Discriminatory Mortgage Lending:

A Sequential Approach

Pranav Krishnan

Department of Economics

Haverford College

Advisor: Professor Richard Ball

Abstract

This thesis approaches discriminatory lending with a holistic approach. Using newly published HMDA data, I construct a sequential model in which I examine the existence of discrimination in both application and pricing outcomes for all FHA and VA insured mortgages originated in 2019.

TABLE OF CONTENTS

1 Introduction

Access to credit is a central component of an individual's financial profile in this country, and this is inclusive of mortgage lending. As seen in Figure 1, outstanding Consumer Mortgage Debt has been growing at a relatively steady rate over the past two decades. In 2019 alone, over 8 million mortgages were issued, representative of a 26% increase from the previous year.[1] Taking a look at this a bit more cumulatively, the value of this outstanding debt currently sits at $16 Trillion. To put this in perspective, the current Gross Domestic Product (GDP) of the United States is about $20.54 Trillion. It should be no surprise that the consumer mortgage market is of this size. According to the US Census Bureau, the home ownership rate in this country currently stands at 65.6%. Close to 63% of those homeowners secured their real estate with a mortgage - this number is only rising moment.[2] As the market continues to grow back to levels comparable to the Great Recession, how does this affect a given borrower's access to this credit source?

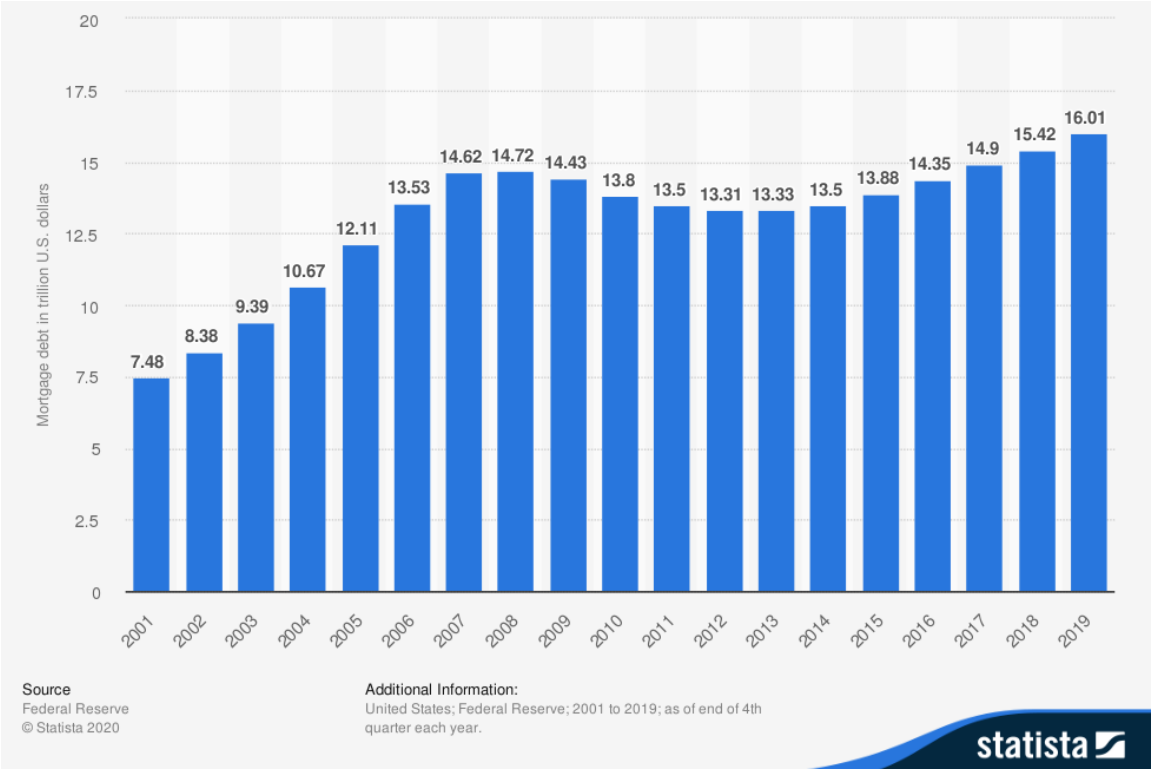

FIGURE 1: ***OUTSTANDING MORTGAGE DEBT FROM 2001-2019***

[1] See CFPB Mortgage Market Activity and Trends, 2019
[2] See 2017 American Community Survey

Even in the aftermath of the subprime lending crisis, discrimination and disparate impact has continued to detriment the livelihood of those marginalized; consumer credit markets are no exception to perpetuating this phenomenon. Historically, we see that the effectiveness of legal mandates has only gone so far to help fix this issue. Early legislature such as the Civil Rights Act of 1968 (also known as the Fair Housing Act) and the Home Mortgage Disclosure Act (1975) are the first attempts in protecting marginalized groups from being denied access to home ownership and credit products as a byproduct of their skin color or geographical location. Seminal legislature such as those mentioned came at a time where racial prejudice was rampant in our country. More recently, the repercussions of the Great Recession in 2008 gave way to the establishment of Dodd Frank and the Consumer Financial Protection Bureau (CFPB) to increase the transparency and regulation of credit markets. More specifically, mandates such as Regulation C and changes to HMDA now force the largest financial and credit institutions to maintain large, publicly disclosed data sets filled with loan-level mortgage information as a way of aiding the effectiveness of policy making.

Despite these implementations, modern examples of disparate impact and discriminatory practices continue to manifest in the markets. In April of 2017, Bellco Credit Union was sued under the accusation that applicants were discriminated based on their sex and familial status. Larger institutions have also failed to uphold fair lending practices, as disparate treatment and predatory lending lawsuits were rather famously tried against Wells Fargo and J.P Morgan in 2017, with the latter being found guilty of violating the Fair Housing Act between the years of 2006 and 2009. We may be out of the 1950s, but financial practices continue to marginalize and segregate the general population. Questions such as the following naturally arise -- is this a systematic market-wide problem? Is there still presence of discriminatory or disparate practices?

Empirical studies that have attempted to answer this question in the past are ambiguous in their entirety. Early studies in the 1990s focus mostly on the mortgage lending decision for a given applicant; that is, it has little to nothing to do with the actual contractual terms of a mortgage loan. Among the first studies to use public HMDA data disclosures, Munnell (1996) found that minority applicants were denied for mortgages more than white applicants in the Metropolitan Statistical Area of Boston. However, HMDA data has historically suffered from omitted variable bias, as disclosure of important applicant-level components of the lending

decision have not been publicly available. HMDA data sets augmented with private external sources have been the fix for many papers in this space and is the case with Munnell (1996). Contrary to Munnell (1996), Tootell (1996), using the same data, found that racial makeup of neighborhood has no significant effects on the lending decision even after controlling for the race of an applicant when analyzing the presence of Redlining.

Conventional lending markets are not the only markets that have been in the scope of early lending decision studies. Among the first papers to include FHA lending into the scope, Gabriel and Rosenthal (1991) find that minorities are more likely to receive an FHA-insured mortgage, suggesting that factors beyond the credit risk of the borrower affect application outcome. The inclusion of federally insured markets raises another important understanding. If the credit risk inherent in a mortgage is completely transferred away from the lender, any additional dispersion in the denial rates of a mortgage applicant by race can be effectively interpreted as evidence of discriminatory practices.

More recent literature is largely comprised of studies surrounding the determination of the contractual elements of a mortgage. Given pricing terms are market determined, there is still scope for inexplicable interest rate and price dispersion at the discretion of the lender after the application outcome. Kau et al. (2012) find that loans in Black areas pay more compared to majority-white neighborhoods in Miami over the course of a 25-year time period. Analyzing mortgage pricing in the subprime era, studies such as Ghent et al. (2014) and Bayer et al. (2017) find consistent evidence of discriminatory price and interest rate dispersion. However, interest rate determination is not the only cost existent in a mortgage contract. Closing mortgages contracts involve election of upfront costs, discount points, and other fees and charges that borrowers are required to meet as a byproduct of the contract's structure. In an attempt to correct for this, Bhutta and Hizmo (2020) include discount points and upfront costs into the determination of discriminatory mortgage pricing. Although they find that minority borrowers pay more for FHA loans, they suggest that the choice to receive less in upfront costs likely offsets any dispersion in interest rate determination.

The ambiguity of mortgage discrimination literature is likely attributable to the face that few studies have accounted for the dual stage nature of the lending process. The decision to accept and originate a mortgage is inherently discretionary. If there is room for disparate impact or discrimination at both stages, one may hypothesize that this is a more effective way of analyzing the state of the market. Among the first of its kind, Bartlett et al. (2019) examine post-crisis algorithmic loan approval and pricing to find that minority borrowers (inclusive of Hispanic and Black borrowers), pay more and are rejected more when compared to their white counterparts. Although novel in its approach, Bartlett et al. (2019) still suffers from some shortcomings, including the limited scope of their pricing model. My thesis attempts to reconcile gaps found in the literature by exploring the existence of mortgage lending discrimination with a sequential approach. Aiming to improve upon the general two-stage approach of Bartlett et al. (2019), I choose to conduct a focused examination of the federally insured Federal Housing Authority (FHA) and Veterans Affairs (VA) lending markets to correct for the limitations of previous studies that are unable to control for the unobservable amount of credit risk that lenders choose to price in even after risk of default is transferred away through methods such as securitization.

In the first stage, I analyze all applications either denied, approved, or originated in a Logit model including the necessary applicant and loan-level controls. I find conclusive evidence of discriminatory practices in approval outcomes for all racial and ethnic minorities. In the second stage, I utilize all loans approved and originated in the first stage and analyze the existence of price dispersion in a similar lens as Bhutta and Hizmo (2020), accounting for the multitude of costs presented to an applicant when mortgage contracts are structured. This study would be the only one of its kind to analyze non-conventional markets in this way. I find that although Asian and American Indian borrowers are not discriminated against, Black and Hispanic borrowers are impacted beyond the objective strength of one's application across all pricing outcomes.

The rest of the paper is structured as follows. Section 2 presents a more in-depth overview of previous literature. Section 3 introduces the institutional and theoretical framework of the mortgage lending process. In Sections 4 and 5, I discuss the data and the methodology in examining mortgage discrimination in both stages of the model. Sections 6 and 7 conclude with a discussion of the results and caveats that arise from the presented models.

2 Background and Literature Review

2.1 Overview

Although the specific process of loan origination is likely to vary between lender institutions, mortgage lending can be broken down broadly into two stages – application approval or rejection, and subsequent pricing. The current standing of lending discrimination literature has historically been divided along these stages, with seminal papers in the 1980s and 1990s mostly focusing on discrimination in the loan approval process and more recent studies focusing on lending during the subprime crisis and discrimination in loan pricing. By focusing on both stages, this novel study analyzes the nature of racial discrimination in the mortgage lending process holistically.

2.2 Loan Application Outcomes

Within the sphere of loan application status, evidence of discrimination is almost entirely conclusive. Black et al. (1978) was one of the first papers to acknowledge the role of race in credit markets. Using a two-part survey on bank mortgage activity and applicant financial standing, the authors conclude an applicant is more likely to be rejected if they identify as being black. However, in addition to its limited sample, the authors fail to control for the possibility of low non-black rejection rates because of applicant prescreening conducted by the banks.

Evolving from initial studies that suffer from limited data, Munnell et al. (1996) uses HMDA data augmented with city specific MSA[3] data in analyzing the lending decision within Boston. HMDA data has historically suffered from omission bias, leaving out key variables such as applicant credit-history, loan-to-value ratio and "other factors considered in making mortgage decisions" (Munnell et al. 1996). The study concludes that minority applicants still get denied 8 percentage points more than white applicants when modeling the decision to lend via a logit model. Contrary to this, Tootell (1996) uses the same data and examines racial composition at the tract-level to uncover the existence of redlining. The study finds that the racial makeup of a neighborhood has an insignificant effect on the lending decision when controlling for the race of an applicant. In other words, the author concludes that redlining is a circumstantial occurrence; if more non-minority applicants were to move into those neighborhoods, lending rates would

[3] Abbreviation for Metropolitan Statistical Area

increase. Although the opposing results may be an interesting phenomenon, both Munnell et al. (1996) and Tootell (1996) suffer from limited application due to the geographic constraints of focusing their studies within the city of Boston.

The existence of racial discrimination can also be found in non-conventional mortgage lending decisions. Unlike conventional lending markets, non-conventional lending effectively eliminates the risk of default a lender may face. This is attributable to the fact that approved non-conventional loans are backed fully by the federal government. Although a large percentage of conventional loans are likely securitized by governmental conservatorship agencies such as Fannie Mae and Freddie Mac, little is said in studies about the extent to which minimal credit risk considerations would still exist in the lending decision even after the incidence of default risk is transferred away. In analyzing whether borrowers choose FHA or conventional mortgages, Gabriel and Rosenthal (1991) find that minorities are more likely to receive FHA-insured loans, suggesting that race affects the lending process beyond reasons related to credit risk of the borrower. In a similar lens, Canner et al. (1991) use a two-stage model to analyze the effect of default risk on lending decisions when controlling for neighborhood and racial influences on credit worthiness. They also find that minority households are less likely to receive conventional mortgages loans when compared to white households.

While the previously mentioned studies analyze a consumer's initial mortgage application, Hawley and Fujii (1991) use the 1983 Survey of Consumer Finances to examine a consumer's credit market experience, inclusive of credit re-applications and discouragement. By deploying a two-stage sequential model similar to that of Canner et al. (1991), they find that white households are significantly less likely to be denied from their first credit application. Furthermore, they find that that white households are also more likely to reapply for credit. With respect to marginalized racial and gender groups, they find that nonwhite and female households are more likely to be discouraged from applying for credit altogether. Though, the study examines consumer credit in a general sense, and does not distinguish between the type of credit that an applicant requests. Although Canner et al. (1991) and Hawley and Fujii (1991) utilize two-stage models, they do not consider contractual determinations after mortgage applications are approved. This study reconciles such issues in addition to using newly published HMDA data

that includes previously omitted variables to gain a better picture of the influence of racism in the mortgage approval process.

2.3 Loan Performance and Mortgage Term Setting

Although literature in loan applications largely supports the existence of racial discrimination in mortgage markets, empirical results in loan performance and pricing are not as conclusive. Berkovec et al. (1994) and Berkovec et al. (1998) are among the first studies to examine default probabilities in FHA lending outcomes. These papers surround a particular theoretical phenomenon of lender bias which suggests that lenders exhibit prejudice by increasing the minimum threshold of loan performance, effectively forcing increased minority performance.[4] The theoretical model presented in Arrow (1973) was originally framed around wage discrimination but is easily transferrable to the credit lending space. Finding the theory inapplicable, Berkovec et al. (1994) identifies that loans granted to minorities contain higher default risk. However, this study suffers from potential omitted variable bias and endogeneity issues, as the model does not entirely control for credit history of a given applicant. In a latter study, Berkovec et al. (1998) attempts to correct for this by deploying a model insensitive to omission bias by instrumenting with a measure of market concentration. More specifically, Berkovec et al. (1998) use the Herfindahl Index, which is indexes the size of a firm in relation to the industry it is in. Also examining FHA-backed loan performance, the study finds no statistically significant evidence of non-economic or prejudicial discrimination. Although default is more likely among black applicants, key interaction terms are insignificant in both the default and mortgage loss models. Opposing results in comparison to earlier literature question whether these studies are doing enough to provide a holistic appraisal of the market.

Besides overall loan performance, a significant portion of the modern discrimination literature has centered around the pricing of approved mortgage loans. Looking at 30-year FRMs originating in Miami from 1975 to 2002, Kau et al. (2012) examines contract rates and outcomes to identify "noncompetitive" discrimination in mortgage lending (Kau et al., 2012). Using a two-stage instrumental variables model, the study finds that loans in black areas are charged 0.4% more compared to white counterparts. However, the study, as others, suffers from omitted

[4] See Arrow (1973)

variable bias, leaving out the inherent possibility that lenders will price in credit risk when determining the terms of a mortgage contract. Cheng, Lin, and Liu (2015) find similar, more economically significant outcomes in their attempt to find disparate impact in pricing outcomes for mortgages originating in both subprime and pre-crisis time periods. Using a novel approach of controlling for consumer behavior in areas such as shopping[5], credit, and education, they find consistent evidence that black borrowers pay anywhere from 29-31 basis points higher for both high and low risk mortgages, with that effect greatly heightened for black women and younger borrowers that are less educated.

Given the Great Recession is primarily centered around the burst of the mortgage bubble, there is also a vast pool of literature that focuses on the presence of discrimination in this time period. Haughwout, Mayer and Tracy (2009) use augmented HMDA data and find that Black and Hispanic borrowers pay lower initial rates and have lower reset margins when looking at 2/28 ARMs.[6] Similar results are observed when the model looks at the cost of loans by location, with mortgage terms being "cheaper" overall in areas where the minority population was high. This result, although somewhat counter-intuitive, is in line with a positive credit supply shock seen in periods leading up to the crisis. Access to credit became easier given lending requirements were greatly eased, likely allowing marginalized ethnic groups to receive relatively cheap loans. Also analyzing pricing outcomes in the subprime era, Ghent et al. (2014) find different outcomes. Examining subprime purchase and refinance mortgages originated in metropolitan areas within California and Florida, the authors find that Black and Hispanic borrowers pay "12 and 29 basis points higher" (Ghent et al., 2014) than their white counterparts. At the tract-level, the difference is significantly smaller, as minority-filled neighborhoods only pay 1.4 basis points more in interest rates for 7 out of the 8 mortgage products they include in the model. This outcome, however, is seen only in the case of non-depository institution-based lending, suggesting that in addition to possible predatory lending practices, the lack of experience on the account of a first-time borrower may be contributing to the results. In a more recent study of lending during the housing crisis, Bayer, Ferreira, and Ross (2017) examine pricing outcomes for high-cost lending. Looking at both prime and subprime loans, they find that

[5] This refers to borrowers that search or "shop" for different rates offered by different lenders.
[6] ARMs, or Adjustable-Rate Mortgages have an initial set period where the interest rate is fixed. After this, the interest is variable month-over-month, usually calculated as a spread on LIBOR over that period.

minority borrowers are much more likely to receive high-cost mortgages for first time home purchases. Though, the authors note that this increase can be attributed to lender-sorting based on foreclosure risk (Bayer et al., 2017).

Situated in the post-crisis period, Bhutta and Hizmo (2020) provide a more accurate appraisal of the modern mortgage pricing stage. Specifically, the study improves upon studies which suffer from omitted variable bias and are mostly one dimensional in their analyses of interest rates. When mortgage rates are set, lenders typically extract premiums through the charging of upfront costs. However, borrowers that are cash-constrained can elect to pay less in discount points, which effectively raises the interest rate charged to lower upfront lump-sum costs associated with the loan. Analyzing FHA loans originated from 2014 and 2015, the authors find that Black and Hispanic borrowers pay interest rates that are 2-3 basis points higher than white borrowers, but also pay less in discount points, suggesting that minority households elect to receive higher interest rates in exchange for cash to cover upfront costs. The authors note that these forces likely offset one another, concluding there is "little evidence" of discrimination in lending (Bhutta and Hizmo, 2020). It is no doubt that to completely expose the state of discrimination, all components of the process must be taken into consideration. Although significantly improving upon previous studies and setting the stage for analysis of post-crisis mortgage lending, Bhutta and Hizmo (2020) do not analyze the loan approval process that precedes the pricing stage. My study attempts to improve upon the approach of Bhutta and Hizmo (2020) by expanding the study of modern post-crisis non-conventional mortgages across both stages of the lending process.

2.4 Rise of FinTech and Modern Lending Methods

Prior to the era of the subprime crisis, most mortgages were issued by traditional "brick and mortar" institutions that relied on in-person lending methods. This includes classic savings and loan associations and large commercial banks such as Wells Fargo, Bank of America, etc. This is no longer the case, as lending methodologies have been digitized with the rise of FinTech and independent mortgage institutions. This is evidenced by recent data collected by the Consumer Financial Protection Bureau, which indicates that Quicken Loans, United Shore, LoanDepot, and Caliber Home Loans (all independent or online-only mortgage institutions) were among the 7

largest organizations in terms of total originations in 2019.[7] Given the application process is less dependent on the physical appearance of the lender, it would be interesting to think about how lending access has been influenced as a byproduct of this trend.

Examining the role of FinTech lenders in mortgage lending, Fuster et al. (2019), although not directly examining lending discrimination, find that there is no conclusive evidence that lenders specifically target marginalized groups to extract greater rents. In their study of machine learning models, Fuster et al. (2017) examine the use of novel statistical models in credit market processes; the outcomes are different, to say the least. As the authors note, the use of newer statistical and predictive technologies always results in a "winning" and "losing" side. In attempting to identify the winners and losers because of this evolving process, the study ultimately finds that the dispersion of rates for minority borrowers is significantly higher than that of their white and non-Hispanic counterparts (Fuster et al., 2017). Additionally, the machine learning model is significantly more accurate in estimating the propensity of default among an individual borrower, even though traditionally marginalized ethnic groups are subject to greater disparate impact, with equilibrium default rates even higher than would be predicted if using classical linear techniques of default estimation.

Shifting to empirical studies in this space, Bartlett et al. (2019) examine algorithmic GSE loan approval and pricing in an effort to decipher the state of statistical lending discrimination in this modern FinTech era. Unlike older mortgage products, loans, such as those under analysis, are backed and securitized by GSE institutions such as Fannie Mae, Freddie Mac, or Ginnie Mae. With reference to the setting of the loan terms, the issuers rely on an algorithmic pricing grid, but have some amount of discretion as to the final rates charged; the authors attempt to analyze the magnitude of these deviations, if at all they exist. As one of the only studies of its kind, Bartlett et al. (2019) examine the extent of discrimination in both the pricing and approval processes. For first-time home purchase mortgages, they find that minority borrowers pay 7.9 basis points higher and are rejected 9.6 percentage points more when compared to white counterparts. When isolating just for FinTech lenders, they find no significant difference at the rate at which minority

[7] These 7 organizations accounted for approximately 14% of the total origination activity in 2019.

loan applications are accepted, but still find a significant increase in the interest rate charged, albeit smaller in magnitude when compared to the larger sample.

Bartlett et al. (2019), although novel in its approach, still suffers from some limitations. Among others, the authors note that only 90% of these loans are sold to GSEs, suggesting that lending institutions continue to hold a fraction of issuances on their balance sheets. Consequently, it is likely that lenders will continue to price in unobservable credit risk when making decisions within the lending process, given the contractual elements are market determined. This thesis will attempt to improve upon the general dual stage approach of Bartlett et al. (2019) by alternatively operating in the space of present-day FHA and VA-insured lending. Additionally, by including considerations of upfront costs as in Bhutta and Hizmo (2020), I aim to present a holistic study that examines the existence of discriminatory lending in post-crisis mortgage markets. Abstracting from the general approach of Hawley and Fujii (1991), I use a sequential two stage model with newly published HMDA data which analyzes loan application outcomes in the first stage, then subsequently analyzes the approved loans to expose the extent to which marginalized ethnic groups are presently discriminated against when they decide to close home-purchase and refinance mortgages.

3 Institutional Framework

The mortgage market, although complicated in its structure, can broadly be broken down into two buckets: conventional and non-conventional mortgages. Conventional mortgages encompass any non-government-insured mortgages originated by typical lender institutions such as credit unions, commercial banks, mortgage companies, etc. The mortgages issued by these firms usually have two outcomes – they are either kept on the lender's balance sheets or bought out and repackaged into derivative securities such as the mortgage-backed security (MBS). Based on the 2018 American Community Survey, about half of the conventional market is bought out by government-sponsored enterprises such as Fannie Mae, Freddie Mac, and Ginnie Mae. About 30% of the issuances in the market are retained on balance sheets and bank portfolios.

Securitization is a method of transferring away the credit risk a lender may face when underwriting a loan. This risk is simply representative of the probability that the holder of the mortgage either defaults on the payments or continually delays payments and becomes

delinquent. Institutions that buy these mortgages then package (or securitize) them into large derivative securities to diversify and pad the individual credit risk of the loan. The securities gain value by pooling together the non-concurrent payment streams as the cash flows, which then reduces the overall risk of the asset. On the end of the lender, given they have sold the loan away to an institution like Freddie or Fannie, they are no longer liable for the risk of default. If this was the outcome of every conventional mortgage in the market, then models focusing on this part of the market could control for the credit history and financial reliability of any given borrower.

Lenders who choose to retain mortgage loans on balance sheets also have ways to guarantee against large losses as a result of default or delinquency. Depending on the size of the down-payment or the financial standing of the borrower, private mortgage insurance (PMI) is usually a part of the mortgage contract at closing. Usually, the cost is lumped with the monthly principal and interest payments associated with the term of the mortgage. However, lenders who choose keep mortgages on their balance sheets still face default and credit risk. The only way that lenders would be completely hedged against default and delinquency is if PMI contracts lasted the entire maturity of a mortgage; this is usually not the case. On the end of the borrower, there are many ways to get out of PMI contracts. The most typical methods include refinances, reappraisals of the property securing the loan, and requesting termination of PMI contracts once the borrower builds 80% equity in the home (i.e., 80% of the loan balance has been paid). In essence, borrowers who are willing to do so can escape PMI obligations quite routinely. Given that lenders cannot protect themselves against default risk for the life of every individual loan, there is likely to be some amount of this consideration when contractual terms are determined for a given application. This is particularly important because the contractual terms of a mortgage are still determined at market and discretionary to the lender institution at some level. As a result, discrimination models that focus primarily on securitized or conventional loans suffer from being unable to fully control for how much lenders price in credit risk when setting the terms of the loan. Thus, such loans are not under the scope of this study.

Non-conventional loans, on the other hand, represent the segment of the mortgage market that is guaranteed by the financial backing of the US government. Specifically, this share of the market is represented by mortgages backed by the Federal Housing Agency (FHA), Veteran's

Affairs (VA), Farm Service Agency (FSA), and the Rural Housing Authority (RHA). If lender institutions and banks meet the minimum requirements, they can issue loans backed by these agencies. Of the roughly 4.5 million non-conventional loan applications in 2019 (see Table 3), FHA-backed loans make up the largest share of the market, followed by VA-backed loans. Together, these insured loans encompass about 95% of the non-conventional lending market. Due to the limited size of the FSA and RHA markets, I choose to omit them from this study. Additionally, given the applicant/borrower population for these loans is predominantly white, there is little-to-no variation to analyze with respect to the racial and ethnic makeup.

Loans such as these are considered non-conventional because of its application and contract setting process. Historically, the FHA and VA have been a source of credit access to lower-to-middle income borrowers with less-than appealing credit history. The average income for households applying for conventional loans in 2019 along is approximately $205,915. In comparison, average household income for FHA borrowers is $71,248 while household income is approximately $85,974 for VA borrowers (Table 5). Given that borrowers applying for FHA and VA loans are significantly less financially endowed than those applying for conventional loans, federally insured loans do come with some benefits. Among many, FHA and VA borrowers are required to pay significantly less in lump-sum upfront and closing costs. Additionally, loan contracts usually require much lower down-payment percentages. To put this in perspective, conventional down-payment requirements usually hover around 15-20% of the principal. Down-payment requirements for FHA and VA loans can be as low as 3%.

In addition to relaxed down-payment and upfront costs, there are requirements that borrowers must meet to qualify for such loans. FHA and VA applicants must meet minimum FICO score conditions and conform to loan limits to qualify for acceptance.[8] Usually, loan limits vary by the income level and cost of living for the specific area in which the application is being requested. Though country-wide, the maximum loan ceiling for both FHA and VA loans is $822,375. To make up for lower lump-sum payment requirements, insured loans usually have higher interest rates and much stricter mortgage insurance premiums. However, unlike conventional loans, PMI contracts are independent of the applicant's financial characteristics and are usually set based on the amount for which the loan is requesting. Additionally, PMI contracts on these loans are near

[8] Minimum FICO Score requirements are usually a credit score of 580 or greater.

impossible to trade-out of, i.e., borrowers are usually required to pay insurance for the entirety of the loan's maturity. All of these prerequisites go towards protecting the lenders from any risk – this is inclusive of credit, default, and prepayment risk. In theory, if these risks are entirely transferred away from the lender, any deviation from what "fair" term setting would be can be effectively interpreted as discrimination. The same thing can be said for any statistically significant dispersion in application approval rates for borrowers of different racial and ethnic backgrounds; lenders who issue non-conventional loans have no incentive to price in any additional credit risk at their discretion, allowing models to fully control for the financial standing of the applicant.

4 Data

4.1 HMDA Data Coverage

This study uses a novel loan-level dataset published by the Home Mortgage Disclosure Act (HMDA) on all mortgage originations for firms that qualify under Regulation C in 2019. Of all the mortgage lending institutions in the US, the Consumer Financial Protection Bureau (CFPB) reports that 2019 origination data encompasses approximately 88% of total mortgage market activity.[9] HMDA institutional coverage is wholly linked to the size and scope of the lending institution, leaning mostly towards larger banks, credit unions, and S&L Associations. Specific to the coverage criteria as recent as 2018, institutions are depository financial institutions under Regulation C only if they meet certain criteria, some of which include size of total asset holdings, being federally insured, and whether the institution originated at least 25 closed-end mortgages and 500 open lines of credit in the preceding year.[10] As aforementioned, I choose to isolate all federally insured mortgages in the sample, leaving out conventional lending markets from consideration.

I note that the use of this dataset to be novel due to the corrections the Consumer Financial Protection Bureau (CFPB) has made to recent data publications. As Bartlett et al. (2019) notes, literature in lending discrimination has historically been limited because of the incomplete nature of public mortgage data. Thus, modern examples such as the aforementioned Bartlett et al.

[9] 2019 CFPB Mortgage Market Activity and Trends Report, 10
[10] For further information, see HMDA Institutional Coverage (Effective 2018-2020).

(2019) and Bhutta & Hizmo (2020) correct for these issues by utilizing merged data sets with private mortgage data. Under the Dodd Frank Act in 2010, many general changes were made to what institutions under the compliance of the CFPB were required to disclosure under HMDA. Additionally, a HMDA rule in 2015 made further additions to data disclosures after 2018 which modified existing data points and added as many as 27 new data points that were reported by compliant institutions[11]; data points which previous literature had previously omitted or sourced from private mortgage databases.

As a result of these changes in HMDA, the dataset utilized in this study provides a conclusive set of applicant, loan, and property characteristics. Previous HMDA disclosures have already had the benefit of measuring applicant and loan-level information such as race, ethnicity, loan-type, and loan-purpose. Its weaknesses are found in the disclosure of specific terms on mortgage term structure, applicant credit, and applicant financial standing. To highlight specific additions, previously omitted measures such as combined LTV, debt-to-income ratio, and income, are now included in the disclosures. Additionally, information on interest rate, rate-spread, discount points, and total upfront costs are also found in the data – all of which will be utilized in the model I present in this paper.

4.2 Formatting the Data

As aforementioned, HMDA data almost encompasses the entire mortgage market in its standing today. In its preliminary state, 2019 data covers 22,066,474 mortgage applications and originations in the country, measured by 99 different variables. Table 3 provides a breakdown of the type of mortgages within the data set. Given the focus of this study is non-conventional lending markets, the 10 million or so conventional mortgages are dropped from the dataset entirely. Applications within the dataset also specify the purpose for which the loan being applied for is. In addition to first-time home purchase mortgages, the dataset is inclusive of home-improvement, refinance, and cash refinance mortgages. However, given the limited representation of some of these categories, I choose to focus solely on home-purchase, refinance, and cash-out refinance mortgages. Much of the variation in loan purpose is largely removed when conventional loans are taken out of consideration. Tables 1 and 2 provide a brief summary

[11] 2019 CFPB Mortgage Market Activity and Trends Report, 4

of loan purpose by racial and ethnic groups. Home-purchase mortgages are the largest category in each of the racial groups followed by cash-out refinances. To distinguish the two from each other, cash-out refinances are very similar to traditional "re-fis" in the sense that an existing mortgage is paid off by the refinance. However, in the case of a cash-out refinance, the amount of the new loan is usually higher than the remaining balance of the existing mortgage. The difference goes to the borrower in cash, at which point they can choose to use the cash for improvements on the home or other financial necessities. It is also pertinent to point out that about close to every mortgage in the dataset under consideration is first lien, i.e., there are little to no applicants who are applying for a second mortgage to secure another.

With respect to the racial breakdown of the sample, 34% of the approximately 1.5 million FHA and VA loans are submitted by minority applicants. Given they make-up a larger percentage of the borrower-population in this segment of the market, there is room for the identification of any dispersion or variation – what I interpret as disparate impact or discrimination. Lenders stand to extract higher premiums by charging higher relative rates and upfront costs to borrowers with limited options due to their weak financial standing. As mentioned before, agencies such as the FHA and VA only insure loans. Mortgages are still originated by lender institutions like how conventional loans would be, suggesting that the potential for approval and pricing dispersion still exists as it would in conventional markets. Dropping conventional and FSA/RHA loans from the dataset, I am left with approximately 3 million mortgages.

At the outset, variables not in the scope of the study are dropped. Pertinent cases to point out include county code and census tract information. I choose to control for lender variation at the state level, given that the mortgage market in its current standing, is dominated by large national institutions. Specific to the data set under analysis, more than half of the mortgages are either originated from nation-wide institutions or online-only lenders who no longer rely on regional branch-based banking to lend to clients. Controlling at the tract or county level, as seen in previous literature such as Kau et al. (2011) and Bartlett et al. (2019), has attempted to control for the possibility that traditional face-to-face lending variation is still present throughout different areas in the country. Given the trend of modern lending methodologies is that of a move

away from traditional mechanisms, I am more interested in how these variations exist in a broader sense.

Given discrimination is the focus of this study, I generally drop all loans that were missing disclosures of sex, ethnicity, and race of the applicant. Mortgage applications missing race are the most material of these considerations, with about 1.4 million loans being dropped in the process. I speak to how racial, ethnic, gender, and age categories are measured and introduced into the model in the next section. Mortgage applications also have a write in option for personal identifiers, and given this is hard to measure quantitatively, observations in which the write in option was used were also dropped. However, this was a small consequence, given these represent an extremely small portion of the data (less than 1%). As one would assume, observations that are missing data on measures of loan contract structure, property attributes, and financials in general are all dropped. The most material of these include observations missing income, debt-to-income ratio, and property value. With specific reference to income, observations that reported household income as being less than zero were also dropped from the sample. The final approval analysis data consists of 1,534,541 mortgages in total.

In constructing the data for the second stage of the model, all loans that are approved and originated are subsequently taken into the second stage data. Consequently, those loans that are denied in the first stage are dropped from the sample. Loans missing information on key variables entering the pricing process are dropped, similar to how variables in the lending outcome decision were dropped in the first stage. The most material of these are inclusive of mortgage applications missing information on the combined loan-to-value ratio, rate spread charged on the mortgage, loan term, and interest rate. Additionally, applications that reported extreme anomalies (i.e., rate spread over 20%) were also dropped from the sample. The final pricing analysis data contains 1,171,656 loans.

4.3 Aggregating Race, Ethnicity, Sex and Age

Given the study is centered around racial, ethnic, and gender backgrounds, I give special attention to how this was treated in the data. At the outset, I define race to be a grouping of people based on geographical location who may share similar physical and behavioral attributes. On the other hand, I define ethnicity to be a marker of someone's cultural background. With

respect to the specific treatment of race and ethnicity in the data, ethnicity is defined as being either Hispanic or Not Hispanic, while racial categories include White, Asian, Black, Native American, and Pacific Islander. Previous literature has mostly treated racial and ethnic backgrounds as being the same (i.e., compare the results of Hispanic households with that of Black or White households), but this is something I avoid while presenting results in Section 6.

On mortgage applicants, households can choose to list co-applicants that represent a second guarantor of the loan in the case the primary applicant is no longer able to honor the loan's recurring dues. The data reports race, ethnicity, sex, and age of the applicant and their co-applicant (if applicable) at an individual level. Given there is the possibility of there being more than one observation of an application's race, sex, ethnicity, and age the measure is disaggregated in a sense, as there is the possibility that applicants are from a different ethnic or racial background. The CFPB derives aggregated categories for all the applications' measures of race, ethnicity, sex, and age as follows: buckets are created for applications with homogeneous applicant and co-applicant racial, ethnic, and sexual identifications. Any time there is a discrepancy (a family with two different racial, ethnic, or sexual backgrounds), the applications are usually placed into a "Joint" category. Dealing with joint sexual markers is a bit more approachable – it is not out of the ordinary for a married couple to joint-apply for a mortgage loan. A similar thing can be said for Age of the members within a single application – I simply deal with this by creating three buckets (representative of Young Adults, Middle Age, and Old Age households) from the derived Age category. With respect to ethnicity and race, this is a bit harder to untangle and is generally uninterpretable. Given the ambiguity of the "Joint" categories, any application that is said to have a derived ethnic or racial category of "joint" is thus dropped from the sample. Going forward, tables and results are presented broken down by both racial and ethnic backgrounds, as opposed to being treated as the same.

4.4 Summary Statistics

Table 4 begins by providing an initial outlook into the financial standing of the average borrower/household by Race and Ethnicity.[12] I categorize all applicants into 6 broad racial groups inclusive of White, Black, Hispanic, Asian, Native American, Pacific Islander, and Dual

[12] For further reference, Figure 2 in Appendix C provides a visual representation of Table 4.

Race (applications with exactly two racial backgrounds). Of these broad groups, White households represent the largest cluster by far, followed by Black, Hispanic, Asian, and borrowers under the "Other" category. At the outset, the data highlights that Asian borrowers, have the highest household incomes in the sample. This is followed by White and Black households, with both having average household incomes approximately 11% and 15% lower than Asian households respectively. It is important to note that across every racial category, those with Hispanic ethnicities all have lower average incomes compared to their non-Hispanic counterparts. Looking at the entire sample, Hispanic household income is about 10% lower than that of their non-Hispanic counterparts. The sample shows that the minority population, specifically Black and Hispanic households are much less financially endowed than the rest, while Asian and White households are more financially stable.

Figure 1 in Appendix C provides a visual representation of income distributions by racial category.[13] After log transformation, White and Black households have the most variable distributions, suggesting that household income within the respective racial buckets is varied a lot. The distribution for all Asian households, on the other hand, has a much lower range, suggesting that more households have incomes at similar levels compared to other racial categories. With reference to the two broad loan categories, FHA loans represent the largest chunk of the data (See Table 5). Those applying for VA loans have higher household incomes on average, with the difference for the average applicant being about 21% (also see Table 5).

Tables 6 and 7 provide an initial breakdown of application rejection rates across racial, ethnic, and loan types. At preliminary glance, we see that mean rejection rates for White and Asian households are among the lowest the data – this holds for both ethnic groupings as well. On the other hand, households identifying as Native American, Pacific Islander, and Black all have significantly elevated rejection rates in comparison. We can see this clearly in the visual representation of average rejection rates by race in Figure 3 (Appendix C). Interestingly enough, the average rejection rate for all Hispanic households is actually lower than that of non-Hispanic households. The finding also holds for each of the racial categories, as average rejection rates are lower in the Hispanic cases. This could be related to the difference in financial standing of Hispanics across the United States. Historically, Hispanic populations in the Southwest of the

[13] Note, these distributions have been log transformed to achieve approximately normal distributions.

country mostly originate from Mexico. On the contrary, those emigrating to areas such as Florida and California are mostly from Latin and South America. The US Census highlights that South American immigrants have historically been more well off than Hispanic immigrants in the Southwest. As of 2016, Latin American median income in the United States was $8,000 dollars higher than African American median income and only 20% lower than the national median. Thus, their increased financial standing may be a contributing factor to a decrease in unfavorable lending outcomes. Across loan types, average rejections are largely similar for both FHA and VA loans, with the mean rejection rate for FHA loans being slightly higher (see Table 7).

Tables 8 and 9 provide a more granular breakdown of lending outcomes in the sample. It is important to note that there is a distinction made between the approval and origination of a loan. The central difference lies in where the mortgage application stands amongst the larger process. Although a mortgage being approved indicates its eventual origination, contractual terms are likely not finalized until steps such as property appraisal are completed. On the end of the borrower, origination is not confirmed until they decide on terms such as election of discount points, lender credits, and total upfront costs. With respect to this particular data set, mortgages that sit in the approved stage of the process encompass an extremely small portion of the data (approximately 4%). Across loan types, the proportions of originations, approvals, and rejections are quite similar. Table 9 paints a similar picture as shown in Table 6 – Black households have the highest ratio of rejections to approvals/origination across both Ethnic groups while Asian and White households boast the lowest rejection-to-origination ratios.

Table 12 breaks down central pricing characteristics under identical race and ethnic categorizations. More specifically, I measure average rate spreads, which are calculated as the difference between the APR and Average Prime Offer Rate. Non-Hispanic Black households pay about 0.129 percentage points higher in average rate spreads, while Asian households pay the lowest amongst the entire sample. Across all racial categories, with the exception Black households, borrowers that identify as Hispanic are charged higher average rate spreads. Taking a step back and looking at the entire sample, Black and Hispanic households are charged the highest rate spreads on average, paying approximately 0.148 and 0.276 percentage points more respectively when compared to their White counterparts. Similar differences are seen in other studies analyzing different time periods (e.g., Bhutta and Hizmo 2020). Figure 4 presents the

distributions of rate-spreads across racial buckets. Although variation and spread are largely the same for most of the racial categories, the variation in the distribution for White households is markedly lower, suggesting that more white borrowers pay lower rate spreads in comparison (i.e., there is significantly more clustering around the peak of the distribution as opposed to clustering around the tails, which would suggest there are significantly more households having to pay high-rate spreads). Across loan types, average rate spreads are largely the same (See Table 13).

Breakdown of total loan costs and origination charges display a similar pattern (Tables 14 and 15). Amongst all the racial categories, Native American and White households pay the lowest in total loan costs on average, while Asian households boast the highest mean. Although the result for White households may be surprising, this result could be a sign that White households may be issued the "cheapest" loans in the sample. Black households pay slightly more in total costs, with difference being as low as .013 percentage points between their White counterparts. Overall, Hispanic households pay significantly more than non-Hispanic households over the course of the loan's maturity on average, with the difference being about 14%. With respect to origination charges, White households are charged the lowest in upfront origination fees on average, while Asian households are the highest in the sample. However, when looking at Hispanic households across racial buckets, the sample shows that Hispanic households are charged more upfront on average compared to non-Hispanic households; this result, although does not hold for Black households and households that identify with two racial categories. The result for White households is somewhat contributory to the story that such households are issued some of the cheapest loans on average. However, we must consider the size of the loan – measures such as total loan costs and origination charges are directly related to these. Looking at central pricing terms as a whole, I find that Asian borrowers pay among the lowest in rate spreads and among the highest in total loan costs and origination charges, on average. All else equal, this would suggest that Asian borrowers are electing to pay more upfront due to their elevated financial standing[14] as I presented above. Part of the fees of issuing a mortgage are at the discretion of the borrower in the sense that borrowers can choose to pay as much or as little to influence the interest rate charged on the loan over the course of its maturity; in this case,

[14] Refer to Table 5; Asian households have among the highest average incomes in the sample.

Asian households pay more upfront to reduce the interest charges over the loan's maturity. Similar results are found in Bhutta & Hizmo (2020). In comparison, minority borrowers (such as Black, Native American, and Hispanic) who are significantly less financially endowed (See Table 5) when applying for mortgages may elect to take on higher rate spreads to decrease the financial strain of a large lump-sum payment at closing. However, as aforementioned, measures of total loan costs and origination charges are likely dependent on the amount for which loans are request. It is also a possibility that loan costs are so low for Black households because their financial standing does not allow them to take out more expensive loans (as seen in Table 10); the contrary can be said for Asian households, who can request significantly larger loans given their elevated financial endowment (also seen in Table 10). This dynamic is an important one to keep in context as I move through the model.

Tables 16 through 19 present a final summary of other loan-level characteristics central to the pricing model. Loan to Value ratios are the highest, on average, for all Hispanic borrowers indicating that loans issued to Hispanic households are likely deemed more high risk than other loans. Unsurprisingly, Asian households have the lowest LTVs on average, suggesting the opposite. Factors such as loan term are largely similar across all racial and ethnic categories (See Table 19). On average, most borrowers in the sample request 30-year fixed-rate mortgages – the market standard. For Asian households, summaries of discount points and lender credits (discretionary costs that can be elected up-front to reduce the interest rate on the mortgage) paint a similar picture as seen with rate spreads, total costs, and origination charges (Tables 17 and 18). However, it is important to note that most of the values for these variables are missing. Given the parts of the data are not entirely public yet, mortgage terms such as discount points and lender credits may not be fully disclosable yet. As a result, there remains little-to-no explanatory power and is left out from the models presented below.

5 Methodology

5.1 Sequential Nature of Mortgage Lending

As aforementioned, the goal of this study is to provide a holistic look into the presence of discrimination in the FHA and VA mortgage markets. Although the process of mortgage lending contains many steps, I aggregate these into two broad stages. Application acceptance/rejection

and subsequent pricing. Abstracted from the approach of Hawley and Fujii (1991), the model follows this sequential process by analyzing all loans in the first stage on the nature of discrimination in the approvals process. In the second stage, I drop all loans that were denied and focus only on the applications which were converted into mortgage originations as a way to introduce continuity into the model.

5.2 Analysis of Approvals

The first stage of the model focuses entirely on the existence of discrimination in the outcome of a mortgage application. HMDA records the action taken for each of the loans, for which there exists 8 possible outcomes. Of those outcomes, the focus is on loans that are originated, applications that are approved, and applications that are denied. I choose to omit applications closed for incompleteness and those withdrawn because they are missing values for key variables which are observed in the model. With respect to applications denied for incompleteness, inclusion in the model is likely to confound results, given that it would be difficult to separate the effect of discrimination from the decision to decline based on an incomplete application. Inclusion of loans withdrawn by the applicant themselves is redundant to the study, given that prejudice is unlikely to affect this decision unless lenders influence this decision through prescreening or preapproval. In this stage, I employ a logit model with a general specification as follows:

$$Rejection_i = \beta_o + \beta_1 RACE_i + \beta_2 ETHNIC_i + \beta_3 SEX_i + \beta_4 Controls_i + \mu_{State} + \epsilon_i$$

where Rejection_i is a binary operator measuring whether the applicant was either rejected or accepted by the lender. Race, Ethnicity, and Sex are all measured with a series of binary operators as well, with the largest categories omitted to account for collinearity.[15] Note that the subscript *i* denotes an individual application in the dataset. The controls are general applicant-level variables that I deem as likely to go into the lender's approval process. This includes the debt-to-income ratio, measured as the ratio of the applicant's monthly debt obligations to the monthly income used to pay off those debts; income, which is measured in thousands of dollars; age of the applicant, represented by binary operators that segment the sample into youth, middle

[15] The largest categories are as follows: Male, White, and Non-Hispanic

age, and senior-citizen populations; loan purpose, also bucketed into binary operators representing first time home mortgages or refinances; loan amount, which measures the amount requested from the loan in dollars; and loan type, controlling for any potential differences that may arise between FHA and VA lending methods.

There are a couple of things to note on how certain variables are dealt with in the model. Given this is a logistic regression model, coefficient estimates are reported as log odds ratios, where the odds ratio is defined as $\frac{p}{q}$, where p is the probability of success (which, in this case represents rejection of the application) and q is simply $1 - p$, or the probability of failure. Thus, the general form for a logistic regression model is:

$$ln(p/q) = a + \beta x$$

To convert coefficient estimates from log odds ratios into an interpretable form, I simply exponentiate both sides of the equation. Thus, we have,

$$p/q = e^{(a+\beta x)}$$

With respect to the actual regression outputs, I exponentiate the coefficient estimates to revert them back to general odds ratios. For independent variables that are binary operators such as those used to measure race, ethnicity, sex, and age, the approach is largely the same. Instead of individual odds ratios as defined above, odds ratios for dummy variables are defined as:

$$\frac{odds_{(black)}}{odds_{(white)}}$$

where, Black is the racial category hypothetically introduced into the model and White is the omitted category. Assuming basic log rules, when introduced into logistic models, this OR becomes a log difference in the odds of the outcome for Black households and the odds of the outcome for the baseline category. Thus, when exponentiated back into basic odds ratios, we can interpret the coefficient estimates on binary independent variables as being the difference in odds ratios for one category versus the baseline.

Specific control variables were also altered when introduced into the model. To normalize its distributions, income and loan amount are log transformed. Additionally, given debt-to-income

ratio is not reported in the data as a continuous variable, I group DI ratios into 6 different binary operators each with a range encompassing the possible values listed within the dataset. In total, there is 6 different buckets created for DI ratios as follows:

1. **DI – 1 is defined as,**
    a. *1, if the debt-to-income ratio of the borrower is < 20%*
    b. *0, if otherwise*
2. **DI – 2 is defined as,**
    a. *1, if the debt-to-income ratio of the borrower is between 20% and 30%*
    b. *0, if otherwise*
3. **DI – 3 is defined as,**
    a. *1, if the debt-to-income ratio of the borrower is between 30% and 36%*
    b. *0, if otherwise*
4. **DI – 4 is defined as,**
    a. *1, if the debt-to-income ratio of the borrower is between 36% and 49%*
    b. *0, if otherwise*
5. **DI – 5 is defined as,**
    a. *1, if the debt-to-income ratio of the borrower is between 50% and 60%*
    b. *0, if otherwise*
6. **DI – 6 is defined as,**
    a. *1, if the debt-to-income ratio of the borrower is > 60%*
    b. *0, if otherwise*

In more enhanced specifications, I include state fixed effects to control for state-based lender variations and analyze the effect of race and ethnicity within states. State fixed effects, however, may be somewhat of a misleading term, as HMDA data also encompasses the United States territories, inclusive of the Virgin Islands and Puerto Rico. Additional controls include lender fixed effects, created by bucketing lender institutions by their unique legal entity identifier (LEI) and its total number of originations in the sample. The largest institution in the sample had 93,800 originations, while any institutions with less than 10,000 originations in the sample were all bucketed into one binary operator. Given the sheer size of this last LEI bucket, this was the one omitted when introduced into the model.

To identify the effects of discrimination in different federally insured mortgage markets, I run two additional approval models in which I isolate FHA loans in one, and VA insured loans in the other. The structure of the models is identical to the main logit specifications described above, apart from loan type controls. I hypothesize that after isolating the two markets, VA markets will attenuate the odds of rejection for minority borrowers given that VA loans are exclusive to past service members. Qualitatively speaking, veterans are treated with an enormous

amount of respect in this country, which would lead me to believe that veterans of all backgrounds are a bit more protected from discriminatory practices. However, I do expect there to be significant dispersions in odds of rejection for minority borrowers in FHA markets, given these are more akin to conventional lending markets in comparison to the VA market.

5.3 Proxying for Credit Score

Before moving onto the approach of the pricing model, I first focus on how I control for the credit history of an applicant in the model. The inclusion of controls for credit scores has been somewhat controversial in the literature. Previous literature has suffered from having little to no data on this measure because of the historical privacy of consumer credit data. Even in the aftermath of the 2008 recession, modern studies such as Bhutta and Hizmo (2020) and Bartlett et al. (2019) use private credit data to include applicant FICO scores in their models. Even though public HMDA data does not disclose the FICO score of an applicant, it does disclose the debt-to-income ratio, which I claim to be a legitimate proxy for credit score in my study.

The calculation of credit score, whether it be an Experian, Equifax, or Transunion-based FICO score follow a similar methodology. Usually the makeup is as follows: 30% of the FICO score is dependent on the amount you owe (i.e., outstanding consumer debt), 35% of the makeup is dependent on previous credit payment history, 10% lies in new credit lines, 15% is based on the length of your current credit lines, and the final 10% of the score is dependent on the mix of credit you hold. Although credit mix and length/age of credit is hard to proxy for, most of the FICO score calculation lies in how much a consumer has extended themselves based on available credit and one's past ability to pay back those credit lines. Although a qualitative argument, it would be not be a far-reaching assumption to say that this is exactly what the "Amount Owed" and "Payment History" categories aim to determine.

I claim that inclusion of the debt-to-income ratio is highly correlated with such measures. As described by the CFPB, the debt-to-income ratio measured in HMDA data represents the ratio of current outstanding debt obligations to the amount of income that is available to the applicant to pay that debt back. Use of the debt-to-income ratio is likely a fair proxy because it measures the ability of a consumer to pay debt and the amount to which the applicant has extended themselves in terms of how much debt they wish to take on. Though, limitations to the use of this proxy still

exist. Although debt-to-income ratio is a good measure of an applicant's current ability, it does not consider the possibility of previous credit delinquency. Applicants with good DI ratios in the present may have had credit events in the past that lenders may worry about when determining contract structures. An argument on the contrary can also be made, especially for applicants that may be in less-than favorable financial standing now but were not in the past.

5.4 Analysis of Mortgage Pricing

In the second stage of the model, I analyze discriminatory practices in the subsequent pricing of these loans via an OLS specification as follows:

$$RateSpread_i = \beta_o + \beta_1 RACE_i + \beta_2 ETHNIC_i + \beta_3 SEX_i + \beta_4 Controls_i + \mu_{Lender} + \epsilon_i$$

I choose rate spread, as opposed to interest rate as the dependent variable because it is a better measure of whether loans are priced higher than others. Rate spread is measured as the difference between a loan's APR (annual percentage rate) and the Average Prime Offer Rate (APOR). The APR is the annual interest rate scaled to the total cost of the loan, inclusive of fees, principal, and mortgage insurance payments. This is especially applicable considering holders of federally insured mortgages are required to pay PMI for the duration of the loan. The APOR, on the other hand, is a general APR that is representative of mortgages offered to the average qualified borrower in the country, with average financial standing and credit history. Given the APR and APOR take into consideration the size and maturity of the loan, this is a more comparable measure of whether one mortgage is more costly than the other.[16] The rule of thumb usually accepted by the market is that first lien mortgages with a rate spread of above 1.5 are priced high, while subordinate or secondary lien mortgages are considered highly priced if rate spreads surpass 3.5 percentage points. Rate spread is also a better indicator of fair pricing because it provides insight into the probability of prepayment. If an applicant receives a mortgage with low-rate spreads over the course of the maturity, it incentivizes against prepayment, i.e., it is less likely that the borrower will consider a refinance. However, applicants stuck with higher cost mortgages (higher rate spreads) are more likely to be shopping for opportunities to refinance at a

[16] This is also dependent on the lien status of the mortgage in consideration.

lower spread and prepay their existing mortgage. Basic interest rates do not really provide much insight into this, as it is less comparable of a pricing measure.

The structure of the pricing model is similar to its first stage counterpart with the addition of a few more controls that enter the lender's considerations when determining contractual elements. These additional controls include loan amount, measured in dollars; combined Loan-to-Value ratio (LTV), which is the total debt amount to the value of the property that is securing the debt; loan term, measured in months; property value, measured in dollars and rounded to the nearest 10,000th interval; and lender fixed effects, measured by a series of binary operators as aforementioned (see Section 5.2). Loan amount, property value, and LTV are all appropriately log transformed to normalize respective distributions and attenuate large outliers. In alternative specifications, controls for origination charges are added, as well as state and loan-type fixed effects. Additionally, I run two other models in which FHA and VA lending markets are separately isolated. If discrimination is present, I hypothesize that compared to White and Non-Hispanic households, minority applicants will be charged higher spreads even after controlling for the financial history and standing of the applicant. However, I expect this result to be muted for Asian households, as their elevated financial endowment is likely to bring down their costs.

Although the rate spread is a good measure of the relative cost of one's mortgage, this reflects solely on a mortgage's monthly dues. To get a better sense of all costs associated with the mortgage over its entire maturity, I run another model in which Total Loan Costs is the dependent variable to see if minority borrowers are also subject to higher costs over the life of a loan. Total loan costs measure exactly what it is labeled as – it accounts for all monetary costs associated with the mortgage, inclusive of principal, interest, upfront, and insurance payments. The model's structure is identical to the Rate Spread model described above, and is as follows:

$$LoanCosts_i = \beta_o + \beta_1 RACE_i + \beta_2 ETHNIC_i + \beta_3 SEX_i + \beta_4 Controls_i + \mu_{Lender} + \epsilon_i$$

After extraordinary outliers are removed from the data set, loan costs are log transformed to normalize its distribution before inclusion in the regression model. It is important to note that controls for origination charges are dropped from this particular model because a primary component of mortgage costs is charges at mortgage issuance. Inclusion of origination charges would thus cause two-way causation. As with the Rate Spread model, alternative specifications

are run where state and loan-type fixed effects are added, and additional models are run where FHA and VA markets are isolated.

As I have previously mentioned, the cost of a loan, is to an extent, at the discretion of the borrower through the election of upfront costs via discount points and lender credits. However, HMDA data does not allow me to examine discount points and lender credits as Bhutta and Hizmo (2020) do given these data points are not fully available to the public. Thus, I use origination charges as the variable of focus, which is simply the total of the "itemized" amounts, in dollars, paid by the borrower at closing.[17] Although I am unable to breakdown these costs by their itemizations, analysis of this variable contributes to the same narrative. The model, which again is identical in structure to the other pricing models presented above, is as follows:

$$OriginationCharges_i = \beta_o + \beta_1 RACE_i + \beta_2 ETHNIC_i + \beta_3 SEX_i + \beta_4 Controls_i + \mu_{Lender} + \epsilon_i$$

Note, that Origination Charges, like total loan costs, is also appropriately log transformed to achieve a normal distribution before inclusion in the model.

Analysis of upfront costs act as robustness checks for the two pricing models presented above. At the outset it could indicate whether borrowers of certain racial or ethnic groups are electing to pay more or less in upfront costs to alter the rate spread assigned to the loan. As a result, it would provide additional insight into whether borrowers are paying more on both fronts due to reasons beyond financials, or whether financial standing of the borrower is affecting one's ability to meet lump-sum payments at closing and thus inflating the rate spreads indirectly.

6 Results

6.1 Discovery of Dispersion in Lending Outcomes

Table 1 in Appendix B presents the results of the logit regressions of Rejection rates on the series of applicant race, ethnicity, sex, and age binary operators. In addition, borrower-level controls as discussed in the previous section are included in each of the specifications. Baseline categories that are omitted to account for collinearity include White households, non-Hispanic households, male and middle age applicants, DI – 4, and the binary operator for all lender

[17] See 'Public HMDA Data Fields with Values and Definitions', *HMDA*, https://ffiec.cfpb.gov/documentation/2019/lar-data-fields/

institutions with less than 10,000 originations in the fiscal year under consideration.[18] Note that coefficient estimates are reported as general odds ratios and have already been exponentiated.

Looking column 1, results exist that are worth pointing out. Coefficient estimates for all ethnic and racial categories are positive and statistically significant at the 99th percentile. Compared to non-Hispanic households, Hispanic households are 1.36 times more likely from being rejected from their loan application. With to respect to race, given all racial dummies are significant and greater than one, compared to White households, all racial categories are more likely to have their loan application rejected in the first stage of the lending process. However, there is more explanatory power in the difference between these racial buckets.[19] Black households and applications listed as "Dual Race" have the highest odds ratios amongst the 5 racial categories. Even after controlling for typical borrower-level financial considerations in the eyes of the lender, the odds of rejection for Black households are approximately 8.1% higher than Asian households, 18% higher than Pacific Islanders, and 25.6% higher than American Indian households. This last result is the most surprising considering that the American Indian population has been among the most marginalized racial groups in this country, likely even comparable to the marginalization of Hispanic and African American population. Interestingly enough, Asian households do not have the lowest odds of rejection compared to White households. Compared to American Indian and Pacific Islander applicants, odds of rejection for Asian households are approximately 16.14% and 9.22% higher, respectively. I find this to be an interesting result given that Asian households boast the highest financial standing in the sample on average. This financial standing is also significantly higher than the average White household in the sample. If loan outcomes were solely based on the financial statistics and objective risk of an applicant, then, all things considered, Asian households should have among the lowest odds of rejection.

With respect to the gender listed on an application, female applicants have virtually the same likelihood of rejection compared to males, while applicants that list their gender as joint have lower odds of rejection. This is somewhat of an intuitive result given that applications that list gender as "Joint" are likely married couples applying and co-signing for loans together. In the

[18] Note that DI -4 includes applications that reported debt-to-income ratios between 36% and 49%.
[19] Note, all these differences were tested and proved statistically significant in every case.

eyes of the lender, this may be deemed as a safer application, especially in the scenario where the married couple has two different income streams. Amongst loan types, I find that both refinance and cash-out refinance mortgage applications are more than 3 times as likely to be rejected compared to conventional home-purchase mortgages.

The most surprising of the results is the coefficient estimate for income. Although one would think that an increase an income would reduce the odds of application rejection, the odds ratio for income is greater than 1; i.e., a percentage increase in household income increases the odds of rejection by about 1.19 times. Although positive and significant the odds are relatively small, as it is only .19 percentage points higher than if an increase in income had no effect on the odds of application rejection. A plausible explanation for this may be related to an observation made earlier in Section 5.3. As I mentioned, although debt-to-income ratio is a sound proxy for credit history of the applicant, there are smaller elements of the FICO score calculation that are still unobservable given the data set being used. Specifically, the age of one's credit and mix of one's credit. Given this is the case, it is a possibility that the coefficient estimates on the DI buckets and income may be picking up these effects and being indirectly altered. A natural robustness check follows as Table 1a in Appendix C runs the same model specifications without controls for the debt-to-income ratio of the applicant. Once omitted, coefficient estimates other variables remain similar while the estimate for income stays significant and drops below 1, which suggests the intuitive result that likelihood of rejection decreases when household income increases.

Column 2 of Table 1 runs the identical model with the addition of loan-type fixed effects. Although results are slightly muted with respect to the magnitudes of the coefficient estimates, the direction of the results are largely the same. This time, Hispanic households are 1.28 times more likely to receive a rejection than their non-Hispanic counterparts; the difference across the two specifications is only .065 percentage points. Dual Race and Black applicants remain the highest likely to be rejected when compared to White households with the odds of rejection increasing for Dual Race applicants across specifications by .021 percentage points. When compared within racial buckets included in the model, differences between odds ratios for Black households and the rest are significantly different; Black households are 23.7% more likely to be rejected than American Indians, 13.4% more likely to be rejected than Asians, and 17.2% more likely to be rejected than applicants that list as Pacific Islanders. With the inclusion of loan-type

fixed effects, the model reports that VA loans in general are 0.61 times less likely to be rejected when compared to FHA loans. This is a result that I hypothesized earlier, given the increased social standing of veterans in this country could aid more favorable credit outcomes. After inclusion of loan-type fixed effects, the odds of application rejection for both female and joint-gendered applications are less than 1 in comparison to male-gendered applications, suggesting that the likelihood of rejection is lower.

At last, Column 3 presents results of the most complete of the three models with the addition of state fixed effects. As seen in Table 1, coefficient estimates are largely the same as they were in Column 2 and contribute to the same narrative as above. There are certain results that hold similar across all three specifications. With respect to age, younger applicants (i.e., those with ages below 35) less likely to be rejected when compared to middle age applicants (the odds ratios are below 1). On the other hand, older aged applicants (ages greater than 55) are more likely to be rejected in comparison to middle aged applicants. The result for the YA category is somewhat surprising given I would expect lenders to think of younger applicants as more risky ones. The result is even more surprising when looking at the average applicant income for the YA bucket in comparison to the middle age category. More specifically, middle-aged applicant income is approximately 23.14% higher than the average income for young-adult applicants. Even considering elevated financial standing, the odds of rejection for younger applicants is lower.

Results for debt-to-income ratio are mostly the same across the three specifications as well. To no surprise, applicants with DI ratios greater than 50% are more likely to be rejected when compared to applicants with DI ratios between 36 and 49%. For applicants with DI ratios greater than 60%, the likelihood of rejection raises above 20 times that of the omitted group. Surprisingly, applications that list DI ratios below 20% (i.e., the coefficient estimate for DI -1) more likely to be rejected than those with DI ratios in the range of the omitted category. More specifically, the odds of rejection are greater than 1.5 times in each of the specifications. This is a counterintuitive result considering that lower debt-to-income ratios are considered a positive thing with respect to someone's financial profile. Coefficient estimates for buckets DI – 2 and DI – 3 are extremely close to 1, suggesting that the odds of rejection are almost the same as if their DI ratio were to be in the omitted category.

Tables 2 and 3 in Appendix B present the results of the loan outcome models when isolating each of the lending markets separately. I focus on the results from Column 2 in each case considering those are the most complete of the two models. Generally, the results in each of the isolated markets are similar. However, there are a couple of interesting results worth addressing. FHA markets are the most like the overall results in presented in Table 1. Black and Dual Race households continue to have the highest odds of rejection across racial groups when compared to White households. The same can be said for VA markets, however, coefficient estimates for Asian households are quite different in the two markets. In FHA markets, Asians are almost 1.5 times more likely to be rejected when compared to White households; as reference, the coefficient estimate of the odds ratio for Black households applying for FHA loans is 1.594, a mere 6.6% difference. The result for Asians in VA markets is what we would expect given their elevated financial standing in the sample – this is assuming that all racial buckets are already at a disadvantage when compared to White households individually. Asian households applying for VA loans are only 1.24 times more likely to be rejected when compared to White households applying for VA loans. More importantly, this odds ratio estimate is the lowest among all racial groups; 8.95% lower than applicants identifying as Pacific Islander and 26.87% lower than Black applicants. This is not to say that VA markets are free from evidence of racial and ethnic dispersion in lending outcomes – Black, Hispanic, and American Indian households are still at higher odds of rejection when compared to White households. However, assuming that racial categories are already at a disadvantage in comparison to White applicants, the results for Asian households in VA markets is what we would expect. This suggests the possibility that specific to Asian households, VA markets are more accessible and boast more favorable lending outcomes in comparison to FHA markets.

Taking a step back, the evidence of dispersion in loan outcomes across racial and ethnic groups is abundant amongst all results presented above. Given the models robustly control for the financial standing and credit risk of an applicant, I am confident in saying that this dispersion can only be interpreted as disparate impact beyond the objective strength of the applicant.

6.2 Discovery of Dispersion in Pricing Outcomes

Table 4 in Appendix C presents the results of the OLS regressions of rate spread on the identical series of racial, ethnic, and gendered binary operators, as well as borrower-level controls that enter the lender's consideration when determining contractual elements of the mortgage loan. Column 1 is the simplest of the specifications, only including lender-fixed effects while columns 2-4 present results for more robust models as evidenced by the significant increase in R^2 values for the subsequent specifications. Columns 3 and 4 are the fullest of the models presented in Table 4, with inclusion of lender, loan-type, and state fixed effects. Additionally, the 4th specification includes controls for origination charges unlike the preceding models.

Compared to non-Hispanic households, Hispanic households are charged higher rate spreads in every specification of the model. In column 1, the coefficient suggests Hispanic applicants pay about 0.26 percentage points more than non-Hispanic households, however, inclusion of loan-type and state fixed effects in subsequent specifications reduce this coefficient by more than half. This is generally the trend with most of the variables in the model – coefficient estimates are reduced in magnitude as the models become more robust. Amongst racial groups, coefficients for Dual Race and Pacific Islander applicants are not significantly different from 0, suggesting on average, people identifying with those racial groups are not charged more or less in spreads when compared to White households. This is contrary to what was found in the previous stage. Interestingly, across all 4 specifications, coefficients for American Indian households are negative and statistically significant. More specifically, American Indian households pay about 0.057 percentage points less in rate spreads when compared to White households (see column 4). The argument that the magnitude of the coefficient is inconsequential is somewhat misleading given the dependent variable in consideration is the interest rate spread over prime. If this were to be basic interest rates, then the argument may hold, but in the case of rate spreads, any statistically significant difference is material.

Observing Asian households in column 1, I find that applicants of this racial category are charged 0.08 percentage points higher when compared to White households. However, when more robust models are analyzed, this result is reversed. In Column 4, which is the most robust

of the models, I find that Asian households are charged lower rate spreads than White households – more specifically, this difference is 0.05 percentage points. This result is significantly more intuitive than the result for American Indian households considering the financial strength of the average Asian in the sample (see section 4.4). In the eyes of an objective lender, for this specific time period, Asian applicants should receive lower rate spreads. Lastly, applicants that identify as Black are the worst off among the racial groups. In every specification, they are charged the highest rate spreads when compared to White applicants. This is also the case when compared to the other racial groups – when the difference is tested, results are significant in every case. Looking at column 4, I find that Black households are charged 0.097 percentage points higher in spreads when compared to their White counterparts. These are the only coefficient estimates among the racial groups that are consistently significant and positive.

Some other pertinent results are worth pointing out in the model. Across specifications 2 through 4, I find that female applicants are charged lower rate spreads in comparison to males while joint applicants are charged higher rate spreads when compared to males. These results are quite surprising considering how the female gender is treated within the world of finance in general and the financial profile of joint applicants. With respect to income, the average woman in the U.S. makes 80.5% of what the average male would make, irrespective of industry.[20] In this data set, average female income is approximately 12% lower than that of the average male counterpart. Given this dispersion in earned income, women would be considered less financially endowed in the eyes of a credit lender. Given that financial standing and credit risk are among the main considerations for a lender's decision, I would think that women would be assigned higher rate spreads on loan applications. Furthermore, applicants or couples that apply joint are likely to have two different income streams, or at least, have two different co-signers for loan applications. In this scenario, lenders have more protection against adverse credit outcomes. This is also supported by the statistics in the data – applicants that list gender as joint have household incomes about 34.8% higher than male applicants on average.

Amongst the borrower-level controls included in the model, the effect of LTV and origination charges on rate spreads is the only counterintuitive result. Given these are continuous

[20] For more, see "The Gender Gap in the Finance Industry" by Janet Berry Johnson, *Accounting Principles*, https://blog.accountingprincipals.com/gender-gap-in-finance-industry

variables, coefficient estimates are converted into elasticities for sake of comparability. Looking at column 4, I find that a percentage increase in the LTV of a loan application results in an approximately 15.8% decrease in the rate spread charged with the associated loan. Considering that loan-to-value ratio is usually used as a measure for the risk factor of a loan application, this result makes little sense. The direction of this result also holds for the other 3 specifications included in the model. Additionally, the model estimates that a percentage increase in origination charges increases the rate spread by approximately 1.85%. Although significant, the magnitude of this result is quite low; an increase of 2% in a rate spread likely amounts to anywhere from 20-50 basis points at most. As one would think, coefficient estimates for Refinances and Cash-out refinance are significant and negative in each of the 4 specifications, considering the only incentive that pushes borrowers to trade out of first-lien home purchase mortgages is the opportunity to receive and pay lower spreads over the remaining maturity of their credit dues. Lastly, when including loan-type fixed effects (see columns 2-4), I find that, in the sample, VA loans are charged lower rate spreads in comparison to FHA loans. More specifically, as seen in column 4, the difference is approximately a full 100 basis points. This particular result is not surprising given that, on average, applicants who apply for VA loans earn 22.8% more in income when compared to FHA applicants.[21]

Tables 5 and 6 in Appendix C present results for the rate spread regression isolating FHA and VA lending markets individually. Note that individually, the markets have lower sample sizes, and thus the explanatory power of the models are greatly reduced – this is easily evidenced when comparing R^2 values between Table 4 and Tables 5-6. As with the lending outcome results in section 6.1, results in general remain largely similar when isolating the two markets. However, there are a few interesting distinctions worth pointing out. As with the results in Table 4, Hispanic and Black households have the worst pricing outcomes amongst the racial and ethnic buckets, receiving rate spreads .07 and 0.1 pp higher than the baseline category in the case of FHA markets, and .009 and .077 pp higher than the baseline in the case of VA markets respectively. Note that the magnitude of the coefficient for Hispanic applicants in VA markets is significantly lower than that of the estimate in FHA markets – the estimate in the case of VA markets essentially rounds to 0. In FHA markets, American Indian applicants, as seen in Table 4,

[21] Note, Income is expressed in thousands of dollars in the sample.

consistently receive lower rate spreads as evidenced by negative coefficients in each of the three specifications. In VA markets on the other hand, the coefficients for American Indians are not significantly different from 0 in each of the three specifications, suggesting that rate spreads charged to a given White and Native Indian applicant applying for VA loans in the sample is not significantly different. Results for Dual Race and Pacific Islander applicants remain the same as seen in Table 4. For Asian applicants, the same can be said for the direction of the results the coefficient estimates imply. However, for a given applicant in FHA markets, the reduction in rate spreads compared to White applicants is lower than it would be if an Asian applicant were to apply for a VA loan, i.e., all else equal an Asian applying for a VA loan would receive lower spreads than an Asian applying for an FHA loan.

As with the Rejection rate models presented in Tables 1-3, evidence of statistical dispersion in this pricing outcome also exists in abundance. Given loans are fully insured by the federal government, and the models robustly control for the objective strength/risk of a borrower, it is likely that these results are explainable by only one reason: discrimination. However, in the case of the pricing stage of credit outcomes, the question of discriminatory practices cannot be singularly addressed by analysis of interest rate charges. Thus, I look to Total Loan Costs and Upfront charges at closing in the ensuing sections.

6.3 Identification of Dispersion in Total Loan Costs

Analysis of Total Loan Costs provides a different angle on the relative cost of one's assigned mortgage contract. As aforementioned, analysis of rate spread only accounts for the monthly dues of a mortgage. However, there are many other costs associated with a mortgage loan, and total loan costs provides a glance into the entire cost of a mortgage contract over its maturity. Tables 7-9 in Appendix C present results for the Loan Cost regressions. As I have done in the past sections, Tables 8 and 9 present results for the Loan Cost models isolating each of the lending markets under consideration.

As evidenced by robust R^2 estimations, specifications in table 7 improve as loan-type and state fixed effects are included in the model. Regressing log transformed loan costs on an identical set of independent variables as the rate spread model presented in the previous section, I find that Hispanic applicants pay more in total loan costs when compared to non-Hispanic

applicants. More specifically, they pay 1.34% more than their non-Hispanic counterparts (see Column 3 of Table 7). Note, given the dependent variable is log transformed, I simply convert coefficient estimates for binary operators into elasticities as follows,

$$elasticity_y = 100 * \left(e^{\beta} - 1\right)$$

where β is the coefficient estimate for the dummy variable. Note that the elasticity is interpreted as a percentage change in the dependent variable when comparing the binary category to the baseline category. Across all three specifications, total loan cost charges are lower for all racial groups relative to the baseline group. Specifically, in comparison to White applicants, Dual Race applicants would pay 6.53% percent less, Native Indians would pay 4.42% less, Black applicants would pay 4.67% less, Asians would pay 1.13% less, and lastly Pacific Islanders would pay 3.32% less in total loan costs. Among the non-omitted racial groups, Asian and Pacific Islanders pay the most in total loan costs, while Dual Race, American Indian, and Black applicants pay the least in aggregated costs.

Given this measure includes all itemized costs over the life of the loan, there are a couple of things to consider when interpreting these results. For Asian households, even though average loan amount requested is among the highest in the sample, they are also the most financially endowed. Given interest is likely to be more expensive over the maturity loan when compared to increased lump-sum costs, it would make sense that, all else equal, an Asian pay less in aggregated loan costs when compared to White households. However, in the case of Black, Pacific Islander, and Native American households, the result is not so clear. Even though their diminished financial standing is likely to hamper the ability to pay lump-sum at closing to reduce rate spreads, households of these racial groups, on average, request less when applying for loans. In the case of applicants that identify as Pacific Islander, we found that the coefficient estimate was not significantly different from 0 in comparison to their White counterparts. When looking at the average size of the loan for a Pacific Islander, I find that this racial category, on average requests larger size loans than White households. If the size of the loan is larger, why wouldn't the aggregate cost also be larger? I resolve this query in the ensuing section on upfront charges.

When we look at rate spread results from Section 6.2, the result for Native American households becomes much more intuitive. Specifically, we found that Native American

applicants are charged lower rate spreads after controlling for loan amount and income. Thus, it would make sense that coefficient estimates for a given Native American household is negative, all else equal. For Black applicants, however, analysis of results in Section 6.2 displayed that across all specifications and lending markets that, ceteris paribus, a given Black applicant would be charged more in spreads when applying for a federally insured mortgage. Even considering the elevated rate spreads and diminished financial standing of Black applicants, it must be that the size of the loan requested causes this coefficient estimate to be lower than 0. Intuitively, applicants who request smaller loans will pay less interest over the time horizon that the loan is disbursed for.

Results for control variables are straightforward, however, there are a couple worth pointing out. When loan-type fixed effects are included, I find that, all else equal, a borrower applying for a VA loan will pay 53.42% less in aggregated costs in comparison to FHA loans. Although this result may see interesting at first glance, it is fairly expected consider VA borrowers are not required to pay PMI over the course of the loan's maturity. PMI constitutes one of the largest parts of the total itemized costs of a mortgage and considering that VA borrowers are free from this obligation, it is no surprise that total costs are significantly diminished. As with the direction of coefficient estimates for LTV in Section 6.1, a 1 percentage increase in the LTV of a loan results in a 0.452% decrease in total loan costs. Similar to the counterintuitive nature of coefficients on income in Section 6.1, it is possible that inclusion of property value and loan amount are confounding the estimate of this coefficient. A simple robustness check follows naturally – when excluding property value and loan amount from column 3 of Table 7, I find that a percentage increase in the LTV of a loan application increases the total cost of the loan by .295%, all else equal.

Tables 8 and 9 present the results of loan cost regressions when FHA and VA markets are individually isolated. Specification 2, in both cases, is the more robust of the models as lender and state fixed effects are included in the model. Within FHA markets, I find that Hispanic applicants pay 3.59% more in total costs when compared to a non-Hispanic applicant, all else equal. With reference to the racial categories, Asian applicants pay 1.34% more in total loan costs and Black applicants pay 3.13% more when compared to a White applicant. For Pacific Islander and Native American households, they pay less than 1% and 1.78% more in total loan

costs when compared to White applicants, respectively. However, it is important to note that the coefficient on Native Indian households is only significant at the 90th percent level. The results seen in VA markets are akin to what was presented in Table 7. All coefficient estimates for racial groups are significant and negative, suggesting aggregated costs for racial minorities in VA markets are lower than a White applicant applying for a VA loan in this year. For Hispanic applicants, when state and lender fixed effects are included in the model, total loan costs are 7.57% lower than non-Hispanic households.

Even considering the fact that VA borrowers are not required to pay PMI contracts for the duration of the loan's maturity, it seems that VA markets provide more favorable pricing outcomes for minority borrowers when compared to FHA markets. This is simply evidenced by the direction of the coefficient estimates for racial and ethnic groups between Tables 8 and 9. With respect to FHA markets, I find that a given applicant in all racial categories is paying more in aggregated costs when compared to a White applicant, which is the opposite of what is found for VA markets. However, I must point out that the explanatory power of these regression models is greatly diminished when VA markets are isolated, given the sample is mostly dominated by FHA loans. Therefore, I caution exercise when stating that VA markets may provide more favorable loan pricing outcomes, all else equal. Although the total cost models supply a better sense for the overall relative cost of a mortgage loan, explanatory power of models presented in Section 6.2 are significantly higher. However, given the explanatory power of results presented in Table 8 is comparable to those preceding it, I am more confident in stating that FHA markets do result in more adverse pricing outcomes for racial and ethnic minorities, with specific reference to Black borrowers, who are consistently paying more in total loan costs as seen in Columns 1 and 2 of Table 8. Yet, as I mentioned at the close of Section 6.2 complete statements cannot be made on the state of discrimination in pricing outcomes until upfront charges are considered.

6.4 The Tradeoff between Upfront Costs and Other Loan Pricing Outcomes

Table 10 present the results of the OLS models running log transformed origination charges on the identical set of ethnic and racial binary operators, as well as identical controls as seen in Tables 4-9. After inclusion of state and loan-type fixed effects in the model, I find that Hispanic borrowers pay 17.94% more in origination charges in comparison to a non-Hispanic borrower,

all else equal. Considering the results from the rate spread and total loan cost models, this result is quite surprising. Naturally, given that at random, a Hispanic borrower would be charged more in rate spreads and would have to pay more over the course of the loan's life, I would think that their lump-sum payments at closing would be lower than non-Hispanic borrowers (see Tables 4 and 7). Moving to Column 3 in Table 10 (which includes the full set of fixed effects and controls), the same result is also seen for Black applicants, who pay significantly more in upfront costs when compared to White borrowers. Although coefficient estimates for Asian borrowers are also positive, this can likely be explained by their elevated financial standing. Given that Asian borrowers are charged less in rate spreads (see Table 4), these results go somewhat hand in hand. As aforementioned, increased financial liquidity allows Asian borrowers to elect higher upfront costs, which in turn reduce interest rate charges over the life of the loan. This is also evidenced by the fact that Asian borrowers pay slightly less than White borrowers in aggregated loan costs (See Table 7). However, I remain skeptical of the explanatory power of the main total cost regression model. Considering these results, it would follow that evidence of disparate impact beyond the strength of the applicant is abundant and complete for Hispanic and Black borrowers when looking at all pricing outcomes in this sample.

For American Indians, the results are as conclusive as they are for Asian borrowers in the sample. In Tables 4 and 7, I found that American Indians were charged less in rate spreads and paid less in aggregate when compared to White borrowers. This result is in line with the coefficient estimate for American Indians in Table 10. Specifically, I find that American Indian applicants, all else equal, pay 11.85% more in upfront costs when compared to White borrowers. This reflects a similar narrative as presented for Asian borrowers. However, the overall result is less intuitive than that of Asian borrowers in the sample considering the financial standing of American Indian borrowers. Compared to other racial groups in the sample, American Indian applicants in the sample have the second lowest income, on average. Although elevated financial standing provides an explanation for Asian households electing to pay more upfront, the same cannot be said of American Indian households in the sample. It is possible that applicants of this racial group may be taking on larger lump-sum costs at closing irrespective of financial liquidity, but this would be hard to separate from the results presented in the models above. For the other racial minority categories, the results are somewhat inconclusive. Given that coefficient estimates for Dual Race and Pacific Islander applicants are insignificant, no conclusions can be

made on the dispersion of upfront costs in comparison to White applicants. As aforementioned, conclusions on the presence of discrimination cannot be answered without wholly encompassing all attributes of loan pricing outcomes. Even though evidence of potential discrimination was found in the case of Dual Race and Pacific Islander applicants in rate spread and aggregate cost determinations, no sound conclusions can be made on the presence of disparate impact on loan pricing outcomes overall.

7 Conclusion and Caveats

Mortgage literature has historically been inconclusive and incomplete in its attempt to identify discriminatory practices in lending markets. Given mortgage data has not been fully available to the public, models attempting to identify dispersion have also been incomplete, mostly suffering from the ability to fully control for the financial standing of the borrower and the amount of credit risk priced in by the lender. This is especially applicable to literature that has focused on conventional credit markets. Even in present times, credit history of the borrower has been near impossible to control for given lender institutions can elect to hold mortgage loans on their balance sheets. As a result, the lenders themselves continue to face and price default and prepayment risk. However, the study of federally insured markets presents an opportunity where such mortgage outcomes are not possible. Given credit risk is completely insured by the faith of the federal entities, what incentive would mortgage issuers have to decline applications at all? Although previous studies have attempted to answer this question, none have done so accounting for the sequential nature of mortgage lender. This thesis contributes to this literature by analyzing the presence of discrimination or disparate impact within federally insured lending markets in a sequential manner, accounting for both lending and pricing outcomes.

Looking at the 2019 HMDA sample, I first analyze dispersion in lending outcomes within and across FHA and VA markets. After controlling for the financial standing and strength of the borrower, I find conclusive evidence that Hispanic and all racial groups (in comparison to White borrowers) are discriminated against beyond the objective strength of the applicant. When isolating FHA and VA markets I find similar results, albeit muted for Asian borrowers within VA markets. Taking loans from approval outcomes to pricing outcomes, I find that when taking into consideration all components of loan pricing, that Hispanic and Black borrowers are subject to disparate impact in the pricing outcomes of federally insured mortgages in the sample. Asian

borrowers, however, do not face the same outcomes in the second stage of mortgage lending as results indicate that, all else equal, pricing terms of mortgage contracts are cheaper for Asian borrowers given their elevated financial standing. More specifically, increased financial liquidity allows them to elect higher upfront costs, therefore reducing the overall interest charges over the life of the loan. Results for American Indians are akin to that of Asians, but significantly less intuitive given their financial standing. In contrast to these conclusive statements, the presence of discrimination for Dual Race and Pacific Islander applicants is entirely clear in the pricing stage of the model. Although they are charged more and face higher aggregate costs over the life of the loan, nothing can be said about the election of upfront costs, given coefficient are insignificant. In general, results of this paper speak to the necessity of increased regulation in all lending markets. Although post-08 regulation has done some to improve the transparency of large mortgage institutions, evidence of discrimination is continually present in modern times. This is even more surprising in the case of federally insured lending markets considering that there is little to no incentive for mortgage issuers to ever decline an application albeit those that do not meet the minimum requirements. Whether this means completely automating lending decision making, or re-considering policies in place, there is no doubt that things need to change.

Some caveats to this study are worthy of addressing. At the outset, the interpretation of these results must be taken with great caution considering they are only applicable to the fiscal year 2019. Although, I find it important to mention that this methodology is likely to be beneficial for credit studies using panel data. Given this study is among the first of its kind, more far-reaching mortgage literature may benefit from such an approach. Additionally, some comments must be made on the continuing debate on how to proxy for credit score credit lending models. Apart from the use of private mortgage data as seen in Bhutta Hizmo (2020) and Bartlett et al. (2019), further study into this subject should investigate how to measure the mix and age of an applicant's credit profile. Although debt-to-income ratio is a sound proxy for the ability to pay back debt obligations, it is not enough of a proxy for overall credit history, as we saw with robustness checks conducted in Section 6.2.

Further questions for future research also come into the spotlight because of this study. Additionally, for studies that continue to tackle the question of discrimination in conventional markets – how can models fully control for the amount which lenders price in credit risk even if

they choose to hold mortgages on their balance sheets? With specific reference to the COVID-19 pandemic, future studies could also benefit from analyzing how pandemic-related mortgage forbearance and lending criterion easing has affected approval and pricing outcomes. Furthermore, recent social justice and freedom movements throughout the country bring into light some of the backward practices still existent in this country. Questions need to be asked as to whether consumer mortgage markets fall under this umbrella.

Appendix A

Table 1
Loan Purpose by Racial Category

	Race					
Loan Purpose	*Dual Race*	*Native American*	*Asian*	*Black*	*Pacific*	*White*
Home Purchase	2,597	8,517	27,026	174,010	3,450	856,453
Refinance	204	855	2,619	15,249	410	84,245
Cash-out Refinance	959	3,416	8,286	54,146	1,924	290,175

Table 2
Loan Purpose by Racial and Ethnic Categories

	Not Hispanic						
Loan Purpose	*DualRace*	*NativeAmerican*	*Asian*	*Black*	*Pacific*	*White*	*Total*
Home Purchase	2,201	6,135	25,942	169,063	2,786	670,688	876,815
Refinance	182	658	2,535	14,937	366	70,852	89,530
Cash-out Refinance	880	2,655	8,015	53,171	1,670	261,016	327,407
	Hispanic						
	DualRace	*NativeAmerican*	*Asian*	*Black*	*Pacific*	*White*	*Total*
Home Purchase	396	2,382	1,084	4,947	664	185,765	195,238
Refinance	22	197	84	312	44	13,393	14,052
Cash-out Refinance	79	761	271	975	254	29,159	31,499

Table 3
Breakdown of Mortgage Inflows – Purchaser and Type

Mortgage Type	**Total**	**Percentage**
Conventional	13,041,150	74.29%
Nonconventional	4,512,662	25.71%
FHA	2,608,152	14.86%
VA	1,695,666	9.66%
FSA/RHA	208,844	1.19%

Table 4
Summary of Income by Racial and Ethnic Category

	Ethnicity		
Race	*Not Hispanic*	*Hispanic*	**Total**
Dual Race	78.0855	75.4930	77.7428
	(72.046)	(54.106)	(69.939)
Native American	72.8024	70.6296	72.2349
	(126.628)	(39.541)	(110.705)
Asian	86.5888	78.2189	86.2713
	(50.952)	(41.514)	(50.651)
Black	73.2156	68.7899	73.1023
	(84.013)	(37.719)	(83.152)
Pacific	81.0840	69.7911	79.2057
	(42.784)	(34.987)	(41.798)
White	78.3465	69.8790	76.7758
	(248.520)	(46.056)	(225.189)
Total	77.6075	69.9223	76.4015
	(222.197)	(45.733)	(204.842)

Note: Standard Deviation reported in parentheses.

Table 5
Summary of Income by Loan Type (Thousands of Dollars)

Loan Type	N	Mean Income	SD Income
FHA	997,527	71.2484	138.9899
VA	537,014	85.9738	289.6179
Total	1,534,541	76.4015	204.8424

Table 6
Rejection Rates by Ethnicity and Race

	Ethnicity		
Race	*Not Hispanic*	*Hispanic*	**Total**
Dual Race	0.24425	0.21328	0.24016
	(0.4297)	(0.4100)	(0.4272)
Native American	0.22851	0.21766	0.22568
	(0.4199)	(0.4127)	(0.4180)
Asian	0.18779	0.18277	0.18760
	(0.3906)	(0.3866)	(0.3904)
Black	0.22965	0.21928	0.22938
	(0.4206)	(0.4138)	(0.4204)
Pacific	0.21443	0.22349	0.21594
	(0.4105)	(0.4168)	(0.4115)
White	0.15197	0.15081	0.15175
	(0.3590)	(0.3579)	(0.3588)
Total	0.16824	0.15412	0.16603
	(0.3741)	(0.3611)	(0.3721)

Note: Standard Deviation Denoted in parentheses.

Table 7
Rejection Rates by Loan Type

Loan Type	N	Mean Rejection Rate	SD Rejection Rate
FHA	997,527	0.1727	0.37800
VA	537,014	0.1536	0.36056
Total	1,534,541	0.1660	0.37210

Table 8
Loan Action Taken by Loan Type

	Action Taken		
Loan Type	**Originated**	**Approved**	**Rejected**
FHA	797,562	27,676	172,289
VA	441,606	12,923	82,485

Table 9
Loan Action Taken by Race and Ethnicity

	Not Hispanic					
Action Taken	*Dual Race*	*Native American*	*Asian*	*Black*	*Pacific*	*White*
Originated	2,375	7,037	28,609	175,243	3,676	825,091
Approved	91	252	1,030	7,462	112	25,111
Rejected	797	2,159	6,853	54,466	1,034	152,354
	Hispanic					
	Dual Race	*Native American*	*Asian*	*Black*	*Pacific*	*White*
Originated	376	2,537	1,145	4,712	721	187,646
Approved	15	76	31	155	26	6,238
Rejected	106	727	263	1,367	215	34,433

Table 10
Loan Amount by Race and Ethnicity

	Ethnicity		
Race	*Not Hispanic*	*Hispanic*	**Total**
Dual Race	268,138.22	255,221.33	266,430.85
	(139,140.38)	(117,821.93)	(136,569.44)
NativeAmerican	208,482.22	242,107.78	217,264.62
	(108,026.38)	(114,489.45)	(110,736.25)
Asian	323,317.99	281,928.42	321,747.78
	(159,318.81)	(136,889.81)	(158,721.34)
Black	232,985.71	234,563.68	233,026.13
	(119,092.83)	(115,649.11)	(119,005.91)
Pacific	306,638.32	254,646.57	297,991.01
	(144,551.36)	(131,055.43)	(143,695.00)
White	230,355.35	243,781.43	232,845.78
	(121,716.96)	(113,999.46)	(120,435.90)
Total	233,679.56	243,814.56	235,269.87
	(123,576.25)	(114,330.42)	(122,227.32)

Note: Standard Deviation noted in parentheses.

Table 11
Property Value by Race and Ethnicity

	Ethnicity		
Race	*Not Hispanic*	*Hispanic*	**Total**
Dual Race	291,000.61	268,299.80	288,000.00
	(160,670.26)	(126,995.27)	(156,808.72)
Native American	228,395.43	264,248.50	237,759.62
	(124,184.51)	(127,727.87)	(126,102.09)
Asian	359,466.73	306,660.88	357,463.42
	(193,416.03)	(155,218.30)	(192,368.59)
Black	250,392.35	248,991.02	250,356.46
	(294,279.09)	(127,377.32)	(291,200.50)
Pacific	339,228.54	276,881.50	328,858.92
	(170,272.96)	(146,210.88)	(168,111.47)
White	253,367.81	261,879.25	254,946.62
	(164,870.24)	(144,640.57)	(161,343.43)
Total	256,047.57	261,919.25	256,968.91
	(196,540.17)	(144,094.63)	(189,286.49)

Note: Standard Deviation noted in parentheses.

Table 12
Rate Spread by Race and Ethnicity

	Ethnicity		
Race	*Not Hispanic*	*Hispanic*	**Total**
Dual Race	0.75573	0.78787	0.76024
	(0.84427)	(0.90780)	(0.85336)
Native American	0.84840	0.98404	0.88731
	(0.81312)	(0.83990)	(0.82313)
Asian	0.78203	0.85293	0.78475
	(0.76109)	(0.83995)	(0.76437)
Black	1.05467	0.99096	1.05304
	(0.82213)	(0.85591)	(0.82307)
Pacific	0.71650	0.88298	0.74451
	(0.81841)	(0.87472)	(0.83038)
White	0.85654	1.18116	0.91739
	(0.81407)	(0.78133)	(0.81791)
Total	0.88746	1.17048	0.93294
	(0.81777)	(0.78644)	(0.81944)

Table 13
Rate Spread by Loan Type

Loan Type	**Mean Rate Spread**	**SD Rate Spread**
FHA	1.32004359	0.646248792
VA	0.168148852	0.540397068
Total	0.932936912	0.819438793

Table 14
Total Loan Costs by Race and Ethnicity

	Ethnicity		
Race	*Not Hispanic*	*Hispanic*	**Total**
Dual Race	7,998.94	7,667.95	7,951.97
	(4913.270)	(4179.177)	(4816.505)
Native American	6,761.22	8,063.82	7,135.48
	(3723.973)	(8701.170)	(5654.798)
Asian	9,381.74	8,297.24	9,339.77
	(9954.294)	(4435.284)	(9800.888)
Black	7,579.24	7,657.72	7,581.26
	(8886.822)	(4229.601)	(8797.732)
Pacific	9,051.71	8,161.89	8,901.68
	(5329.418)	(4766.549)	(5248.757)
White	7,265.36	8,457.24	7,486.54
	(11653.814)	(7605.480)	(11025.292)
Total	7,380.93	8,429.82	7,548.13
	(11136.252)	(7531.660)	(10650.662)

Note: Standard Deviation noted in parentheses.

Table 15
Summary of Origination Charges by Race and Ethnicity

	Ethnicity		
Race	*Not Hispanic*	*Hispanic*	**Total**
Dual Race	1,909.74	1,714.83	1,882.30
	(2260.340)	(1971.346)	(2222.586)
Native American	1,833.92	2,017.53	1,886.45
	(1914.313)	(2115.542)	(1975.588)
Asian	2,117.59	1,949.83	2,111.17
	(3282.916)	(4141.720)	(3319.917)
Black	1,919.63	1,833.14	1,917.45
	(2710.728)	(2134.272)	(2697.731)
Pacific	2,349.06	2,135.78	2,313.68
	(2644.674)	(2557.033)	(2631.206)
White	1,755.11	2,014.76	1,803.25
	(2781.644)	(3123.049)	(2849.808)
Total	1,795.48	2,010.14	1,829.67
	(2780.359)	(3096.241)	(2834.120)

Note: Standard Deviation denoted in parentheses.

Table 16
Summary of Combined to Loan to Value Ratios

	Ethnicity		
Race	*Not Hispanic*	*Hispanic*	**Total**
DualRace	98.485	96.077	98.147
	(207.8230)	(7.2504)	(192.7056)
NativeAmerican	93.539	93.864	93.632
	(10.6955)	(10.2476)	(10.5694)
Asian	92.564	94.119	92.623
	(11.2030)	(9.1708)	(11.1358)
Black	101.981	95.378	101.812
	(2450.9869)	(7.9102)	(2419.4870)
Pacific	92.656	93.730	92.837
	(19.0315)	(9.8589)	(17.8253)
White	95.677	171.881	109.962
	(2002.7331)	(31609.5348)	(13804.1702)
Total	96.647	168.221	108.147
	(2047.7278)	(30850.7872)	(12507.6472)

Note: Standard Deviations denoted in parentheses.

Table 17
Discount Points by Race and Ethnicity

	Ethnicity		
Race	*Not Hispanic*	*Hispanic*	**Total**
DualRace	2,282.42	2,072.68	2,254.98
	(2232.747)	(2084.946)	(2214.061)
NativeAmerican	1,889.77	1,998.74	1,918.05
	(1921.225)	(1915.293)	(1920.019)
Asian	2,478.49	2,122.19	2,464.42
	(2659.939)	(2216.774)	(2644.662)
Black	2,139.87	2,046.13	2,137.57
	(2133.732)	(2017.905)	(2130.998)
Pacific	2,743.77	2,507.56	2,705.82
	(2511.326)	(2640.779)	(2533.228)
White	1,924.66	1,811.02	1,904.12
	(1932.535)	(1870.500)	(1921.963)
Total	1,979.53	1,823.96	1,955.47
	(1997.218)	(1881.754)	(1980.601)

Note: Standard Deviation denoted in parentheses.

Table 18
Lender Credits by Race and Ethnicity

	Ethnicity		
Race	*Not Hispanic*	*Hispanic*	**Total**
DualRace	1,282.88	665.71	1,200.41
	(1994.909)	(1016.913)	(1904.909)
NativeAmerican	754.88	1,070.38	844.91
	(1325.701)	(1879.494)	(1511.051)
Asian	1,593.20	1,366.56	1,584.94
	(2496.487)	(1980.964)	(2479.872)
Black	1,112.60	1,107.00	1,112.47
	(2136.679)	(1754.181)	(2128.211)
Pacific	1,365.42	1,393.09	1,369.89
	(2190.600)	(2365.793)	(2219.030)
White	931.06	1,185.64	981.26
	(1597.959)	(1910.174)	(1667.252)
Total	984.45	1,183.43	1,017.67
	(1745.415)	(1907.673)	(1775.083)

Note : Standard Deviation denoted in parentheses.

Table 19
Loan Term by Race and Ethnicity

	Ethnicity		
Race	*Not Hispanic*	*Hispanic*	**Total**
DualRace	356.989	357.303	357.033
	(21.695)	(20.460)	(21.522)
NativeAmerican	355.222	356.412	355.564
	(27.983)	(24.068)	(26.923)
Asian	357.006	356.936	357.004
	(22.255)	(23.680)	(22.311)
Black	356.281	357.514	356.312
	(24.785)	(20.286)	(24.681)
Pacific	357.320	356.085	357.112
	(20.627)	(26.868)	(21.803)
White	355.527	358.010	355.992
	(27.147)	(19.020)	(25.837)
Total	355.704	357.963	356.067
	(26.611)	(19.191)	(25.578)

Note: Standard Deviation denoted in Parentheses

Appendix B

Table 1
Loan Outcome Regressions

$Rejection_i$	(1)	(2)	(3)
Hispanic	1.363***	1.280***	1.228***
	(0.00971)	(0.00917)	(0.00950)
DualRace	1.682***	1.718***	1.728***
	(0.0752)	(0.0768)	(0.0775)
AmericanIndian	1.326***	1.336***	1.370***
	(0.0328)	(0.0331)	(0.0343)
Asian	1.540***	1.458***	1.446***
	(0.0239)	(0.0226)	(0.0227)
Black	1.665***	1.653***	1.641***
	(0.0109)	(0.0108)	(0.0112)
Pacific	1.410***	1.411***	1.441***
	(0.0529)	(0.0531)	(0.0549)
Female	1.014**	0.920***	0.918***
	(0.00627)	(0.00580)	(0.00580)
Joint	0.766***	0.755***	0.754***
	(0.00465)	(0.00458)	(0.00459)
YA	0.846***	0.842***	0.846***
	(0.00547)	(0.00545)	(0.00549)
OldAge	1.096***	1.192***	1.184***
	(0.00668)	(0.00739)	(0.00738)
DI - 1	1.747***	1.870***	1.871***
	(0.0218)	(0.0234)	(0.0235)
DI - 2	1.046***	1.092***	1.095***
	(0.00902)	(0.00948)	(0.00952)
DI - 3	0.952***	0.974***	0.976***
	(0.00784)	(0.00805)	(0.00808)
DI - 5	1.452***	1.411***	1.405***
	(0.00918)	(0.00897)	(0.00896)
DI - 6	21.94***	22.50***	22.37***
	(0.233)	(0.236)	(0.235)
ln(Income)	1.178***	1.160***	1.155***
	(0.00511)	(0.00502)	(0.00503)
ln(Loan Amount)	0.514***	0.564***	0.557***
	(0.00282)	(0.00320)	(0.00352)
Refinance	3.180***	3.048***	3.057***
	(0.0301)	(0.0288)	(0.0290)
Cash-out Refinance	3.141***	3.257***	3.290***
	(0.0200)	(0.0209)	(0.0216)
LEI_1	1.316***	1.294***	1.292***
	(0.0125)	(0.0124)	(0.0124)
LEI_2	0.662***	0.648***	0.643***
	(0.0105)	(0.0102)	(0.0102)
LEI_3	0.506***	0.702***	0.702***
	(0.0108)	(0.0154)	(0.0154)
LEI_4	1.878***	1.927***	1.921***
	(0.0231)	(0.0237)	(0.0237)
LEI_5	0.623***	0.622***	0.633***
	(0.0129)	(0.0129)	(0.0131)
VA Loan		0.608***	0.617***
		(0.00415)	(0.00425)
Constant	135.8***	53.52***	58.99***
	(8.489)	(3.440)	(5.532)
Observations	1,534,541	1,534,541	1,534,541
Lender FE	YES	YES	YES
Loan Type FE		YES	YES
State FE			YES

Robust seeform in parentheses
**** p<0.01, ** p<0.05, * p<0.1*

Table 1A
Loan Outcome Regressions - Robustness Checks

$Rejection_i$	(1)	(2)	(3)
Hispanic	1.393***	1.302***	1.228***
	(0.00943)	(0.00889)	(0.00902)
DualRace	1.738***	1.772***	1.773***
	(0.0732)	(0.0747)	(0.0750)
AmericanIndian	1.319***	1.331***	1.369***
	(0.0310)	(0.0313)	(0.0325)
Asian	1.662***	1.570***	1.543***
	(0.0241)	(0.0228)	(0.0227)
Black	1.724***	1.706***	1.686***
	(0.0107)	(0.0106)	(0.0108)
Pacific	1.434***	1.431***	1.457***
	(0.0502)	(0.0502)	(0.0519)
Female	1.033***	0.934***	0.933***
	(0.00605)	(0.00560)	(0.00561)
Joint	0.816***	0.806***	0.804***
	(0.00468)	(0.00463)	(0.00463)
YA	0.820***	0.817***	0.821***
	(0.00509)	(0.00508)	(0.00511)
OldAge	1.129***	1.223***	1.214***
	(0.00645)	(0.00712)	(0.00711)
ln(Income)	0.875***	0.859***	0.854***
	(0.00283)	(0.00278)	(0.00279)
ln(Loan Amount)	0.676***	0.736***	0.727***
	(0.00339)	(0.00385)	(0.00427)
Refinance	3.270***	3.178***	3.201***
	(0.0285)	(0.0277)	(0.0281)
Cash-out Refinance	3.192***	3.318***	3.398***
	(0.0193)	(0.0202)	(0.0212)
LEI_1	1.249***	1.225***	1.220***
	(0.0116)	(0.0114)	(0.0114)
LEI_2	0.713***	0.694***	0.687***
	(0.0110)	(0.0107)	(0.0107)
LEI_3	0.463***	0.638***	0.637***
	(0.00958)	(0.0135)	(0.0135)
LEI_4	1.778***	1.820***	1.808***
	(0.0208)	(0.0213)	(0.0212)
LEI_5	0.615***	0.610***	0.625***
	(0.0125)	(0.0125)	(0.0128)
VA Loan		0.617***	0.628***
		(0.00389)	(0.00400)
Constant	21.58***	9.488***	10.66***
	(1.267)	(0.575)	(0.941)
Observations	1,534,541	1,534,541	1,534,541
Lender FE	YES	YES	YES
Loan Type FE		YES	YES
State FE			YES

Robust seeform in parentheses
**** p<0.01, ** p<0.05, * p<0.1*

Table 2
Rejection Rate Regressions - FHA Only

$Rejection_i$	(1)	(2)
Hispanic	1.273***	1.228***
	(0.0104)	(0.0110)
DualRace	1.761***	1.750***
	(0.106)	(0.106)
AmericanIndian	1.273***	1.308***
	(0.0395)	(0.0411)
Asian	1.512***	1.496***
	(0.0276)	(0.0276)
Black	1.603***	1.594***
	(0.0129)	(0.0133)
Pacific	1.477***	1.459***
	(0.0713)	(0.0725)
Female	0.928***	0.926***
	(0.00667)	(0.00667)
Joint	0.787***	0.784***
	(0.00606)	(0.00606)
YA	0.818***	0.822***
	(0.00616)	(0.00621)
OldAge	1.269***	1.260***
	(0.00984)	(0.00981)
DI -1	2.041***	2.044***
	(0.0332)	(0.0334)
Di - 2	1.136***	1.140***
	(0.0124)	(0.0124)
DI - 3	1.002	1.004
	(0.0101)	(0.0101)
DI - 5	1.385***	1.381***
	(0.0103)	(0.0103)
DI - 6	44.16***	44.02***
	(0.690)	(0.688)
Ln(Income)	1.181***	1.179***
	(0.00731)	(0.00735)
Ln(Loan Amount)	0.563***	0.552***
	(0.00399)	(0.00435)
Refinance	3.072***	3.085***
	(0.0321)	(0.0324)
Cash-out Refinance	3.097***	3.125***
	(0.0245)	(0.0253)
LEI_1	1.322***	1.322***
	(0.0148)	(0.0148)
LEI_2	0.714***	0.709***
	(0.0124)	(0.0124)
LEI_3	0.604***	0.614***
	(0.0804)	(0.0819)
LEI_4	1.768***	1.762***
	(0.0269)	(0.0269)
LEI_5	0.625***	0.633***
	(0.0147)	(0.0149)
Constant	48.07***	67.84***
	(3.743)	(8.409)
Observations	997,527	997,527
Lender FE	YES	YES
State FE		YES

Robust seeform in parentheses
**** p<0.01, ** p<0.05, * p<0.1*

Table 3
Rejection Rate Regressions - VA Only

$Rejection_i$	(1)	(2)
Hispanic	1.264***	1.197***
	(0.0211)	(0.0208)
DualRace	1.625***	1.639***
	(0.106)	(0.107)
AmericanIndian	1.443***	1.467***
	(0.0602)	(0.0615)
Asian	1.260***	1.241***
	(0.0384)	(0.0385)
Black	1.704***	1.697***
	(0.0198)	(0.0205)
Pacific	1.321***	1.363***
	(0.0777)	(0.0810)
Female	0.957***	0.960***
	(0.0145)	(0.0146)
Joint	0.706***	0.705***
	(0.00705)	(0.00709)
YA	0.912***	0.918***
	(0.0122)	(0.0123)
OldAge	1.080***	1.072***
	(0.0112)	(0.0113)
DI - 1	1.600***	1.602***
	(0.0324)	(0.0326)
DI - 2	0.993	0.997
	(0.0144)	(0.0144)
DI - 3	0.905***	0.906***
	(0.0131)	(0.0131)
DI - =5	1.557***	1.549***
	(0.0193)	(0.0192)
DI - 6	10.41***	10.30***
	(0.170)	(0.170)
Ln(Income)	1.123***	1.116***
	(0.00821)	(0.00821)
Ln(Loan Amount)	0.560***	0.564***
	(0.00578)	(0.00650)
Refinance	2.939***	2.938***
	(0.0747)	(0.0750)
Cash-out Refinance	3.493***	3.532***
	(0.0390)	(0.0406)
LEI_1	1.196***	1.183***
	(0.0228)	(0.0226)
LEI_2	0.443***	0.439***
	(0.0171)	(0.0170)
LEI_3	0.650***	0.649***
	(0.0145)	(0.0145)
LEI_4	2.178***	2.171***
	(0.0451)	(0.0451)
LEI_5	0.616***	0.627***
	(0.0280)	(0.0286)
Constant	46.50***	36.76***
	(5.634)	(5.769)
Observations	537,014	537,014
Lender FE	YES	YES
State FE		YES

Robust seeform in parentheses
**** p<0.01, ** p<0.05, * p<0.1*

Table 4
Rate Spread Regressions

Rate Spread	(1)	(2)	(3)	(4)
Hispanic	0.257***	0.0902***	0.0686***	0.0680***
	(0.00184)	(0.00154)	(0.00159)	(0.00157)
DualRace	-0.0224	0.0131	0.00631	0.00207
	(0.0141)	(0.0108)	(0.0108)	(0.0108)
AmericanIndian	-0.0674***	-0.0538***	-0.0520***	-0.0573***
	(0.00769)	(0.00659)	(0.00649)	(0.00646)
Asian	0.0809***	-0.0388***	-0.0487***	-0.0509***
	(0.00407)	(0.00311)	(0.00314)	(0.00314)
Black	0.143***	0.0964***	0.0989***	0.0966***
	(0.00192)	(0.00157)	(0.00161)	(0.00161)
Pacific	0.0113	0.00212	-0.0118	-0.0102
	(0.0113)	(0.00895)	(0.00884)	(0.00894)
Female	0.166***	-0.0261***	-0.0264***	-0.0279***
	(0.00167)	(0.00142)	(0.00141)	(0.00140)
Joint	0.0605***	0.0269***	0.0264***	0.0258***
	(0.00153)	(0.00121)	(0.00120)	(0.00120)
YA	-0.0267***	-0.0384***	-0.0388***	-0.0365***
	(0.00152)	(0.00122)	(0.00121)	(0.00120)
OldAge	-0.211***	0.000498	-0.00517***	-0.0109***
	(0.00179)	(0.00148)	(0.00147)	(0.00147)
Di - 1	-0.293***	-0.0791***	-0.0733***	-0.0714***
	(0.00433)	(0.00331)	(0.00330)	(0.00329)
DI - 2	-0.207***	-0.0898***	-0.0837***	-0.0833***
	(0.00230)	(0.00182)	(0.00180)	(0.00179)
DI - 3	-0.107***	-0.0454***	-0.0423***	-0.0420***
	(0.00203)	(0.00161)	(0.00160)	(0.00159)
DI - 5	0.0365***	-0.0253***	-0.0287***	-0.0282***
	(0.00159)	(0.00131)	(0.00130)	(0.00129)
DI - 6	-0.184***	0.0392***	0.0346***	0.0318***
	(0.00601)	(0.00439)	(0.00443)	(0.00430)
Ln(Income)	0.0717***	0.0361***	0.0386***	0.0366***
	(0.00122)	(0.000937)	(0.000949)	(0.000930)
Ln(Loan Amount)	0.111***	0.401***	0.407***	0.384***
	(0.0260)	(0.0229)	(0.0230)	(0.0206)
Ln(Property Value)	-0.688***	-0.730***	-0.763***	-0.744***
	(0.0259)	(0.0226)	(0.0227)	(0.0205)
Loan Term	-0.00113***	-0.00188***	-0.00197***	-0.00199***
	(3.25e-05)	(2.85e-05)	(2.84e-05)	(2.68e-05)
Ln(LTV)	-0.444***	-0.160***	-0.161***	-0.147***
	(0.0233)	(0.0180)	(0.0181)	(0.0163)
Ln(Origination Charges)				0.0173***
				(0.000177)
Refinance	-0.119***	-0.227***	-0.227***	-0.237***
	(0.00291)	(0.00241)	(0.00242)	(0.00240)
Cash-Out Refinance	-0.243***	-0.110***	-0.106***	-0.125***
	(0.00241)	(0.00204)	(0.00207)	(0.00202)
VA Loan		-1.028***	-1.023***	-0.994***
		(0.00131)	(0.00132)	(0.00134)
Constant	10.24***	6.662***	6.822***	6.699***
	(0.108)	(0.0830)	(0.0845)	(0.0766)
Observations	1,171,656	1,171,656	1,171,656	1,109,577
R-squared	0.290	0.538	0.548	0.565
Lender FE	YES	YES	YES	YES
Loan Type FE		YES	YES	YES
State FE			YES	YES

Robust standard errors in parentheses
**** p<0.01, ** p<0.05, * p<0.1*

Table 5
Rate Spread Regression - FHA Markets

Rate Spread$_i$	(1)	(2)	(3)
Hispanic	0.102***	0.0719***	0.0705***
	(0.00178)	(0.00187)	(0.00184)
DualRace	0.0289*	0.0289*	0.0218
	(0.0156)	(0.0156)	(0.0153)
AmericanIndian	-0.0866***	-0.0792***	-0.0872***
	(0.00882)	(0.00861)	(0.00855)
Asian	-0.0329***	-0.0382***	-0.0411***
	(0.00396)	(0.00399)	(0.00400)
Black	0.0987***	0.102***	0.100***
	(0.00198)	(0.00203)	(0.00201)
Pacific	0.00571	0.00391	0.00196
	(0.0126)	(0.0122)	(0.0125)
Female	-0.0203***	-0.0202***	-0.0215***
	(0.00166)	(0.00163)	(0.00162)
Joint	0.0314***	0.0307***	0.0309***
	(0.00167)	(0.00165)	(0.00164)
YA	-0.0415***	-0.0411***	-0.0392***
	(0.00152)	(0.00150)	(0.00148)
OldAge	-0.00798***	-0.0122***	-0.0183***
	(0.00211)	(0.00210)	(0.00209)
DI - 1	-0.0315***	-0.0275***	-0.0129**
	(0.00554)	(0.00549)	(0.00548)
DI - 2	-0.0793***	-0.0704***	-0.0683***
	(0.00262)	(0.00259)	(0.00257)
DI - 3	-0.0400***	-0.0353***	-0.0340***
	(0.00219)	(0.00216)	(0.00215)
DI - 5	-0.0344***	-0.0393***	-0.0397***
	(0.00160)	(0.00159)	(0.00157)
DI - 6	0.163***	0.157***	0.150***
	(0.00950)	(0.00956)	(0.00914)
Ln(Income)	0.0555***	0.0552***	0.0504***
	(0.00145)	(0.00147)	(0.00143)
Ln(Loan Amount)	0.396***	0.405***	0.373***
	(0.0260)	(0.0260)	(0.0235)
Ln(Property Value)	-0.741***	-0.777***	-0.746***
	(0.0258)	(0.0258)	(0.0234)
Loan Term	-0.00147***	-0.00160***	-0.00164***
	(4.40e-05)	(4.39e-05)	(4.07e-05)
Ln(LTV)	-0.211***	-0.210***	-0.186***
	(0.0182)	(0.0183)	(0.0163)
Ln(Origination Charges)			0.0212***
			(0.000264)
Refinance	-0.274***	-0.275***	-0.282***
	(0.00301)	(0.00303)	(0.00300)
Cash-Out Refinance	-0.247***	-0.239***	-0.249***
	(0.00361)	(0.00364)	(0.00349)
Constant	6.897***	6.918***	6.723***
	(0.0872)	(0.0902)	(0.0809)
Observations	777,908	777,908	737,447
R-squared	0.166	0.189	0.204
Lender FE	YES	YES	YES
State FE		YES	YES

Robust standard errors in parentheses
**** p<0.01, ** p<0.05, * p<0.1*

Table 6
Rate Spread Regressions - VA Markets

Rate Spread$_i$	(1)	(2)	(3)
Hispanic	0.0153***	0.00909***	0.00958***
	(0.00286)	(0.00289)	(0.00293)
DualRace	-0.00291	-0.0126	-0.0160
	(0.0139)	(0.0138)	(0.0143)
AmericanIndian	0.00322	0.00199	0.000667
	(0.00891)	(0.00889)	(0.00878)
Asian	-0.0496***	-0.0666***	-0.0669***
	(0.00463)	(0.00468)	(0.00468)
Black	0.0779***	0.0789***	0.0771***
	(0.00246)	(0.00251)	(0.00254)
Pacific	-0.00463	-0.0193	-0.0127
	(0.0118)	(0.0118)	(0.0116)
Female	-0.0201***	-0.0206***	-0.0219***
	(0.00276)	(0.00275)	(0.00279)
Joint	0.00715***	0.00956***	0.00947***
	(0.00162)	(0.00162)	(0.00163)
YA	-0.0451***	-0.0477***	-0.0454***
	(0.00190)	(0.00189)	(0.00191)
OldAge	-0.00798***	-0.0154***	-0.0163***
	(0.00190)	(0.00190)	(0.00192)
DI - 1	-0.141***	-0.135***	-0.137***
	(0.00373)	(0.00372)	(0.00371)
DI - 2	-0.112***	-0.109***	-0.109***
	(0.00228)	(0.00227)	(0.00228)
DI - 3	-0.0584***	-0.0576***	-0.0577***
	(0.00219)	(0.00217)	(0.00218)
DI - 5	0.0139***	0.0131***	0.0140***
	(0.00210)	(0.00209)	(0.00211)
DI - 6	-0.00408	-0.00238	-0.00284
	(0.00466)	(0.00466)	(0.00462)
Ln(Income)	0.0291***	0.0347***	0.0333***
	(0.00126)	(0.00126)	(0.00126)
Ln(Loan Amount)	0.209***	0.196***	0.220***
	(0.0413)	(0.0420)	(0.0449)
Ln(Property Value)	-0.514***	-0.531***	-0.557***
	(0.0414)	(0.0421)	(0.0451)
Loan Term	-0.00232***	-0.00236***	-0.00237***
	(2.80e-05)	(2.78e-05)	(2.79e-05)
Ln(LTV)	-0.0495	-0.0390	-0.0618
	(0.0407)	(0.0414)	(0.0443)
Ln(Origination Charges)			0.00641***
			(0.000236)
Refinance	-0.0982***	-0.0931***	-0.102***
	(0.00452)	(0.00448)	(0.00451)
Cash-Out Refinance	0.0457***	0.0395***	0.0302***
	(0.00210)	(0.00212)	(0.00219)
Constant	4.980***	5.157***	5.257***
	(0.190)	(0.194)	(0.208)
Observations	393,748	393,748	372,130
R-squared	0.264	0.274	0.287
Lender FE	YES	YES	YES
State FE		YES	YES

Robust standard errors in parentheses
**** p<0.01, ** p<0.05, * p<0.1*

Table 7
Total Loan Cost Regressions

Ln(Total Loan Costs)	(1)	(2)	(3)
Hispanic	0.185***	0.0599***	0.0133***
	(0.00231)	(0.00215)	(0.00235)
DualRace	-0.0739***	-0.0467*	-0.0676***
	(0.0255)	(0.0241)	(0.0241)
AmericanIndian	-0.0559***	-0.0453***	-0.0452***
	(0.0115)	(0.0108)	(0.0109)
Asian	0.0985***	0.00863	-0.0114*
	(0.00643)	(0.00611)	(0.00613)
Black	-0.0128***	-0.0479***	-0.0478***
	(0.00299)	(0.00282)	(0.00289)
Pacific	0.0133	0.00792	-0.0338**
	(0.0173)	(0.0165)	(0.0167)
Female	0.146***	0.00133	-0.000382
	(0.00218)	(0.00201)	(0.00201)
Joint	0.0491***	0.0237***	0.0206***
	(0.00232)	(0.00220)	(0.00220)
YA	0.0400***	0.0318***	0.0302***
	(0.00213)	(0.00201)	(0.00201)
OldAge	-0.0982***	0.0615***	0.0554***
	(0.00280)	(0.00273)	(0.00273)
DI - 1	-0.241***	-0.0772***	-0.0707***
	(0.00623)	(0.00596)	(0.00593)
DI - 2	-0.0831***	0.00426	0.0135***
	(0.00332)	(0.00319)	(0.00319)
DI - 3	-0.0509***	-0.00527*	-0.000288
	(0.00297)	(0.00284)	(0.00283)
DI - 5	0.0646***	0.0178***	0.0108***
	(0.00228)	(0.00216)	(0.00215)
DI - 6	-0.298***	-0.121***	-0.135***
	(0.0128)	(0.0124)	(0.0124)
Ln(Income)	0.0455***	0.0186***	0.0170***
	(0.00180)	(0.00169)	(0.00171)
Ln(Loan Amount)	1.133***	1.372***	1.378***
	(0.0292)	(0.0251)	(0.0252)
Ln(Property Value)	-0.707***	-0.757***	-0.795***
	(0.0293)	(0.0251)	(0.0252)
Loan Term	0.000504***	-5.42e-05	-0.000116***
	(3.85e-05)	(3.64e-05)	(3.65e-05)
Ln(LTV)	-0.661***	-0.452***	-0.452***
	(0.0302)	(0.0238)	(0.0238)
Refinance	0.0938***	0.00885***	0.0132***
	(0.00364)	(0.00329)	(0.00328)
Cash-Out Refinance	0.180***	0.282***	0.295***
	(0.00308)	(0.00302)	(0.00306)
VA Loan		-0.771***	-0.764***
		(0.00317)	(0.00318)
Constant	6.022***	3.343***	3.756***
	(0.142)	(0.113)	(0.117)
Observations	1,124,632	1,124,632	1,124,632
R-squared	0.097	0.185	0.191
Lender FE	YES	YES	YES
Loan Type FE		YES	YES
State FE			YES

Robust standard errors in parentheses
**** p<0.01, ** p<0.05, * p<0.1*

Table 8
Total Loan Costs Regressions - FHA Markets

Ln(Total Loan Costs)	(1)	(2)
Hispanic	0.0671***	0.0353***
	(0.00114)	(0.00121)
DualRace	0.0245**	0.0180
	(0.0124)	(0.0122)
AmericanIndian	0.00692	0.00873*
	(0.00521)	(0.00520)
Asian	0.0238***	0.0133***
	(0.00267)	(0.00266)
Black	0.0330***	0.0308***
	(0.00125)	(0.00127)
Pacific	0.0386***	0.0175**
	(0.00711)	(0.00710)
Female	0.00543***	0.00358***
	(0.00105)	(0.00104)
Joint	-0.00203**	-0.00403***
	(0.00104)	(0.00102)
YA	-0.0135***	-0.0146***
	(0.000959)	(0.000944)
OldAge	0.0384***	0.0339***
	(0.00139)	(0.00139)
DI - 1	-0.0599***	-0.0572***
	(0.00396)	(0.00393)
DI - 2	-0.00427**	0.00225
	(0.00167)	(0.00164)
DI - 3	-0.00722***	-0.00337**
	(0.00138)	(0.00136)
DI - 5	0.00968***	0.00497***
	(0.000991)	(0.000976)
DI - 6	-0.0549***	-0.0633***
	(0.00767)	(0.00762)
Ln(Income)	0.0230***	0.0216***
	(0.00101)	(0.00102)
Ln(Loan Amount)	1.179***	1.186***
	(0.0285)	(0.0286)
Ln(Property Value)	-0.535***	-0.574***
	(0.0283)	(0.0283)
Loan Term	0.000142***	9.73e-05**
	(4.16e-05)	(4.19e-05)
Ln(LTV)	-0.378***	-0.380***
	(0.0230)	(0.0230)
Refinance	0.00853***	0.0129***
	(0.00259)	(0.00259)
Cash-Out Refinance	0.0828***	0.0921***
	(0.00345)	(0.00348)
Constant	2.588***	2.895***
	(0.105)	(0.106)
Observations	744,579	744,579
R-squared	0.479	0.493
Lender FE	YES	YES
State FE		YES

Robust standard errors in parentheses
**** p<0.01, ** p<0.05, * p<0.1*

Table 9
Total Loan Cost Regressions - VA Markets

Ln(Total Loan Costs)	(1)	(2)
Hispanic	-0.00183	-0.0787***
	(0.00950)	(0.00983)
DualRace	-0.150***	-0.185***
	(0.0506)	(0.0506)
AmericanIndian	-0.138***	-0.136***
	(0.0289)	(0.0289)
Asian	-0.0481**	-0.0796***
	(0.0188)	(0.0189)
Black	-0.242***	-0.239***
	(0.00833)	(0.00852)
Pacific	-0.0475	-0.106***
	(0.0372)	(0.0375)
Female	-0.0478***	-0.0521***
	(0.00936)	(0.00934)
Joint	0.0557***	0.0474***
	(0.00530)	(0.00530)
YA	0.135***	0.139***
	(0.00666)	(0.00669)
OldAge	0.0702***	0.0659***
	(0.00606)	(0.00609)
DI - 1	-0.0434***	-0.0327***
	(0.0111)	(0.0111)
DI - 2	0.0240***	0.0356***
	(0.00713)	(0.00711)
DI - 3	0.00474	0.0109
	(0.00709)	(0.00707)
DI - 5	0.0277***	0.0163**
	(0.00716)	(0.00714)
DI - 6	-0.0512***	-0.0761***
	(0.0168)	(0.0167)
Ln(Income)	0.00532	0.00228
	(0.00380)	(0.00383)
Ln(Loan Amount)	3.288***	3.301***
	(0.319)	(0.319)
Ln(Property Value)	-2.715***	-2.746***
	(0.319)	(0.320)
Loan Term	-1.82e-05	-9.79e-05
	(6.77e-05)	(6.78e-05)
Ln(LTV)	-2.467***	-2.469***
	(0.328)	(0.329)
Refinance	-0.0760***	-0.0743***
	(0.0134)	(0.0134)
Cash-Out Refinance	0.518***	0.549***
	(0.00596)	(0.00611)
Constant	12.33***	12.62***
	(1.520)	(1.524)
Observations	380,053	380,053
R-squared	0.101	0.108
Lender FE	YES	YES
State FE		YES

Robust standard errors in parentheses
**** p<0.01, ** p<0.05, * p<0.1*

Table 10
Upfront Cost Regressions

Ln(Origination Charges)	(1)	(2)	(3)
Hispanic	0.429***	0.156***	0.165***
	(0.00779)	(0.00764)	(0.00815)
DualRace	-0.0128	0.0420	0.0551
	(0.0620)	(0.0592)	(0.0592)
AmericanIndian	0.147***	0.165***	0.112***
	(0.0313)	(0.0303)	(0.0305)
Asian	0.261***	0.0660***	0.0451**
	(0.0190)	(0.0185)	(0.0186)
Black	0.112***	0.0346***	0.124***
	(0.00823)	(0.00799)	(0.00826)
Pacific	0.131***	0.113**	0.0460
	(0.0481)	(0.0465)	(0.0469)
Female	0.359***	0.0465***	0.0432***
	(0.00709)	(0.00700)	(0.00698)
Joint	0.105***	0.0503***	0.0424***
	(0.00671)	(0.00654)	(0.00653)
YA	-0.0393***	-0.0592***	-0.0771***
	(0.00668)	(0.00646)	(0.00644)
OldAge	-0.0887***	0.259***	0.254***
	(0.00759)	(0.00760)	(0.00761)
DI - 1	-0.466***	-0.109***	-0.0739***
	(0.0188)	(0.0186)	(0.0185)
DI - 2	-0.169***	0.0216**	0.0359***
	(0.00987)	(0.00966)	(0.00963)
DI - 3	-0.106***	-0.00722	0.00118
	(0.00872)	(0.00847)	(0.00844)
DI - 5	0.0969***	-0.00350	0.00354
	(0.00705)	(0.00688)	(0.00687)
DI - 6	-0.572***	-0.188***	-0.149***
	(0.0290)	(0.0283)	(0.0283)
Ln(Income)	0.0997***	0.0414***	0.0673***
	(0.00549)	(0.00551)	(0.00555)
Ln(Loan Amount)	-0.138***	0.381***	0.415***
	(0.0398)	(0.0332)	(0.0334)
Ln(Property Value)	0.0616	-0.0500	-0.232***
	(0.0396)	(0.0331)	(0.0335)
Loan Term	-0.00109***	-0.00229***	-0.00243***
	(9.43e-05)	(9.23e-05)	(9.31e-05)
Ln(LTV)	-0.630***	-0.191***	-0.239***
	(0.0351)	(0.0288)	(0.0291)
Refinance	0.773***	0.596***	0.587***
	(0.0122)	(0.0120)	(0.0120)
Cash-Out Refinance	1.190***	1.413***	1.357***
	(0.00844)	(0.00848)	(0.00856)
VA Loan		-1.684***	-1.648***
		(0.00742)	(0.00748)
Constant	9.338***	3.596***	5.683***
	(0.179)	(0.154)	(0.174)
Observations	1,109,577	1,109,577	1,109,577
R-squared	0.137	0.181	0.188
Lender FE	YES	YES	YES
Loan Type FE		YES	YES
State FE			YES

Robust standard errors in parentheses
**** p<0.01, ** p<0.05, * p<0.1*

Appendix C

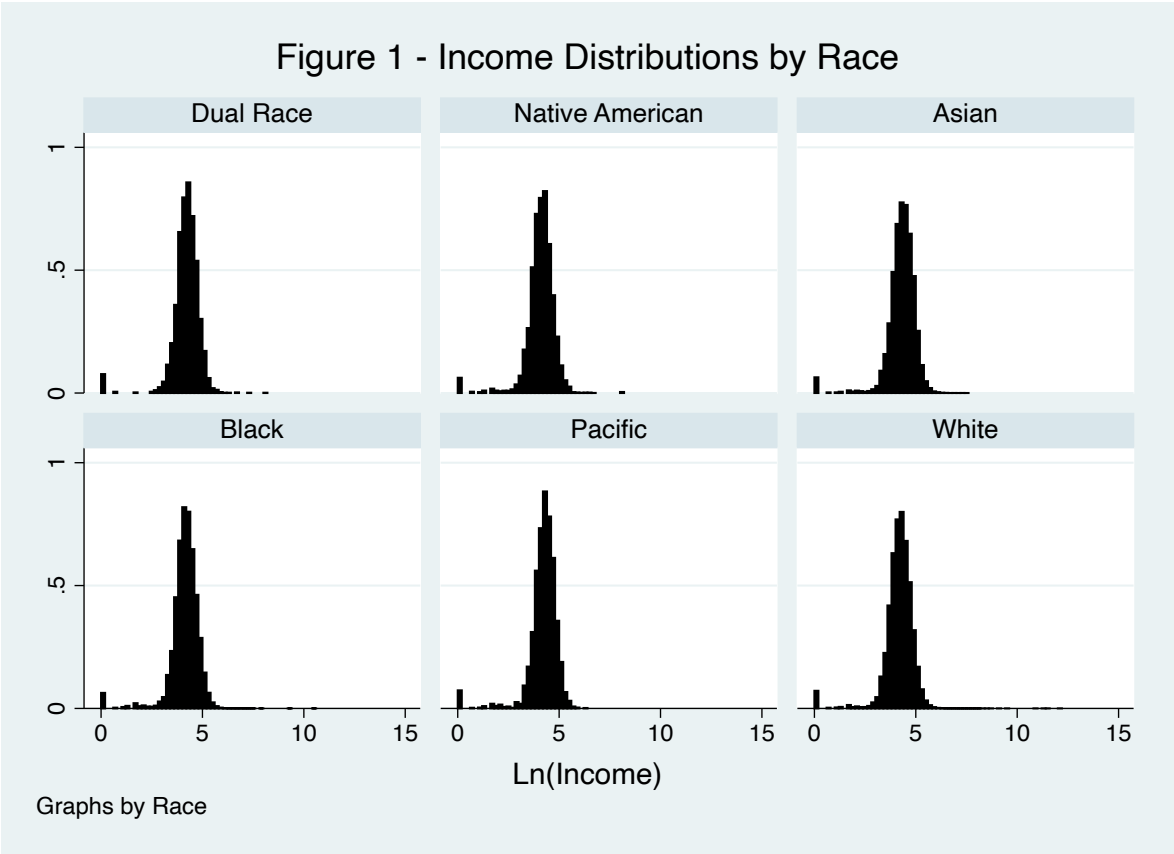
Figure 1 - Income Distributions by Race
Dual Race
Native American
Asian
Black
Pacific
White
0
.5
1
0
5
10
15
Ln(Income)
Graphs by Race

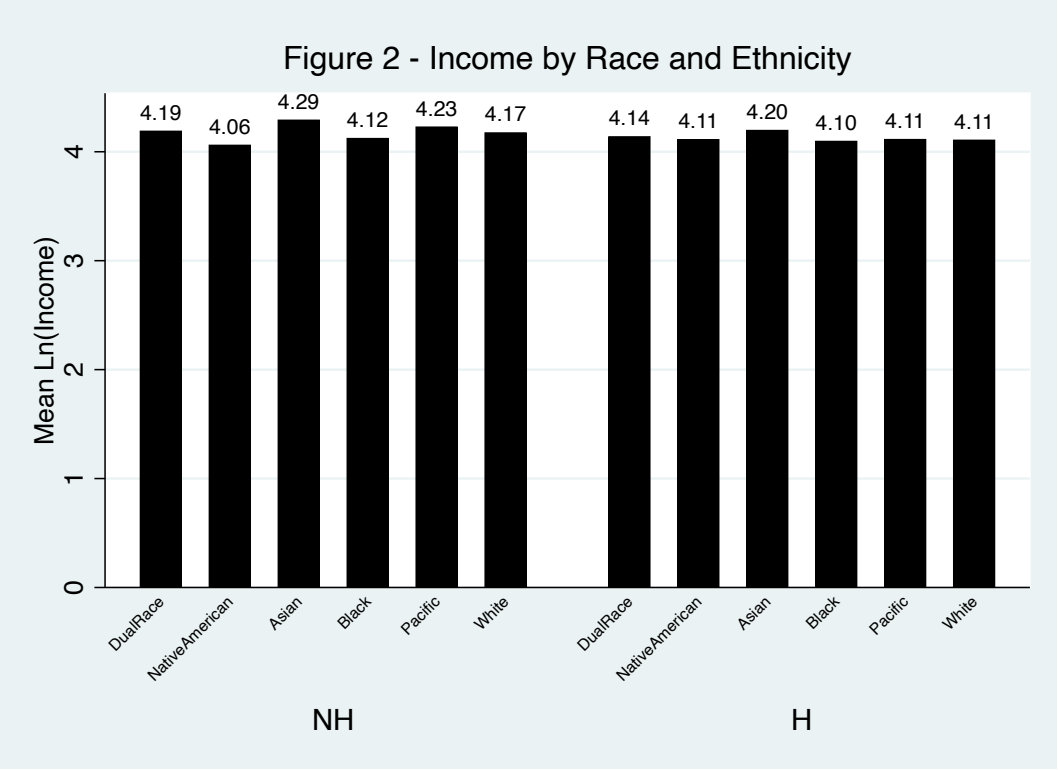
Figure 2 - Income by Race and Ethnicity
Mean Ln(Income)
0
1
2
3
4
4.19
4.06
4.29
4.12
4.23
4.17
4.14
4.11
4.20
4.10
4.11
4.11
DualRace
NativeAmerican
Asian
Black
Pacific
White
DualRace
NativeAmerican
Asian
Black
Pacific
White
NH
H

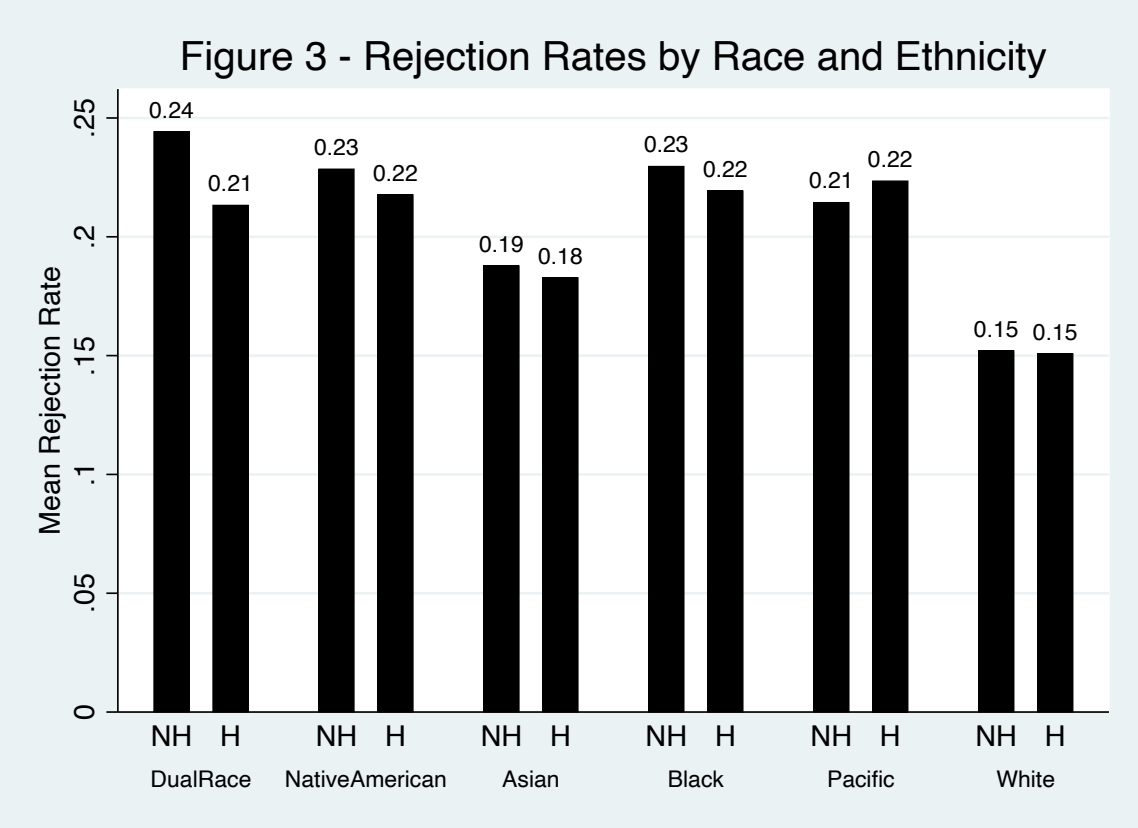
Figure 3 - Rejection Rates by Race and Ethnicity
Mean Rejection Rate
0
.05
.1
.15
.2
.25
0.24
0.21
0.23
0.22
0.19
0.18
0.23
0.22
0.21
0.22
0.15
0.15
NH H
DualRace
NativeAmerican
Asian
Black
Pacific
White

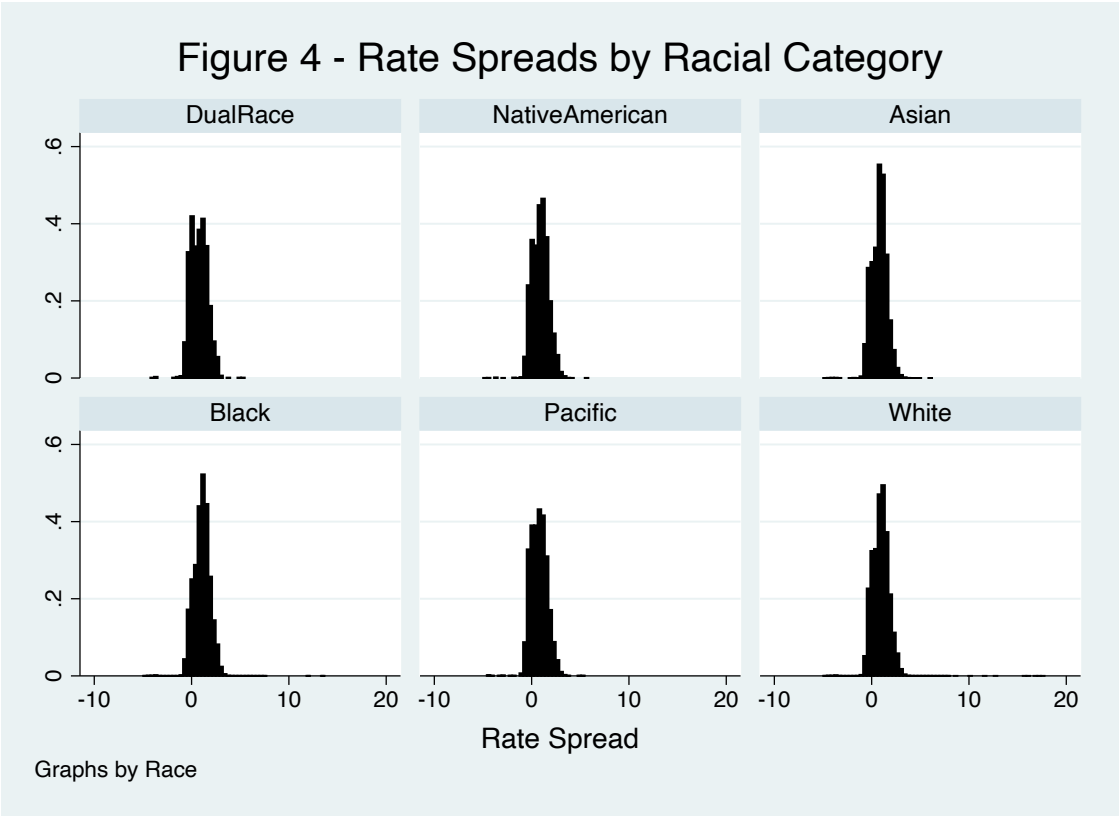
Figure 4 - Rate Spreads by Racial Category
DualRace
NativeAmerican
Asian
Black
Pacific
White
0
.2
.4
.6
-10
0
10
20
Rate Spread
Graphs by Race

References

Arrow, Kenneth J. (1973). The Theory of Discrimination. *Discrimination in Labor Markets,* 3(10), 3–33

Bartlett, R., Morse, A., Stanton, R., & Wallace, N. (2019). Consumer-Lending Discrimination in the FinTech Era. *NBER Working Paper Series*, 44.

Bayer, Patrick., Ferreira, Fernando., & Ross, S.L (2017). 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

Berkovec, J.A., Canner, G.B., Gabriel, S.A. & Hannan, T.H (1994). Race, redlining, and residential mortgage loan performance. *Journal of Real Estate Finance and Economics*, 9, 263–294.

Berkovec, J.A., Canner, G.B., Gabriel, S. A., & Hannan, T.H. (1998). Discrimination, Competition, and Loan Performance in FHA Mortgage Lending. *Review of Economics and Statistics*, *80*(2), 241–250.

Bhutta, N., & Hizmo, A. (2020). Do Minorities Pay More for Mortgages? *The Review of Financial Studies*, 28.

Black, B. H., Schweitzer, R. L., & Mandell, L. (1978). Discrimination in Mortgage Lending. *The American Economic Review*, 68(2), 7.

Canner, G. B., Gabriel, S. A., & Woolley, J. M. (1991). Race, Default Risk Mortgage Lending: A Study of the FHA and Conventional Loan Markets. *Southern Economic Journal*, 58(1), 249-262.

Cheng, P., Lin, Z. & Liu, Y. Racial Discrepancy in Mortgage Interest Rates. *Journal of Real Estate Finance and Economics*, 51, 101–120 (2015)

Fuster, Andreas., Goldsmith-Pinkham, Paul., Ramadorai, Tarun., & Walther, Ansgar (2017). Predictably Unequal? The Effects of Machine Learning on Credit Markets. *CEPR Discussion Paper No. DP12448.*

Fuster, Andreas., Plosser, Matthew., Schnabl, Philipp., & Vickery, James (2019). The Role of Technology in Mortgage Lending. *The Review of Financial Studies,* 32(5), 1854-1899.

Gabriel, S. A., & Rosenthal, S. S. (1991). Credit Rationing, Race, and the Mortgage Market. *Journal of Urban Economics*, 29(3), 371-379.

Ghent, A. C., Hernandez-Murillo, R., & Owyang, M. T. (2014). Differences in Subprime Loan Pricing across Races and Neighborhoods. *Regional Science and Urban Economics*, 48, 199-215.

Geoffrey M. B. Tootell, Redlining in Boston: Do Mortgage Lenders Discriminate Against Neighborhoods?, *The Quarterly Journal of Economics*, Volume 111, Issue 4, November 1996, Pages 1049–1079

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/Comments. *Brookings-Wharton Papers on Urban Affairs*, 33-63.

Hawley, C. B., & Fujii, E. T. (1991). Discrimination in Consumer Credit Markets. *Eastern Economic Journal*, *17*(1), 21–30. JSTOR.

Johnson, J. B. (2019, October 30). The Gender Gap in the Finance Industry. Accounting Principals Blog. https://blog.accountingprincipals.com/gender-gap-in-finance industry/#:~:text=In%20the%20U.S.%2C%20the%20average,workers%20partially%20cause%20this%20disparity.&text=The%20median%20earnings%20for%20a,for%20a%20female%20financial%20manager.

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.

Munnell, A., Geoffrey M. B. Tootell, Browne, L., & McEneaney, J. (1996). Mortgage Lending in Boston: Interpreting HMDA Data. *The American Economic Review*, 86(1), 25-53.

Macias, J. (2018, October 16). Racial Wealth Snapshot: Latino Americans. Racial Wealth Snapshot: Latino Americans | Prosperity Now. https://prosperitynow.org/blog/racial-wealth-snapshot-latino-americans.

Quarterly Residential Vacancies and Homeownership, First Quarter 2021. (2021). US Census, CB21.