An Unequal Dream:
The Mortgage Rate Premium Paid by Black Communities

Michael Nicholson

Professor Emma Rasiel, Faculty Advisor
Professor Kent Kimbrough, Honors Seminar Instructor

Duke University
Durham, North Carolina
2020

Acknowledgements

Abstract

Introduction

Few ideas embody the spirit of the United States as thoroughly as the American Dream, the idea that anyone with enough time, effort, and grit can be successful, no matter their background or upbringing. However, over the years, the core principals of this ideal have been called into question, as many find its tenants largely out of reach (Goodman et al., 2018; Shlay, 2006). This begs the question, is there inequality in the American Dream?

To provide greater clarity to this question, this paper analyzes loan pricing discrimination against predominately black communities in US mortgage markets. Building on previous literature, this paper posits that pricing discrimination has reduced significantly since the pre-crisis boom in housing, but ceteris paribus, prominently black neighborhoods still face economically significant discrimination in mortgage pricing. These results have significant ramifications for wealth and wellbeing in black communities with homeownership serving as a key pillar of the American Dream and mortgage attainment being vital to financing a home.

Background

The realized ambitions of the American Dream have varied over the years, but since the late 1900s, homeownership has been a key component to the American Dream. The Clinton and Bush administrations made homeownership central to their campaigns and precipitated a historic rise in homeownership rates. George W. Bush's pledge in an October 2004 speech to create an "ownership society" epitomizes these efforts. The push for homeownership led to an expansion in housing credit from the late 1990s to mid-2000s[1] that gave rise to historically high homeownership rates for all races and ethnicities, including a peak of 49.7% in 2004 for black households (Bayer, 2016). As a result of its association with the American Dream,

---

[1] See https://fred.stlouisfed.org/series/RHORUSQ156N

homeownership has become engrained in American society and has become a key component of

wellbeing in the United States (Pager & Shepard, 2018). This is in addition to the broad positive

wealth effects of homeownership.

However, despite the efforts of various administrations and interest groups, homeownership

for black populations has long lagged national averages by a wide margin[2] (Rohe et al., 2002).

Several factors contribute to this gap, but the literature and resulting legislation focus primarily

on credit availability since most homes are purchased with a loan. Discrimination in the

attainment of a loan arises from two main factors, denial rates and charged interest rates. These

factors capture the ability to obtain a mortgage for a home purchase and the cost in order to pay

that loan. Both of these factors introduce the potential for discrimination on the basis of race.

Based on fears of discrimination in mortgage lending, the Fair Housing Act was enacted in

1968 by the Office of Fair Housing and Equal Opportunity[3] within the U.S. Department of

Housing and Urban Development. This act explicitly prohibited lenders from discriminating on

the basis of race. Subsequently, the Home Mortgage Disclosure Act[4] (HMDA) and Community

Reinvestment Act (CRA) were passed in 1975 and 1977, respectively, to monitor discrimination

against disadvantaged borrowers (Delis & Papadopoulos, 2019). This legislation was passed on

fears of redlining[5] against minority and low-income neighborhoods (Bayer et al. 2016). It was

widely believed that such discrimination was an explicit part of lenders' policy (Ladd, 1998;

Munnel et al., 1996). In order to support the enforcement of the Fair Housing Act, these two acts

intended to provide the public with relevant information on the lending practices of large

---

[2] See https://www.census.gov/housing/hvs/data/charts/fig08.pdf
[3] The Office of Fair Housing and Equal Opportunity is responsible for overseeing the enforcement of Federal laws prohibiting discrimination on the basis of race, color, religion, national origin, sex, disability, and familial status
[4] The analysis within this paper is largely based on current iterations of disclosures required by this act
[5] Discrimination on the basis of geographical characteristics, primarily referring to discrimination on the basis of race

financial institutions and gave public officials critical information to enact more targeted policy. The implementation of these acts led to the creation of the HMDA dataset.

The HMDA requires that most financial institutions[6] that originate loans disclose loan-level data for all applications they receive in a given year. The Federal Financial Institutions Examination Council compiles and releases the data publicly on an annual basis. The creation of this database largely founded the current literature on origination and pricing discrimination in mortgage markets and is widely used today. However, despite its extensive scope, the original dataset suffered from several well-documented deficiencies, most notably the absence of vital borrower characteristics, that limited its use in isolation (Avery et al., 2008; Horne, 1997). To address these concerns, several reforms have been implemented to the disclosure requirements for lenders. These revisions and empirical strategies to correct for the past omitted variable problem of the dataset will be discussed more thoroughly in the next section.

Most recently, implementation of the Dodd-Frank legislation led to a significant expansion in the required disclosures for HMDA compliant institutions. These changes largely address previously omitted variables and, for the first time, include loan pricing data for all loans in the sample. These changes came into effect beginning with the 2018 dataset, which was released in 2019. For the first time, the HMDA dataset allows for comprehensive analysis to be conducted on loan pricing data in mortgage markets for the nation as a whole and will be the basis for the empirical work in this paper.

---

[6] Institutions that originate less than 100 loans, do not have branches in a Metropolitan Statistical Area (MSA), or have total assets under certain thresholds are exempt

Methodology and Structure

Using the significantly expanded HMDA data released in 2018, this paper will analyze whether predominately black neighborhoods face discrimination in loan pricing[7] after accounting for borrower, mortgage, and geographic characteristics. The study will conduct this analysis using aggregated tract-level data from the HMDA dataset from 2018. This analysis will include several variables, including measures of income, debt-to-income[8] (DTI), loan-to-value[9] (LTV), lender market penetration, racial and ethnic demographics, and the novel use of denials on the basis of credit as a proxy for the strength of credit for each tract. Fixed effects will also be analyzed on the county in which each tract resides and for the most prominent lender within each tract. The primary independent variable of interest is a binary variable indicating whether over 50% of the applicants in a tract are black. Rate spread is utilized to measure loan pricing as the dependent variable. Rate spread is the difference in the interest rate charged to the borrower and the relevant average prime offer rate. The average prime offer rate is derived from a survey administered by the Federal Financial Institutions Examination Council. The coefficient on this variable measures the differential in rate spreads attributable to race in predominantly black neighborhoods and will be the primary measure of discrimination in this study. In short, this paper seeks to calculate the difference in rate spreads paid by predominantly black tracts verse non-predominantly black tracts for 2018 after accounting for borrower, mortgage, and geographic characteristics.

In contrast to previous literature which largely focusses on 30-year mortgages, this paper analyzes all first-lien home loans[10] of any maturity to better capture the full breadth of loan types

---

[7] Rate spread will be utilized to measure loan pricing differentials
[8] DTI is a ratio comparing a borrower's income to total debt payments
[9] LTV is a ratio computing a borrower's leverage, the amount of the mortgage over the value of the property
[10] The population of loans included in the regression analysis will be described more fully in the Data section

taken by borrowers. This choice is predicated on the fact that the distribution of loan types differs by race, for example, black applicants are much more likely to seek financing for a manufactured home than white applicants and looking solely at conventional 30-year fixed-rate mortgages may hide differences in the availability and pricing of loans actually taken out by predominately black communities.

The work in this paper is further differentiated from previous literature by its use of the expanded HMDA dataset, its focus on rate spreads at the tract-level, and its use of credit denials as a proxy for credit. Much of the current literature attempts to account for omitted variables from past iterations of the HMDA data by matching records with localized or private datasets (Bayer et al., 2018; DeLoughy, 2012; Ghent et al., 2014; Rugh & Massey, 2010). While this often increases the available scope of analysis, this opens up the potential for additional idiosyncratic omitted variables from the chosen regions of study. Idiosyncratic factors in these areas have even been used as a defense by lenders facing discrimination suits from the National Association for the Advancement of Colored People (Delis & Papadopoulos, 2019). The use of the HMDA data allows for an analysis of a significant portion of all loan applications in the country, reducing idiosyncratic factors.

Further, these local data sets are typically for large metropolitan areas, for example, New York and Boston, reducing their inference to the country as a whole. Large urban centers have housing markets that operate differently than smaller metropolitan areas and more rural regions. For example, the large concentration of people may allow for the support of a greater number of lenders to operate in these areas allowing for greater competition. These areas are also more diverse than other parts of the country, also potentially biasing the results. Though the HMDA

data has issues of its own[11], it allows for a much more complete analysis of loans in the country as a whole.

Building on previous literature, this paper posits that pricing discrimination has reduced significantly since the pre-crisis boom in housing, but ceteris paribus, prominently black neighborhoods still face economically significant discrimination in loan pricing. Building to this conclusion, the following sections will first conduct a Literature Review of the research concerning discrimination in loan availability and loan pricing, focusing primarily on discrimination on the basis of race. The Data section will describe the available variables as well as the potential strengths and weaknesses of using HMDA data. It will also describe the calculation of the tract-level variables used in the regression analysis. The Empirical Specification section will describe the model used as well as the expected outputs for the most relevant independent variables. Next, the Results section will present the empirical estimations of the model therein described. The paper will then conclude with a summary of this study's findings.

Literature Review

With its ties to homeownership and the American Dream, the literature on discrimination in mortgage markets is deep and active. This research has largely been predicated on the public availability of the HMDA dataset. For the past few decades, the annual HMDA dataset has proved instrumental to analyzing loan discrimination and has been used at least in part by most studies since its enactment (Avery et al., 2008; Bayer et al., 2018; Bocian et al., 2008; Courchane, 2007; Delis & Papadopoulos, 2019; Delought, 2012; Ghent et al., 2014; Haughwout et al., 2009; Rugh & Massey, 2010; Wheeler et al., 2015). However, early implementations of

---

[11] These will be addressed in the Data section

the dataset lacked key borrower metrics negating causal interpretation of loan acceptance rate differentials as evidence of discrimination. This section will provide an overview of the various methodologies and additional datasets used to address these issues and give a short history of the HMDA dataset and relevant literature.

In 1989, the HMDA dataset was greatly expanded, precipitating a surge of research on discrimination on loan denial rates (Delis & Papadopoulos, 2019). The expanded data included highly relevant and previously excluded characteristics, including loan outcome, location, and the applicant's race and gender. For the first time, a comprehensive national dataset of loan-level application data was made available to the public allowing investigation into potential discrimination and the lenders at fault. However, the HMDA dataset still lacked several vital characteristics to determine the quality of the applicant, most notably DTI, LTV, and credit scores. Despite these deficiencies, the expanded loan characteristics in the dataset brought to light large gaps in the approval rates between different races and ethnicities (Avery et al., 2008). However, the lack of critical borrower characteristics negated the ability of this dataset to prove discriminatory practices. Although, it did allow for a comprehensive comparison of lenders' actions towards minority groups that was revolutionary to the research community.

In light of the omitted variables in the HMDA data, a common approach was, and broadly still is, to match the data to local datasets to allow for a more complete, though localized, analysis (Bayer et al., 2018; DeLoughy, 2012; Ghent et al., 2014). Most famously, the Boston Fed's 1990 dataset included 38 additional variables and spawned several papers attempting to account for the omitted variable problems of the HMDA data (Day et al., 1998; Munell et al., 1996), though it has been argued that these added variables are not sufficient to account for all sources of omitted variable bias (Horne, 1997).

In the early 2000s, during the housing boom that preceded the financial crisis, it is largely agreed upon that discrimination shifted from primarily affecting loan approval rates to loan pricing, realized as higher rate spreads[12] for black and Hispanic populations (Faber, 2013; Ghent et al., 2014; Williams et al., 2005). Along with the dramatic rise in housing seen in the early 2000s[13], credit standards were largely relaxed, allowing for many previously barred borrowers to access the housing market[14]. Lower bars to access credit gave many black and Hispanic Americans access to the mortgage market for the first time, but this access largely came at the cost of higher interest rates (Rugh & Massey, 2010).

Utilizing high-cost loans did increase homeownership rates but also had the effect of increasing monthly housing payments and thus constraining housing budgets. Higher rates also increase the hurdle rate required to break even on an investment, potentially leading to long-term reductions in accumulated wealth. In addition to increasing payments, high-cost mortgages have been associated with a six-percentage-point increase in subsequent foreclosure notices (Bayer et al., 2016). In combination, these effects left black populations particularly vulnerable entering into the Financial Crisis.

In 2002, the Federal Reserve Board revised the HMDA disclosure requirements, substantially increasing the depth and coverage of the dataset. Most notably, from the 2004 release onward, pricing characteristics were required for loans with rate spreads over a certain threshold and several mortgage characteristics[15] were added (Avery et al., 2008). While previous literature had explored the potential of loan pricing discrimination using survey data and other small datasets,

---

[12] Rate spread refers to the difference in the interest rate charged to the borrower and a benchmark interest rate for the same or similar maturity, US treasuries and the prime offer rate are common benchmarks
[13] See https://fred.stlouisfed.org/series/USSTHPI
[14] This created a circular effect which helped perpetuate the rise in housing prices as larger populations had access to the market and relatively cheap credit
[15] These additions included lien status, designation of manufactured housing, HOEPA status, etc.

the expanded HMDA guidelines led to a significant increase in literature focusing on rate spreads at a comprehensive national level. However, the data still lacked important characteristics relating to applicant credit risks and the type of loan extended (Avery et al., 2008).

The newly expanded HMDA data was released just as the housing market was peaking, giving insight into the effect of the housing crash on black and other minority populations. Black and Hispanic populations were found to be particularly impacted by the crisis, with both reduced credit availability (Avery et al., 2011) and higher foreclosure rates (Edmiston, 2009; Reid et al., 2009). These effects were largely contributed to high-cost lending predominantly concentrated in minority neighborhoods (Chan et al., 2015; Mayer & Sherlund, 2008). In addition to the trauma of enduring the Financial Crisis, these disadvantaged populations faced significant reductions in wealth and carried foreclosure on their credit histories for years (Bayer et al., 2016).

In the years following the crisis, research has continued on loan pricing discrimination and has resulted in several U.S. Department of Justice cases[16] (Bayer et al., 2018). However, most analysis is conducted on data up to or preceding 2013. In 2018, as part of the Dodd-Frank Wall Street Reform and Consumer Protection Act, the HMDA data was expanded to include several previously omitted variables. Most notably, these include credit score[17], age, LTV, DTI, origination charges, points and fees, loan term, interest rate, and rate spread for all loans[18]. To my knowledge, this expanded dataset has yet to be exploited by any published work and will be the basis for the analysis in this paper.

---

[16] Defendants include Wells Fargo, HSBC, CitiMortgage, SunTrust, JP Morgan, First Horizon, Long Beach Mortgage Company, and Bear Sterns among others
[17] Credit score must be reported but is redacted from public data releases on the grounds of preserving the anonymity of the borrower
[18] Rate spreads were previously only included for loans with a spread exceeding a determined threshold

In addition to the use of the updated HMDA dataset, this paper differentiates from the literature by analyzing the loan pricing disparities for predominantly black neighborhoods in place of individual loans. The literature on loan-level racial discrimination is deep, but much less research has been done on neighborhood-level racial effects. This analysis allows for a deeper understanding of whether entire communities face discrimination on the basis of race and relevant demographic effects on mortgage rates in these neighborhoods. Regarding community level price discrimination, the limited research that has been conducted utilized limited survey data (Nothaft & Perry, 2002) or localized data (Kau et al., 2012), which constrains the extension of their respective results to the national level. Additionally, the existing literature analyzes data that pre-dates the financial crisis, whereas subsequent studies have shown that mortgage discrimination did not shift from loan origination to loan prices until the years immediately preceding the financial crisis (Faber, 2013; Ghent et al., 2014; Williams et al., 2005). This paper's use of 2018 data allows for conclusions that are more relevant to current market dynamics.

Data

HMDA Data

The analysis in this paper uses the 2018 HMDA dataset. Released on an annual basis, this dataset includes loan applications for most lenders in the United States. The coverage is robust with approximately 80% of all mortgage applications included in the dataset (Avery et al., 2008; Wheeler et al., 2015). As previously mentioned, the HMDA disclosure requirements were enacted as part of the Home Mortgage Disclosure Act passed in 1975. It is intended to give researchers and public officials a means with which to monitor and police discrimination against minority and disadvantaged groups. Historically, the dataset has suffered from several omitted

13

variables, but as part of the Dodd-Frank legislation, the data released for 2018 onward includes a number of the most pressing missing variables. Most relevant to this study, rate spread has been added for all loans.

In 2018, the dataset included 15,119,625 total data points with 99 included variables, though a number of these are redundant. It includes data from all 50 states as well as for Puerto Rico, Guam, and the Virgin Islands. The variables largely categorize as borrower characteristics, loan characteristics, application decision variables, lender and geographical information, and appended census data for the relevant tract[19]. A list of these variables can be found in the Appendix.

In order to focus the study solely on loans that contribute to homeownership and to homogenize the data, this study only uses a subset of the application data. Non-home purchase applications are removed to excluded refinancing and other loans that do not directly increase homeownership. Only first-lien loans are considered to limit the idiosyncrasies of subordinate lien loans. Only primary resident, non-commercial loans are considered to focus the study on loans intended for first/primary homes since owning additional homes or business properties does not affect overall homeownership. The study also excludes preapproval applications. These constraints reduce the size of the dataset to 6,613,302 applications.

The expanded HMDA dataset was chosen for this paper for its comprehensive nature and robust data. No other publicly available dataset includes the breadth of mortgage applications contained within this data. This dataset is used in isolation to allow for inference on the country as a whole. Matching to localized data significantly reduces the scope of the analysis limiting conclusions at the national level and potentially introducing idiosyncratic variables (Delis &

---

[19] The full variable specification can be found here: https://ffiec.cfpb.gov/documentation/2019/lar-data-fields/

Papadopoulos, 2019). While this was largely necessary in the past to account for the significant number of omitted variables in the HMDA data, the recently expanded dataset now encompasses most of these variables, notably rate spread and DTI.

That being said, the HMDA data still suffers from several weaknesses[20]. While the data covers the vast majority of relevant mortgage applications, it systematically underrepresents applications from rural areas. Lenders with assets underneath certain thresholds, these differ based on the type of institution, or those that do not have branch offices in MSAs do not have to report their loan applications. There is also an exception for lenders that originate less than 100 loans in a given year. With these limitations, the analysis of this paper is best interpreted as solely applying to tracts housed within MSAs. Even with these constraints, it has been estimated that approximately 80% of all mortgage loans are included in the sample (Avery et al., 2008; Wheeler et al., 2015).

Another weakness of the HMDA data in its current form is the redaction of credit scores. Credit scores are a key component in loan decisions and must be accounted for to estimate the presence of discrimination accurately. While the data does not include credit scores, it does include the applications rejected on the grounds of poor credit. Since this paper analyzes aggregated tract-level data, credit denials can be added as a credit metric to the regression model. Though a noisy and imperfect measure of credit, the use of credit denials as a proxy for the strength of credit within a tract introduces credit effects to the model.

In order to analyze discrimination against black neighborhoods, the loan-level data is aggregated at the tract-level. For all continuous variables, the mean[21] is used as the tract-level variable; for example, income is calculated as the average income of borrowers in the tract. This

---

[20] A more comprehensive analysis of the HMDA data can be found in Avery et al. 2008
[21] To account for typos in some applications, the 1% trimmed mean is used

aggregation was conducted for income, loan amount, LTV, origination charges, discount points, and loan term. DTI is reported as a binned variable. As such, the median of borrowers in a tract was taken in place of a mean. Categorical and binary variables were calculated as a proportion of borrowers in a tract with that characteristic. These calculations were conducted for race, gender, ethnicity, purchaser, denial reason, manufacture housing, and conventional loan variables.

In addition to these aggregated variables, tract characteristics not directly available from the individual loans were found as well. To account for the level of lending competition within a tract, the percent market share of the most prominent lender was calculated as the number of applications from the most prominent lender for a given tract over the total number of applications in that tract. To allow lender fixed effects in the model, the Legal Entity Identifier of the most prominent lender was found for each tract. To allow for county fixed effects in the model, the county Federal Information Processing Standards (FIPS) code was derived from the tract census code. Lastly, two binary variables were calculated from the data for if the percent of black applicants in a tract was greater than .25 and if the percent is greater than .5. This last variable is the measure of whether a tract is predominantly black.

For the tract-level dataset, there are 72,253 tracts with 41 aggregated or appended variables. With 74,134 tracts[22] in the 2010 US Census, these account for 97.5% of all tracts in the United States. This dataset is far more comprehensive than past rate spread literature using the Mortgage Interest Rate Survey or Survey of Consumer Finances datasets (Cheng et al., 2015; Nothaft & Perry, 2002). For literature utilizing the HMDA data, rate spread was previously only available for a small subset of loans (Bayer et al., 2018; Delis & Papadopoulos, 2019).

---

[22] See https://transition.fcc.gov/form477/Geo/more_about_census_tracts.pdf

Table 1
Descriptive statistics for aggregated and appended HMDA data variables

	All Loan Terms		30 Year Loans	
	Mean	SD	Mean	SD
Tract Applicant Average				
Rate spread	0.82	0.59	0.72	0.41
Interest rate	4.92	0.47	4.80	0.35
Income	90.57	43.70	90.35	43.60
Loan amount (10,000s)	23.94	14.81	24.65	14.79
Total loan costs (1,000s)	4.96	4.54	5.01	4.74
Total points and fees (1,000s)	2.62	3.01	3.33	3.14
Origination charges (1,000s)	1.62	0.84	1.62	0.85
Discount points (1,000s)	1.61	1.15	1.61	1.17
Lender credits (1,000s)	0.87	1.63	0.88	1.64
Loan term (months)	344.80	21.83	360.00	0.00
DTI*	38.33	4.50	38.64	4.44
LTV	86.85	7.36	87.88	7.31
Proportion of Tract Applicants/Loans				
Black	0.09	0.17	0.08	0.17
White	0.70	0.23	0.71	0.23
Asian	0.06	0.13	0.06	0.13
Pacific Islander	0.00	0.02	0.00	0.02
Native American	0.01	0.04	0.01	0.03
Hispanic	0.12	0.21	0.12	0.21
Female	0.24	0.13	0.24	0.14
Fannie Mae purchased	0.13	0.09	0.13	0.09
Ginnie Mae purchased	0.12	0.10	0.13	0.11
Freddie Mac purchased	0.10	0.08	0.10	0.08
Private purchased	0.01	0.02	0.01	0.02
Conventional loan	0.67	0.22	0.62	0.25
Manufactured	0.07	0.15	0.02	0.07
Denied for DTI	0.03	0.05	0.02	0.04
Denied for credit	0.03	0.06	0.05	0.21
Denied for employment	0.00	0.01	0.17	0.13
Percent denied	0.10	0.11	0.09	0.09
HOEPA status	0.00	0.01	0.00	0.01
Meet conforming loan limit	0.05	0.13	0.05	0.13
HMDA Appended Census Data				
Tract minority population	37.26	30.38	37.14	30.30
MSA median income	73.06	16.79	73.10	16.79
Tract to MSA median income	1.01	0.43	1.02	.43
Other Variables				
Lender market share	0.17	0.13	.17	.13
Majority black	0.05	0.21	0.03	0.06
Quarter black	0.10	0.31	0.10	0.30

This table presents means and standard deviations for the aggregated or appended tract-level variables. In the first panel, the average value of the applicants in the tract is presented, (*) with the exception of DTI, which is the median. The second panel shows the proportion of applicants/applicant loans that meet the given criteria. The third panel shows appended census data that was present in the original HMDA data. Lastly, the fourth panel shows the market share of the most prominent lender and the two binary variables marking if a given tract has over 50% black applicants and 25% black applicants, respectively.

Table 1 shows descriptive statistics for the tract-level variables calculated from the loan-level HMDA data. In order to compare the applicant data used in this paper with 30-year samples used in previous literature, statistics are provided for both 30-year loan applications and the full home mortgage applicant pool. As seen above, interest rates in the sample are largely similar for the full sample and 30-year loans. Additionally, these levels are comparable with those reported by the GSEs[23] and Federal Reserve[24]. Regarding race, black and white borrowers comprise approximately 8% and 70% of applicants, respectively. These levels are comparable to demographic distributions from the 2018 census[25].

Overall, the full and 30-year loan pools are nearly identical; however, rate spreads and interest rates are approximately ten basis points higher for the sample including all loans. This is likely driven by the significantly higher proportion of manufactured housing in the full sample. Manufactured housing units often face higher rate spreads as they are considered inferior collateral to site-built properties, especially in cases where the land the unit resides on does not secure the loan.

Census Data

In addition to the variables in the HMDA dataset described above, tract-level data from the 2017 American Community Survey (ACS) 5-year data was collected from the Census Bureau[26]. Unlike the 1-year datasets, which have certain population thresholds, the 5-year dataset includes data on every tract in the country. This survey includes data on a range of social, economic, demographic, and housing characteristics and is released to the public in the form of aggregate

---

[23] See http://www.freddiemac.com/pmms/
[24] See https://fred.stlouisfed.org/graph/?g=NUh
[25] See https://www.census.gov/quickfacts/fact/table/US/PST045218
[26] A complete description of the methods and variables of the ACS can be found here: https://www.census.gov/data/developers/data-sets/acs-5year.2017.html

counts, for example, the number of unemployed residents in a tract. This data was appended to the HMDA dataset to account for potential omitted variable bias at the tract level.

From the ACS, counts of the residents who are unemployed, do not have health insurance, receive food stamps or Supplemental Nutrition Assistance Program (SNAP) benefits, rent their housing, and have completed a bachelor's degree, respectively, were added. These variables were then converted to percentage terms using the total population of the tract. In addition, the number of housing units that are vacant was retrieved and also converted to percentage terms using the aggregate housing unit count.

Empirical Specification

Using the data described above, this paper utilizes a multivariate regression model with fixed effects to estimate the rate spread differential for predominantly black neighborhoods in the United States. The primary dependent variable is the average rate spread for each tract represented in the dataset. This variable was chosen over interest rate in order to account for the variation in benchmark interest rates across the year[27] and to account for the difference in average loan terms in the tracts. The primary independent variable the paper seeks to estimate is the binary variable representing if the majority of applicants in a tract are black. The coefficient on this variable estimates the rate spread differential attributable to the majority of the applicants being black, ceteris paribus. Based on previous literature, I expect this differential to be positive, indicating an increase in rate spreads, with a magnitude of approximately 25% of the unadjusted differential between predominantly and non-predominantly black neighborhoods. This estimation represents discrimination in loan pricing and the core result of this analysis.

---

[27] In 2018, US ten-year treasury yields opened at 2.46% and closed the year at 2.69%, ranging from 2.37% to 3.24% See https://www.treasury.gov/resource-center/data-chart-center/interest-rates/Pages/TextView.aspx?data=yieldYear&year=2018

Various other independent variables are included to account for potential omitted variable bias in the model. Several factors go into the approval of a loan and the resulting interest rate charged. Many of these, such as credit score, are well known to differ across races (Hanson et al., 2016). These can bias the results since the rate spread differential found may well be attributed to worse borrower characteristics for black neighborhoods, not discrimination by the lender. Thus, in order to correctly estimate the discrimination faced by black neighborhoods, these characteristics must be included. The included variables can largely be divided into three categories: aggregate borrower characteristics, aggregate mortgage characteristics, and geographic characteristics.

The aggregate borrower characteristics in the model include several metrics designed to capture the quality of the applicants in a tract. These include the average income, percentage of applicants that are female[28], and the percent of applicants that are Asian, Native American, Hispanic, or Pacific Islander[29] for the given tract. Additionally, the proportion of applicants denied for credit reasons is included. Average income is included to give a measure of the financial strength of the applicant. It also shows the capacity of a borrower to make their monthly payment. I expect the coefficient on this variable to be negative, though extremely high incomes are often paired with jumbo loans that exceed the maximum size for conventional mortgages and thus can face higher rate spreads. Percent female is included as previous studies have shown that female applicants face higher rate spreads, though this effect has been attributed to shopping behavior in place of discrimination on the basis of gender[30] (Cheng et al., 2011). The breakdown of race and ethnicity accounts for the potential of other minorities to be driving the

---

[28] Percent male is excluded to avoid multicollinearity
[29] Each race and ethnicity is included as a separate variable
[30] It has been shown women are more likely than men to use a lender recommended by a friend or family member and are also less likely to shop around for the best rate (Cheng et al., 2011)

higher spreads in prominently black neighborhoods. Many minority groups face loan pricing discrimination, with active literature on black, Hispanic, and Native populations (Bayer et al., 2018; Cheng et al., 2015; Delis & Papadopoulos, 2019).

The novel inclusion of credit denials from a tract intends to serve as a proxy for unavailable credit variables. Though this variable is for denied applicants and rate spread is solely included for approved borrowers, the inclusion of this variable gives a measure of the overall strength of credit in a tract. It would be expected that tracts with higher rates of credit denials, for example, have lower credit scores as a whole. This would not have to be the case in tracts where there are large populations of both extremely strong and extremely weak applicants, but at the granularity of the tract, where overall wealth and financial security are largely correlated, large inequalities across borrowers is unlikely.

The aggregate mortgage characteristics include a number of metrics designed to capture the desirability of the mortgage itself to a lender and the ability for an applicant to pay such a mortgage. These variables include the average loan amount, median DTI, average LTV, average origination charges, average discount points paid, the proportion of conventional loans, average loan term, and proportion of loans that are for manufactured housing. Loan amount captures the overall size of the loan and reflects both the lender's capacity to make such a loan and the applicant's ability to repay. DTI and LTV are relative measures that more directly capture the applicant's ability to take on more debt. High LTV loans are more highly leveraged and thus riskier for the lender. Thus, I expect its coefficient to be positive. DTI directly captures whether an applicant can make their monthly payment with higher DTIs being riskier for the lender. Thus, I expect the coefficient on this variable to be positive as well.

Average origination charges and average discount points paid capture outside costs apart from the interest rate that may affect the rate spread of a loan. Discount points directly lower the paid interest rate by acting as prepaid interest. Discount points have also been cited as a source of potential omitted variable bias in past literature (Cheng et al., 2015). Conventional loans are those that are not explicitly backed by the government, though many of these loans are still conforming[31] and are often sold to one of the GSEs. Loan term varies greatly depending on the type of loan being originated, though approximately 90% of loans in the dataset are 30-year loans. A shorter loan term can reflect several loan characteristics, including manufactured housing, adjustable-rate mortgages, or 15-year mortgages. Since data includes non-standard 30-year loans, manufactured housing is an important variable to consider. Manufactured properties are less desirable in foreclosure and are often present in less desirable areas. These factors make them riskier for lenders and thus have an expected positive effect on rate spreads. Black applicants are also significantly more likely to apply for a mortgage for a manufactured property than white applicants, introducing the potential for omitted variable bias if the variable is not included.

The included geographical variables measure various tract-level factors that may influence the realized rates spreads. These variables include the proportion of loans purchased by Fannie Mae, Freddie Mac, or Ginnie Mae[32], the percentage the average tract income is of the corresponding average metropolitan statistical area (MSA) income, and the market share of the most prominent lender. Fannie Mae, Freddie Mac, and Ginnie Mae are the three largest government-sponsored entities (GSEs) and are all chartered by the government with the intent of

---

[31] Conforming indicates meeting the standards of government-sponsored entities (GSEs) like Fannie Mae and Freddie Mac, which both have an implicit backing by the government and are currently in conservatorship
[32] The proportion of each GSE was separated into individual variables. Other GSEs are included in the HMDA data but these all have extremely small market shares and as such were excluded as independent variables

increasing homeownership by increasing mortgage availability. All three purchase or insure originated loans in order to increase liquidity in mortgage markets and to free up capital for lenders. Thus, if a larger share of loans in a given tract is purchased by one of the GSEs, this should have the effect of making loans cheaper by lowering the rate spread. By purchasing loans, the GSEs essentially increase the demand for mortgages, which should push down interest rates in a competitive market.

The tract to MSA average income variable captures the relative income of a tract in relation to its surrounding area. This allows the model to capture the presence of lower and higher-income neighborhoods, adjusting for the level of wages in a certain area. The economics of lower and higher-income markets may differ. For example, the GSEs have certain mandates that stipulate they must purchase a certain percentage of low-income loans. Potential bias derives from the fact that lower-income neighborhoods are often riskier for lenders since they are weaker borrowers on average, resulting in higher rate spreads. Additionally, predominantly black neighborhoods have lower average incomes than non-predominately black neighborhoods introducing the potential for omitted variable bias. The relative measure of income in comparison to the rest of an MSA's is more informative than absolute income measures since cost of living differs greatly from one region of the country to another. It would hardly be beneficial to compare incomes from neighborhoods in New York City and Detroit.

The model also includes a measure of competition measured by the market share of the most prominent lender. Depending on the population as well as the geographic and financial characteristics of a tract, lending competition may vary greatly. The inclusion of this variable accounts for any difference that is attributable to the ability to shop around to achieve the highest rate. Though an imperfect measure, this variable captures a portion of the realized distribution of

lending in a tract. For example, it can be expected that lenders with near 100% market share of applicants have little competition or that residents in a particular neighborhood have a strong preference for a certain lender. On the other extreme, the largest lender may have very little market share, indicating a greater level of competition and rate shopping. Assuming that this variable is a viable measure of competition, I would expect the coefficient to be positive, indicating that rate spreads rise as the market share of the largest lender increases.

Additionally, the analysis separately looks at two sets of potential fixed effects, county fixed effects and lender fixed effects. County fixed effects account for the variation of rate spreads attributable to the geographic region in which the tract resides. Housing market dynamics vary widely, potentially biasing the results. Several geographical effects are taken into account with these fixed effects, including population and GDP growth, population migration, the strength of the housing market, racial demographics, relative cost of living, etc. For example, Detroit has a languishing housing market with thousands of homes sitting empty as well as a relatively high black population, whereas San Francisco has a hot housing market with a relatively low black population. County fixed effects take into account these and other geographical differences. Past studies have found accounting for these effects lowers the rate spread differential attributable to race, and thus I would expect their inclusion to reduce the coefficient on the variable representing predominantly black neighborhoods (Bayer et al., 2016; Delis & Papadopoulos, 2019). The country fixed effects model with HMDA variables will be considered the baseline model of the study.

Lender fixed effects partially take into account the shopping behavior of borrowers, the potential for high-cost lenders, and access to traditional lenders (Bayer et al., 2018). Certain neighborhoods may lack access to traditional lenders and thus sort into higher-cost lenders.

Further, previous literature has shown that black and Hispanic borrowers systematically use lenders that are more likely to issue high-cost loans (Bayer et al., 2018; Bhutta, 2015). This effect represents a difference in shopping behavior, not explicit discrimination by lenders. Based on past results, I expect the inclusion of fixed effects on the most prominent lender to reduce the rate spread attributable to race.

In the last specification, the appended tract-level census variables are added to the baseline model. From the ACS, the proportions of the residents in a tract who are unemployed, do not have health insurance, receive food stamps or Supplemental Nutrition Assistance Program (SNAP) benefits, rent their housing, and have completed a bachelor's degree are included. The proportion of vacant housing units is included as well. These variables account for a number of potential omitted tract-level variables in the HMDA dataset.

Employment represents the strength of the employment market in a tract and serves as another proxy of financial stability. Health insurance correlates with overall health outcomes (Hadley, 2003), and uninsured health costs may also play a role in the budget constraints of the borrowers from each tract. Food stamps represent a proxy for government subsidies in a tract, which will also affect the budget constraints of borrowers. This variable will also be highly correlated with tracts with higher proportions of government-subsidized housing. The attainment of a bachelor's degree is correlated with financial literacy (Lusardi & Mitchell, 2014), which may contribute to the ability of a borrower to obtain the best interest rate on their mortgage. The rent variable controls for the distribution of owners and renters across tracts, which may impact rate spreads. Tracts with higher proportions of vacant housing likely have weaker housing markets and represent less desirable collateral from the perspective of lenders. Additionally, unemployment, vacancy rates, and education attainment all correlate with crime rates in an area

collectively helping to control for this metric as well (Cui & Walsh, 2015; Lochner, 2020;

Raphael & Winter-Ebmer, 2011).

Results

County Fixed Effects

It should be noted that in the absence of counterfactuals, no estimation of racial pricing

differentials within this paper can be interpreted as purely discriminatory as there is always the

possibility of omitted variable bias and endogeneity within the model. With this caveat aside, this

study will henceforth refer to the estimated racial differentials as the approximate estimation of

lender pricing discrimination.

Table 2 shows the results of the rate spread regression models using the aggregated HMDA

data. The first three specifications can be characterized by the equation:

$$y_i = \sum_{all\ z} \alpha_z X_{i,z} + \epsilon_i \qquad (1)$$

where *y* is the average rate spread in tract *i* and *X* represents the tract-level variables in the model

described above. As previously discussed, rate spread is the difference in the interest rate

charged to the borrower and the relevant average prime offer rate. The first specification only

includes the binary variable Majority Black, which indicates that over half the applicants in a

tract are black. This is included to showcase the rate gap before adjusting for borrower and loan

characteristics. The second specification includes borrower and loan characteristics, for example,

applicant race and loan term. The difference in the estimated coefficient for this specification and

specification one can be attributed to weaker borrower characteristics and taste for riskier loan

types, like manufactured housing, for black borrowers. The third specification further includes

tract characteristics like the prominence of the GSEs, the market share of the largest lender, and

the credit proxy variable measuring the proportion of applicants denied on the basis of credit.

Lastly, the fourth specification includes county-level fixed effects. This specification will henceforth be the baseline model and can be characterized by the equation:

$$y_{i,j} = \sum_{all\ z} \alpha_z X_{i,z} + w_j + \epsilon_{i,j} \tag{2}$$

where X and y are the same as equation (1) and with w_j representing the county fixed effects. These fixed effects capture the idiosyncratic differences across geographical regions, and their effect on rate spreads.

As seen in Table 2, the rate gap between predominantly and non-predominantly black neighborhoods is 52.5 basis points. This corresponds to average rate spreads of 132 and 80 basis points in these neighborhoods. Of note, the average interest rate for predominantly and non-predominantly black neighborhoods is 5.22% and 4.91%, respectively. This constitutes a large differential in mortgage payments and interest expense for black households, constraining budgets and eroding long term wealth accumulation. Assuming the average loan size and interest rate from Table 1, an increase in rate spreads by 52.5 basis points corresponds to approximately $950 of higher interest payments a year for residents of predominately black neighborhoods. For the 25th and 75th percentile of loans, this corresponds to approximately $500 and $1,200 in higher annual payments. Of this gap, approximately 65-70%[33] is attributable to weaker borrower characteristics for black neighborhoods and tract-level characteristics, like lender market share. Approximately 7-10% is attributable to geographical characteristics at the county level. This leaves 11.8 basis points of rate differential that this study predominately attributes to discrimination against black neighborhoods. With an average rate spread of 79.6 basis points in

---

[33] These results hold for any order of adding these factors to the model.

Table 2
Regression models of rate spread on aggregated HMDA variables
Dependent variable: Rate Spread

Variables	Race	+ Borrower	+ Tract	+ County FE
	(1)	(2)	(3)	(4)
Majority Black	0.525***	0.212***	0.175***	0.118***
	(0.01)	(0.01)	(0.01)	(0.01)
Income (1,000s)		-0.001***	-0.001***	-0.001***
		(0.00)	(0.00)	(0.00)
Loan amount (10,000s)		-0.008***	-0.009***	-0.009***
		(0.00)	(0.00)	(0.00)
Female		0.227***	0.170***	0.139***
		(0.01)	(0.01)	(0.01)
DTI		0.007***	0.008***	0.004***
		(0.00)	(0.00)	(0.00)
LTV		0.012***	0.012***	0.009***
		(0.00)	(0.00)	(0.00)
Origination charges (1,000s)		0.068***	0.072***	0.068***
		(0.00)	(0.00)	(0.00)
Discount points (1,000s)		-0.035***	-0.037***	-0.027***
		(0.00)	(0.00)	(0.00)
Conventional loan		-0.211***	-0.127***	-0.350***
		(0.01)	(0.01)	(0.01)
Loan term (months)		-0.003***	-0.002***	-0.002***
		(0.00)	(0.00)	(0.00)
Manufactured		1.909***	1.696***	1.617***
		(0.01)	(0.01)	(0.01)
Asian		0.092***	0.061***	-0.017
		(0.01)	(0.01)	(0.01)
Native American		0.206***	0.204***	0.227***
		(0.05)	(0.05)	(0.05)
Hispanic		0.387***	0.357***	0.343***
		(0.01)	(0.01)	(0.01)
Pacific Islander		-0.563***	-0.531***	-0.162**
		(0.08)	(0.08)	(0.08)
Fannie Mae purchased			-0.389***	-0.390***
			(0.02)	(0.02)
Freddie Mac purchased			-0.303***	-0.481***
			(0.02)	(0.02)
Ginnie Mae purchased			-0.187***	-0.278***
			(0.02)	(0.02)
Denied for credit			0.547***	0.115***
			(0.04)	(0.03)
Tract to MSA median income			-0.004***	-0.002***
			(0.00)	(0.00)
Lender market share			-0.058***	0.245***
			(0.01)	(0.01)
Constant	0.796***	0.455***	0.424***	
	(0.00)	(0.04)	(0.04)	
Observations	71,232	68,139	68,139	68,139
R^2	0.03	0.70	0.71	0.61
Adjusted R^2	0.03	0.70	0.71	0.59

The table on the previous page presents OLS estimates for 4 models of rate spread. Table 1 includes descriptive statistics for all variables seen above. Column (1) solely includes the binary variable of whether over 50% of the applicants are black. Column (2) adds borrower and loan characteristics. Column (3) additionally includes tract characteristics. Column (4) includes the previous variables and adds fixed effects on county with standard errors clustered at the county level. Standard errors are shown in parentheses.
Note: $^*p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

non-predominately black neighborhoods, 11.8 basis points of pricing discrimination corresponds

to 14.8% higher rate spreads for predominately black neighborhoods. Notably, this rate premium

is paid by all borrowers in these tracts.

Further, the coefficients of all controlling factors match the expected direction discussed in

the Empirical Specification section. In line with previous research, the coefficient on female is

positive, indicating communities with larger female applicant pools face higher-priced mortgages

(Cheng et al. 2011). The coefficients on Hispanic and Native American variables are positive as

well, confirming results from previous studies indicating these populations face pricing

discrimination (Bayer et al., 2018; Cheng et al., 2015; Delis & Papadopoulos, 2019). Lastly, the

coefficient on credit denials, the novel variable to measure credit strength across tracts, is

positive and significant. This suggests that the credit denial variable is a valid, though imperfect,

means of capturing credit strength in the estimation of pricing discrimination. Together, these

results suggest the model is well specified.

The estimate of pricing discrimination derived in this study for predominantly black

communities is also in line with estimates of pricing discrimination found in previous literature,

which have estimated pricing discrimination to range between 5 and 25 basis points for black

borrowers (Cheng et al., 2015; Ghent et al., 2014). However, with the more comprehensive

nature of the data used in this study, these findings can be extended to the nation as a whole in

contrast to the localized interpretations required of matched datasets. Additionally, with the

exclusion of rate spread from past HMDA datasets, much of the past literature has focused on the

incidence of high-cost loans in contrast to the more granular price differences found in this study (Bayer et al., 2018; DeLoughy, 2012). The availability of rate spread in place of the binary incidence of a high-cost loan allows for a deeper understanding of the discrimination faced by all black applicants, not just the weaker borrowers who would be targets for high-cost loans. These results suggest it is not just the weakest of black borrowers that face discrimination in the market through the higher incidence of high-cost loans, but that entire communities face higher rate spreads because of the proportion of black residents in that area.

Lender Fixed Effects

In order to account for the significant impact of shopping behavior and choice of lender by applicants shown by Bhutta and Ringo (2014) and Bayer et al. (2018), this section analyzes the effect of adding lender fixed effects to the models used in Table 2. These fixed effects account for the fact that black applicants are more likely to take out loans with higher cost lenders. This study also adds an additional interaction term to capture the potential for a lender's impact on rate spreads to increase as their market share increases.

Specification five, which uses the same variables as specification three and adds fixed effects on lender, can be represented by the equation:

$$y_{i,s} = \sum_{all\ z} \alpha_z X_{i,z} + l_s + \epsilon_{i,s} \tag{3}$$

where y and X are the same as previous specifications and l represents lender fixed effects.

Specification six, which includes fixed effects on county and lender, can be represented by the equation:

$$y_{i,j,s} = \sum_{all\ z} \alpha_z X_{i,z} + l_s + w_j + \epsilon_{i,j,s} \tag{4}$$

where $y, X, w,$ and l are the same as previous specifications. Lastly, specification seven includes the interaction of l and m, which represents the market share of the most prominent lender with the fixed effect of that lender. This allows the effects of lenders to be scaled by their penetration in a tract. This specification can be represented by the equation:

$$y_{i,j,s} = \sum_{all\ z} \alpha_z X_{i,z} + l_s + w_j + m_i l_s + \epsilon_{i,j,s} \tag{5}$$

In contrast to previous literature analyzing loan-level data, including lender fixed effects has a small impact on rate spread differentials in the model. Including the interacted market share and lender fixed effects variables only accounts for an additional .7 basis points in rate spreads. In practice, .7 basis points has little economic impact indicating the inclusion of these effects is inconsequential to the model. Additionally, the coefficients of the county fixed model are nearly identical to the model including both fixed effects. In contrast, the coefficients in the model with lender fixed effects differ significantly. These results would suggest that when including geographic effects for the aggregated tract-level model, lender fixed effects are not omitted variables in the baseline model.

Considering the identical rate spread differentials in specification four and six, it is likely that geographic fixed effects largely encapsulate the lender effects seen in previous studies. This may come as a result of aggregating the loans at the tract level. Whereas lenders vary widely at the national level, and thus for the individual applicants, at the local level, only a subset of all national lenders will operate. These results suggested that while individual black applicants are more likely to use lenders that have a higher incidence of high-cost loans, black neighborhoods do not show the same aggregate shopping behavior once the geographic distribution of lenders is accounted for. Also, in contrast to this paper, Bhutta and Ringo (2014) used a 1% matched dataset of credit statistics, and Bayer et al. (2018) used a matched dataset for seven large MSAs

Table 3
Regression models of rate spread on aggregated HMDA variables with fixed effects
Dependent variable: Rate Spread

Variables	County FE	Lender FE	Both FE	+ Interaction
	(4)	(5)	(6)	(7)
Majority Black	0.118***	0.157***	0.118***	0.111***
	(0.01)	(0.01)	(0.01)	(0.01)
Income (1,000s)	-0.001***	-0.001***	-0.001***	-0.001***
	(0.00)	(0.00)	(0.00)	(0.00)
Loan amount (10,000s)	-0.009***	-0.010***	-0.009***	-0.001***
	(0.00)	(0.00)	(0.00)	(0.00)
Female	0.139***	0.144***	0.127***	0.112***
	(0.01)	(0.01)	(0.01)	(0.01)
DTI	0.004***	0.006***	0.004***	0.003***
	(0.00)	(0.00)	(0.00)	(0.00)
LTV	0.009***	0.011***	0.009***	0.007***
	(0.00)	(0.00)	(0.00)	(0.00)
Origination charges (1,000s)	0.068***	0.065***	0.062***	0.060***
	(0.00)	(0.00)	(0.00)	(0.00)
Discount points (1,000s)	-0.027***	-0.026***	-0.025***	-0.024***
	(0.00)	(0.00)	(0.00)	(0.00)
Conventional loan	-0.350***	-0.210***	-0.345***	-0.393***
	(0.01)	(0.01)	(0.01)	(0.01)
Loan term (months)	-0.002***	-0.002***	-0.001***	-0.001***
	(0.00)	(0.00)	(0.00)	(0.00)
Manufactured	1.617***	1.436***	1.403***	1.145***
	(0.01)	(0.02)	(0.02)	(0.02)
Asian	-0.017	-0.001	-0.038***	-0.024**
	(0.01)	(0.01)	(0.01)	(0.01)
Native American	0.227***	0.274***	0.246***	0.148***
	(0.05)	(0.04)	(0.05)	(0.05)
Hispanic	0.343***	0.402***	0.317***	0.299***
	(0.01)	(0.01)	(0.01)	(0.01)
Pacific Islander	-0.162**	-0.420***	-0.238***	-0.231***
	(0.08)	(0.08)	(0.08)	(0.08)
Fannie Mae purchased	-0.390***	-0.402***	-0.363***	-0.352***
	(0.02)	(0.02)	(0.02)	(0.02)
Freddie Mac purchased	-0.481***	-0.421***	-0.486***	-0.484***
	(0.02)	(0.02)	(0.02)	(0.02)
Ginnie Mae purchased	-0.278***	-0.260***	-0.276***	-0.293***
	(0.02)	(0.02)	(0.02)	(0.02)
Denied for credit	0.115***	0.403***	0.084**	-0.010
	(0.03)	(0.04)	(0.03)	(0.03)
Tract to MSA median income	-0.017***	-0.011***	-0.015***	-0.003***
	(0.00)	(0.00)	(0.00)	(0.00)
Lender market share	0.245***	0.200***	0.268***	0.18
	(0.01)	(0.01)	(0.01)	(0.36)
Observations	68,139	68,139	68,139	68,139
R^2	0.61	0.53	0.64	0.68
Adjusted R^2	0.59	0.52	0.62	0.64

The table on the previous page presents OLS estimates for 4 models of rate spread. Table 1 includes descriptive statistics for all variables seen above. Column (4) is the same as specification 4 in Table 2. Column (5) includes all HMDA variables included in specification 3 and adds fixed effects on the most prominent lender in each tract with standard errors clustered at the lender level. Column (6) includes the variables in Column (4) and Column (5) with fixed effects on both county and lender. Column (7) includes all variables in Column (6) but adds the interaction between market share and the fixed effects on lender. Bootstrapped standard errors are presented in Column (6) and (7). Standard errors are in parenthesis.
Note: *p<0.1; **p<0.05; ***p<0.01

for their analysis. Matching the HMDA data with smaller subsets limits the overall scope of the analysis and constrains the use of geographic fixed effects. The use of a national dataset allows for county-level effects to be accounted for, which may encapsulate the lender effects found by Bhutta and Ringo (2014) and Bayer et al. (2018).

Additionally, previous studies used the binary incidence of a high-cost loan in place of continuous rate spreads. It may be that certain lenders are more willing to give high-cost loans, but on average, do not charge significantly higher rate spreads than other lenders. It is also possible that lenders are now more hesitant to offer high-cost loans. As mentioned previously, a number of lawsuits have been filed in recent years, and this may have had the effect of reducing the incidence of high-cost loans, even in cases where they were not inherently discriminatory.

Census Data

To account for potential omitted variable bias, specification 8 includes tract-level census variables from the ACS dataset. Table 4 shows the results of the multivariate model including both HMDA variables and the appended ACS variables. Specification 8 can be represented by the equation:

$$y_{i,j} = \sum_{all\ z} \alpha_z X_{i,z} + \sum_{all\ p} \beta_p C_{i,p} + w_j + \epsilon_{i,j} \tag{6}$$

where $y, X,$ and w are the same as previous specifications and C represents the appended ACS variables for each tract.

As seen in Table 4, the model estimates predominately black neighborhoods pay 10 basis points higher spreads when including the ACS data, 1.8 basis points lower than when only utilizing HMDA data. Overall, this difference is small but significant, suggesting there is a small amount of omitted variable bias when solely utilizing the HMDA data to estimate pricing discrimination. This result likely stems from the fact black neighborhoods in the data, in comparison to non-predominantly black neighborhoods, have lower rates of bachelor's degree attainment (17.3% vs. 31.8%), higher rates of uninsured residents (31.8% vs. 13.1%), significantly higher rates of government subsidies (29.7% vs. 12.8%), higher unemployment rates (5.7% vs. 3.0%), and higher vacancy rates (16.9% vs. 11.5%). Considering the signs of their coefficients in Table 4 and matching the economic intuition discussed in the Empirical Specification, these factors are all associated with higher rate spreads and thus represent omitted variables from the HMDA model.

Overall, the omitted variable bias is small, with loan pricing discrimination of 10 basis points indicating a 12.6% higher rate spreads for black neighborhoods, in comparison to 14.8% when solely using HMDA data, but showcases the significant geographic effects even at the granularity of the tract. As previously discussed, county-level effects account for approximately 10% of the higher rate spreads faced by black neighborhoods, and tract-level characteristics approximately account for a further 1.8 basis points in rate differentials. These results suggest that neighborhood and geographic characteristics are considered by lenders in the aggregate, adding validity to the notion that entire communities face different lending standards based on the areas' underlying geographic characteristics. As these characteristics may correlate with the

Table 4
Regression models of rate spread with appended Census variables
Dependent variable: Rate Spread

Variables	County FE	+ Census Variables
	(4)	(8)
Majority Black	0.118***	0.100***
	(0.01)	(0.01)
Income (1,000s)	-0.001***	-0.001***
	(0.00)	(0.00)
Loan amount (10,000s)	-0.009***	-0.009***
	(0.00)	(0.00)
Female	0.139***	0.168***
	(0.01)	(0.01)
DTI	0.004***	0.004***
	(0.00)	(0.00)
LTV	0.009***	0.009***
	(0.00)	(0.00)
Origination charges (1,000s)	0.068***	0.068***
	(0.00)	(0.00)
Discount points (1,000s)	-0.027***	-0.026***
	(0.00)	(0.00)
Conventional loan	-0.350***	-0.310***
	(0.01)	(0.01)
Loan term (months)	-0.002***	-0.002***
	(0.00)	(0.00)
Manufactured	1.617***	1.577***
	(0.01)	(0.01)
Asian	-0.017	-0.006
	(0.01)	(0.01)
Native American	0.227***	0.182***
	(0.05)	(0.05)
Hispanic	0.343***	0.297***
	(0.01)	(0.01)
Pacific Islander	-0.162**	-0.144*
	(0.08)	(0.09)
Fannie Mae purchased	-0.390***	-0.345***
	(0.02)	(0.02)
Freddie Mac purchased	-0.481***	-0.435***
	(0.02)	(0.02)
Ginnie Mae purchased	-0.278***	-0.210***
	(0.02)	(0.02)
Denied for credit	0.115***	0.071**
	(0.03)	(0.03)
Tract to MSA median income	-0.017***	0.028***
	(0.00)	(0.01)
Lender market share	0.245***	0.270***
	(0.01)	(0.01)
Bachelor's Degree		-0.145***
		(0.01)
Vacancy Rates		0.02
		(0.01)
Unemployment		0.196***
		(0.07)
Food Stamps		0.002***
		(0.00)
Rentals		-0.097***
		(0.01)
Uninsured		0.169***
		(0.02)
Observations	68,139	67,332
R^2	0.61	0.63
Adjusted R^2	0.59	0.61

The table on the previous page presents OLS estimates for 2 models of rate spread. Column (4) is the same as specification 4 in Table 2. Specification 8 includes all variables in Column (4) and adds the appended census variables with standard errors clustered at the county level. Standard errors are in parenthesis.
Note: $^*p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

ability of a borrower to make timely payments, this differentiation is not inherently

discriminatory. However, when race becomes a factor in pricing, the line between differentiating

between neighborhoods based on economic factors and prejudice against racial characteristics

has been crossed.

Limitations

Though the expanded HMDA dataset corrects for a number of the omitted variables of past

releases of the dataset, a number of variables are still omitted and thus absent in this study.

Principally these variables include foreclosure rates, prepayments, and a direct credit variable. It

has long been speculated that black applicants face higher rate spreads and denial rates because

of a higher tendency to default than white borrowers. The Financial Crisis appears to support this

argument since minority borrowers did face higher rates of foreclosure, but this view is

simplistic and does not consider predatory tactics that put minorities in high-cost loans which are

more likely to default (Bayer et al., 2016; Chan et al., 2015; Mayer & Sherlund, 2008).

Accounting for these practices and the weaker baseline financial characteristics of these

borrowers, Kau et al. (2011) does not find any significant difference in default rates for black

applicants. Further, they do find a significant impact on prepayment rates, but these effects lower

the probability of prepayment, which should increase the desirability of loans originated to black

applicants.

Unlike most other fixed-income investments, such as US treasury bonds, investors in

mortgages take on prepayment risk. This is the risk that borrowers prepay more rapidly when

rates fall and prepay more slowly when rates rise. Both of these behaviors are disadvantageous to

mortgage investors. The findings of Kau et al. (2011) indicate that black borrowers are less reactive to changes in interest rate, which should make these loans more attractive to lenders. In short, if race is going to be taken into account on the basis of default and prepayment risk, black borrowers should receive a rate spread discount for their expected behavior, not a rate premium. In the context of this study, this would suggest that omitting foreclosure and prepayment variables underestimates the pricing discrimination faced by black communities.

In regard to credit variables, this study does include a measure of credit approximated by the credit denial rates within a tract. However, while this does serve to add credit effects to the model, this measure of credit is likely noisy and potentially does not fully capture the effect of credit on the model. This would predominantly be an issue in tracts with significant disparity between rejected and accepted applicant pools. At the level as granular as the tract, which is typically a small geographical area with approximately 4,000 residents[34], this disparity is likely small but, in some cases, may still bias the estimated strength of credit in an area.

Further, the inclusion of county and lender fixed effects also helps account for inefficiencies in this study's measure of credit. County fixed effects encapsulate the strength of credit for residents in that region, and lender fixed effects encapsulate the variation in credit present in the pools of tracts that predominantly use that lender. This is especially true for lenders that cater to weaker borrowers. In aggregate, the use of credit denials and fixed effects likely do not fully capture the variation in credit between black and non-black neighborhoods, but together these variables likely significantly reduce the potential credit bias in the estimate for loan pricing discrimination.

---

[34] See https://www2.census.gov/geo/pdfs/education/CensusTracts.pdf

Robustness

In order to compare the results of this study with those focused on 30-year loans, the baseline

model was additionally run utilizing only applicants applying for 30-year loans and non-30-year

loans. As can be seen in Table 5, the estimated pricing discrimination for predominately black

neighborhoods is approximately one basis point lower for 30-year terms than the sample as a

whole. This suggests that focusing on 30-year loans does underestimate loan pricing

discrimination, but the overall bias is small since over 90% of loans are 30-year loans. Notably,

pricing discrimination for non-30-year loans is 4.2 basis points higher for non-30-year loans than

traditional 30-year loans. A number of factors could be contributing to this result. While

manufactured housing and loan term variables are accounted for, other features of non-30-year

loans that are not included in the model and could be driving the difference in coefficients. For

example, adjustable-rate mortgages have loan terms under 30 years and are not separately

Table 5[35]
Regression models of rate spread by loan term
Dependent variable: Rate Spread

Variables	All Terms	30 Year	Non-30 Year
	(4)	(9)	(10)
Majority Black	0.118***	0.109***	0.151***
	(0.01)	(0.01)	(0.03)
Observations	68,139	67,792	30,143
R^2	0.61	0.58	.51
Adjusted R^2	0.59	0.55	.47

This table presents a subset of OLS estimates for three models of rate spread. Column (4) is the baseline
regression discussed in previous sections and utilizes all loan terms. For this specification, the number of
observations counts the number of tracts with at least one loan of any term. Column (9) uses the same
variable specification but is only run on data from applicants seeking 30-year loans. For this specification,
the number of observations counts the number of tracts with at least one 30-year loan. Column (10) uses
the same variable specification but is only run on data from applicants seeking non-30-year loans. For this
specification, the number of observations counts the number of tracts with at least one non-30-year loan.
Standard errors are clustered at the county level for all models. Standard errors are in parenthesis.
Note: *p<0.1; **p<0.05; ***p<0.01

---

[35] The full regression table can be found in the Appendix

identified in the model. Given the idiosyncratic differences of these and other less prevalent loan types, a portion of the estimated pricing discrimination in non-30-year loans may instead be reflecting a taste for higher-cost mortgage types by prominently black neighborhoods.

However, pricing discrimination may still be the root cause of this disparity. Since most research focusses on discrimination in 30-year-loans and public datasets have historically excluded many non-traditional loan features, lenders may have shifted discriminatory practices to non-30-year loans to mask their behavior. This could result from direct pricing discrimination in these loans or by steering applicants in black neighborhoods into higher-cost loans. Whether the difference in estimates for non-30-year loans and 30-year-loans is reflective of taste-based mortgage selection from black applicants or discrimination on the part of the lender, the net effect is still that black neighborhoods face higher-cost mortgages. With weaker economic characteristics already hindering mortgage attainment, higher rate spreads in non-traditional loans for black communities illustrate yet another hurdle to homeownership for black populations in the United States.

Moving from the population of loans utilized in the model, robustness for the chosen measure of race is analyzed here. Since only five percent of the tracts in the data are predominantly black, making the Majority Black variable heavily skewed, the model was rerun with a binary variable indicating that at least 25 percent of the applicants in the tract were black and a continuous variable measuring the proportion of black applicants in a tract. Ten percent of the tracts in the sample had a quarter or more black applicants. Ten percent is also the approximate distribution of the black population in the country as a whole[36].

---

[36] See https://www.census.gov/quickfacts/fact/table/US/PST045218

As can be seen in Table 6, tracts with at least a quarter of black applicants faced an average of 8 basis points of pricing discrimination, 4 basis points lower than the original model. In the presence of loan pricing discrimination, these results follow economic intuition. Neighborhoods with large but not predominately black populations still face pricing discrimination but to a lesser degree than those with larger black populations. Additionally, the estimated coefficient for the model utilizing the proportion of black applicants is positive and significant, suggesting that loan pricing discrimination increases as the proportion of black applicants in a tract grows larger,

Table 6[37]
Regression models of rate spread with different measures of race
Dependent variable: Rate Spread

Variables	Majority Black	Quarter Black	Proportion Black
	(4)	(11)	(12)
Majority Black	0.118***		
	(0.01)		
Quarter Black		0.080***	
		(0.00)	
Proportion of applicants by race:			
Black			0.224***
			(0.01)
Asian	-0.017	-0.020*	-0.010
	(0.01)	(0.01)	(0.01)
Native American	0.227***	0.232***	0.251***
	(0.05)	(0.05)	(0.05)
Hispanic	0.343***	0.337***	0.358***
	(0.01)	(0.01)	(0.01)
Observations	68,139	68,139	68,139
R^2	0.61	0.61	0.61
Adjusted R^2	0.59	0.59	0.59

This table presents a subset of OLS estimates for three models of rate spread. Column (4) is the baseline regression discussed in previous sections and utilizes Majority Black as its measure of race. This binary variable is one when at least 50 percent of the applicants in a tract are black. Column (11) utilizes the same controlling factors but includes Quarter Black as its measure of race. This binary variable is one when at least 25 percent of the applicants in a tract are black. Column (12) utilizes the same controlling factors but includes the proportion of black applicants as its measure of race. Standard errors are clustered at the county level for all models. Standard errors are in parenthesis.
Note: *p<0.1; **p<0.05; ***p<0.01

---

[37] The full regression table can be found in the Appendix

ceteris paribus. The coefficient on this specification can be interpreted as the expected increase

in rates spreads if a tract went from 0 to 100% black applicants. Identically, a 10 percentage-

point increase in black applicants is associated with a 2.24 basis point increase in rate spreads.

Together, these results illustrate the model is robust despite the use of a skewed independent

variable as the measure of race.

Table 7[38]
Regression models of rate spread with white applicant independent variables
Dependent variable: Rate Spread

Variables	Majority White	Quarter White	Proportion White
	(13)	(14)	(15)
Majority White	-0.051***		
	(0.00)		
Quarter White		-0.096***	
		(0.01)	
Proportion White			-0.159***
			(0.01)
Observations	68,139	68,139	68,139
R^2	0.61	0.61	0.61
Adjusted R^2	0.59	0.59	0.59

This table presents a subset of OLS estimates for two models of rate spread. Column (13) is the baseline regression discussed in previous sections but utilizes Majority White in place of Majority Black. This binary variable is one when at least 50 percent of the applicants in a tract are white. Column (14) utilizes the same controlling factors but includes Quarter White as its measure of race. This binary variable is one when at least 25 percent of the applicants in a tract are white. Column (15) utilizes the same controlling factors but includes the proportion of white applicants as its measure of race. Standard errors are clustered at the county level for all models. Standard errors are in parenthesis.
Note: *p<0.1; **p<0.05; ***p<0.01

To test the validity of the model under different conditions, the specifications in Table 6 were

repeated using white applicant independent variables in place of black. As seen in Table 7, the

coefficients on these variables are negative, indicating that white tracts receive a rate discount in

comparison to tracts with more significant minority populations. This matches economic

intuition that white borrowers receive preferential treatment compared to other races.

Additionally, as seen in Table 6, the coefficients on Hispanic and Native American variables are

positive matching findings of discrimination against these groups in past studies (Bayer et al.,

---

[38] The full regression table can be found in the appendix.

2018; Cheng et al., 2015; Delis & Papadopoulos, 2019). Together these results illustrate the model is robust across races and matches the findings of previous literature.

Conclusion

In summary, this study finds that predominately black communities face approximately 10-12 basis points of pricing discrimination in mortgage loans. This estimation of pricing discrimination comes after accounting for a number of factors, including geographic and lender effects, borrower quality, tract-level characteristics, and loan type. This estimate of increased rates spreads corresponds to 12.6-14.8% higher rate spreads for black communities. This illustrates a non-trivial economic burden placed on black households that seek to buy a home and very likely contributes to the lower homeownership rates witnessed for this population. These results confirm past findings of pricing discrimination against black communities and illustrate yet another financial barrier for black households in this country. Considering previous evidence that black applicants are more likely to be denied a mortgage loan, regardless of price, and are less likely to receive a response from mortgage brokers (Hanson et al., 2016), it would appear that discrimination takes on a number of forms in mortgage markets, adding to the hurdles of homeownership.

In addition to discrimination by lenders, this study suggests that geographic characteristics also inhibit black borrowers from getting and paying for home loans. Including county fixed effects and tract-level variables accounts for a significant proportion of the rate premium paid by black communities, suggesting that separate from any discrimination, black communities are often present in weaker housing markets. A notable example of this is the fact that black communities take out loans for manufactured properties at nearly twice the rate of non-predominantly black communities. These kinds of properties are less likely to appreciate in value

over time and are often in areas with languishing housing markets. These effects showcase just one of the many structural hurdles faced by black households in purchasing a home. Any attempt to add equality to mortgage markets must address the wide scope of structural barriers faced by black households. Focusing in on one particular issue can obscure the larger structural barriers at hand that drive these results.

In regard to policy, legislators and regulators must be cognizant of the indirect effects of any attempt to legislate change. The push for homeownership by the Bush and Clinton administrations certainty put more black households in their own home but also came at the cost of increased pricing discrimination and novel predatory tactics by lenders. For example, the results of this study do suggest that expanding the penetration of the GSEs would have a positive impact on loan prices, but the effects of such policy on overall borrower quality must be considered. Regulation and policy intended to lower loan prices cannot be considered in a vacuum. Focusing policy on targeting the lenders that showcase the highest rates of loan and pricing discrimination would likely result in both reducing the prevalence of discrimination and limit the distortions to the competitive market. Additionally, an emphasis must be placed on the communities that black applicants reside in. Communities with weak housing markets face hurdles to financing homes that extend beyond the strength of individual borrowers. Structural change will not come focusing solely at the loan-level. Geographic effects are large and must be considered.

Even with findings of pricing discrimination in this study, these results are not entirely negative. Though even one basis point of pricing discrimination is too much, the estimate of 10-12 basis points of discrimination in combination with the most recent literature showcases that lender discrimination is likely falling (Delis & Papadopoulos, 2019). Especially considering the

scope and audacity of predatory practices in the run-up to the Financial Crisis, these results

suggest that the most egregious examples of mortgage discrimination are largely in the past.

With significantly lower homeownership rates, higher denial rates, and higher rate spreads, there

is certainly still work to be done to both reduce discrimination in mortgage markets and to reduce

the disparities of wealth and financial strength in this country. Large scale change is often best

accomplished in small steps, and at least since the passing of the Financial Crisis, reductions in

mortgage discrimination appear to be moving in the right direction.

References

Avery, R. B., Brevoort, K. P., Canner, G. B., Cooper, C. R., Gibbs, C. N., Tsang, R., & Wallace, S. (2008). *The 2007 HMDA data*. Washington: Board of Governors of the Federal Reserve System.

Avery, R. B., Bhutta, N., Brevoort, K. P., Canner, G. B., Henning, N. W., & Stolarsky, S. E. (2011). *The mortgage market in 2010: Highlights from the data reported under the home mortgage disclosure act*. Washington: Board of Governors of the Federal Reserve System.

Bayer, P., Ferreira, F., & Ross, S. L. (2016). The vulnerability of minority homeowners in the housing boom and bust. *American Economic Journal: Economic Policy, 8*(1), 1-27. doi:10.1257/pol.20140074

Bayer, P., Ferreira, F., & Ross, S. L. (2018). What drives racial and ethnic differences in high-cost mortgages? the role of high-risk lenders. *The Review of Financial Studies, 31*(1), 175-205. doi:10.1093/rfs/hhx035

Bhutta, N. (2015). The ins and outs of mortgage debt during the housing boom and bust. *Journal of Monetary Economics, 76*, 284-298. doi:10.1016/j.jmoneco.2015.02.005

Bhutta, N., & D. Ringo. (2014). The 2013 home mortgage disclosure act data. *Federal Reserve Bulletin, 100*, 1–37.

Bocian, D. G., Ernst, K. S., & Li, W. (2008). Race, ethnicity and subprime home loan pricing. *Journal of Economics and Business, 60*(1), 110-124. doi:10.1016/j.jeconbus.2007.10.001

Cheng, P., Lin, Z., & Liu, Y. (2011). Do women pay more for mortgages? *The Journal of Real Estate Finance and Economics, 43*(4), 423-440. doi:10.1007/s11146-009-9214-y

Cheng, P., Lin, Z., & Liu, Y. (2015). Racial discrepancy in mortgage interest rates. *The Journal of Real Estate Finance and Economics, 51*(1), 101-120. doi:10.1007/s11146-014-9473-0

Courchane, M. J. (2007). The pricing of home mortgage loans to minority borrowers: How much of the APR differential can we explain? *The Journal of Real Estate Research, 29*(4), 399-440.

Cui, L., & Walsh, R. (2015). Foreclosure, vacancy and crime. *Journal of Urban Economics, 87*, 72-84. doi:10.1016/j.jue.2015.01.001

Delis, M. D., & Papadopoulos, P. (2019). Mortgage lending discrimination across the U.S.: New methodology and new evidence. *Journal of Financial Services Research, 56*(3), 341-368. doi:10.1007/s10693-018-0290-0

DeLoughy, S. T. (2012). Risk versus demographics in subprime mortgage lending: Evidence from three Connecticut cities. *The Journal of Real Estate Finance and Economics, 45*(3), 569-587. doi:10.1007/s11146-010-9281-0

Edmiston, K. D. (2009). Characteristics of high-foreclosure neighborhoods in the tenth district. *Economic Review (Kansas City), 94*(2), 51.

Faber, J. W. (2013). Racial Dynamics of Subprime Mortgage Lending at the Peak. *Housing Policy Debate*, *23*(2), 328-349. https://doi.org/10.1080/10511482.2013.771788

Ghent, A. C., Hernández-Murillo, R., & Owyang, M. T. (2014). Differences in subprime loan pricing across races and neighborhoods. *Regional Science and Urban Economics, 48*, 199-215. doi:10.1016/j.regsciurbeco.2014.07.006

Goodman, L. S., & Mayer, C. (2018). Homeownership and the american dream. *Journal of Economic Perspectives, 32*(1), 31-58. doi:10.1257/jep.32.1.31

Hadley, J. (2003). Sicker and Poorer—The Consequences of Being Uninsured: A Review of the Research on the Relationship between Health Insurance, Medical Care Use, Health, Work, and Income. *Medical Care Research and Review, 60*(2), 3S-75S. https://doi.org/10.1177/1077558703254101

Hanson, A., Hawley, Z., Martin, H., & Liu, B. (2016). Discrimination in mortgage lending: Evidence from a correspondence experiment. *Journal of Urban Economics, 92*, 48-65. doi:10.1016/j.jue.2015.12.004

Haughwout, A., Mayer, C., Tracy, J., Jaffee, D. M., & Piskorski, T. (2009). Subprime mortgage pricing: The impact of race, ethnicity, and gender on the cost of borrowing. *Brookings-Wharton Papers on Urban Affairs, 2009*, 33-63.

Horne, D. K. (1997). Mortgage lending, race, and model specification. *Journal of Financial Services Research, 11*(1), 43-68. doi:10.1023/A:1007927123582

Kau, J. B., Keenan, D. C., & Munneke, H. J. (2012). Racial discrimination and mortgage lending. *The Journal of Real Estate Finance and Economics, 45*(2), 289-304. doi:10.1007/s11146-011-9330-3

Ladd, H. F. (1998). Evidence on discrimination in mortgage lending. *The Journal of Economic Perspectives, 12*(2), 41-62. doi:10.1257/jep.12.2.41

Lochner, L. (2020). Education and crime. *The Economics of Education* (pp. 109-117). Academic Press.

Lusardi, A., & Mitchell, O. S. (2014). The economic importance of financial literacy: Theory and evidence. *Journal of Economic Literature, 52*(1), 5-44. doi:http://dx.doi.org/10.1257/jel.52.1.5

Mayer, C., Pence, K., & Sherlund, S. M. (2009). The rise in mortgage defaults. *The Journal of Economic Perspectives, 23*(1), 27-50. doi:10.1257/jep.23.1.27

Munnell, A. H., Geoffrey M. B. Tootell, Browne, L. E., & McEneaney, J. (1996). Mortgage lending in boston: Interpreting HMDA data. *The American Economic Review, 86*(1), 25-53.

Nothaft, F. E., & Perry, V. G. (2002). Do mortgage rates vary by neighborhood? implications for loan pricing and redlining. *Journal of Housing Economics, 11*(3), 244-265. doi:10.1016/S1051-1377(02)00103-1

Pager, D., & Shepherd, H. (2008). The sociology of discrimination: Racial discrimination in employment, housing, credit, and consumer markets. *Annual Review of Sociology, 34*(1), 181-209. doi:10.1146/annurev.soc.33.040406.131740

Raphael, S., & Winter-Ebmer, R. (2001). Identifying the effect of unemployment on crime. *The Journal of Law & Economics, 44*(1), 259-283. doi:10.1086/320275

Reid, C. K., Bocian, D., Li, W., & Quercia, R. G. (2017). Revisiting the subprime crisis: The dual mortgage market and mortgage defaults by race and ethnicity. *Journal of Urban Affairs, 39*(4), 469-487. doi:10.1080/07352166.2016.1255529

Rohe, W. M., Van Zandt, S., & McCarthy, G. (2002). Home ownership and access to opportunity. *Housing Studies, 17*(1), 51-61. doi:10.1080/02673030120105884

Rugh, J. S., & Massey, D. S. (2010). Racial segregation and the american foreclosure crisis. *American Sociological Review, 75*(5), 629-651. doi:10.1177/0003122410380868

Shlay, A. B. (2006). Low-income homeownership: American dream or delusion? *Urban Studies, 43*(3), 511-531. doi:10.1080/00420980500452433

Williams, R., Nesiba, R., & Mcconnell, E. D. (2005). The changing face of inequality in home mortgage lending. *Social Problems, 52*(2), 181-208. doi:10.1525/sp.2005.52.2.181

Appendix

HMDA Variables:

Borrower characteristics include ethnicity, race, sex, income, and age.

Loan characteristics include: whether the loan is conforming[39], whether the loan is a preapproval, loan type (conventional, Federal Housing Administration (FHA) insured, Veterans Affairs guaranteed (VA), USDA Rural Housing Service or Farm Service Agency guaranteed), loan purpose (home purchase, home improvement, refinancing, cash-out refinancing, other), lien status (first or second lien), whether the loan is a reverse mortgage, whether the application is for an open-end line of credit, whether the loan is for a business or commercial purpose, loan amount, LTV, interest rate, rate spread, HOEPA status[40] of the loan, total loan costs, total points and fees, origination charges, discount points paid, lender credits, prepayment penalty term, intro rate period, property value securing the loan, construction method of the property (site built or manufactured), occupancy type (principal residence, second residence, or investment property), total units of the property, and DTI.

Application decision variables include action taken (loan originated, application approved but not accepted, application denied, application withdrawn, application closed for incompleteness, etc.), purchaser type (Fannie Mae, Ginnie Mae, Freddie Mae, commercial bank, etc.), and denial reason (DTI, employment, credit, collateral, insufficient cash, unverifiable info, etc.).

Lender and geographical information includes Legal Entity Identifier of lender, MSA, state, county, and census tract.

---

[39] Conforming refers to loans that meet the GSE conforming loan limit
[40] Whether the loan is a high-cost loan as designated by the FHA

Appended census data for the relevant tract includes population, percent of the population that is minority, median family income, percentage of tract median family income compared to MSA median family income, number of dwellings that are lived in by the owner, and median age of homes.

Table 5

Regression models of rate spread on aggregated HMDA variables with fixed effects

Dependent variable: Rate Spread

Variables	All Terms	30 Year	Non-30 Year
	(4)	(9)	(10)
Majority Black	0.118***	0.109***	0.151***
	(0.01)	(0.01)	(0.03)
Income	-0.001***	-0.0001	-0.0002*
	(0.00)	(0.00)	(0.00)
Loan amount (10,000s)	-0.009***	-0.010***	-0.013***
	(0.00)	(0.00)	(0.00)
Female	0.139***	0.124***	-0.016
	(0.01)	(0.01)	(0.02)
DTI	0.004***	0.005***	0.003***
	(0.00)	(0.00)	(0.00)
LTV	0.009***	0.007***	0.004***
	(0.00)	(0.00)	(0.00)
Origination charges (1,000s)	0.068***	0.066***	0.049***
	(0.00)	(0.00)	(0.00)
Discount points (1,000s)	-0.027***	-0.028***	-0.020***
	(0.00)	(0.00)	(0.00)
Conventional loan	-0.350***	-0.470***	-0.093**
	(0.01)	(0.01)	(0.04)
Loan term (months)	-0.002***		0.001***
	(0.00)		(0.00)
Manufactured	1.617***	0.377***	2.460***
	(0.01)	(0.02)	(0.02)
Asian	-0.017	0.020**	-0.120***
	(0.01)	(0.01)	(0.02)
Native American	0.227***	-0.084*	-0.003
	(0.05)	(0.04)	(0.11)
Hispanic	0.343***	0.312***	0.287***
	(0.01)	(0.01)	(0.03)
Pacific Islander	-0.162**	-0.092	-0.240
	(0.08)	(0.06)	(0.16)
Fannie Mae purchased	-0.390***	-0.210***	-0.216***
	(0.02)	(0.01)	(0.02)
Freddie Mac purchased	-0.481***	-0.333***	-0.243***
	(0.02)	(0.01)	(0.02)
Ginnie Mae purchased	-0.278***	-0.174***	-0.077
	(0.02)	(0.01)	(0.06)
Denied for credit	0.115***	0.198***	-0.550***
	(0.03)	(0.03)	(0.05)
Tract to MSA median income	-0.017***	-0.004***	-0.066***
	(0.00)	(0.00)	(0.01)
Lender market share	0.245***	-0.023**	0.388***
	(0.01)	(0.01)	(0.02)
Observations	68,139	67,792	30,143
R^2	0.61	0.58	.51
Adjusted R^2	0.59	0.55	.47

This table presents a subset of OLS estimates for two models of rate spread. Column (4) is the baseline regression discussed in previous sections and utilizes all loan terms. For this specification, the number of observations counts the number of tracts with at least one loan of any term. Column (9) uses the same variable specification but is only run on data from applicants seeking 30-year loans. For this specification, the number of observations counts the number of tracts with at least one 30-year loan. Column (10) uses the same variable specification but is only run on data from applicants seeking non-30-year loans Standard errors are clustered at the county level for both models. For this specification, the number of observations counts the number of tracts with at least one non-30-year loan. Standard errors are in parenthesis. Note: *p<0.1; **p<0.05; ***p<0.01

Table 6
Regression models of rate spread on aggregated HMDA variables with fixed effects
Dependent variable: Rate Spread

Variables	Majority Black	Quarter Black	Proportion Black
	(4)	(11)	(12)
Majority Black	0.118***		
	(0.01)		
Quarter Black		0.080***	
		(0.00)	
Proportion Black			0.224***
			(0.01)
Income	-0.001***	-0.001***	-0.001***
	(0.00)	(0.00)	(0.00)
Loan amount (10,000s)	-0.009***	-0.009***	-0.009***
	(0.00)	(0.00)	(0.00)
Female	0.139***	0.141***	0.117***
	(0.01)	(0.01)	(0.01)
DTI	0.004***	0.004***	0.004***
	(0.00)	(0.00)	(0.00)
LTV	0.009***	0.009***	0.008***
	(0.00)	(0.00)	(0.00)
Origination charges (1,000s)	0.068***	0.069***	0.068***
	(0.00)	(0.00)	(0.00)
Discount points (1,000s)	-0.027***	-0.028***	-0.028***
	(0.00)	(0.00)	(0.00)
Conventional loan	-0.350***	-0.351***	-0.331***
	(0.01)	(0.01)	(0.01)
Loan term (months)	-0.002***	-0.002***	-0.002***
	(0.00)	(0.00)	(0.00)
Manufactured	1.617***	1.617***	1.630***
	(0.01)	(0.01)	(0.01)
Asian	-0.017	-0.020*	-0.010
	(0.01)	(0.01)	(0.01)
Native American	0.227***	0.232***	0.251***
	(0.05)	(0.05)	(0.05)
Hispanic	0.343***	0.337***	0.358***
	(0.01)	(0.01)	(0.01)
Pacific Islander	-0.162**	-0.183**	-0.176**
	(0.08)	(0.08)	(0.08)
Fannie Mae purchased	-0.390***	-0.387***	-0.376***
	(0.02)	(0.02)	(0.02)
Freddie Mac purchased	-0.481***	-0.475***	-0.462***
	(0.02)	(0.02)	(0.02)
Ginnie Mae purchased	-0.278***	-0.276***	-0.265***
	(0.02)	(0.02)	(0.02)
Denied for credit	0.115***	0.111***	0.080**
	(0.03)	(0.03)	(0.03)
Tract to MSA median income	-0.017***	-0.016***	-0.015***
	(0.00)	(0.00)	(0.00)
Lender market share	0.245***	0.250***	0.237***
	(0.01)	(0.01)	(0.01)
Observations	68,139	68,139	68,139
R^2	0.61	0.61	0.61
Adjusted R^2	0.59	0.59	0.59

This table presents a subset of OLS estimates for two models of rate spread. Column (4) is the baseline regression discussed in previous sections and utilizes "Majority Black" as its measure race. This binary variable is one when at least 50 percent of the applicants in a tract are black. Column (11) utilizes the same controlling factors but includes "Quarter Black" as its measure of race. This binary variable is one when at least 25 percent of the applicants in a tract are black. Column (12) utilizes the same controlling factors but includes the proportion of black applicants as its measure of race. Standard errors are clustered at the county level for all models. Standard errors are in parenthesis. Note: *p<0.1; **p<0.05; ***p<0.01

Table 7

Regression models of rate spread on aggregated HMDA variables with fixed effects

Dependent variable: Rate Spread

Variables	Majority White	Quarter White	Proportion White
	(13)	(14)	(15)
Majority White	-0.051***		
	(0.00)		
Quarter White		-0.096***	
		(0.01)	
Proportion White			-0.159***
			(0.01)
Income	-0.001***	-0.001***	-0.001***
	(0.00)	(0.00)	(0.00)
Loan amount (10,000s)	-0.009***	-0.009***	-0.009***
	(0.00)	(0.00)	(0.00)
Female	0.158***	0.155***	0.144***
	(0.01)	(0.01)	(0.01)
DTI	0.005***	0.005***	0.004***
	(0.00)	(0.00)	(0.00)
LTV	0.009***	0.009***	0.009***
	(0.00)	(0.00)	(0.00)
Origination charges (1,000s)	0.070***	0.068***	0.070***
	(0.00)	(0.00)	(0.00)
Discount points (1,000s)	-0.027***	-0.027***	-0.028***
	(0.00)	(0.00)	(0.00)
Conventional loan	-0.365***	-0.356***	-0.349***
	(0.01)	(0.01)	(0.01)
Loan term (months)	-0.002***	-0.002***	-0.002***
	(0.00)	(0.00)	(0.00)
Manufactured	1.611***	1.611***	1.624***
	(0.01)	(0.01)	(0.01)
Asian	-0.098***	-0.100***	-0.161***
	(0.01)	(0.01)	(0.01)
Native American	0.175***	0.163***	0.090*
	(0.05)	(0.05)	(0.05)
Hispanic	0.328***	0.338***	0.349***
	(0.01)	(0.01)	(0.01)
Pacific Islander	-0.245***	-0.226***	-0.319***
	(0.08)	(0.08)	(0.08)
Fannie Mae purchased	-0.393***	-0.395***	-0.380***
	(0.02)	(0.02)	(0.02)
Freddie Mac purchased	-0.485***	-0.488***	-0.470***
	(0.02)	(0.02)	(0.02)
Ginnie Mae purchased	-0.287***	-0.285***	-0.280***
	(0.02)	(0.02)	(0.02)
Denied for credit	0.137***	0.138***	0.104***
	(0.03)	(0.03)	(0.03)
Tract to MSA median income	-0.016***	-0.017***	-0.014***
	(0.00)	(0.00)	(0.00)
Lender market share	0.255***	0.249***	0.240***
	(0.01)	(0.01)	(0.01)
Observations	68,139	68,139	68,139
R^2	0.61	0.61	0.61
Adjusted R^2	0.59	0.59	0.59

This table presents a subset of OLS estimates for two models of rate spread. Column (13) is the baseline regression discussed in previous sections but utilizes Majority White in place of Majority Black. This binary variable is one when at least 50 percent of the applicants in a tract are white. Column (14) utilizes the same controlling factors but includes "Quarter White" as its measure of race. This binary variable is one when at least 25 percent of the applicants in a tract are white. Column (12) utilizes the same controlling factors but includes the proportion of white applicants as its measure of race. Standard errors are clustered at the county level for all models. Standard errors are in parenthesis.

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$