirect effects, in interaction with public and private actors, further limited the opportunity of accumulating wealth and extracting capital from the redlined neighborhoods.

This study further suggests that without bringing historical context, applying only fairness criteria in algorithmic decision-making in mortgage lending in the US is unsatisfactory and unfair. For instance, mortgage lending developers are prohibited from using protected attributes like race and gender [78]. This is in line with the idea that individuals who have similar characteristics, and credit-worthiness in the loan context, should be treated similarly. However, in this case, such a loan algorithm would miss low-wealth applicants’ counterfactual characteristics, such as whether applicants or their parents had been affected by racially discriminatory housing policies and practices. Identifying such historical bias is crucial to bring a reparative approach to addressing the historical inequalities in the loan application data and thinking beyond fairness criteria.

There are limitations and considerations of this study. Although this study relied on causal inference methods, isolating the effect of D-grade neighborhood is limited because the selection on observable methods (matching/weighting) relies on the assumption that there are no confounders outside the covariates we obtained. However, this assumption is challenging to fully achieve because the availability of historical data in this context is scarce and limited. Instead, the results of this study can be understood in relation to numerous other studies that try to identify the structural factors that have contributed to widening the racial wealth gap [21, 23, 54]. It also suggests future research to develop a more nuanced understanding of how housing discrimination has operated through residential segregation, gentrification, and predatory lending, all of which are contributing to the racial wealth gap.

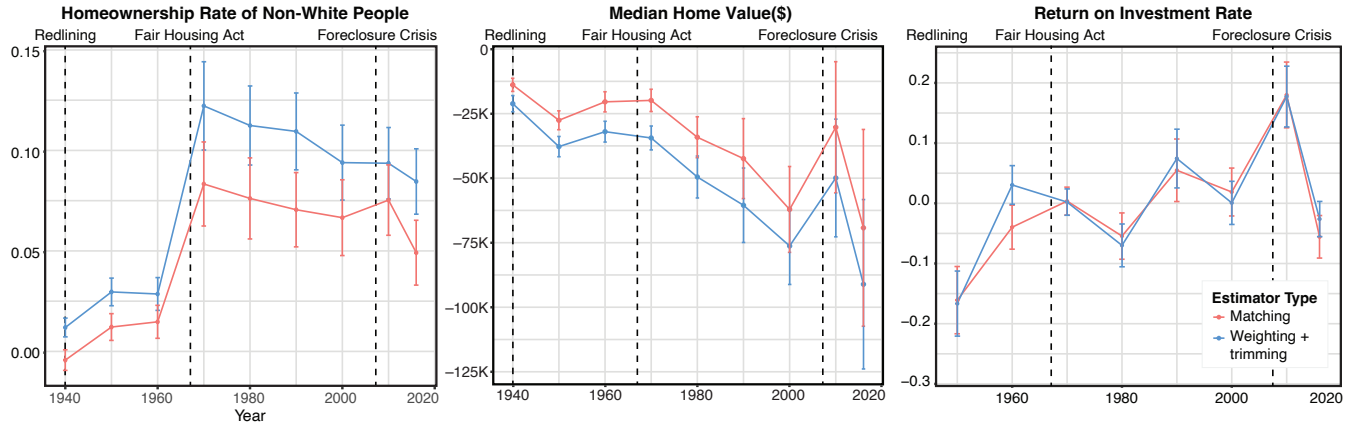


Figure 3: Average Treatment Effects on the Treated (D-grade neighborhood) on three housing-related measures. Error bar shows 95% CI.

3.2 Study 2: Measuring the intervention: cost of housing reparations

Given the existence of historical discrimination and the limitations of applying fairness criteria, how can algorithmic systems support a reparative approach to mortgage lending discrimination and broader policy formulation? The second study estimates financial resources for supporting today’s Black borrowers. We draw on Barabas et al.’s argument that an algorithmic system should be used for intervening into an unequal system, not simply predicting risks [4]. Using the previous literature on actionable recourse [47, 75, 76] and the HMDA data, this study quantifies how much money would be needed in order for Black borrowers who were previously rejected from a prime loan to flip the decision.

Using this method, for instance, this study can examine the effectiveness of the Restorative Housing program in Evanston, IL. The city of Evanston offered \$25,000 per Black person for housing reparations, and it was designated to be used for housing-related expenses. If Black borrowers who had previously been rejected from receiving loans used that additional money to apply for new housing, would the current lending algorithm accept them? If the \$25,000 down payment assistance was not helpful in terms of getting access to mortgage loans, the city of Evanston might need to increase the amount of down payment assistance or need to devise creative policies to forgive the debt (if their debt was one of the reasons they were denied a loan). Also, it is particularly important to assess whether or not they could get a loan with prime interest rates because previous research studying FinTech shows that there are racial disparities in granting subprime loans and higher interest loans are harmful and related to foreclosure [7]. Overall, this study contributes to the estimation of the overall financial scale of the federal-level housing reparations program and also to evaluating local housing reparations programs that a few cities in the US have been preparing [8].

3.2.1 Data and Methods. This study uses the 2020 Home Mortgage Disclosure Act (HMDA) data. HMDA is a public dataset that requires lending institutions in the US to document every mortgage loan decision. Using the latest HMDA data is necessary because of the recent inclusion of loan-to-value ratio and debt-to-income ratio

which are the two most important metrics of mortgage loan decision making. The descriptive statistics of the dataset can be found in the appendices. For the estimation, we first excluded second lien loans to exclude the effects of other loans attached to a property. Then, we separated conventional loans and FHA loans for analysis because the distribution of loan-to-value ratio varies depending on the types of loans—conventional loans typically require 20% down payment whereas FHA loans only require 3.5% down payment. For the estimation of the cost of flipping loan decisions, this study relies on the epistemology of algorithmic recourse [76], particularly the methodology of the actionable recourse paper [75]. The paper provides specific ways of actionable recourse if one was negatively categorized from a linear classifier by testing a grid of action sets of features. “Action sets” comprise guidelines for whether or not one can manipulate the features to flip the algorithm’s decision. We set the loan-to-value ratio and debt-to-income ratio as actionable in the context of the housing reparations program because, for instance, borrowers can improve the loan-to-value ratio by getting financial assistance from governments in the form of reparations. However, housing reparations programs cannot manipulate borrowers’ income, loan terms, or property value.

The overall process was the following. First, we trained a logistic regression classifier using the HMDA data. We set the outcome variable as a dummy variable of whether or not a borrower gets a prime loan (which is defined by whether or not the rate spread is less than 1.5 percentage points [29]). We incrementally trained the model state by state using Stochastic Gradient Descent (SGD) classifier in the Scikit-learn package. The protected features such as race and gender of borrowers were not included in the training data because usually modelers of loan algorithms are prohibited from accessing these variables [78]. Then, we constructed the action sets using the coefficient of the model. Given that housing reparations programs can financially support borrowers, We set the action sets that we can decrease either the loan-to-value ratio or the debt-to-income ratio of each borrower. Then we calculated the cost to flip the decision by estimating the cost of filling the gap of each ratio if the borrower’s race was Black or African American. For instance, when a Black borrower applied a \$85,000 loan to buy a

property valued \$100,000 with \$15,000 down-payment, the loan-to-value ratio is 85%. If the actionable recourse algorithm recommends decreasing the loan-to-value ratio to 80%, then the estimated cost of assistance is \$5,000. We also included the cost of decreasing the debt-to-income ratio by suggesting that the reparations program could support multiple years to mitigate the debt if borrowers need to decrease their debt-to-income ratio. However, the specific ways of supporting financial means to decrease the debt-to-income ratio should be further studied because it is hard to estimate the scope of assistance. For instance, suppose a borrower has a \$40,000 auto loan and needs to pay \$200 per month. In this case, even if a reparations program supported the borrower to pay the auto loan a year (\$2,400), that would not fully cancel the debt. Nevertheless, knowing the amount of monthly debt relief that could flip the decision would be helpful for policymakers to devise a housing reparations program.

3.2.2 Results. For conventional loans, the total number of Black borrowers who were denied a prime conventional loan in the model was 41,830. Among them, 39,028 people could flip the decision by decreasing the loan-to-value ratio or debt-to-income ratio but 2,802 people could not even if a housing reparations program supported either decreasing loan-to-value ratio or debt-to-income ratio. Decreasing borrowers' loan-to-value ratio is related to providing down payment assistance. It is relatively easy to formulate a housing reparations program by providing down payment assistance because it is a one-time payment. The total cost for decreasing the loan-to-value ratio was \$891M. However, it is challenging to devise a program for decreasing borrowers' debt-to-income ratio because it is unclear to know the total debt of each borrower by only checking the debt-to-income ratio. Nevertheless, A total of \$3.6M per month to decrease the debt-to-income ratio to flip the decisions. On average, one denied Black borrower who applies for a conventional loan needs to have \$41,256 down payment assistance and \$205 monthly support for the debt. Fig. 4 shows the compensation needed for conventional loans by states—Blue bar graphs show the compensation needed to decrease the loan-to-value ratio and red bar graphs show the compensation needed to decrease the debt-to-income ratio (per month).

For FHA-backed loans, because they require only a 3.5% down payment, there was no way of flipping loan decisions by decreasing the loan-to-value ratio. Thus, we only found ways of decreasing the debt-to-income ratio to flip the decisions. The total number of Black borrowers who were denied a prime FHA-backed loan in the model was 52,973. Among them, 45,055 people could flip the decision by decreasing the debt-to-income ratio but 7,918 people could not even if a housing reparations program supported decreasing the debt-to-income ratio. The total cost for decreasing the debt-to-income ratio was \$26M per month. On average, one denied Black borrower who applies for an FHA-backed loan would need to have \$578 monthly support for debt relief. Fig. 5 shows the compensation needed for FHA-backed loans by states—Red bar graphs show the compensation needed to decrease the debt-to-income ratio (per month).

The results of this study contribute to formulating and evaluating local housing reparations. For instance, it can be used for evaluating the effectiveness of the housing reparations program in the city of Evanston, which offered \$25,000 for housing-related expenses.

In Cook County, where Evanston is located, on average, a Black borrower who applied for a conventional loan but could not get a prime loan needed to get a \$33,289 down payment assistance and \$173 monthly support for paying the debt to flip the decision. Alternately, if the borrower applied for an FHA-backed loan, then they needed to get a \$591 monthly support for the debt payment on average. Among 1,843 borrowers who applied for conventional loans, only 500 borrowers (27.1%) could flip the decision if they used the \$25,000 for lowering the loan-to-value ratio. Among 2,266 borrowers who applied for FHA-backed loans, 1,513 people (66.7%) could flip the decision if they spend \$25,000 for two years paying down their debt. This estimation suggests that the city of Evanston might need to increase the amount of down payment assistance and also need to formulate a policy of lowering down the debt-to-income ratio, potentially in collaboration with local lending institutions. Because lending institutions do not know the total debt of each borrower, the logic of forgiving the debt-to-income ratio should be thoroughly structured in collaboration with lending institutions and governments to account for the support to decrease the debt-to-income ratio. For instance, local lending institutions could devise a unique way of calculating the debt-to-income ratio if one applies for a loan through the housing reparations program with assistance.

There are limitations and considerations of this study. First, we could not obtain credit histories or credit scores, which is one of the most crucial determinants for granting a loan, but they are not visible on the HMDA data. Although it would affect the accuracy of the statistical modeling, nevertheless, it is impossible to directly manipulate credit histories in the scenario of a housing reparations program. Second, since every lending institution has its own lending algorithm for decision-making, there would be variations from this estimation. Therefore, the results of the study can be understood as an algorithmic system that reflects today's lending conditions. Lastly, since we only calculated the recourse of Black borrowers who applied but failed to get a prime loan, we miss potential borrowers if such a reparative plan is enacted. Therefore, this estimation should be understood as the minimum number for covering active borrowers, and further research is needed to figure out the ways in which potential borrowers can be included. Additionally, some critics might argue that this reparative approach could contribute to another foreclosure crisis because it only opens the entry point of owning a home but does not intervene in the later steps of paying mortgages. Although there is a study that demonstrates that subprime conditions produce more foreclosures, not individual conditions [25], these potential critiques suggest keeping an eye on other sectors that could be additional barriers to improving material conditions of securing a mortgage, including employment, education, and residential segregation.

4 CONCLUSION

Through two case studies, we have envisioned what reparative algorithms could be in the context of housing discrimination in the US. The first case study was undertaken to justify the development of reparative policies by identifying and characterizing historical bias. This is not a single study that proves the causal relationship between housing discrimination and the racial wealth



Figure 4: Compensation needed to flip conventional loan decisions by states. The blue bar graphs show the amount of money needed to decrease the loan-to-value ratio and the red bar graphs show the amount of money needed to decrease the debt-to-income ratio.

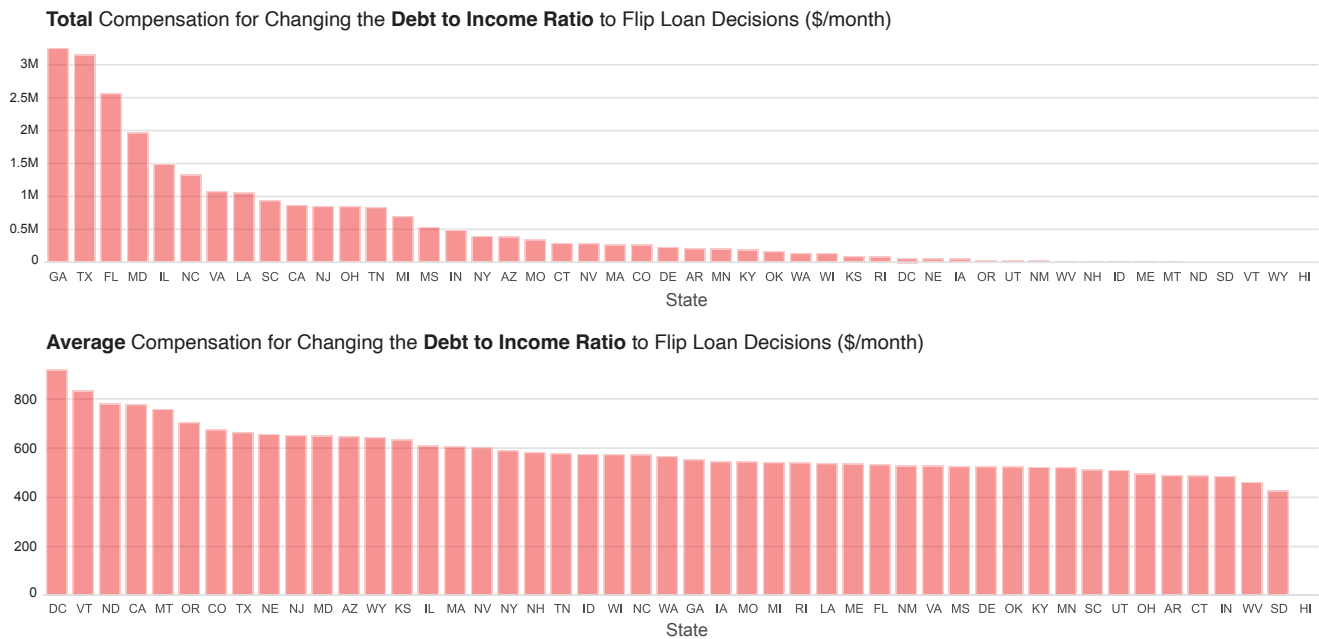


Figure 5: Compensation to flip FHA loan decisions by states. The red bar graph shows the amount of money needed to decrease the debt-to-income ratio.

gap. Rather, with numerous other studies identifying such historical bias, it should be understood as a method of corroborating existing evidence of historical bias. In conjunction with existing literature, it confirms that we need to take a different approach than conventional fairness in the context of racial disparities in contemporary mortgage lending. The second case study exemplified how an algorithmic system can contribute to evaluating reparative policies in the context of housing discrimination in the US. Relying on the literature of actionable recourse [75], it estimates the amount of money that governments need to provide in order for today’s Black borrowers to secure a prime mortgage loan, thus providing a method for examining and evaluating current housing reparations policies.

This work connects to important insights in regards to how an algorithmic system can contribute to redressing past harms [22] and broadly pursue substantive fairness [35], rather than developing “fairer” risk assessment and prediction. This work bridges between local housing-based reparations programs—and broadly restorative justice literature—and algorithmic accountability by developing ways to measure the cost of housing-based reparations. Ultimately, this work contributes to how to operationalize reparations because many loan decisions rely on automated decision-making through algorithms. However, we also emphasize algorithmic intervention cannot do everything. Rather, we argue that algorithmic systems should work in concert with legal, social and political reforms to undertake the work of “reworking institutional incentive structures and power arrangements” [22]. Only when we act beyond fairness and devise algorithmic processes grounded in historical contexts, can we achieve substantial improvement and converge differences that occurred because of historical and structural discrimination.

ACKNOWLEDGMENTS

MIT-IBM Watson AI Lab Exploratory award generously funds this research. The authors are solely responsible for the accuracy of the statements and interpretations contained in this publication.

REFERENCES

- [1] Daniel Aaronson, Daniel A. Hartley, and Bhashkar Mazumder. 2020. *The Effects of the 1930s Holc ‘Redlining’ Maps*. SSRN Scholarly Paper ID 3038733. Social Science Research Network, Rochester, NY. <https://papers.ssrn.com/abstract=3038733>
- [2] Alekh Agarwal, Alina Beygelzimer, Miroslav Dudik, John Langford, and Hanna Wallach. 2018. A Reductions Approach to Fair Classification. *arXiv:1803.02453 [cs]* (July 2018). <http://arxiv.org/abs/1803.02453> arXiv: 1803.02453.
- [3] Julia Angwin, Jeff Larson, Surya Mattu, and Lauren Kirchner. 2016. Machine Bias. *ProPublica* (May 2016). <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>
- [4] Chelsea Barabas, Madars Virza, Karthik Dinakar, Joichi Ito, and Jonathan Zittrain. 2018. Interventions over Predictions: Reframing the Ethical Debate for Actuarial Risk Assessment. In *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*. PMLR, 62–76. <https://proceedings.mlr.press/v81/barabas18a.html> ISSN: 2640-3498.
- [5] Solon Barocas, Moritz Hardt, and Arvind Narayanan. 2019. *Fairness and Machine Learning*. fairmlbook.org.
- [6] Solon Barocas and Andrew D. Selbst. 2016. Big Data’s Disparate Impact. *SSRN Electronic Journal* (2016). <https://doi.org/10.2139/ssrn.2477899>
- [7] Robert Bartlett, Adair Morse, Richard Stanton, and Nancy Wallace. 2019. Consumer-Lending Discrimination in the FinTech Era. *NBER Working Paper Series* (June 2019), 51. <https://doi.org/10.3386/w25943>
- [8] Adam Beam. 2021. 11 US mayors pledge to pay reparations for slavery to small groups of Black residents. <https://www.usatoday.com/story/news/nation/2021/06/19/reparations-slavery-pledged-11-us-mayors-pilot-program/7753319002/>
- [9] R. K. E. Bellamy, K. Dey, M. Hind, S. C. Hoffman, S. Houde, K. Kannan, P. Lohia, J. Martino, S. Mehta, A. Mojsilović, S. Nagar, K. Natesan Ramamurthy, J. Richards, D. Saha, P. Sattigeri, M. Singh, K. R. Varshney, and Y. Zhang. 2019. AI Fairness 360: An extensible toolkit for detecting and mitigating algorithmic bias. *IBM Journal of Research and Development* 63, 4/5 (July 2019), 4:1–4:15. <https://doi.org/10.1147/JRD.2019.2942287> Conference Name: IBM Journal of Research and Development.

- [10] Ruha Benjamin. 2019. *Race After Technology: Abolitionist Tools for the New Jim Code* (1st edition ed.). Polity, Medford, MA.
- [11] Jason R. Bent. 2020. Is Algorithmic Affirmative Action Legal? *Georgetown Law Journal* 108 (2020), 803–853. <https://papers.ssrn.com/abstract=3372690>
- [12] Boris I. Bittker. 2003. *The case for Black reparations*. Beacon Press, Boston, MA.
- [13] Eduardo Bonilla-Silva. 1997. Rethinking Racism: Toward a Structural Interpretation. *American Sociological Review* 62, 3 (1997), 465–480. <https://doi.org/10.2307/2657316> Publisher: [American Sociological Association, Sage Publications, Inc.].
- [14] Eduardo Bonilla-Silva. 2018. *Racism without racists: color-blind racism and the persistence of racial inequality in America* (fifth edition ed.). Rowman & Littlefield, Lanham.
- [15] Anupam Chander. 2017. The Racist Algorithm? *Michigan Law Review* 115, 6 (2017), 1023.
- [16] Alexandra Chouldechova and Aaron Roth. 2020. A snapshot of the frontiers of fairness in machine learning. *Commun. ACM* 63, 5 (April 2020), 82–89. <https://doi.org/10.1145/3376898>
- [17] Ta-Nehisi Coates. 2014. The Case for Reparations. <https://www.theatlantic.com/magazine/archive/2014/06/the-case-for-reparations/361631/> Section: Business.
- [18] Sam Corbett-Davies and Sharad Goel. 2018. The Measure and Mismeasure of Fairness: A Critical Review of Fair Machine Learning. *arXiv:1808.00023 [cs]* (Aug. 2018). <http://arxiv.org/abs/1808.00023> arXiv: 1808.00023.
- [19] William Darity. 2008. Forty Acres and a Mule in the 21st Century. *Social Science Quarterly* 89, 3 (Sept. 2008), 656–664. <https://doi.org/10.1111/j.1540-6237.2008.00555.x>
- [20] William A. Darity and A. Kirsten Mullen. 2020. *From here to equality: reparations for black Americans in the twenty-first century*. The University of North Carolina Press, Chapel Hill.
- [21] William Darity Jr, Darrick Hamilton, Mark Paul, Alan Aja, Anne Price, Antonio Moore, and Caterina Chiopris. 2018. *What We Get Wrong About Closing The Racial Wealth Gap*. Technical Report. Samuel DuBois Cook Center on Social Equity. 67 pages. <https://socialequity.duke.edu/wp-content/uploads/2019/10/what-we-get-wrong.pdf>
- [22] Jenny L. Davis, Apryl Williams, and Michael W. Yang. 2021. Algorithmic reparation. *Big Data & Society* 8, 2 (July 2021), 20539517211044808. <https://doi.org/10.1177/20539517211044808> Publisher: SAGE Publications Ltd.
- [23] Ellora Derenoncourt, Chi Hyun Kim, Moritz Kuhn, and Moritz Schularick. 2021. The racial wealth gap, 1860–2020. (Aug. 2021), 62. https://egc.yale.edu/sites/default/files/Economic%20History%20Conferences/2021-09/DKKS_NBER_2021%20ada-ns.pdf
- [24] Matthew Desmond. 2019. American Capitalism Is Brutal. You Can Trace That to the Plantation. *The New York Times* (Aug. 2019). <https://www.nytimes.com/interactive/2019/08/14/magazine/slavery-capitalism.html>; <https://www.nytimes.com/interactive/2019/08/14/magazine/slavery-capitalism.html>
- [25] Lei Ding, Roberto Quercia, Wei Li, and Janneke Ratcliffe. 2011. Risky Borrowers or Risky Mortgages Disaggregating Effects Using Propensity Score Models. *Journal of Real Estate Research* 33, 2 (Jan. 2011), 245–278. <https://doi.org/10.1080/10835547.2011.12091305> Publisher: Routledge _eprint: <https://doi.org/10.1080/10835547.2011.12091305>.
- [26] Richard Edwards. 2019. African Americans and the Southern Homestead Act. *Great Plains Quarterly* 39, 2 (2019), 103–129. <https://doi.org/10.1353/gpq.2019.0018> Publisher: University of Nebraska Press.
- [27] Jacob W. Faber. 2013. Racial Dynamics of Subprime Mortgage Lending at the Peak. *Housing Policy Debate* 23, 2 (April 2013), 328–349. <https://doi.org/10.1080/10511482.2013.771788> Publisher: Routledge _eprint: <https://doi.org/10.1080/10511482.2013.771788>.
- [28] Jacob William Faber. 2018. Segregation and the Geography of Creditworthiness: Racial Inequality in a Recovered Mortgage Market. *Housing Policy Debate* 28, 2 (March 2018), 215–247. <https://doi.org/10.1080/10511482.2017.1341944> Publisher: Routledge _eprint: <https://doi.org/10.1080/10511482.2017.1341944>.
- [29] Federal Financial Institutions Examination Council. 2016. FFIEC rate spread calculator. <https://www.ffiec.gov/ratespread/newcalchelp.aspx>
- [30] Federal Housing Administration. 1938. *Underwriting manual: Underwriting and valuation procedure under title II of the National Housing Act*. Technical Report. U.S. Government Printing Office. <https://www.huduser.gov/portal/sites/default/files/pdf/Federal-Housing-Administration-Underwriting-Manual.pdf>
- [31] Price Fishback, 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*. Technical Report w28146. National Bureau of Economic Research, Cambridge, MA. w28146 pages. <https://doi.org/10.3386/w28146>
- [32] Price V. Fishback, Jessica LaVoice, Allison Shertzer, and Randall Walsh. 2020. *Race, Risk, and the Emergence of Federal Redlining*. Working Paper 28146. National Bureau of Economic Research. <https://doi.org/10.3386/w28146> Series: Working Paper Series.
- [33] Andreas Fuster, Matthew Plosser, Philipp Schnabl, and James Vickery. 2019. The Role of Technology in Mortgage Lending. *The Review of Financial Studies* 32, 5 (May 2019), 1854–1899. <https://doi.org/10.1093/rfs/hhz018>
- [34] Ben Green. 2020. The false promise of risk assessments: epistemic reform and the limits of fairness. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT* '20)*. Association for Computing Machinery, New York, NY, USA, 594–606. <https://doi.org/10.1145/3351095.3372869>
- [35] Ben Green. 2021. Impossibility of What? Formal and Substantive Equality in Algorithmic Fairness. *arXiv:2107.04642 [cs]* (July 2021). <http://arxiv.org/abs/2107.04642> arXiv: 2107.04642.
- [36] Nina Grgic-Hlaca, Muhammad Bilal Zafar, Krishna P Gummadi, and Adrian Weller. 2016. The Case for Process Fairness in Learning: Feature Selection for Fair Decision Making. https://people.mpi-sws.org/~gummadi/papers/process_fairness.pdf
- [37] Nikole Hannah-Jones. 2020. From the Magazine: 'It Is Time for Reparations'. *The New York Times* (June 2020). <https://www.nytimes.com/interactive/2020/06/24/magazine/reparations-slavery.html>
- [38] Cheryl I Harris. 1993. Whiteness as Property. *Harvard Law Review* 106, 8 (June 1993), 1710–1791. <https://harvardlawreview.org/1993/06/whiteness-as-property/>
- [39] Tyler Hauptert. 2020. The Racial Landscape of Fintech Mortgage Lending. *Housing Policy Debate* 0, 0 (Nov. 2020), 1–32. <https://doi.org/10.1080/10511482.2020.1825010> Publisher: Routledge _eprint: <https://doi.org/10.1080/10511482.2020.1825010>.
- [40] Tyler Hauptert. 2022. New Technology, Old Patterns: Fintech Lending, Metropolitan Segregation, and Subprime Credit. *Race and Social Problems* (Jan. 2022). <https://doi.org/10.1007/s12552-021-09353-0>
- [41] A. Leon Higginbotham Jr. 1980. *In the Matter of Color: Race and the American Legal Process: The Colonial Period* (4th edition ed.). Oxford University Press, Oxford.
- [42] Amy E. Hillier. 2003. Redlining and the Home Owners' Loan Corporation. *Journal of Urban History* 29, 4 (May 2003), 394–420. <https://doi.org/10.1177/0096144203029004002>
- [43] Daniel E. Ho and Alice Xiang. 2020. Affirmative Algorithms: The Legal Grounds for Fairness as Awareness. <https://lawreviewblog.uchicago.edu/2020/10/30/aa-ho-xiang/>
- [44] Jackelyn Hwang, Michael Hankinson, and Kreg Steven Brown. 2015. Racial and Spatial Targeting: Segregation and Subprime Lending within and across Metropolitan Areas. *Social Forces* 93, 3 (March 2015), 1081–1108. <https://doi.org/10.1093/sf/sou099>
- [45] Kenneth T. Jackson. 1980. Race, Ethnicity, and Real Estate Appraisal: The Home Owners Loan Corporation and the Federal Housing Administration. *Journal of Urban History* 6, 4 (Aug. 1980), 419–452. <https://doi.org/10.1177/009614428000600404>
- [46] Jonathan Kaplan and Andrew Valls. 2007. Housing Discrimination as a Basis for Black Reparations. *Public Affairs Quarterly* 21, 3 (2007), 255–273. <https://www.jstor.org/stable/40441462> Publisher: [North American Philosophical Publications, University of Illinois Press].
- [47] Amir-Hossein Karimi, Bernhard Schölkopf, and Isabel Valera. 2021. Algorithmic Recourse: from Counterfactual Explanations to Interventions. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAcCT '21)*. Association for Computing Machinery, New York, NY, USA, 353–362. <https://doi.org/10.1145/3442188.3445899>
- [48] Ana Hernandez Kent, Lowell Ricketts, and Ray Boshara. 2019. What Wealth Inequality in America Looks Like: Key Facts & Figures. <https://www.stlouisfed.org/open-vault/2019/august/wealth-inequality-in-america-facts-figures>
- [49] Jon Kleinberg, Sendhil Mullainathan, and Manish Raghavan. 2017. Inherent Trade-Offs in the Fair Determination of Risk Scores. *Proceedings of Innovations in Theoretical Computer Science (ITCS)* (2017). <http://arxiv.org/abs/1609.05807> arXiv: 1609.05807.
- [50] Steven Manson, Jonathan Schroeder, David Van Riper, Tracy Kugler, and Steven Ruggles. 2021. National Historical Geographic Information System: Version 16.0. <https://doi.org/10.18128/D050.V16.0> Version Number: 16.0 Type: dataset.
- [51] Scott Markley. 2021. Tabulating HOLC Area Description Sheet Data. (Sept. 2021). <https://doi.org/10.17605/OSF.IO/QYTJ8> Publisher: OSF.
- [52] Douglas S. Massey and Nancy A. Denton. 2003. *American apartheid: segregation and the making of the underclass* (10. print ed.). Harvard Univ. Press, Cambridge, Mass.
- [53] Thomas McCarthy. 2004. Coming to Terms with Our Past, Part II: On the Morality and Politics of Reparations for Slavery. *Political Theory* 32, 6 (2004), 750–772. <https://www.jstor.org/stable/4148144> Publisher: Sage Publications, Inc..
- [54] Ana Patricia Munoz, Marlene Kim, Mariko Chang, Regine Jackson, Darrick Hamilton, and William A. Darity. 2015. *The Color of Wealth in Boston*. SSRN Scholarly Paper ID 2630261. Social Science Research Network, Rochester, NY. <https://doi.org/10.2139/ssrn.2630261>
- [55] Arvind Narayanan. 2018. 21 Fairness Definitions and Their Politics. <https://www.youtube.com/watch?v=jXUuYdnyyk>
- [56] Garrett Dash Nelson. 2021. Redlining, Race, and the Color of Money. https://www.dissentmagazine.org/online_articles/redlining-race-and-the-color-of-money
- [57] Robert K. Nelson, Ladale Winling, Richard Marciano, and Nathan Connolly. 2017. Mapping Inequality. <https://dsl.richmond.edu/panorama/redlining/>

- [58] Michael Omi and Howard Winant. 2015. *Racial formation in the United States* (third edition ed.). Routledge/Taylor & Francis Group, New York.
- [59] Samir Passi and Solon Barocas. 2019. Problem Formulation and Fairness. In *Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT* '19)*. Association for Computing Machinery, New York, NY, USA, 39–48. <https://doi.org/10.1145/3287560.3287567>
- [60] Rashida Richardson. 2021. *Racial Segregation and the Data-Driven Society: How Our Failure to Reckon with Root Causes Perpetuates Separate and Unequal Realities*. SSRN Scholarly Paper 3850317. Social Science Research Network, Rochester, NY. <https://papers.ssrn.com/abstract=3850317>
- [61] Jacob S. Rugh, Len Albright, and Douglas S. Massey. 2015. Race, Space, and Cumulative Disadvantage: A Case Study of the Subprime Lending Collapse. *Social Problems* 62, 2 (May 2015), 186–218. <https://doi.org/10.1093/socpro/spv002>
- [62] Jacob S. Rugh and Douglas S. Massey. 2010. Racial Segregation and the American Foreclosure Crisis. *American Sociological Review* 75, 5 (Oct. 2010), 629–651. <https://doi.org/10.1177/0003122410380868> Publisher: SAGE Publications Inc.
- [63] Saint Paul City Council. 2021. Reparations Legislative Advisory Committee. <https://www.stpaul.gov/department/city-council/reparations-legislative-advisory-committee>
- [64] Marc Schlossberg. 2003. GIS, the US Census and Neighbourhood Scale Analysis. *Planning Practice and Research* 18, 2-3 (May 2003), 213. <https://doi.org/10.1080/0269745032000168269>
- [65] Trian Williams Shanks and Michael Sherraden. 2005. The Homestead Act: A Major Asset-Building Policy in American History. In *Inclusion in the American Dream: Assets, Poverty, and Public Policy*. Oxford University Press.
- [66] Jennifer L. Skeem and Christopher T. Lowenkamp. 2016. Risk, race, and recidivism: Predictive bias and disparate impact. *Criminology* 54, 4 (Nov. 2016), 680–712. <https://doi.org/10.1111/1745-9125.12123>
- [67] Neil Smith. 1996. *The new urban frontier: gentrification and the revanchist city*. Routledge, London ; New York.
- [68] Justin P. Steil, Len Albright, Jacob S. Rugh, and Douglas S. Massey. 2018. The social structure of mortgage discrimination. *Housing Studies* 33, 5 (July 2018), 759–776. <https://doi.org/10.1080/02673037.2017.1390076> Publisher: Routledge _eprint: <https://doi.org/10.1080/02673037.2017.1390076>
- [69] Harini Suresh and John V. Guttag. 2021. A Framework for Understanding Sources of Harm throughout the Machine Learning Life Cycle. *Equity and Access in Algorithms, Mechanisms, and Optimization* (Oct. 2021), 1–9. <https://doi.org/10.1145/3465416.3483305> arXiv: 1901.10002.
- [70] Gregory Svirnovskiy. 2021. A bill on studying reparations is getting a House vote 30 years in the making. <https://www.vox.com/2021/4/14/22384080/hr-40-black-reparations-vote-house-judiciary-committee>
- [71] Keeanga-Yamahtta Taylor. 2019. *Race for Profit: How Banks and the Real Estate Industry Undermined Black Homeownership*. University of North Carolina Press, Chapel Hill. https://www.jstor.org/stable/10.5149/9781469653686_taylor
- [72] Ashley Traynum-Carson. 2021. Asheville City Council makes initial \$2.1 million in reparations funding appropriation. <https://www.ashevilenc.gov/news/asheville-city-council-makes-initial-2-1-million-in-reparations-funding-appropriation/>
- [73] Rachel Treisman. 2021. In Likely First, Chicago Suburb Of Evanston Approves Reparations For Black Residents. *NPR* (March 2021). <https://www.npr.org/2021/03/23/980277688/in-likely-first-chicago-suburb-of-evanston-approves-reparations-for-black-reside>
- [74] Urban institute. 2019. Reducing the Racial Homeownership Gap. <https://www.urban.org/policy-centers/housing-finance-policy-center/projects/reducing-racial-homeownership-gap>
- [75] Berk Ustun, Alexander Spangher, and Yang Liu. 2019. Actionable Recourse in Linear Classification. *Proceedings of the Conference on Fairness, Accountability, and Transparency* (Jan. 2019), 10–19. <https://doi.org/10.1145/3287560.3287566> arXiv: 1809.06514.
- [76] Suresh Venkatasubramanian and Mark Alfano. 2020. The philosophical basis of algorithmic recourse. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT* '20)*. Association for Computing Machinery, New York, NY, USA, 284–293. <https://doi.org/10.1145/3351095.3372876>
- [77] Rinaldo Walcott. 2021. *On Property: Policing, Prisons, and the Call for Abolition*. Biblioasis.
- [78] Mark Weber, Mikhail Yurochkin, Sherif Botros, and Vanio Markov. 2020. Black Loans Matter: Distributionally Robust Fairness for Fighting Subgroup Discrimination. *arXiv:2012.01193 [cs]* (Nov. 2020). <http://arxiv.org/abs/2012.01193> arXiv: 2012.01193.
- [79] Samuel C. Wheeler. 1997. Reparations Reconstructed. *American Philosophical Quarterly* 34, 3 (1997), 301–318. <https://www.jstor.org/stable/20009901> Publisher: [North American Philosophical Publications, University of Illinois Press].
- [80] LaDale C Winling and Todd M Michney. 2021. The Roots of Redlining: Academic, Governmental, and Professional Networks in the Making of the New Deal Lending Regime. *Journal of American History* 108, 1 (June 2021), 42–69. <https://doi.org/10.1093/jahist/jaab066>
- [81] Wenfei Xu. 2018. The Legacy of Redlining: Residential Segregation in 147 American Cities. <https://www.wenfeixu.com/redliningmap/>
- [82] Wenfei Xu. 2022. Legacies of Institutionalized Redlining: A Comparison Between Speculative and Implemented Mortgage Risk Maps in Chicago, Illinois. *Housing Policy Debate* 32, 2 (March 2022), 249–274. <https://doi.org/10.1080/10511482.2020.1858924> Publisher: Routledge _eprint: <https://doi.org/10.1080/10511482.2020.1858924>

A APPENDICES

Variable	Description	Mean	SD
Treatment Variable			
<i>treat</i>	Binary variable indicating whether a neighborhood was D-grade (<i>D</i>).		
Unit			
<i>holc_id</i>	Unique ID of a HOLC neighborhood (<i>i</i>).		
Time			
<i>year</i>	Year of the time (1940-2016). Ten-year intervals except 2010 and 2016 (<i>t</i>).		
Covariates			
<i>population</i>	Population of a neighborhood.	6,064.4	12,750.6
<i>pop_density</i>	Population density of a neighborhood.	2,696.3	3,745.5
<i>white_perc</i>	Percentage of White people in a neighborhood.	0.94	0.13
<i>dissimilarity_index_weighted</i>	Spatially-weighted dissimilarity index for measuring segregation.	0.19	0.25
<i>median_adj_rent</i>	Median rent adjusted to 2019 dollars.	556.0	248.7
<i>mid_age</i>	Median age of buildings of a neighborhood.	19.0	16.7
Outcome Variables			
<i>median_adj_homevalue</i>	Median home value adjusted to 2019 dollars.	198,750.2	209,093.5
<i>roi_homevalue</i>	Return on Investment rate of given time. Calculated as $\frac{\text{median_homevalue}_t - \text{median_homevalue}_{t-10}}{\text{median_homevalue}_t}$. Therefore, the ROI values are calculated after $t = 1950$.	0.13	0.37
<i>nonwhite_homeownership</i>	Home ownership rate of non-White people in a neighborhood.	0.44	0.33

Table 1: Variables and descriptive statistics for the analysis of case study 1.

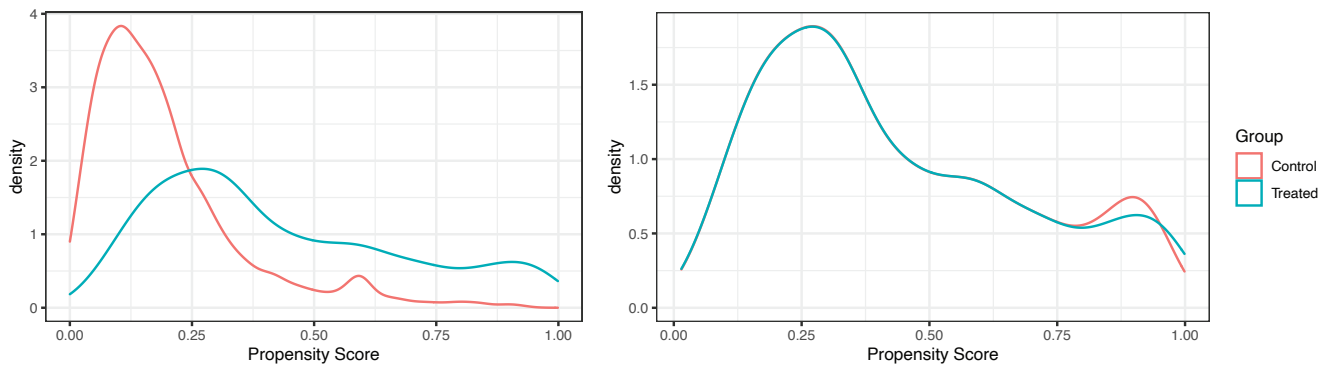


Figure 6: Propensity score distribution before (left) and after (right) matching. Treated are neighborhoods that were D-graded.

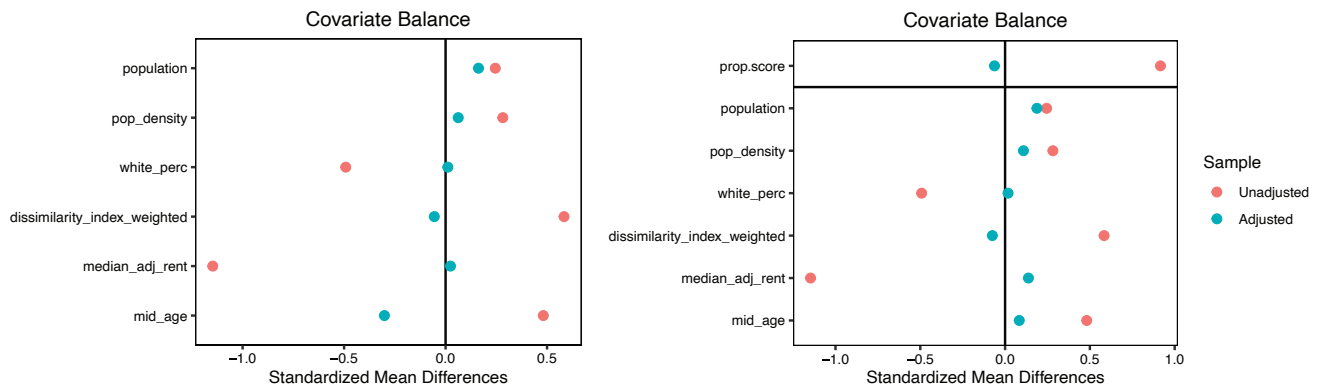


Figure 7: Covariate balance results after matching (left) and weighting (right).

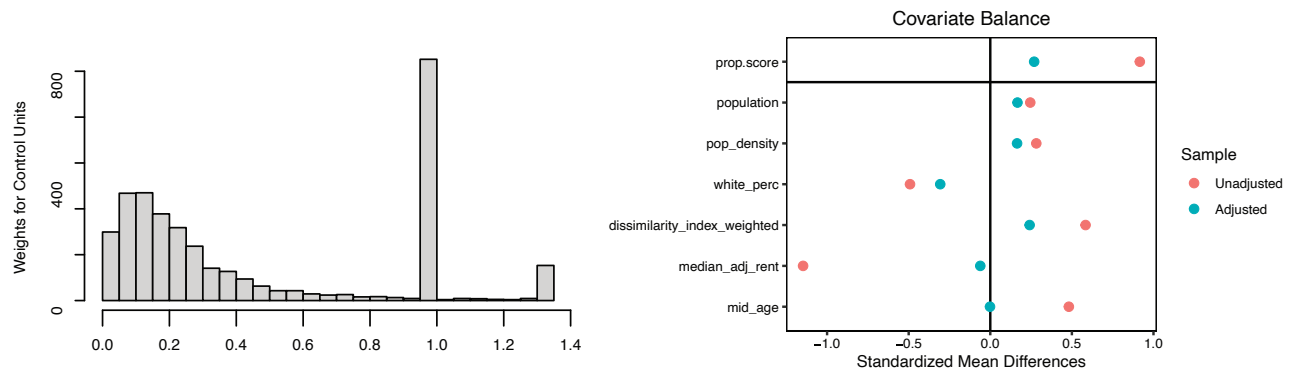


Figure 8: Histogram of control weights after trimming the 5th and 9th percentile weights (left) and the result of covariate balance using trimmed weights (right).

Feature Name	Actionable	Flip Direction	Grid Size	Step Type	Step Size	Lower Bound	Upper Bound
<i>income</i>	False	-1	101	relative	0.01	-0.570	3.127
<i>loan_term</i>	False	1	101	relative	0.01	-4.696	0.214
<i>loan_to_value_ratio</i>	True	-1	102	relative	0.01	-3.479	1.312
<i>loan_amount</i>	False	1	102	relative	0.01	-1.089	3.861
<i>property_value</i>	False	-1	101	relative	0.01	-0.924	3.732
<i>tract_minority_population_percent</i>	False	1	101	relative	0.01	-0.991	3.851
<i>tract_to_msa_income_percentage</i>	False	1	101	relative	0.01	-2.063	2.895
<i>tract_owner_occupied_units</i>	False	1	101	relative	0.01	-2.083	2.129
<i>tract_one_to_four_family_homes</i>	False	-1	101	relative	0.01	-2.199	2.493
<i>tract_median_age_of_housing_units</i>	False	1	101	relative	0.01	-1.777	1.487
<i>linearized_debt_to_income_ratio</i>	True	-1	102	relative	0.01	-1.688	2.393
<i>interest_only_payment_1</i>	False	0	2	absolute	1.0	0.0	1.0
<i>interest_only_payment_2</i>	False	0	2	absolute	1.0	0.0	1.0
<i>balloon_payment_1</i>	False	0	2	absolute	1.0	0.0	1.0
<i>balloon_payment_2</i>	False	0	2	absolute	1.0	0.0	1.0
<i>other_nonamortizing_features_1</i>	False	0	2	absolute	1.0	0.0	1.0
<i>other_nonamortizing_features_2</i>	False	0	2	absolute	1.0	0.0	1.0
<i>conforming_loan_limit_C</i>	False	0	2	absolute	1.0	0.0	1.0
<i>conforming_loan_limit_NC</i>	False	-1	2	absolute	1.0	0.0	1.0
<i>conforming_loan_limit_U</i>	False	0	2	absolute	1.0	0.0	1.0
<i>conforming_loan_limit_NA</i>	False	0	2	absolute	1.0	0.0	1.0

Table 2: Action sets for finding recourse of the logistic regression classifier trained for loan decisions. The lower and upper bound shows the data of Massachusetts.

	Conventional First-Lien Loan Applications		FHA-backed First-Lien Loan Applications	
Total #	3,104,084		826,231	
	%		%	
Approved	96.27		97.07	
Subprime	4.42		21.39	
Applicant Demographics				
	%		%	
Non-Hispanic White	62.40		45.55	
Black or African American	4.34		15.34	
Hispanic or Latino	8.39		21.47	
Asian	6.93		2.13	
Native American	0.39		0.76	
Other Race	0.27		1.01	
Male	31.76		36.55	
Female	22.24		29.68	
Age is over 62	11.71		3.69	
Loan Characteristics				
	%		%	
Jumbo Loan	6.1		0.15	
Interest Only Payment	2.4		0.32	
Balloon Payment	0.32		0.004	
Other Non-amortizing Features	0.09		0.07	
	Mean	SD	Mean	SD
Income (1,000\$)	137.89	707.3	74.42	119.60
Loan Term (month)	345.24	48.52	359.57	8.49
Loan Amount (\$)	327,787.3	285,892	234,620.75	106,604.41
Property Value (\$)	419,111.29	487,631.51	243,137.63	111,855.9
loan-to-value ratio	82.31	14.76	96.73	5.36
debt-to-income ratio	34.58	8.71	42.92	9.28
Census Tract Characteristics				
	Mean	SD	Mean	SD
Minority Population (%)	27.67	22.78	37.28	27.52
Family Income to MSA (%)	118.88	44.68	101.03	32.62
# of Owner Occupied Units	1510.31	962.87	1492.66	931.58
# of 1-4 Family Houses	2042.55	1212.28	2097.86	1148.58
Median Age of Buildings	33.96	17.90	34.49	17.91

Table 3: Descriptive statistics of conventional and FHA-backed loan applications used for the case study 2. Data: 2020 HMDA data.