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

Michael Nicholson

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

Honors Thesis submitted in partial fulfillment of the requirements for Graduation with Distinction in Economics in Trinity College of Duke University.

Duke University
Durham, North Carolina
2020

Acknowledgements

Abstract

Introduction

Few ideas embody the spirit of the United States as fully 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 (Shlay 2006; Goodman et al. 2018). This begs the question, is there inequality in the American Dream?

The realized ambitions of this 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 to create an "ownership society" in an October 2004 speech 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, homeownership has become engrained in American society and has become a key component of wellbeing in the United States (Pager and 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

---

[1] See https://fred.stlouisfed.org/series/RHORUSQ156N
[2] See https://www.census.gov/housing/hvs/data/charts/fig08.pdf

charged interest rates. These factors capture the ability to obtain a mortgage for a home purchase and the cost in order to pay for 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 the 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 (Munnel et al. 1996; Ladd 1998). 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 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 must 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

---

[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
[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 except.

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 (Horne 1997; Avery et al. 2008). 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 discourses 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 for 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.

Using the significantly expanded HMDA data released in 2018, this paper will analyze whether predominately black neighbors 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, DTI, LTV, lender market penetration, racial and ethnic proportions, 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 within which each tract resides and for the most prominent lender within each tract.

---

[7] Rate spread will be utilized to measure loan pricing differences

Rate spread is utilized to measure loan pricing as the independent variable. A binary variable indicating whether over 50% of the applicants in a tract were black will be included in the regression. 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 of 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[8] of any maturity to better capture the full breadth of loan types 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 differentiated from previous literature by its use of the expanded HMDA dataset, its focus on rate spreads at the tract-level, its use of credit denials as a proxy for credit, and its inclusion of non-30-year loans. 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 (Rugh & Massey 2010; DeLoughy 2012; Ghent et al. 2014; Bayer et al. 2018). 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

---

[8] The population of loans included in the regression analysis will be described more fully in the Data section

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[9], it allows for a much more complete analysis of loans in the country as a whole.

Building on previous literature, this paper posits that price discrimination has reduced significantly since the pre-crisis boom in housing, but caeteris 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. Finally, the Empirical Specification section will describe the model used as well as the expected outputs for the most relevant independent variables. This section will conclude with the empirical

---

[9] These will be addressed in the Data section

estimations of the model therein described. The paper will then conclude with a summary of the results.

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 (Courchane 2007; Avery et al. 2008; Bocian et al. 2008; Haughwout et al. 2009; Rugh & Massey 2010; Delought 2012; Ghent et al. 2014; Wheeler et al. 2015; Bayer et al. 2018; Delis & Papadopoulos 2019). However, early implementations of 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 instigation 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 debt-to-income[10] (DTI), loan-to-value[11] (LTV), and credit scores. Despite these deficiencies,

---

[10] DTI is a ratio comparing a borrower's income to total debt payments
[11] LTV is a ratio computing a borrower's leverage, the amount of the mortgage over the value of the property

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

borrow 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[12] (DeLoughy 2012; Ghent et al. 2014; Bayer et al. 2018). 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 (Munell et al.

1996; Day et al. 1998), 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[13] for black and Hispanic populations (Williams et al.

2005; Faber 2013; Ghent et al. 2014). Along with the dramatic rise in housing seen in the

early 2000s[14], credit standards were largely relaxed, allowing for many previously barred

borrowers to access the housing market[15]. 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 and Massey 2010).

---

[12] A more complete review of this literature is conducted in the Literature Review section

[13] The rate spread refers to the difference in the interest rate charged to the borrow and a benchmark interest rate for the same or similar maturity, US treasuries are a common such benchmark

[14] See https://fred.stlouisfed.org/series/USSTHPI

[15] 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

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[16] 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 going 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[17] were added (Avery et al. 2008). While previous literature had explored the potential of loan pricing discrimination using survey data and other small datasets, 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 (Mayer & Sherlund 2008; Chan et al.

---

[16] High cost loans are defined as those with a rate spread of 3% or greater
[17] These additions included lien status, designation of manufactured housing, HOEPA status, etc.

2015). 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[18] (Bayer et al. 2018). However,

most analysis is conducted on data up to or preceding 2013. 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 scored[19], age, LTV,

DTI, origination charges, points and fees, loan term, interest rate, and rate spread for all

loans[20]. 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.

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.

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

---

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

subsequent studies have shown that mortgage discrimination did not shift from loan origination to loan prices until the years preceding the financial crisis (Williams et al. 2005; Faber 2013; Ghent et al. 2014). This paper's use of 2018 data allows for conclusions that are more relevant to current market dynamics.

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 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[21]. A list of these variables can be found in the Appendix.

---

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

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 & 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[22]. 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

---

[22] A more comprehensive analysis of the HMDA data can be found in Avery et al. 2008

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 accurately investigate the presence of discrimination. 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 for borrowers 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[23] is used as the tract-level variable; for example, income is calculated as the average income of borrowers in the tract. This 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

---

[23] To account for typos in some applications, the 1% trimmed mean is used

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	11.59	4.80	6.69
Income	90.57	43.70	90.35	43.60
Loan amount (1,000s)	239.48	147.97	246.47	147.94
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	101.88	42.92	102.03	42.85
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.

applications in that tract. To allow lender fixed effects in the model, the lei 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[24] 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 (Nothaft & Perry

2002; Cheng et al. 2015). 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).

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, statics 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[25] and Federal Reserve[26]. 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[27].

---

[24] See: https://transition.fcc.gov/form477/Geo/more_about_census_tracts.pdf
[25] See: http://www.freddiemac.com/pmms/
[26] See: https://fred.stlouisfed.org/graph/?g=NUh
[27] See: https://www.census.gov/quickfacts/fact/table/US/PST045218

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.

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[28] and to account for the difference in average maturities 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 were black. The coefficient on this variable estimates the rate spread differential attributable to the majority of the applicants being black, caeteris 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 the discriminant in loan pricing and the core result of this analysis.

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

---

[28] 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

charged. Many of these, such as credit scores, 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 characters 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[29], and the percent of applicants that are Asian, Native American, Hispanic, or Pacific Islander[30] for the given tract. Additionally, the proportion of applicants denied for credit or employment reasons is included as well. 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 over discrimination on the basis of gender[31] (Cheng et al. 2011). The breakdown of race and ethnicity account for the potential for the presence of other minorities to be driving the higher spreads in place of the black population. Many minority

---

[29] Percent male is excluded to avoid multicollinearity
[30] Each race and ethnicity is included as a separate variable
[31] 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

groups face loan price discrimination, with active literature on black, Hispanic, and Native

populations (Cheng et al. 2015; Bayer et al. 2018; Delis & Papadopoulos 2019).

The novel inclusion of credit and employment denials from a tract intends to serve as a proxy

for unavailable credit and employment variables. Though these two variables are for denied

applicants and rate spread is solely included for approved borrowers, the inclusion of these

variables gives a measure of the overall strength of credit and employment 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 is largely correlated, large inequalities across borrowers are

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 DTI 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 essentially 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[32] 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 characterizes, 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 significant 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[33], 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

---

[32] Meet 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

[33] 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 and 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 prices 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 and 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.

Lastly, the analysis separately looks at two sets of potential fixed effects, county fixed effects, and lender fixed effects[34]. County fixed effects account for the variation of rate spreads attributable to the geographic region the tract resides in. 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).

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.

---

[34] Note that these are two separate models

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

Results

Table 2 shows the results of the rate spread regression model. The first three specifications can be characterized by the equation:

$$y_i = \alpha X_i + \epsilon_i \qquad (1)$$

where *y* is the average rate spread in tract *i,* and *X* represents the tract-level variables in the model described above. 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 statistical discrimination on the basis of 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.

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

Variables	Race	Borrower	Tract	FE
	(1)	(2)	(3)	(4)
Majority Black	0.525***	0.212***	0.175***	0.118***
	(0.01)	(0.01)	(0.01)	(0.01)
Income		-0.001***	-0.001***	-0.001***
		(0.00)	(0.00)	(0.00)
Loan amount (1,000s)		-0.001***	-0.001***	-0.001***
		(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			-0.0004***	-0.0002***
			(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
R²	0.03	0.70	0.71	0.61
Adjusted R²	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. Standard errors are shown in parentheses.
Note: *p<0.1; **p<0.05; ***p<0.01

Lastly, the fourth specification includes county-level fixed effects. This specification can be characterized by the equation:

$$y_{i,j} = \alpha X_i + w_j + \epsilon_{i,j} \tag{2}$$

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 constitutes a large differential in loan payments for black households constraining budgets and eroding long term wealth accumulation. Of this gap, approximately 60% is attributable to statistical discrimination on the weaker borrower characteristics of black neighborhoods. A further 7% is attributable to tract-level characteristics and 11% to geographical characteristics. This leaves 11.8 basis points of rate differential that we attributable to discrimination against black neighborhoods. These results are in line with the estimates of rate premia paid by individual borrowers found in previous literature (Ghent et al. 2014; Cheng et al. 2015). 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 (DeLoughy 2012; Bayer et al. 2018).

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. (2016). What drives racial and ethnic differences in high cost mortgages? the role of high risk lenders.

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

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.

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

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

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

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

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

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

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

HDMA Variables:

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

Loan characteristics include: whether the loan is conforming[35], 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[36] 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.

---

[35] Conforming refers to loans that meet the GSE conforming loan limit
[36] 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.